# TMA
# 4101
# 4106
# 4111
# 4121

# Innhold

# Kapittel 1

# Introduksjon

Realfag er en slags lek med følgende regler:

1 Gjør grunnleggende antagelser.

2 Utled konsekvenser av antagelsene og fremsett hypoteser som lar seg teste.

3 Test hypotesene ved repeterbare eksperimenter.

4 Skriv ned hva du har gjort, slik at andre kan repetere eksperimentene og dobbeltsjekke at du ikke har gjort noen feil.

I noen fag kan det være så vanskelig å presisere de grunnleggende antagelsene at utledningen av konsekvenser blir tilnærmet meningsløs. Matematikk sitter helt i den andre enden av skalaen. Man kan alltid skrive opp antagelsene så og si dønn presist. Gale hypoteser kan sables ned med penn og papir, og spørsmålet om repeterbarhet faller bort.

I matematikk er det derfor et primært fokus på utledning. Mange mennesker som liker matematikk vil påstå at matematikk i bunn og grunn handler om å lære seg å tenke. I dette kompendiet har jeg imidlertid fokusert på lesbarhet heller enn stringens, og derfor er det viktig å slå opp i alternative kilder. I begynnelsen av hvert kapittel kommer det en liten grå boks med henvisning til (utgavenummer i parentes):

- Walter Rudin:
  Principles of Mathematical Analysis (3)
- Erwin Kreyszig:
  Advanced Engineering Mathematics (10)
- Robert Adams/Christopher Essex:
  Calculus, A complete course (7)
- Lars Ahlfors:
  Complex analysis (2)
- Elias Stein/ Rami Shakarchi:
  Fourier Analysis (1)
- Tom Lindstrøm:
  Kalkulus (3)
- Tom Lindstrøm:
  `https://www.uio.no/studier/emner/matnat/math/MAT1110/v09/FVLAbok.pdf`

Hvis du liker stringens, les Rudin, Ahlfors og Stein. Disse regnes som klassikere. Hvis du lurer mest på hva all matematikken skal brukes til og har sterk rygg, les Kreyszig. (Den er en ordentlig murstein.) Hvis du liker gode og lettleste amerikanske forklaringer med fine figurer, les Adams. Lindstrøms bøker er ryddige, stringente og fantastisk skrevet, men Adams har flere figurer og Kreyszig har flere anvendelser. Ingen av bøkene dekker alt vi skal gjennom. Derav dette kompendiet.

I blå bokser finner du definisjoner:

**Det vakreste**

En *fourierrekke* er en rekke

$$h(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

der

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt$$

I røde bokser finner du utledete setninger. Som oftest vil du finne utledning, eller skisse av utledning, eller henvisning til hvor man kan lese mer, i teksten rundt boksen.

**Det pene konvergensteoremet**

La $f : \mathbb{R} \to \mathbb{R}$ være en kontinuerlig deriverbar $2\pi$-periodisk funksjon, og la

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt.$$

Da er

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

for alle $t$.

I grønne bokser finner du all slags mulige andre ting, for eksempel viktige spørsmål.

**Det store spørsmålet**

Hvorfor er fourierkoeffisientene gitt ved

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt \quad ?$$

# Aksiomer

All vitenskap har en ting til felles:

**Et grunnleggende prinsipp**

Det går ikke an å gjøre deduksjon uten å legge et eller annet til grunn.

Det er nødvendig gjøre noen grunnleggende antagelser for å komme igang. Disse antagelsene har litt forskjellig navn, alt etter fagfelt - i matematikk kalles de aksiomer, og fysikerne sier postulat. I dagligtale snakker man om premisser som ligger til grunn for en diskusjon, og når man er litt usikker på noe, jobber man gjerne under en midlertidig arbeidshypotese.

Felles for begrepene antagelse, aksiom, postulat, premiss og hypotese, er at de alle betyr noe sånt som "et fundament man kan gjøre deduksjon på". Man utsetter inntil videre spørsmålet om fundamentet er trygt å bygge på, og studerer heller fundamentets logiske konsekvenser. Når man har studert konsekvense tilstrekkelig, formulerer man en spådom som kan testes, og hvis spådommen bommer, må man revurdere fundamentet.

**Eksempel 1.1.** Den spesielle relativitetsteorien postulerer at

1. Fysikkens lover er identiske i alle inertialsystemer.
2. Lysets hastighet i vakuum er den samme i alle intertialsystemer.

Dette høres selvfølgelig helt sykt ut, men fysiske eksperimenter bekrefter visst at det er sånn verden er skrudd sammen. △

**Eksempel 1.2.** Kochs postulater er et sett av kriterier som må være oppfylt for at man skal kunne fastslå at en mikroorganisme er årsak til en gitt sykdom.

1. Mikroorganismen skal kunne isoleres og fremstilles i renkultur i alle sykdomstilfeller.
2. Mikroorganismen må ikke forekomme ved andre sykdommer.
3. Den isolerte mikroorganismen må være i stand til å fremkalle sykdommen hos forsøksdyr, og mikroorganismen må kunne gjenfinnes hos disse.

(Kilde: Store medisinske leksikon.) △

**Eksempel 1.3.** Darwins postulater for evolusjon går som følger:

1. Det er variasjon mellom individer i en art.
2. Noen av disse variasjonene videreføres til avkom.
3. Det fødes mer avkom enn det er plass til.
4. Det avkommet som tilfeldigvis er best tilpasset omgivelsen har et konkurransefortrinn, og produserer mer avkom som selv overlever til reproduktiv alder.

Konsekvensene av disse postulatene er at arter vil tilpasse seg omgivelsene over tid, og omvendt. Darwin tilbrakte tid på Galapagos, og oppdaget at halsen på skilpaddene på de forskjellige øyene så ut til å være tilpasset vegetasjonen på de forskjellige øyene. På en av de minste øyene, som snart synker i havet (alle øyene på Galapagos er midlertidige vulkanske øyer) finnes det en fink som lever av å suge blod fra stjerten til hekkende sjøfugl. △

**Eksempel 1.4.** Første Mosebok postulerer at jorden er omtrent seks tusen år gammel. Hvis dette er sant, så er det noe galt med vår forståelse av grunnstoffet karbon.

Karbon forekommer i tre isotoper som kalles C-12, C-13 og C-14. Alle har en atomkjerne som inneholder seks protoner, men kjernen kan i tillegg inneholde seks, syv eller åtte nøytroner. C-14 er et ustabilt isotop, og halveringstiden er på omtrent 5700 år. Men andelen C-14 i atmosfæren ligger allikevel (på grunn av kosmisk stråling og et par atombombeprøvesprengninger) stabilt på omtrent ett atom per $10^{12}$.

Siden du puster, utveksler du hele tiden karbon med atmosfæren, og derfor er andelen C-14 i kroppen din og i atmosfæren omtrent den samme. Men når du dør, blir du avskåret fra denne utvekslingen, og etter 5700 år vil C-14-andelen i kroppen din være omtrent halvert. Siden vi har fossiler av levende vesener som har vesentlig lavere C-14-andel enn femti prosent, må jorden antakelig være mye eldre enn 6000 år.

Med andre ord: om du tror at jorden er 6000 år gammel, kan du ikke samtidig tro på fysikk. Du er nødt til å velge en av dem. △

I dette kompendiet skal vi studere litt aksiomer fra tid til annen. Det er ikke sikkert at akkurat du trenger å kunne noe særlig om aksiomene for å sende en rakett til månen eller bygge en oscillator, men jeg tar det med fordi det er en viktig del av det vitenskapelige paradigmet.

# Mengder

Det er en god investering å lære seg foran og bak på mengdesymbolene. [1] Det finnes aksiomer for mengdelære, men de er kompliserte.

En *mengde* er en samling med ting, kalt *elementer*. I grunnleggende matematikk er elementene gjerne tall, mens i sannsynlighetsregning er elementene utfall. Vi kan også ha en punktmengde i planet eller i rommet, en sirkelskive eller et kuleskall, eller noe helt annet. For eksempel er

$$A = \{1, 2, 3, 4, 5\}$$

[1] Statistikk og sannsynlighetsregning blir veldig mye enklere om man er komfortabel med mengdelære!

en mengde med fem elementer, og

$$B = \{1, 3, 5, \pi\}$$

en mengde med fire elementer. Vi skriver

$$\pi \in B$$

for å uttrykke at $\pi$ er et av elementene i $B$. Av og til bruker vi synonymene "samling", "familie" eller "klasse" istedet for mengde, fordi litt språklig variasjon kan gjøre det lettere å lese. Hvis man for eksempel trenger å definere en mengde av mengder, kan det lenger ned i teksten bli kløne te å referere til denne, siden "mengde" da både kan bety både "en av mengdene i mengden av mengder" og "selve mengden av mengder". Da er det like greit å si "en samling av mengder", slik at man senere har mulighet til å skille dem uten å finne opp masse notasjon.

Vi kan sette sammen mengder til nye mengder. To vanlige operasjoner på mengder, er *union*:

$$\cup$$

og *snitt*:

$$\cap$$

Unionen mellom $A$ og $B$ er er en mengde som inneholder alle elementer som er med i enten $A$ eller i $B$ eller i begge:

$$A \cup B = \{1, 2, 3, 4, 5, \pi\}$$

Snitt er en mengde som inneholder de elementene som er i både $A$ og i $B$:

$$A \cap B = \{1, 3, 5\}$$

Dersom alle elementer i en mengde $A$ er inneholdt i en større mengde $B$, skriver vi

$$A \subset B,$$

som er er det samme som at

$$A \cap B = A.$$

Dersom det hersker litt usikkerhet om hvorvidt $A \subset B$ eller $A = B$ skriver vi

$$A \subseteq B.$$

Vi skriver

$$A - B$$

for mengden av alle elementer som er i $A$, men ikke i $B$, og

$$\overline{A}$$

(leses "ikke $A$") for alle elementer som ikke er i $A$. En mengde uten elementer kalles "den tomme mengden", og ser slik ut:

$$\emptyset$$

Denne er praktisk å ha, for nå kan vi skrive

$$A \cap B = \emptyset$$

for å uttrykke at $A$ og $B$ ikke har noen felles elementer, og

$$A \neq \emptyset$$

for å uttrykke at det finnes noen elementer i $A$.

I matematikk er det trygt å si at størrelsen teller:

**Ordnet mengde**

En mengde er ordnet dersom det finnes en relasjon $<$ slik at

- kun en av

  $$x < y, \quad x = y, \quad y < x$$

  er sann
- dersom $x < y$ og $y < z$ er $x < z$

Den neste definisjonen er litt teknisk. Ordet "skranke" er ikke mye brukt blant ungdommen nå til dags, så her er Store Norske Leksikons definisjon: `https://snl.no/skranke`

**Skranker**

En ordnet mengde $A \subset B$ er begrenset ovenfra dersom det finnes $y \in B$ slik at $x \leq y$ for alle $x \in A$. Vi sier at $y$ er en øvre skranke for $A$. Dersom ingen $z < y$ er en øvre skranke for $A$, kalles $y$ en minste øvre skranke for $A$.

Tilsvarende definisjon finnes for nedre skranke og største nedre skranke. Alt dette kan fremstå som litt trukket ut av hatten, men skranker er en sentral komponent i vår presise forståelse av reelle tall.

**Eksempel 1.5.** Mengden av alle tall $x$ slik at $x^2 < 2$ har mange skranker, for eksempel $y = 3$, $y = 100$ og $y = \sqrt{3}$. Følg med i neste bolk for informasjon om en eventuell minste øvre skranke. △

# Viktige algebraiske strukturer

Tallmengder har antagelig vært i bruk lengre enn skriftsystemer. De første rudimentære skriftsystemer oppsto i forbindelse med handelsbokføring i Mesopotamia for litt over fem tusen år siden. Streker ble skrevet med en butt penn for å holde styr på antall, mens andre symboler (kalt piktografer, og skrevet med skarpere penn), ble funnet opp for det vi idag kaller substantiv - korn, øltønner, fisk, og så videre. Idag dominerer tall omgivelsene våre. Dersom du løfter blikket fra denne teksten er det ikke en eneste ting du ser hvis konstruksjon eller drift ikke involverer tall på et eller annet vis, enten det er radaren til flyet du sitter i, eller maskineriet som har saget opp alle panelbordene på hytteveggen din i lik bredde. Bokstavene i denne teksten er bare en haug med nullere og enere inni datamaskinen din.

Tallmengdene du har jobbet med siden barneskolen er eksempler på noe som kalles algebraiske strukturer.

**Algebraiske strukturer**

En algebraisk struktur består av

- en mengde med noen elementer i
- et sett med regler for å kombinere elementene i mengden til nye elementer i mengden
- et sett med aksiomer som reglene må tilfredsstille

En regel for å kombinere elementer kalles gjerne en binær operasjon, siden den er noe som tar inn to elementer. Tallmengdene du har regnet på siden barneskolen er alle algebraiske strukturer, med addisjon og multiplikasjon som binære operasjoner. Når læreren lærte deg å regne, var det en blanding av aksiomatiske regneregler, for eksempel

$$a + b = b + a$$

og regneregler som var utledet fra aksiomene, slik som

$$a^n \cdot a^m = a^{m+n}.$$

Læreren nevnte mest sannsynlig ikke at noen av disse er aksiomer og noen er utledete, men henspilte på diverse former for sunt bondevett for å argumentere for reglenes plausibilitet.

Tallmengder er de mest grunnleggende eksemplene på algebraiske strukturer, men det finnes andre viktige eksempler. Boolsk algebra brukes til å modellere det som skjer i datamaskinen din, vektorrom brukes til å systematisere kunnskap om lineære differensialligninger, og grupper brukes i kryptografi og til å analysere symmetri i molekyler. Listen er lang.

De viktigste tallmengdene for en ingeniør er:

- De naturlige tallene $\mathbb{N}$
- De hele tallene $\mathbb{Z}$
- De rasjonale tallene $\mathbb{Q}$
- De reelle tallene $\mathbb{R}$
- De komplekse tallene $\mathbb{C}$
- Kvaternionene $\mathbb{H}$

Disse sitter inni hverandre som matrjosjkadukker:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \subset \mathbb{H}.$$

De tre første, ($\mathbb{N}$, $\mathbb{Z}$ og $\mathbb{Q}$), jobbet du med allerede på barneskolen. Den fjerde ($\mathbb{R}$) ble ikke gjennomskuet ordentlig av menneskeheten før andre halvdel av det nittende århundre. [2] Moderne fysikk, kjemi, elektroteknikk og kybernetikk er utenkelig uten den femte ($\mathbb{C}$), og den sjette ($\mathbb{H}$) får du bruk for dersom du går kyb og skal jobbe med satelitter.

Alle unntatt $\mathbb{N}$ er eksempler på en type algebraisk struktur som kalles "ring", mens $\mathbb{Q}$, $\mathbb{R}$ og $\mathbb{C}$ er eksempler på noe som kalles "kropp". [3] En kropp er en algebraisk struktur med to *operasjoner*, addisjon ($+$) og og mutliplikasjon ($\cdot$). Det er vanlig å sløyfe gangetegnet, og skrive

$$x \cdot y = xy.$$

La $F$ være en kropp. Følgende aksiomer skal være tilfredsstilt for addisjon:

**Addisjon**

1 $F$ er *lukket* under addisjon:
$x, y \in F \implies x + y \in F$

2 Addisjonen er *assositativ*:
$(x + y) + z = x + (y + z)$

3 Addisjonen er *kommutativ*:
$x + y = y + x$

4 *Additiv identitet*:
Det finnes et element 0 slik at $x + 0 = x$

5 *Additiv invers*:
For hver $x$ finnes $y$ slik at $x + y = 0$

Det additive inverselement til $x$ skrives $-x$. Merk at 0 er sin egen additive invers. Følgende aksiomer skal være tilfredsstilt for multiplikasjon:

**Multiplikasjon**

6 $F$ er *lukket* under multiplikasjon:
$x, y \in F \implies x \cdot y \in F.$

7 Mutliplikasjonen er *assositativ*:
$(x \cdot y) \cdot z = x \cdot (y \cdot z)$

8 Multiplikasjonen er *kommutativ*:
$x \cdot y = y \cdot x.$

9 *Multiplikativ identitet*:
Det finnes et elelement 1 slik at $1 \cdot x = x.$

10 *Multiplikativ invers*:
For hver $x \neq 0$ finnes $y$ slik at $x \cdot y = 1.$

Det additive inverselement til $x$ skrives

$$1/x \quad \text{eller} \quad \frac{1}{x}.$$

[2] Flere tusen år etter at Arkimedes skjønte at arealet under grafen til $y = x^2$ går som $x^3/3$.

[3] Kropp heter 'field' på engelsk. En artigere oversettelse ville kanskje vært 'felt', 'åker', 'eng', eller 'bane'.

Til slutt er det et aksiom for rekkefølgen på regneoperasjonene.

**Det distributive aksiomet**

11 Operasjonene er *distributive*:
$(x+y)\cdot z = xz + yz$.

Fra disse aksiomene kan vi gå i gang med å utlede flere regneregler.

**Eksempel 1.6.** Hvis du går på fest og sier du er matematiker, spør folk gjerne "hvorfor er $(-1)\cdot(-1) = 1$, egentlig?"

Merk først at

$$y \cdot x = y \cdot (x+0) = y \cdot x + y \cdot 0$$

Dersom vi legger til $-y\cdot x$ på begge sider, står vi igjen med

$$0 = y \cdot 0,$$

så det at null ganger noe må være null, følger av aksiomene. Vi kan nå bruke at

$$\begin{aligned} 0 &= 0 \cdot (-1) \\ &= (1-1)\cdot(-1) \\ &= 1\cdot(-1) + (-1)\cdot(-1) \\ &= -1 + (-1)\cdot(-1) \end{aligned}$$

og konkludere med at $-1$ og $(-1)\cdot(-1)$ må summere til null. Altså er de hverandres additive invers, og derfor må $(-1)\cdot(-1) = 1$. △

**Eksempel 1.7.** Du opererer med et tallmengde basert på $\mathbb{Z}$ når du ser på klokken. Dersom klokken er ti om morgenen, og du skal gå hjem fra gløs om syv timer, får du det relevante klokkeslettet ved å regne ut

$$10 + 7 = 5.$$

Dette kalles *klokkearitmetikk*, og tallmengdet heter $\mathbb{Z}_{12}$. Dette er ikke en kropp, men $\mathbb{Z}_n$ er dersom $n$ er et primtall. Den minste kroppen er $\mathbb{Z}_2$, og den har kun elementene 0 og 1. Datamaskinen din er helt avhengig av vår forståelse av $\mathbb{Z}_2$ for å fungere. △

**Ordnet kropp**

En *ordnet kropp* er en kropp som er en ordnet mengde og i tillegg tilfredsstiller

- $x + y < x + z$ dersom $y < z$
- $xy > 0$ når $x > 0$ og $y > 0$

La oss nå ta en liten runde på de seks tallmengdene.

## Naturlige tall

De naturlige tallene består av de tallene du bruker når du teller:

$$\mathbb{N} = \{1, 2, 3...\}$$

Studiet av $\mathbb{N}$ kalles tallteori. Det finnes mange gode bøker om tallteori, men tallteori har ikke tradisjonelt vært pensum i den grunnleggende sivingmatematikken på Gløshaugen. De naturlige tallene utgjør en ordnet mengde, men ikke en kropp, for alle elementer i dette tallmengden mangler additiv invers, og de aller fleste mangler multiplikativ invers.

Levende vesener har nok brukt de første elementene i dette systemet siden lenge før H. Sapiens så dagens lys for omtrent tre hundre tusen år siden. [4] Havørnen kan visst telle til tre. Johan Willgohs, zoolog og pioner innen rovfuglvern, oppdaget på sekstitallet at dersom havørnen observerte en, to eller tre mennesker gå inn i et observasjonstelt i nærheten av redet sitt, fór den derifra, betraktet teltet fra lang avstand, og kom ikke tilbake til redet før den hadde observert det samme antall mennesker forlate teltet. Hvis derimot fire mennesker gikk inn i teltet, gikk havørnen i surr, og vendte tilbake så snart den hadde sett tre mennesker forlate teltet. Dette kunne man utnytte til å lure havørnen, slik at man fikk observert den. Det sies at Yamamanostammen i Brazil også kun opererer med de tre tallene "en", "to" og "mange". [5]

Et *primtall* er et tall som kun er delelig med 1 og seg selv, og som er større enn eller lik 2. En morsom sannhet som kan bevises (slå opp i en tilfeldig bok om grunnleggende tallteori) er:

**Aritmetikkens fundamentalteorem**

Alle hele tall kan faktoriseres i primtall på en entydig måte.

Det er fornuftig å definere at 1 ikke er et primtall, for ellers hadde ikke dette vært sant. Det er også mulig å bytte om på faktorenes orden, men dette ser vi åpenbart vekk fra.

**Eksempel 1.8.** $96 = 2\cdot2\cdot2\cdot2\cdot2\cdot3$ △

**Eksempel 1.9.** $51 = 3\cdot17$ △

**Eksempel 1.10.** $78 = 2\cdot3\cdot13$ △

**Eksempel 1.11.** $91354 = 2\cdot45677$ △

I matematikk er det stort sett slik at det ikke går an å forklare for lekfolk hva man forsker på om man driver med forskning. Å forstå de uavklarte spørsmålene er ikke mulig uten mange års trening. Men tallteori er til dels et unntak. Det er for eksempel ingen som vet om alle partall kan skrives som en sum av to primtall.

[4] Vi er den eneste gjenlevende arten i vår genus ("Homo"), og noen av de andre artene (for eksempel H. neanderthalensis eller H. florensienis) har antagelig hatt høyt utviklede hjerner (neandertalerne hadde større hjerner enn oss), men ikke vårt velutviklede strupehode og effektive tarmsystem. Derfor har de ikke kunnet kommunisere like presist som vi gjør, og følgelig har vi vært i stand til å utrydde dem, slik vi antagelig har gjort med så og si all megafauna utenfor det afrikanske kontinentet.

[5] Når David Attenborough i sin 13-episoders BBC-klassiker "Life on Earth" besøker hittil ukontaktede stammefolk på Papua Ny Guinea, filmes det når de forklarer hvordan de teller. De bruker fingrene på venstre hånd for en til og med fem, og så legger de høyre pekefinger på forskjellige steder på venstre arm for tall over dette - håndleddet betyr seks, underarmen syv, albuen åtte, øverst på biceps ni, skulder ti, og midt på halsen elleve. På samme måte som vi ville gjort, ler de og himler med øynene over den hvite manns dumskap når han ikke umiddelbart skjønner systemet og må få det forklart.

## Hele tall

Det er ofte slik at nye tallmengder lages dersom man mangler noen tall i det tallmengdet man har. De hele tallene

$$\mathbb{Z} = \{0, 1, -1, 2, -2...\}$$

er et godt eksempel på dette.

**Eksempel 1.12.** Amund har høns og kyr på gården sin. Vrang av vane vil han ikke si hvor mange dyr han har, men opplyser heller at de har 382 bein og 141 hoder. Hvor mange høns og hvor mange kyr har Amund? Dersom $x$ er antall kyr og $y$ antall høns, sier Amund at

$$\begin{aligned} 4x + 2y &= 382 \\ x + y &= 141 \end{aligned}$$

som gir $x = 50$ og $y = 91$. △

Men hva om late mennesker med sans for systematikk begynner å studere slike likningssystemer? Eidsnes, matematikklæreren min på gymnaset, var en lat og effektiv person. Han var glad i lage prøver med gamle oppgaver og nye tall, men han likte ikke å kontrollregne prøvene selv på forhånd [6], og dersom Amund i en slik oppgave fra hans hånd hadde opplyst at dyrene hadde for eksempel nitti hoder og fire hundre bein, ville likningssystemet blitt

$$\begin{aligned} 4x + 2y &= 400 \\ x + y &= 90 \end{aligned}$$

og løsningen vært $x = 110$ og $y = -20$. Her har vi et matematisk problem der alle parametre i det oppstilte problemet er naturlige tall, mens løsningen ikke er det.

De hele tallene er ikke så gamle, og oppsto visst på liknende vis i forbindelse med løsning av polynomlikninger. For eksempel har polynomlikningen

$$x^2 + 3x + 2 = 0$$

positive koeffisienter, men ingen positive løsninger. Akkurat som $\mathbb{N}$, er $\mathbb{Z}$ ordnet, men ikke en kropp, for de aller fleste elementene mangler multiplikativ invers.

## Rasjonale tall

De rasjonale tallene, $\mathbb{Q}$, altså alle brøker

$$\frac{m}{n}$$

der $m$ og $n$ er hele tall, er mye eldre enn de hele tallene. Å dele tre kjøttstykker likt mellom fem mennesker har nok vært en problemstilling lenge før skriftsystemer oppsto. Denne tallmengden er en ordnet kropp, men har noen alvorlige defekter: Det mangler noen viktige tall!

[6] En gang gav han oss ved en feiltagelse en oppgave om en trekant der hjørnene lå på en rett linje.

**Eksempel 1.13.** Det finnes ingen hele tall slik at

$$\sqrt{2} = \frac{m}{n}.$$

Dette kan vi bevise som følger. Anta at det finnes hele tall $m$ og $n$ slik at likningen over holder. Vi kan anta at $m$ og $n$ ikke har noen felles faktorer, for hvis de hadde hatt det, kunne vi bare forkortet brøken til de ikke lenger hadde det. Vi ganger nå hele likningen med $n$, og kvadrerer, slik at vi får

$$2n^2 = m^2.$$

Av denne likningen ser vi at $m^2$ må være et partall. Men dersom $m^2$ skal være et partall, må jo $m$ være et partall, og dette betyr at $m^2$ må være delelig med fire. Dette betyr at $m^2/2$ er et partall, og hvis vi skriver

$$n^2 = \frac{m^2}{2},$$

ser vi at $n^2$ er et partall, på da må $n$ være et partall. Men nå har vi oppnådd en selvmotsigelse. Vi vet jo at det må være mulig å velge $m$ og $n$ uten felles faktorer, og nå viser det seg at 2 må være en felles faktor allikevel. Med andre ord er det noe som er galt her. Det som er galt, er antagelsen om at man i det hele tatt kan skrive

$$\sqrt{2} = \frac{m}{n}$$

for hele tall $m$ og $n$. △

Eksemplet over illustrerer at dersom du har et kvadrat med sidekant 1, finnes det ikke noe rasjonalt tall som beskriver diagonalens lengde. Majoriteten av alle mennesker på jorden trenger matematikk først og fremst som et redskap for presis kvantifisering av verden rundt oss, så dette er rett og slett ikke bra nok. Eksemplet over har vært kjent for menneskeheten i flere tusen år. Det er lett å utvide argumentet til å vise at for eksempel $\sqrt[3]{2}$ også ikke kan skrives som en brøk. De aller fleste tall kan faktisk ikke skrives som en brøk, se Rudin kap. 1.

## Reelle tall

Vi er enige om at det burde finnes et tall for $\sqrt{2}$, siden det er diagonalen i et kvadrat med sidekant en. Men i $\mathbb{Q}$ finnes det altså ikke noe tall for denne lengden.

Georg Cantor var først ute med å publiserte en ordentlig definisjon av $\mathbb{R}$ i 1871, og vi skal ta en titt på hans konstruksjon i kapitlet om følger og rekker. Richard Dedekind gjorde det samme på en litt annen måte, og er i sin artikkel "Stetigkeit und irrationale Zahlen" (1872) veldig spesifikk på at han fikk det til den 24. november 1858 etter lang tids tenking. Både Cantor og Dedekinds konstruksjoner er relativt intuitive, men det er allikevel på sin plass med et sitat fra Walter Rudins klassiker fra 1953, "Principles of Mathematical Analysis":

" Experience has convinced me that it is pedagogically unsound (though logically correct) to start off with the construction of the real numbers from

the rational ones. ”

Nå var Walter Rudin en ganske grundig type, og i hans tid leste man bøker fra perm til perm. Denne boken er ikke skrevet for en publikum som avkrever noen som helst form for stringens, så det går fint å hoppe over dette avsnittet om man synes det er uforståelig. Det er et interessant faktum at Isaac Newton publiserte Principia i 1687, nesten to hundre år før noen skjønte hva reelle tall egentlig var. [7]

La oss nå ta et illustrerende eksempel.

**Eksempel 1.14.** Vi studerer mengden av alle positive rasjonale tall $p$ slik at $p^2 < 2$. Denne mengden har ikke noe største element, for dersom du tar et tilfeldig element $p$ i mengden, og definerer

$$q = p + \frac{2 - p^2}{a}$$

er

$$\begin{aligned} q^2 &= \left(p + \frac{2-p^2}{a}\right)^2 \\ &= p^2 + 2p\frac{2-p^2}{a} + \left(\frac{2-p^2}{a}\right)^2 \\ &= p^2 + \frac{1}{a}\left(2p(2-p^2) + \frac{(2-p^2)^2}{a}\right) \end{aligned}$$

og vi ser at dersom $a$ velges stor nok, er $q^2 < 2$.

Mengden har heller ingen minste øvre skranke. Alle rasjonale tall slik at $p^2 > 2$ er øvre skranker for mengden, men vi kan gjenta resonnementet over, og danne

$$q = p - \frac{p^2 - 2}{a}$$

slik at

$$\begin{aligned} q^2 &= \left(p - \frac{p^2-2}{a}\right)^2 \\ &= p^2 - 2p\frac{p^2-2}{a} + \left(\frac{p^2-2}{a}\right)^2 \\ &= p^2 - \frac{1}{a}\left(2p(p^2-2) - \frac{(p^2-2)^2}{a}\right) \end{aligned}$$

og likeledes se at dersom $a$ velges stor nok, får vi $q^2 > 2$. Med andre ord: hvis du har en øvre skranke, går det alltid an å finne en ny øvre skranke som er mindre. △

Dersom vi hadde hatt tilgang på $\sqrt{2}$ blant de rasjonale tallene, hadde mengden i eksemplet over hatt en minste øvre skranke, nemlig $\sqrt{2}$.

Grunnideen i Dedekinds konstruksjon er å identifisere $\sqrt{2}$ med et par av mengder $A$ og $B$ der

$$A = \text{alle positive rasjonale tall } p \text{ slik at } p^2 < 2$$

og

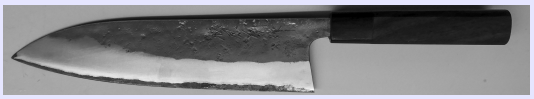
$$B = \text{alle positive rasjonale tall } q \text{ slik at } q^2 > 2$$

Vi tenker på $\sqrt{2}$ som et punkt der vi kapper den positive rasjonale tallinjen [8] i to, og skriver

$$\mathcal{A} = A|B.$$

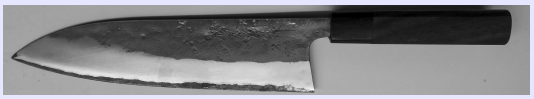
Et slikt par med mengder kalles et dedekindsk snitt [9] og er definert som følger:

**Dedekindske snitt**



Et snitt $\mathcal{A} = A|B$ er to rasjonale tallmengder $A$ og $B$ som tilfredsstiller

- $A \neq \emptyset$ og $B \neq \emptyset$
- $A \cap B = \emptyset$
- $A \cup B = \mathbb{Q}$
- $a \in A$ og $b \in B$ impliserer $a < b$
- $A$ har ikke noe største element



Siden $A$ er en del av $\mathbb{Q}$ og $B$ er resten av $\mathbb{Q}$ og alle nå har skjønt hva et snitt er, kaller vi opp snittet $\mathcal{A}$ etter mengden $A$ i paret $A|B$; referansen til $B$ er overflødig. Vi er også kun interessert i hvorvidt et rasjonalt tall er et element i $A$, så $B$ er egentlig bare med helt i starten for å med for å hjelpe intuisjonen igang.

Dersom $\mathcal{S}$ og $\mathcal{T}$ er snitt, kan vi definere $\mathcal{S} + \mathcal{T}$ på følgende vis. Det er ikke så vanskelig å vise at mengden

$$\begin{aligned} V = &\text{ alle rasjonale tall på formen } v = s + t \\ &\text{der } s \in S \text{ og } t \in T \end{aligned}$$

tilfredsstiller kriteriene for den nederste mengden i definisjonen av snitt. Vi kan derfor definere

$$\mathcal{S} + \mathcal{T} = \mathcal{V}$$

og tilsvarende for multiplikasjon. Vi sier at $\mathcal{S} < \mathcal{T}$ dersom $S \subset T$. Det er nå en litt kjedelig rutineoperasjon å sjekke at alle aksiomene for ordnede kropper er tilfredsstilt. Det som derimot er morsomt, er å si at $\mathbb{R}$ er mengden av alle dedekindsnitt og at

**Minste-øvre-skranke-egenskapen!**

Alle øvre begrensede delmengder av $\mathbb{R}$ har en minste øvre skranke. Alle nedre begrensede delmengder av $\mathbb{R}$ har en største nedre skranke.

[7] Så og si alle stringente beviser i differensial- og integralregning bygger bygger på en presis definisjon av $\mathbb{R}$, men Isaac Newton og Gudefred Leibniz gjennomskuet mye av differensial- og integralregningen lenge før noen visste hva $\mathbb{R}$ var.

[8] Man konstruerer de positive reelle tallene først, og så bekymrer man seg for de negative når man har konstruert de positive.

[9] Snitt er et litt gammeldags ord, og et mer moderne valg ville kanskje vært “klipp”, “kutt” eller “kapp”.

Dette følger så og si av definisjonen av snitt. Anta at vi har en samling snitt som er begrenset ovenfra. Det er nå lett å konstruere en minste øvre skranke. La $S$ være unionen av alle rasjonale tall i alle nedre mengder i alle snitt. Hvis du klarer å se at $\mathcal{S}$ er en minste øvre skranke for samlingen, er du i mål.

**For de interesserte**

Du finner hele konstruksjonen i Rudin og her :
`https://www.math.ntnu.no/emner/TMA4101/2020h/diverse/pugh.pdf`
I Rudin finnes et par andre referanser til bøker som også går gjennom hele greia i detalj. Her er en skikkelig rar måter å gjøre det på:
`https://www.math.ntnu.no/seminarer/perler/2016-04-15_getreal.pdf`

Til slutt to formaliteter. Dersom $a < b$, skriver vi

$$(a, b)$$

for mengden av alle reelle tall større enn $a$ og mindre enn $b$. Dette kalles et *åpent intervall*. Vi skriver

$$[a, b]$$

for mengden av alle relle tall større enn eller lik $a$ og mindre enn eller lik $b$. Dette kalles et *lukket intervall*.

## Komplekse tall

Komplekse tall er gjennomgått i mange bøker. Noen eksempler er:

- Ahlfors kap. 1
- Adams appendiks 1
- Kreyszig kap. 13

De komplekse tallene, $\mathbb{C}$, er et tallmengde som i likhet med negative tall oppsto i forbindelse med løsning av polynomlikninger. Likningen

$$x^2 + 1 = 0$$

har ingen løsning blant noen av de foregående tallmengdene, og derfor har man kommet til at det er best å definere et helt nytt tall:

**Den imaginære enheten $i$**

$$i^2 = -1$$

Det kunne vært fristende å 'løse' likningen $x^2 + 1 = 0$ for $x$, og definere

$$i = \sqrt{-1}.$$

Men vi må være litt forsiktige med denne strategien. Regneregelen

$$\sqrt{ab} = \sqrt{a}\sqrt{b}$$

gjelder ikke når både $a$ og $b$ er negative tall. Dette vises av følgende klassiske eksempel:

$$\begin{aligned} 1 &= (-1) \cdot (-1) \\ &= \sqrt{(-1) \cdot (-1)} \\ &= \sqrt{-1} \cdot \sqrt{-1} = i^2 = -1. \end{aligned}$$

Men vi skal allikevel tillate oss å bruke det nye tallet $i$ til å skrive kvadratroten av negative tall på en pen måte:

$$\sqrt{-4} = \sqrt{4 \cdot (-1)} = \sqrt{4} \cdot \sqrt{(-1)} = \pm 2i.$$

**Eksempel 1.15.** Løser vi likningen

$$x^2 + x + 1 = 0$$

gir annengradsformelen

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

at

$$x = \frac{-1 \pm \sqrt{-3}}{2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2} i. \qquad \triangle$$

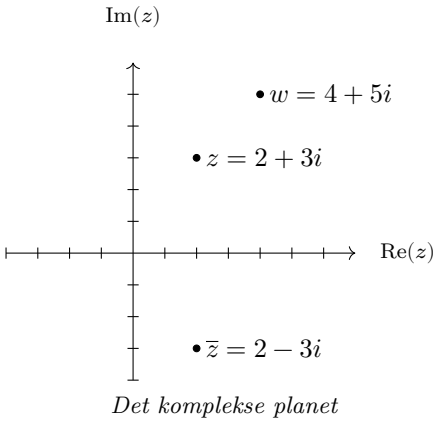
Eksemplet over inspirerer oss til å definere

**Komplekse tall på kartesisk form**

$$z = a + bi$$

Her er $a$ og $b$ reelle tall. De kalles henholdsvis *realdelen* og *imaginærdelen* til $z$, og skrives gjerne $\mathrm{Re}(z)$ og $\mathrm{Im}(z)$. Mengden av alle komplekse tall skrives $\mathbb{C}$. De reelle tallene er en delmengde av de komplekse tallene,for dersom $b = 0$, er $z$ reell. De komplekse tallene danner en kropp, men den er ikke ordnet, for det går ikke an å lage en regel som tilfredsstiller aksiomene for orden. Litt folkelig kan man si at det ikke går an å sortere komplekse tall etter størrelse på en fornuftig måte.

Et komplekst tall likner på mange måter en vektor i $\mathbb{R}^2$. Vi kan tenke at realdelen $a$ og imaginærdelen $b$ er komponenter i en vektor, og avmerke $z$ i *det komplekse planet*.



*Det komplekse planet*

La $z = a + bi$ og $w = c + di$ være komplekse tall. Regneregler for komplekse tall følger regnereglene for

reelle tall, men du må huske at $i^2 = -1$. Vi ganger sammen komplekse tall slik:

$$\begin{aligned} z \cdot w &= (a+bi)(c+di) \\ &= ac + bci + adi + bdi^2 \\ &= ac - bd + (bc + ad)i \end{aligned}$$

og deler dem slik:

$$\begin{aligned} \frac{z}{w} &= \frac{a+bi}{c+di} = \frac{a+bi}{c+di} \cdot \frac{c-di}{c-di} \\ &= \frac{ac + bd + (bc - ad)i}{c^2 + d^2} \end{aligned}$$

Addisjon og subtraksjon er trivielle operasjoner.

**Regneregler for kartesisk form**

La $z = a + bi$ og $w = c + di$ være komplekse tall. Vi har

$$\begin{aligned} z + w &= a + c + (b + d)i \\ z - w &= a - c + (b - d)i \\ z \cdot w &= ac - bd + (bc + ad)i \\ \frac{z}{w} &= \frac{ac + bd + (bc - ad)i}{c^2 + d^2} \end{aligned}$$

Merk at komplekse tall legges sammen komponentvis akkurat som vektorer i $\mathbb{R}^2$. Multiplikasjon og divisjon har ingen tilsvarende operasjoner i $\mathbb{R}^2$ i 'vanlig bruk'.

**Eksempel 1.16.** La $z = 2 + 3i$ og $w = 4 + 5i$.

$$z + w = 2 + 4 + (3 + 5)i = 6 + 8i$$

$$z - w = 2 - 4 + (3 - 5)i = -2 - 2i$$

$$\begin{aligned} z \cdot w &= (2 + 3i) \cdot (4 + 5i) \\ &= 2 \cdot 4 + 3 \cdot 4i + 2 \cdot 5i + 3 \cdot 5i^2 \\ &= 8 - 15 + (12 + 10)i = -7 + 22i. \end{aligned}$$

$$\begin{aligned} \frac{z}{w} &= \frac{2+3i}{4+5i} = \frac{2+3i}{4+5i} \cdot \frac{4-5i}{4-5i} \\ &= \frac{8 + 15 + (12 - 10)i}{16 + 25} = \frac{22}{41} + \frac{2}{41}i. \end{aligned}$$

△

Når vi deler et komplekst tall på $z = a + bi$, ganger vi oppe og nede med $z$ *konjugert*:

$$\overline{z} = a - bi$$

Merk at $z\overline{z} = a^2 + b^2$ er et reelt tall.

**Regneregler for konjugert**

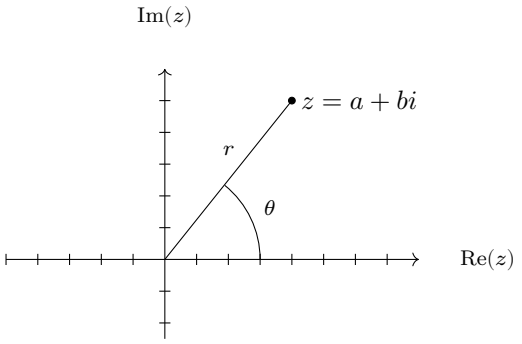
La $z = a + bi$ og $w$ være komplekse tall. Noen regneregler er:

$$\begin{aligned} \overline{z + w} &= \overline{z} + \overline{w} & \overline{z - w} &= \overline{z} - \overline{w} \\ \overline{z \cdot w} &= \overline{z} \cdot \overline{w} & \overline{z/w} &= \overline{z}/\overline{w} \\ z + \overline{z} &= 2a & z - \overline{z} &= 2bi \end{aligned}$$

La $r$ være avstanden fra det komplekse tallet $z = a + bi$ til origo, og la $\theta$ være vinkelen $z$ gjør med den reelle aksen. Noen enkle geometriske betraktninger gir

**Den ene veien**

$$\begin{aligned} a &= \mathrm{Re}(z) = r\cos\theta \\ b &= \mathrm{Im}(z) = r\sin\theta \end{aligned}$$




*Polare koordinater*

De inverse trigonometriske funksjonene kan bli litt forvirret på om $z$ ligger til høyre eller venstre for den imaginære aksen. Derfor må man være litt forsiktig når man går motsatt vei. Merk også at vi kan legge til vilkårlige multipler av $2\pi$ overalt, samt at for $z = 0$ er ikke $\theta$ definert.

Vi skriver ellers

$$|z| = r = \sqrt{a^2 + b^2} = \sqrt{z\overline{z}}$$

for avstanden fra $z$ til origo. Dette tallet kalles gjerne *absoluttverdi* eller *modulus* til $z$. Det er praktisk å bruke den samme notasjonen som for absoluttverdien til et reelt tall, siden $|z|$ blir den reelle absoluttverdien

$$|z| = \begin{cases} z & z \geq 0 \\ -z & z < 0 \end{cases}$$

dersom $z$ er reell.

Vinkelen

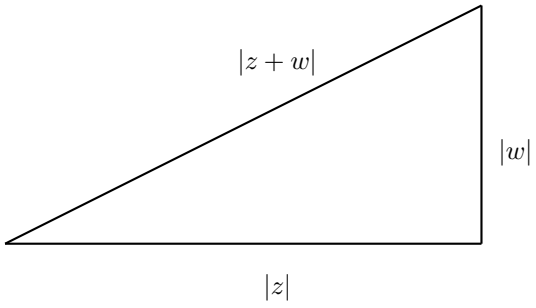
$$\theta = \arg z$$

kalles *vinkelen* eller *argumentet* til $z$. Følgende ulikhet kalles *trekantulikheten*, siden de involverte størrelsene er sidene i en rettvinklet trekant.

**Trekantulikheten**

La $z$ og $w$ være komplekse tall. Da gjelder at

$$|z + w| \leq |z| + |w|$$



### $\mathbb{R}^n$ og $\mathbb{C}^n$

Vi definerer $\mathbb{R}^n$ som mengden av alle vektorer $(x_1, x_2, \cdots, x_n)$ der $x_k \in \mathbb{R}$, og $\mathbb{C}^n$ som mengden av alle slike vektorer med $x_k \in \mathbb{C}$.

Skalarproduktet mellom vektorer i $\mathbb{R}^n$ er

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k$$

og den pytagoreiske lengden er gitt ved

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

For $\mathbb{C}^n$ er vi mest interessert i

$$\overline{\mathbf{x}} \cdot \mathbf{y} = \sum_{k=1}^{n} \overline{x_k} y_k.$$

Mer om dette senere.

## Kvaternioner

## Grupper

# Funksjoner

Euler mente at en funksjon var det samme som en formel, for eksempel

$$y = x^2$$

eller noe liknende. Han aksepterte ikke slikt som for eksempel

$$f(x) = \begin{cases} x & x > 0 \\ 1 + x & x \leq 0 \end{cases}$$

men han modererte seg visst på sine eldre dager.

**Viktig!**

En funksjon $f : A \to B$ er en regel som for hvert element i mengden $A$ tilordner ett og bare ett element i mengden $B$.

Vi snakker gjerne om en funksjon "på $A$". Uttrykket $A \to B$ leses "$A$ til $B$", slik som byens busselskap. Her er et par kommentarer:

- Vi skriver $y = f(x)$ for å signalisere at $y \in B$ er elementet som korresponderer til $x \in A$.
- Mengden $A$ kalles *definisjonsmengden* til $f$.
- Mengden $B$ kalles *verdimengden* til $f$.
- De elementene i $B$ som kommer ut av $f$ når du stapper inn hele $A$ skrives $f(A)$ og kalles *bildet* til $f$. Vi kan ha $f(A) \subset B$ eller $f(A) = B$.
- Ett og bare ett element i $B$ skal spesifiseres for hvert element i $A$!

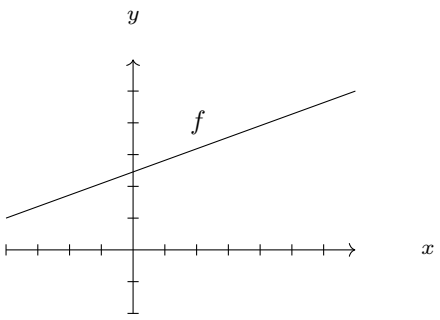
Så og si all matematikk du skal lære de neste to årene, handler om funksjoner. Funksjonsregelen spesifiseres av en matematiske likning eller en tabell eller en graf.

**Eksempel 1.17.** Funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved

$$f(x) = x^2$$

tar inn et reelt tall og kvadrerer det. △

**Eksempel 1.18.** Koordinatsystemet vi bruker i dag, med $x$- og $y$-aksen, kalles *det kartesiske koordinatsystem*, etter Rene Descartes, som innførte det i en bok i 1637. En funksjon kan beskrives av en graf i et koordinatsystem, men det er noen føringer på hvordan grafen kan se ut. Siden det kun skal være en funksjonsverdi for hvert element i definisjonsmengden, kan ikke grafen skjære en gitt vertikal linje mer enn en gang. △




**Eksempel 1.19.** En artigklassiker er $f : \mathbb{R} \to \mathbb{R}$ gitt ved

$$f(x) = \begin{cases} 1 & \text{rasjonale } x \\ 0 & \text{irrasjonale } x \end{cases}$$

△

Definisjonsmengden kan ikke inneholde elementer som regelen ikke greier å tilordne en funksjonsverdi til. Merk forskjellen på funksjonen $f$, som er en pakke med definisjonsmengde, verdimengde og funksjonsregel, og funksjonsverdien $f(x)$, som er det elementet i $B$ som korresponderer til $x$.

**Eksempel 1.20.** Det gir ingen mening å snakke om funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved

$$f(x) = \frac{1}{x}$$

for denne regelen klarer ikke spesifisere et element til $x = 0$. Det går derimot fint å prate om funksjonen $f : \mathbb{R} - \{0\} \to \mathbb{R}$ gitt ved

$$f(x) = \frac{1}{x}$$

eller funksjonen $f : (0, 1) \to \mathbb{R}$ gitt ved

$$f(x) = \frac{1}{x}$$

△

Dersom ikke definisjonsmengden er eksplisitt definert, er det underforstått hva den er. og det hender vi bruker teknisk gale uttrykksmåter, som "funksjonen $f(x)$".

**Eksempel 1.21.** Funksjonen $f(x) = \frac{1}{x}$ tar inn et reelt tall og inverterer det. Nå er det underforstått at definisjonsmengden er et eller annet som ikke inneholder 0. △

**Eksempel 1.22.** Et polynom av orden $n$ er en funksjon med funksjonsuttrykk

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0.$$

Studiet av polynomer er et av de eldste i matematikkens historie, antagelig noen tusen år gammelt. René Descartes oppfant standardnotasjonen vi bruker i dag, med koeffisienter hentet fra begynnelsen av alfabetet, og variable fra slutten av alfabetet. △

Stort sett alle funksjoner du studerte på skolen, var av typen fra $\mathbb{R} \to \mathbb{R}$. Vi skal studere mange andre typer funksjoner.

**Eksempel 1.23.** Funksjonen $f : \mathbb{N} \to \mathbb{N}$ gitt ved

$$f(n) = n! = n \cdot (n-1) \cdots 3 \cdot 2$$

vokser fryktelig fort med $n$. Vi skal få bruk for denne funksjonen flere ganger dette semesteret. △

**Eksempel 1.24.** Funksjonen $f : \mathbb{R}^m \to \mathbb{R}^n$ gitt ved

$$f(\mathbf{x}) = A\mathbf{x}$$

der $A$ er en $m \times n$-matrise og $\mathbf{x}$ er en $n \times 1$-søylevektor, er et eksempel på noe som kalles en lineæravbildning. Dette betyr at dersom $\mathbf{x}$ og $\mathbf{y}$ er vektorer og $a$ og $b$ tall, er

$$f(a\mathbf{x} + b\mathbf{y}) = af(\mathbf{x}) + bf(\mathbf{y})$$ △

**Eksempel 1.25.** Funksjonen $z : [0, 2\pi) \to C$ gitt ved

$$f(\mathbf{x}) = \cos t + i \sin t$$

er en parametrisering av enhetssirkelen i det komplekse planet. △

I disse eksemplene var definisjonsmengden en tallmengde. Men definisjonsmengden kan også være andre ting, for eksempel en mengde av funksjoner.

**Eksempel 1.26.** Det ubestemte integralet

$$\int_a^b f(x)\, dx$$

tar inn er i seg selv en funksjon som tar inn en funksjon og gir ut et tall. Denne typen funksjon kalles gjerne funksjonal, for å i språket kunne holde styr på hva som er funksjonen (det bestemte integralet) og elementene i definisjonsmengden (funksjonen som integreres). △

**Eksempel 1.27.** Vi kan også definere en funksjonal som tar inn en funksjon og gir ut funksjonens stigningstall i et punkt

$$T(f) = f'(a)$$ △

**Eksempel 1.28.** Verdimengden kan også være en mengde av funksjoner. Det bestemte integralet

$$\int_0^x f(t)\, dt$$

der $x$ er en variabel, tar inn en funksjon og gir ut en annen. Denne typen funksjon kalles gjerne operator. Derivasjonsoperatoren

$$T(f) = f'(x)$$

er et annet eksempel. △

**Eksempel 1.29.** Definisjonsmengden kan også være noe helt annet, for eksempel utfall i et utfallsrom $S$. Sannsynlighetsfunksjonen er en benevningsløs funksjon $P$ som tilordner en sannsynlighet $p$ til hvert utfall $A$ i $S$:

$$P(A) = p$$

For en sannsynlighetsfunksjon setter vi opp

**Kolmogorovs tre aksiomer**

- $0 \leq P(A) \leq 1$ for alle $A \in S$
- $P(\bigcup_i A_i) = \sum_i P(A_i)$ for alle $A_i \in S$
- $P(S) = P\left(\bigcup_I A_i\right) = \sum_i P(A_i) = 1$

Disse tre kravene sier at sannsynligheter må være mellom null og en, at sannsynligheter for disjunkte utfall kan legges sammen, og at alle sannsynligheter i utfallsrommet må summere til en (noe må skje). Fra disse aksiomene kan man for eksempel bevise at

$$P\left(\overline{A}\right) = 1 - P(A)$$

og at

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

dersom $A$ og $B$ ikke er disjunkte. △

## Injektivitet og surjektivitet

I mange tilfeller er vi interessert i å vite om det la seg gjøre å finne en funksjon som går motsatt vei av $f$.

**Spørsmål**

Dersom vi har

$$f : A \to f(A)$$

gitt ved for eksempel $y = f(x)$, er det mulig å finne en funksjon

$$g : f(A) \to A$$

som sender alt i retur, altså at

$$x = g(y) \quad \text{for alle} \quad y \in f(A)\ ?$$

Denne kalles (dersom den eksisterer) den inverse funksjonen, og vi skriver [10]

$$g = f^{-1}$$

Siden $x = g(y)$ må være entydig bestemt, ser vi umiddelbart at det ikke er mulig å definere $f^{-1}$ dersom $f(x_1) = f(x_2)$ for to elementer $x_1$ og $x_2$ i $A$. Dette motiverer følgende definisjon.

**Injektivitet**

En funksjon $f : A \to B$ er *injektiv* dersom $f(x_1) = f(x_2)$ impliserer at $x_1 = x_2$.

[10] ikke forveksle med $1/f$, som er noe helt annet unntatt når $f(x) = x$

Dersom $f : A \to f(A)$ er injektiv, kan vi alltid løse likningen $y = f(x)$ for $x$. Da får vi en likning på formen $x = g(y)$. Dersom vi setter $g$ inn i $f$ eller $f$ inn i $g$, får vi

$$g(f(x)) = f(g(x)) = x.$$

**Eksempel 1.30.** La $f : \mathbb{R} \to \mathbb{R}$ være gitt ved

$$f(x) = x^3.$$

Da er

$$f^{-1}(x) = \sqrt[3]{x}. \qquad \triangle$$

**Eksempel 1.31.** La $f : \mathbb{R} \to \mathbb{R}$ være gitt ved

$$f(x) = x^2.$$

Denne funksjonen har ingen invers. Den er ikke injektiv, siden $f(-x) = f(x)$ for alle $x \in \mathbb{R}$ $\triangle$

**Eksempel 1.32.** La $f : \mathbb{R} \to \mathbb{R}$ være gitt ved

$$f(x) = x^2 + 1.$$

Denne funksjonen har ingen invers. Den er ikke injektiv, for $f(-x) = f(x)$ for alle $x$. $\triangle$

Hvorvidt $f$ injektiv, avhenger både av funksjonsuttrykket og definisjonsmengden.

**Eksempel 1.33.** La $f : [0, 1] \to [1, 2]$ være gitt ved

$$f(x) = x^2 + 1.$$

Denne funksjonen har den inverse funksjonen $g : [1, 2] \to [0, 1]$ gitt ved

$$g(x) = \sqrt{x - 1}. \qquad \triangle$$

**Surjektivitet**

En funksjon $f : A \to B$ er *surjektiv* dersom det for hver $y \in B$ finnes en $x$ slik at $f(x) = y$. Dersom $f$ er både injektiv og surjektiv, sier vi at $f$ er *bijektiv*.

**Eksempel 1.34.** Funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved

$$f(x) = x^2 + 1$$

er ikke surjektiv, siden $f(x) \geq 1$ for alle $x$. $\triangle$

**Eksempel 1.35.** Funksjonen $f : [0, 1] \to \mathbb{R}$ gitt ved

$$f(x) = x^2 + 1$$

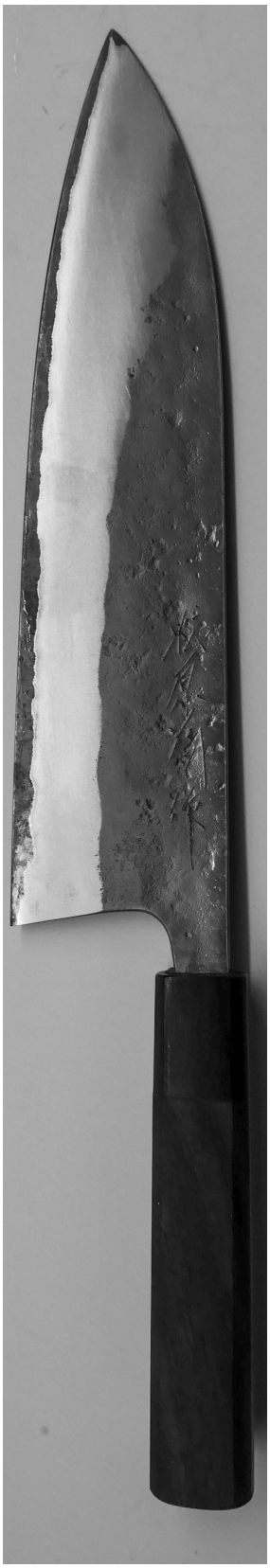
er heller ikke surjektiv. $\triangle$

**Eksempel 1.36.** Funksjonen $f : [0, 1] \to [1, 2]$ gitt ved

$$f(x) = x^2 + 1$$

er derimot surjektiv. $\triangle$

**Eksempel 1.37.** En funksjon $f : A \to f(A)$ er per definisjon surjektiv. $\triangle$



# Kapittel 2

# Konkret lineæralgebra

Det er lurt å begynne matematikkstudiet med litt grunnleggende lineæralgebra. Dette er både nyttig å kunne, og ikke så vanskelig å lære seg, sammenliknet med differensial- og integralregning. Flervariabel kalkulus blir dessuten mye lettere å lære om man kan lineæralgebra.

Alternative kilder:

- Lindstrøm FVLA kap. 1 og 3
- Kreyszig kap. 7 og 8

## Gausseliminasjon

Standardmetoden for å løse lineære likningssystemer kalles gausseliminasjon, til tross for at metoden ikke er oppfunnet av Carl Friedrich Gauss. Metoden består i å forenkle likningssystemet steg for steg uten å endre løsningsmengden.

Det er lettest å lære denne metoden gjennom et regneeksempel. Vi skal løse

$$\begin{aligned} 2x + 3y + 4z &= 4 \\ 3x + 4y + 5z &= 5 \\ 4x + 5y + 7z &= 3 \end{aligned}$$

Fra videregående skole husker du kanskje at hver av disse likningene beskriver et plan, og geometrisk sett leter likningssystemet etter et punkt $(x, y, z)$ som ligger på alle tre plan samtidig. Hvis tar tre A4-ark og fikler litt med dem, bør det være klart at et slikt likningssystem kan ha enten en, mange, eller ingen løsninger.

Før vi løser likningssystemet, skal vi bytte notasjon. Det er bedre å skrive

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ 3x_1 + 4x_2 + 5x_3 &= 5 \\ 4x_1 + 5x_2 + 7x_3 &= 3 \end{aligned}$$

for dette er mer praktisk når likningssystemene blir store. Et likningssystem i en moderne anvendelse kan fint ha en milliard likninger, men det lengste moderne alfabetet inneholder bare 74 tegn:
`https://en.wikipedia.org/wiki/Khmer_script`

**Eksempel 2.1.** Det første vi gjør, er å observere at dersom vi ganger den første likningen med 3 og den andre likningen med 2 og trekker dem fra hverandre:

$$\begin{aligned} 3(2x_1 + 3x_2 + 4x_3 &= 4) \\ -2(3x_1 + 4x_2 + 5x_3 &= 5) \end{aligned}$$

får vi likningen

$$x_2 + 2x_3 = 2$$

Det neste er å observere at en eventuell løsning av likningssystemet også må tilfredsstille denne nye likningen, siden den er en konsekvens av to likninger i det opprinnelige systemet. Hvis vi bytter ut den opprinnelige likning 2 med den nye likningen, får vi

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ x_2 + 2x_3 &= 2 \\ 4x_1 + 5x_2 + 7x_3 &= 3 \end{aligned}$$

Det er nå også slik at likningen vi hev ut er en konsekvens av de to første likningene i det nye systemet, så det nye systemet må pent ha de samme løsningene som systemet vi startet med.

Fordelen med det nye systemet er at det er litt enklere, en av $x_1$-ene er blitt borte. Kanskje det er slik at vi kan skrelle av så mye greier at løsningen til slutt står skrevet svart på hvitt på siden? Slik er det. La oss nå ta for oss likning 1 og 3, og regne ut

$$\begin{aligned} 2(2x_1 + 3x_2 + 4x_3 &= 4) \\ -(4x_1 + 5x_2 + 7x_3 &= 3) \end{aligned}$$

som blir

$$x_2 + x_3 = 5$$

og bytte ut likning 3:

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ x_2 + 2x_3 &= 2 \\ x_2 + x_3 &= 5 \end{aligned}$$

Hvis vi nå trekker likning 3 fra likning 2 i dette systemet og bytter ut likning 3, får vi

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ x_2 + 2x_3 &= 2 \\ x_3 &= -3 \end{aligned}$$

og nå har vi oppnådd en form på systemet som er spesielt gunstig, nemlig trappeform. Verdien til $x_3$ kommer nå tydelig frem av likningssystemet, og vi kan jobbe oss oppover og beregne

$$x_2 = 2 - 2 \cdot (-3) = 8$$

og

$$x_1 = (4 - 3 \cdot 8 - 4 \cdot (-3))/2 = -4$$

som er den korrekte løsningen. △

Hvis man har utført denne prosedyren noen ganger, blir man sliten i hånden, og lei av å skrive ned $x_2$ og $=$ og så videre. Følgende kompakte notasjon er bedre:

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 7 & 3 \end{array}\right)$$

Vi bruker symbolet $\sim$ for å signalisere at vi har utført et gausseliminasjonssteg som ikke endrer løsningsmengden, slik at beregningen over kan skrives

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 7 & 3 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 2 \\ 4 & 5 & 7 & 3 \end{array}\right) \sim$$
$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 1 & 5 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & -3 \end{array}\right)$$

Med litt trening er dette lettere å lese.

**Eksempel 2.2.** La oss nå ta et med komplekse tall:

$$\begin{aligned} (1-i)z + 3w &= 2 - 3i \\ iz + (1+2i)w &= 1 \end{aligned}$$

Likningssystemet har totalmatrise

$$\left(\begin{array}{cc|c} 1-i & 3 & 2-3i \\ i & 1+2i & 1 \end{array}\right).$$

Vi ønsker å kvitte oss med $i$-en til venstre i den andre raden. Den første raden ganget med $\frac{i}{1-i}$ er

$$\left(\begin{array}{cc|c} i & \frac{3i}{1-i} & \frac{3+2i}{1-i} \end{array}\right).$$

Vi trekker dette fra den andre raden og erstatter den andre raden med resultatet:

$$\left(\begin{array}{cc|c} 1-i & 3 & 2-3i \\ 0 & 1+2i-\frac{3i}{1-i} & 1-\frac{3+2i}{1-i} \end{array}\right).$$

Jeg tror vi ganger den andre raden med $1-i$ for å rydde litt:

$$\left(\begin{array}{cc|c} 1-i & 3 & 2-3i \\ 0 & 3-2i & -2-3i \end{array}\right)$$

Vi er nå klare for å beregne $w$ og $z$:

$$\begin{aligned} w &= \frac{-2-3i}{3-2i} = \frac{-2-3i}{3-2i} \cdot \frac{3+2i}{3+2i} = -i \\ z &= \frac{2-3i-3(-i)}{1-i} = 1+i \end{aligned}$$ △

# Matriser og vektorer

Men lineære likningssystemer kan ha flere løsninger. La oss løse systemet

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 6 & 6 \end{array}\right)$$

Vi kan stort sett bruke de samme eliminasjonsstegene som i forrige runde, men resultatet blir at en hel likning forsvinner:

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 6 & 6 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 2 \\ 4 & 5 & 6 & 6 \end{array}\right) \sim$$
$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 2 & 2 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Likningssystemet

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ 3x_1 + 4x_2 + 5x_3 &= 5 \\ 4x_1 + 5x_2 + 6x_3 &= 6 \end{aligned}$$

har altså de samme løsningene som likningssystemet

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ x_2 + 2x_3 &= 2 \end{aligned}$$

Hvis vi nå velger $x_3 = s$, må $x_2 = 2-2s$ og $x_1 = s-1$. Det spiller ingen rolle hva $s$ er, det blir en løsning uansett. Prøv selv med $s = 2$ eller $s = 1+i$ eller hva som helst.

Men nå er det på tide å innføre litt mer notasjon. En kolonnevektor er

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

og en radvektor er

$$\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}$$

De to viktigste regnereglene for vektorer er skalarmultiplikasjon

$$a\mathbf{x} = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix}$$

og vektoraddisjon

$$\mathbf{x} + \mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

(Tilsvarende for radvektorer.) En sammensetning av operasjonene

$$a\mathbf{x} + b\mathbf{y} = a \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + b \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ \vdots \\ ax_n + by_n \end{pmatrix}$$

kalles en *lineærkombinasjon*. Skalarene $a$ og $b$ kalles vekter. Hvis vi har $m$ vektorer $\mathbf{x}_k$, definerer vi *det lineære spennet*, eller

$$\mathrm{Sp}\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$$

som alle lineærkombinasjoner av vektorene, altså alle vektorer på formen

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_m\mathbf{x}_m.$$

**Eksempel 2.3.**

$$3\begin{pmatrix}1\\2\\3\end{pmatrix}+2\begin{pmatrix}4\\5\\6\end{pmatrix}=\begin{pmatrix}11\\16\\21\end{pmatrix}. \qquad \triangle$$

**Eksempel 2.4.** Spennet til vektorene i eksemplet over, er alle vektorer på formen

$$a\begin{pmatrix}1\\2\\3\end{pmatrix}+b\begin{pmatrix}4\\5\\6\end{pmatrix}. \qquad \triangle$$

La oss først introdusere:

**Matrise**

En $m \times n$-matrise $A$ er et tableau med tall

$$A=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}$$

Disse her

$$\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{pmatrix} \quad \begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{m2}\end{pmatrix} \quad \cdots \quad \begin{pmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{mn}\end{pmatrix}$$

kalles matrisens kolonner, og disse

$$\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\end{pmatrix}$$
$$\begin{pmatrix}a_{21} & a_{22} & \cdots & a_{2n}\end{pmatrix}$$
$$\vdots$$
$$\begin{pmatrix}a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}$$

kalles matrisens rader.

Legg merke til at venstresiden i hver likning i et likningssystem er et skalarprodukt. Likningen

$$2x_1+3x_2+4x_3=4$$

sier for eksempel at skalarproduktet mellom vektorene $(2,3,4)$ og $(x_1,x_2,x_3)$ skal være lik 4. Vi definerer derfor

**Matriseprodukt**

Dersom $A$ og $B$ er matriser med dimensjoner $m \times n$ og $n \times p$, er elementet $c_{ij}$ i produktet $C = AB$ gitt ved

$$c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj}$$

altså skalarproduktet mellom rad nummer $i$ i $A$ og kolonne nummer $j$ i $B$.

Dette ser forferdelig ut ved første øyekast, men er ikke så ille når man blir varm i trøyen.

**Eksempel 2.5.**

$$\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}\begin{pmatrix}7\\8\\9\end{pmatrix}=\begin{pmatrix}50\\122\end{pmatrix} \qquad \triangle$$

**Eksempel 2.6.**

$$\begin{pmatrix}0&1\\2&3\end{pmatrix}\begin{pmatrix}4&5&6\\7&8&9\end{pmatrix}=\begin{pmatrix}7&8&9\\29&34&39\end{pmatrix} \qquad \triangle$$

**Eksempel 2.7.**

$$\begin{pmatrix}2&3&4\\3&4&5\\4&5&6\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}2x_1+3x_2+4x_3\\3x_1+4x_2+5x_3\\4x_1+5x_2+6x_3\end{pmatrix} \qquad \triangle$$

Vi ser nå at likningssystemet

$$\left(\begin{array}{ccc|c}2&3&4&4\\3&4&5&5\\4&5&6&6\end{array}\right)$$

kan skrives

$$\begin{pmatrix}2&3&4\\3&4&5\\4&5&6\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}4\\5\\6\end{pmatrix}$$

og dette åpner opp for veldig mange nye og interessante greier. For det første kan vi nå definere

$$A=\begin{pmatrix}2&3&4\\3&4&5\\4&5&6\end{pmatrix}$$

som gjerne kalles systemmatrisen, og

$$\mathbf{x}=\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$$

og

$$\mathbf{b}=\begin{pmatrix}4\\5\\6\end{pmatrix}$$

og skrive hele systemet veldig kompakt:

$$A\mathbf{x}=\mathbf{b}$$

For det andre kan vi nå tenke på venstresiden i likningssystemet som en funksjon

$$T(\mathbf{x})=A\mathbf{x}$$

og dette er veldig viktig.

Vel vel. La oss gå tilbake til eksemplet

$$\left(\begin{array}{ccc|c}2&3&4&4\\3&4&5&5\\4&5&6&6\end{array}\right)\sim\left(\begin{array}{ccc|c}2&3&4&4\\0&1&2&2\end{array}\right)$$

Vi ble enige om at $x_3 = s$, må $x_2 = 2-2s$ og $x_1 = s-1$, eller

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}s-1\\2-2s\\s\end{pmatrix}=\begin{pmatrix}-1\\2\\0\end{pmatrix}+s\begin{pmatrix}1\\-2\\1\end{pmatrix}$$

om du vil. Uttrykket

$$\mathbf{x}(s)=\begin{pmatrix}x_1(s)\\x_2(s)\\x_3(s)\end{pmatrix}=\begin{pmatrix}-1\\2\\0\end{pmatrix}+s\begin{pmatrix}1\\-2\\1\end{pmatrix}$$

kalles gjerne løsningens parametrisering. Du kjenner kanskje igjen parametriseringen for en rett linje fra videregående?

# Lineær uavhengighet

Nå lurer du kanskje på hvorfor likningssystemet

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 7 & 3 \end{array}\right)$$

har entydig løsning, mens likningssystemet

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 6 & 6 \end{array}\right)$$

har uendelig mange. Det beste begrepet for å uttrykkene følelsene sine om dette på en presis måte kalles lineær uavhengighet. Men la oss først se litt på venstresiden av likningssystemene på enda en måte.

Vi husker fra skolen at vi kan gange vektorer med skalarer:

$$a\mathbf{x} = \begin{pmatrix} ax_1 \\ ax_2 \\ \vdots \\ ax_n \end{pmatrix}$$

og legge vektorer sammen:

$$\mathbf{x}+\mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_n+y_n \end{pmatrix}$$

Kombinerer vi disse operasjonene, får vi noe som kalles en lineærkombinasjon:

$$a\mathbf{x}+b\mathbf{y} = a\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + b\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} ax_1+by_1 \\ ax_2+by_2 \\ \vdots \\ ax_n+by_n \end{pmatrix}$$

Skalarene $a$ og $b$ kalles vekter. Hvis vi har $m$ vektorer $\mathbf{x}_k$, definerer vi *det lineære spennet*, eller

$$\mathrm{Sp}\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m\}$$

som alle lineærkombinasjoner av vektorene, altså alle vektorer på formen

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + ... + a_m\mathbf{x}_m.$$

Bruker vi dette, ser vi at vi kan tenke på venstresiden i systemet

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 6 & 6 \end{array}\right)$$

som lineærkombinasjonen av kolonnene i matrisen, med de ukjente som vekter:

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} + x_2\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} + x_3\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

Den store forskjellen mellom systemene

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 7 & 3 \end{array}\right)$$

og

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 4 \\ 3 & 4 & 5 & 5 \\ 4 & 5 & 6 & 6 \end{array}\right)$$

ligger nå i det geometriske forholdet mellom kolonnene i venstresidene. I det andre systemet ligger kolonnevektorene i samme plan, men i det første peker de litt mer i "hver sin retning". Følgende definisjon formaliserer hva vi mener med "hver sin retning".

**Viktig I**

Vi sier at vektorer $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er lineært uavhengige dersom

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

impliserer at

$$c_1 = c_2 = \cdots = c_n = 0.$$

Dersom man ønsker å sjekke om kolonnene i en matrise er lineært uavhengige, må vi altså løse systemet

$$A\mathbf{x} = \mathbf{0}$$

og sjekke om $\mathbf{x} = \mathbf{0}$ er den eneste løsningen.

**Eksempel 2.8.** Vi sjekker om kolonnene i

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 7 \end{pmatrix}$$

er lineært uavhengige. Siden

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 0 \\ 3 & 4 & 5 & 0 \\ 4 & 5 & 7 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

ser vi umiddelbart at $x_1 = x_2 = x_3 = 0$ er eneste løsning, og kolonnene er følgelig lineært uavhengige. △

**Eksempel 2.9.** Hva med

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{pmatrix} \quad ?$$

Siden

$$\left(\begin{array}{ccc|c} 2 & 3 & 4 & 0 \\ 3 & 4 & 5 & 0 \\ 4 & 5 & 6 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} 2 & 3 & 4 & 0 \\ 0 & 1 & 2 & 0 \end{array}\right)$$

har vi uendelig mange løsninger på formen

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = s\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

Det finnes altså et valg av $\mathbf{x} \neq \mathbf{0}$ slik at

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} + x_2\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} + x_3\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Med andre ord er kolonnene lineært avhengige. △

Følgende faktaopplysning er ofte nyttig.

**Viktig II**

Dersom $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er lineært uavhengige og

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{w}$$

er koeffisientvektoren $\mathbf{c}$ entydig bestemt.

Dette er relativt lett å se. Anta at både

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{w}$$

og

$$d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n = \mathbf{w}$$

Dersom vi trekker disse to likningene fra hverandre, får vi

$$(c_1-d_1)\mathbf{v}_1+(c_2-d_2)\mathbf{v}_2+\cdots+(c_n-d_n)\mathbf{v}_n = \mathbf{w}-\mathbf{w} = \mathbf{0}$$

og siden $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er lineært uavhengige må

$$c_1 - d_1 = c_2 - d_2 = \cdots = c_n - d_n = 0$$

**Eksempel 2.10.** Siden

$$\left(\begin{array}{cccc|c} 8 & -7 & 0 & 0 \\ -8 & -7 & 3 & 0 \\ -4 & 5 & -8 & 0 \\ -6 & 6 & -4 & 0 \end{array}\right) \sim \left(\begin{array}{cccc|c} 8 & -7 & 0 & 0 \\ 0 & -14 & 3 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

bør det kanskje ikke komme som noe sjokk at systemet

$$\left(\begin{array}{cccc|c} 8 & -7 & 0 & -3 \\ -8 & -7 & 3 & -7 \\ -4 & 5 & -8 & -3 \\ -6 & 6 & -4 & 0 \end{array}\right)$$

har entydig løsning. (Prøv selv.) Systemet

$$\left(\begin{array}{cccc|c} 8 & -7 & 0 & -3 \\ -8 & -7 & 3 & -7 \\ -4 & 5 & -8 & -3 \\ -6 & 6 & -4 & 1 \end{array}\right)$$

har ingen løsning, for høyresiden kan ikke skrive som en lineærkombinasjon av kolonnene i matrisen på venstresiden. Dersom likningssystemet ikke har noen løsninger, hjelper det ikke at kolonnene er lineært uavhengige! △

Merk til slutt at en lineært uavhengig vektormengde ikke kan inneholde nullvektoren, samt at lineær uavhengighet er ekvivalent med parallellitet dersom vi bare sjekker to vektorer. Konseptet lineær uavhengighet føles nok for de fleste litt rart i begynnelsen, og det er standardprosedyre å innføre dette under matriseregningen. Senere skal vi anvende det på helt andre ting, for eksempel differensiallikninger.

# Determinanter

Dersom det lineære likningssystemet er kvadratisk, finnes det en rask måte å sjekke om kolonnene i systemmatrisen er lineært uavhengige. Det kalles determinanten. For en $2 \times 2$-matrise

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

er den gitt ved

$$\det A = a_{11}a_{22} - a_{21}a_{12}$$

Vi kjenner igjen dette som arealet av parallellogrammet utspent av kolonnevektorene i $A$. Dersom $\det A = 0$ er kolonnene parallelle, og dersom $\det A \neq 0$ er de ikke. Siden parallelle vektorer er lineært uavhengighet (og omvendt så lenge vi bare har to vektorer i kikkerten), har vi egentlig rundet spillet for $2 \times 2$-matriser.

For en $3 \times 3$-matrise

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

er formelen

$$\begin{aligned} \det A = &a_{11}\left(a_{11}a_{22} - a_{21}a_{12}\right) \\ &-a_{12}\left(a_{21}a_{33} - a_{23}a_{31}\right) \\ &+a_{13}\left(a_{21}a_{32} - a_{22}a_{31}\right) \end{aligned}$$

Fulgte du godt med på skolen, kjenner du igjen dette som volumet av parallellepipedet utspent av kolonnene i $A$. Dersom du tar tre kulepenner og fikler litt med dem, vil du nok klare å se geometrisk at de tre kolonnene er lineært uavhengige hvis og bare hvis $\det A \neq 0$.

Dette kan generaliseres til $n \times n$-matriser, men det er ikke helt trivielt. De fleste ingeniører går nok gjennom livet uten å kjenne til denne utledningen, men kanskje jeg får plass til å skrive det ned siden.

# Invers

Dersom det lineære likningssystemet er kvadratisk, og kolonnene i systemmatrisen er lineært uavhengige, kan man beregne noe som kalles invers. La oss ta en ny titt på systemet

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \\ 7 \end{pmatrix}$$

Matrisen

$$\begin{pmatrix} -3 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 1 \end{pmatrix}$$

har noen artige egenskaper. For det første:

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 7 \end{pmatrix} \begin{pmatrix} -3 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Matrisen på høyre side kalles identitetsmatrisen, og du kan sjekke selv at denne oppfører seg som tallet en - når du ganger den med noe, skjer det ingening. For det andre:

$$\begin{pmatrix} -3 & 1 & 1 \\ 1 & 2 & -2 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} -4 \\ 8 \\ -3 \end{pmatrix}$$

Hvis du har god tallhukommelse, kjenner du kanskje igjen høyresiden som løsningen til systemet.

Det som har skjedd her, er at vi startet med et system

$$A\mathbf{x} = \mathbf{b},$$

ganget med den inverse $A^{-1}$ på begge sider, og fikk

$$\mathbf{x} = A^{-1}A\mathbf{x} = A^{-1}\mathbf{b}.$$

Inversmatrisen er altså en slags generalisering av delingsoperasjonen, og notasjonen $A^{-1}$ er ikke tilfeldig. Dette er analogt til å løse likningen

$$ax = b$$

ved å dele ut $a$ på begge sider:

$$x = \frac{1}{a}b = a^{-1}b$$

Hvordan beregner vi $A^{-1}$? Enkelt. Vi må ha

$$AA^{-1} = I$$

der $I$ er identitetsmatrisen. Den som er dreven i matrisemultiplikasjon ser at dette gir opphav til $n$ likningssystemer på formen

$$A\mathbf{x} = \mathbf{e}_i$$

der $\mathbf{e}_i$ er enhetsvektorene i $\mathbb{R}^n$. Nå kan vi utvide notasjonen for likningssystemer enda litt, og skrive $AA^{-1} = I$ som

$$\left[\begin{array}{ccc|ccc} 2 & 3 & 4 & 1 & 0 & 0 \\ 3 & 4 & 5 & 0 & 1 & 0 \\ 4 & 5 & 7 & 0 & 0 & 1 \end{array}\right]$$

og gausseliminere seg frem til

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -3 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right]$$

Det er også slik at dersom $A^{-1}$ eksisterer, er $AA^{-1} = I = A^{-1}A$. Dette er ikke helt innlysende, men trikset er å skjønne dersom man er enig i at nøyaktig de samme gauss-stegene som i eliminasjonen over vil produsere eliminasjonen

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 3 & 4 \\ 0 & 1 & 0 & 3 & 4 & 5 \\ 0 & 0 & 1 & 4 & 5 & 7 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} -3 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & -2 & 0 & 1 & 0 \\ 1 & -2 & 1 & 0 & 0 & 1 \end{array}\right]$$

kan vi skrive om til

Det finnes også formler for inversmatrisen basert på determinant, men ting som er basert på determinant er stort sett ikke så relevante i 2021. Man kan stort sett beregne ting på mer effektive måter.

**Eksempel 2.11.** Inversen til

$$\begin{pmatrix} 1-i & 3 \\ i & 1+2i \end{pmatrix}$$

er

$$\frac{1}{3-2i}\begin{pmatrix} 1+2i & -3 \\ -i & 1-i \end{pmatrix}$$

siden

$$\frac{1}{3-2i}\begin{pmatrix} 1-i & 3 \\ i & 1+2i \end{pmatrix}\begin{pmatrix} 1+2i & -3 \\ -i & 1-i \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

△

# Egenvektorer

Her er en annen matrise:

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 6 \end{pmatrix}$$

La oss gange noen vektorer inn fra høyre i denne. Vi prøver først

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 10 \\ 10 \end{pmatrix}$$

Ikke stort å melde om. Men hva med denne?

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 9 \\ 18 \\ 18 \end{pmatrix} = 9\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

Noen vektorer har den egenskap at når man ganger dem inn i matrisen, kommer det ut en skalarmultippel av den samme vektoren. Dette ser tilforlatelig ut, men er utrolig viktig, og kalles egenvektorer. Skaleringsfaktoren kalles egenverdi. Merk at $A$ må være kvadratisk, for ellers gir det ingen mening å si at matrisen skalerer vektorer.

Vi leter altså etter vektorer slik at

$$A\mathbf{x} = \lambda\mathbf{x}.$$

I eksemplet over er

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

og $\lambda = 9$. Vi sier at $\mathbf{x}$ er en egenvektor med tilhørende egenverdi $\lambda$.

La oss nå lage en systematisk fremgangsmåte for å finne egenvektorer og egenverdier. Likningen

$$A\mathbf{x} = \lambda\mathbf{x}.$$

$$A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$$

og har vi litt trening i enhetsmatrisen, ser vi at vi kan faktorisere dette som

$$(A - \lambda I)\,\mathbf{x} = \mathbf{0}$$

Nå vet du at dersom kolonnene i $A-\lambda I$ er lineært uavhengige, har dette problemet kun en løsning, nemlig nullvektoren. Nullvektoren vil vi ikke skal klassifisere som egenvektor, for hvis den gjorde det, ville alle tall klassifisert som egenverdier, siden

$$A\mathbf{0} = \lambda\mathbf{0}$$

uansett hva $\lambda$ er. Vi må altså kreve at kolonnene i $A - \lambda I$ er lineært uavhengige, og det er de hvis og bare hvis $\det(A - \lambda I) = 0$.

La oss prøve å finne egenverdiene til matrisen $A$ over. Vi beregner

$$\begin{aligned}\det(A-\lambda I) &= \det\begin{pmatrix} 1-\lambda & 2 & 2 \\ 2 & 6-\lambda & 2 \\ 2 & 2 & 6-\lambda \end{pmatrix} \\ &= (1-\lambda)\left((6-\lambda)^2-4\right) \\ &\quad -2\left(2(6-\lambda)-4\right) \\ &\quad +2\left(4-2(6-\lambda)-4\right) \\ &= \lambda^3-13\lambda^2+36\lambda \\ &= \lambda(\lambda-4)(\lambda-9)\end{aligned}$$

slik at $\lambda_1 = 0$, $\lambda_2 = 4$ og $\lambda_3 = 9$ er de tre egenverdiene. Polynomet $p(\lambda) = \lambda(\lambda-4)(\lambda-9)$ kalles matrisens karakteristiske polynom. Husk nå fra forrige kapittel at algebraens fundamentalteorem sier at et polynom

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$$

kan alltid faktoriseres

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = \prod_{i=1}^{n}(\lambda-\lambda_i),$$

der der $\lambda_i \in \mathbb{C}$ er løsninger av likningen

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = 0.$$

Det karakteristiske polynomet er alltid monisk (ser du hvorfor?), og følgelig er det karakteristiske polynomet alltid av orden $n$. På folkemunne sier vi gjerne at en matrise alltid har "n" egenverdier, dersom du teller riktig, altså antall lineære faktorer i det karakteristiske polynomet. Men det er ikke nødvendigvis enkelt å finne egenverdiene bare fordi de finnes. Niels Henrik Abel beviste i 1824 at det finnes ingen generell formel for å løse polynomlikninger med høyere orden enn 5:
`https://en.wikipedia.org/wiki/Abel-Ruffini_theorem`

Vi vet altså at egenverdiene til $A$ er $\lambda_1 = 0$, $\lambda_2 = 4$ og $\lambda_3 = 9$. Egenvektorene er løsninger av likninger

$$(A-\lambda_k I)\,\mathbf{x} = \mathbf{0}$$

for hver av de tre egenverdiene. For eksempel er egenvektoren til $\lambda_2 = 4$ gitt ved alle skalarmultipler av

$$\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

siden

$$\left(\begin{array}{ccc|c} 1-4 & 2 & 2 & 0 \\ 2 & 6-4 & 2 & 0 \\ 2 & 2 & 6-4 & 0 \end{array}\right) \sim \left(\begin{array}{ccc|c} -3 & 2 & 2 & 0 \\ 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

De andre finnes på samme måte. Alle skalarmultipler (unntatt nullvektoren) av en egenvektor er også egenvektorer. Derfor er det vanlig å definere noe som kalles egenrommet til en egenverdi - dette er alle egenverdiens egenvektorer samt nullvektoren.

**Eksempel 2.12.** Matrisen

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

har visst egenverdier allikevel. Det karakteristiske polynomet er

$$\lambda^2 + 1,$$

så egenverdiene er $\pm i$. △

**Eksempel 2.13.** Matrisen

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

har karakteristisk polynom

$$(3-\lambda)((1-\lambda)^2+1) = (3-\lambda)(2-2\lambda+\lambda^2).$$

Den ene egenverdien er åpenbart $\lambda = 3$, mens andregradspolynomet $2-2\lambda+\lambda^2$ har røtter

$$\lambda = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i.$$

Her er det altså en reell egenverdi 3, og to komplekse egenverdier $1+i$ og $1-i$. △

Fra en oppgave i forrige kapittel ser vi at

> Egenverdiene til en reell matrise kommer i komplekskonjugerte par.

**Eksempel 2.14.** Matrisen

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

har karakteristisk likning

$$(2-\lambda)(1-\lambda)^2 = 0,$$

med en enkel egenverdi $\lambda = 2$, og en dobbel egenverdi $\lambda = 1$. △

Det karakteristiske polynomet til en $n \times n$-matrise kan alltid spaltes i $n$ lineære faktorer, men det finnes ikke alltid $n$ lineært uavhengige egenvektorer.

**Eksempel 2.15.** Matrisen

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

har egenverdier

$$\lambda = \pm i.$$

Egenrommet til $-i$ er nullrommet til

$$\begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix}.$$

Vi vet at denne matrisen ikke er inverterbar, og da må radene være skalarmultipler av hverandre (i dette

tilfellet er den nederste $i$ ganger den øverste), så vi kan egentlig bare stryke den nederste, og se at

$$ix_1 - x_2 = 0,$$

slik at en egenvektor til $-i$ blir

$$\begin{pmatrix} 1 \\ i \end{pmatrix}$$

Likeledes blir en egenvektor til $i$

$$\begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Vi dobbeltsjekker

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} = i \begin{pmatrix} i \\ 1 \end{pmatrix}. \qquad \triangle$$

**Eksempel 2.16.** Vi beregner egenrommet til matrisen

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}$$

sin egenverdi $\lambda = 1 - i$. Dette er nullrommet til

$$\begin{pmatrix} 2+i & 0 & 0 \\ 0 & i & -1 \\ 0 & 1 & i \end{pmatrix}.$$

Den øverste raden forteller at $x_1 = 0$. De to nedereste ligner mistenkelig på forrige eksempel, så en egenvektor blir

$$\begin{pmatrix} 0 \\ 1 \\ i \end{pmatrix}.$$

Vi dobbeltsjekker

$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ i \end{pmatrix} = \begin{pmatrix} 0 \\ 1-i \\ 1+i \end{pmatrix} = (1-i) \begin{pmatrix} 0 \\ 1 \\ i \end{pmatrix}. \qquad \triangle$$

**Eksempel 2.17.** Matrisen

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

har dobbel egenverdi $\lambda = 1$. Det tilhørende egenrommet er nullrommet til

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Dette gir at $x_2 = x_3 = 0$, så en egenvektor til $\lambda = 1$ er

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

Her er egenrommet endimensjonalt, mens egenverdien hadde multiplisitet 2. $\triangle$

Egenrommet har dimensjon mindre enn eller lik multiplisiteten til egenverdien. Dersom et egenrom har lavere dimensjon enn multiplisiteten til egenverdien, sier vi at egenverdien er *defekt*.

# Ortogonale matriser

Skalarproduktet mellom vektorer i $\mathbb{R}^n$ er:

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

og vi sier at $\mathbf{x}$ og $\mathbf{y}$ er ortogonale om $\mathbf{x} \cdot \mathbf{y} = 0$. Det går an å definere lengde på mange fornuftige måter, for eksempel

$$\|\mathbf{x}\|_1 = \sum_{k=1}^{n} |x_k|$$

eller

$$\|\mathbf{x}\|_\infty = \max_k |x_k|$$

men den vanligste er nok den pytagoreiske lengden

$$\|\mathbf{x}\|_2 = \sqrt{\mathbf{x} \cdot \mathbf{x}}$$

siden dette er den fysiske lengden til $\mathbf{x}$ for $n = 2$ og $n = 3$. Vi sier at denne lengden er indusert av skalarproduktet, og dersom det ikke er noen subindeks, er det underforstått at det er denne vi mener.

Hvordan kan vi utvide alt dette til vektorer i $\mathbb{C}^n$? I lys av avsnittet over, er det lurt å definere

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} \overline{x_k} y_k = \overline{x_1} y_1 + \overline{x_2} y_2 + \cdots + \overline{x_n} y_n$$

siden indreproduktet av $\mathbf{x}$ med seg selv da blir en sum av de kvadrerte absoluttverdiene til komponentene til $\mathbf{x}$ og dette virker som et fornuftig mål på lengden til $\mathbb{C}^n$. Siden dette er noe vi kan komme til å gjøre ofte, er det funnet opp en spesiell notasjon som tar i bruk matrisemultiplikasjon.

**Adjungert**

La

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

være en $m \times n$-matrise. Den *adjungerte* av $A$ er $n \times m$-matrisen

$$A^* = \begin{pmatrix} \overline{a_{11}} & \overline{a_{21}} & \cdots & \overline{a_{m1}} \\ \overline{a_{12}} & \overline{a_{22}} & \cdots & \overline{a_{m2}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \cdots & \overline{a_{mn}} \end{pmatrix}$$

der radene og kolonnene i $A$ er byttet om. Dersom $A$ er reell, sier vi istedet transponert og skriver $A^T$.

Siden $A^* = \overline{A}$ dersom $A$ er en $1 \times 1$-matrise, ser vi at vi nå har to notasjoner for komplekskonjugert, $\overline{z}$ og $z^*$. Noen fagfelt foretrekker den ene, og noen foretrekker den andre, så det er greit å vite om begge. Vi kan nå skrive skalarproduktet som:

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^* \mathbf{y}$$

Vi sier at en vektormengde er innbyrdes ortogonale dersom

$$\mathbf{x}^* \mathbf{y} = 0$$

for alle $\mathbf{x} \neq \mathbf{y}$ i mengden. Dersom alle vektorene i tillegg har lengde 1, blir det enda penere.

**Ortogonale matriser**

Vi sier at en matrise $A$ er ortonormal dersom

$$A^*A = I$$

Det er ikke så vanskelig å se at en ortogonal matrise har ortogonale kolonner. Siden ortogonale vektorer åpenbart må være lineært uavhengige, ser vi også at $A^* = A^{-1}$ dersom $A$ er kvadratisk.

# Diagonalisering

La $A$ være en $n \times n$-matrise med $m$ lineært uavhengige egenvektorer $\mathbf{v}_1$, $\mathbf{v}_2 \ldots \mathbf{v}_m$ og tilhørende egenverdier $\lambda_1$, $\lambda_2$,$\cdots$, $\lambda_m$, . For hver egenvektor gjelder

$$A\mathbf{v}_k = \lambda_k \mathbf{v}_k.$$

Disse $m$ ligningene kan like gjerne organiseres i en matriseligning

$$AV = VD,$$

der

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_m \end{pmatrix}$$

og

$$V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_m \end{pmatrix}.$$

Siden $\mathbf{v}_1$, $\mathbf{v}_2 \ldots \mathbf{v}_m$ er lineært uavhengige, er matrisen $V$ invertibel dersom $m = n$. Vi ganger med $V^{-1}$ fra venstre og får

$$V^{-1}AV = D.$$

Denne operasjonen kalles å *diagonalisere* $A$. Man kan også gå motsatt vei. Dersom

$$V^{-1}AV = D$$

for en inverterbar $n \times n$-matrise $V$, kan vi gange fra venstre med $V$, og få

$$AV = VD$$

Vi ser av denne likningen at kolonnene til $V$ utgjør $n$ lineært uavhengige egenvektorer for $A$. Vi sier at en $n \times n$-matrise med $n$ lineært uavhengige egenvektorer er diagonaliserbar.

**Diagonaliserbarhet**

Vi kan skrive

$$V^{-1}AV = D$$

hvis og bare hvis $A$ har $n$ lineært uavhengige egenvektorer.

**Eksempel 2.18.** Vi diagonaliserer matrisen

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

Egenvektorene er

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

med respektive egenverdier 3 og 1. Vi definerer

$$V = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

og beregner

$$V^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Vi dobbeltsjekker ved å beregne produktet

$$\begin{aligned} V^{-1}AV &= \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} = D \end{aligned}$$

△

**Eksempel 2.19.** Matrisen

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

har bare to lineært uavhengige egenvektorer, og er følgelig ikke diagonaliserbar. △

**Eksempel 2.20.** Vi kan faktorisere matrisen

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

som

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

△

I eksemplet over spiller ikke plasseringen av $V$ og $V^{-1}$ noen rolle. Mer om dette under.

La $\mathbf{v}$ være en kolonnevektor i $\mathbb{C}^n$ og $A$ en $n \times n$-matrise. En kvadratisk form er et uttrykk på formen

$$\mathbf{v}^*A\mathbf{v}$$

Et slikt uttrykk er spesielt interessant dersom $\mathbf{v}$ er en egenvektor med lengde 1 og egenverdi $\lambda$:

$$\mathbf{v}^*A\mathbf{v} = \mathbf{v}^*\lambda\mathbf{v} = \lambda\mathbf{v}^*\mathbf{v} = \lambda.$$

**Eksempel 2.21.** Vi lar

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

og

$$\mathbf{v} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

slik at

$$\mathbf{v}^*A\mathbf{v} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = 3.$$

△

**Symmetrisk matrise**

En kompleks matrise sies å være *symmetrisk* dersom $A = A^*$.

Dersom $A$ er reell er $A^* = A^T$, slik at kravet for en symmetrisk matrise er at $A = A^T$. I litteraturen er det vanlig å reservere begrepet *symmetrisk* for reelle matriser der $A = A^T$, mens komplekse matriser kalles *hermittiske* hvis $A = A^*$. Jeg har aldri sett noen transponere en kompleks matrise uten å kompleks-konjugere den, så man kunne i bunn og grunn bare sagt symmetrisk.

**Eksempel 2.22.** Matrisen

$$\begin{pmatrix} 1 & 1+i & -i \\ 1-i & 0 & 2-i \\ i & 2+i & 2 \end{pmatrix}$$

er symmetrisk. Merk at en symmetrisk matrise må ha reelle diagonalelementer. △

Dersom $A$ har $n$ ortonormale egenvektorer, sier vi at $A$ er ortogonalt diagonaliserbar. Dersom vi setter egenvektorene inn som kolonner i en $n \times n$-matrise $V$, ser vi at

$$V^*AV = V^*VD = D$$

siden $V^*V = I$. Motsatt ser vi at dersom

$$V^*AV = D,$$

for en ortonormal matrise $V$, utgjør $V$ sine kolonner $n$ ortonormale egenvektorer for $D$.

**Ortogonal diagonaliserbarhet**

Vi kan skrive

$$V^*AV = D$$

hvis og bare hvis $A$ har $n$ ortogonale egenvektorer.

Følgende teorem er ikke så vanskelig å bevise, bare skriv ut matrisemultiplikasjonen og se nøye på elementene.

**Om den kvadratiske formen**

Dersom $A = A^*$ er $\mathbf{x}^*A\mathbf{x}$ reell.

La $\mathbf{v}$ være en normalisert egenvektor med egenverdi $\lambda$. Vi vet at

$$\mathbf{v}^*A\mathbf{v} = \lambda,$$

og venstresiden er reell, så da må også $\lambda$ være det.

**Om symmetriske matriser**

En symmetrisk matrise har reelle egenverdier.

La $\mathbf{v}_1$ og $\mathbf{v}_2$ være to egenvektorer med egenverdier $\lambda_1$ og $\lambda_2$. Vi beregner (husk at $\lambda_1$ og $\lambda_2$ er reelle)

$$\begin{aligned}\lambda_1\mathbf{v}_1^*\mathbf{v}_2 &= (\lambda_1\mathbf{v}_1)^*\mathbf{v}_2 = (A\mathbf{v}_1)^*\mathbf{v}_2 \\ &= \mathbf{v}_1^*A^*\mathbf{v}_2 = \mathbf{v}_1^*(A\mathbf{v}_2) = \mathbf{v}_1^*(\lambda_2\mathbf{v}_2) = \lambda_2\mathbf{v}_1^*\mathbf{v}_2\end{aligned}$$

Vi vet altså at

$$0 = \lambda_1\mathbf{v}_1^*\mathbf{v}_2 - \lambda_2\mathbf{v}_1^*\mathbf{v}_2 = (\lambda_1 - \lambda_2)\mathbf{v}_1^*\mathbf{v}_2,$$

og hvis vi bruker at $\lambda_1$ og $\lambda_2$ er forskjellige, må vi ha $\mathbf{v}_1^*\mathbf{v}_2 = 0$.

**Om ortogonale egenvektorer**

Egenvektorene til to distinkte egenverdier er ortogonale for symmetriske matriser.

Til slutt kommer den vanlige historien om et par teoremer som er nyttige, men for vanskelige å bevise for oss.

**Om normale matriser**

En symmetrisk matrise er ortogonalt diagonaliserbar.

**Eksempel 2.23.** Matrisen

$$\begin{pmatrix} 1 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 6 \end{pmatrix}$$

har egenverdier 4, 9 og 0. Egenvektorerer er henholdsvis

$$\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} -4 \\ 1 \\ 1 \end{pmatrix}.$$

Merk at alle vektorer er innbyrdes ortogonale, og matrisen er følgelig ortogonalt diagonaliserbar, med

$$V = \begin{pmatrix} 0 & \frac{1}{3} & -\frac{4}{3\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{2}{3} & \frac{1}{3\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{2}{3} & \frac{1}{3\sqrt{2}}. \end{pmatrix}$$

og

$$D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Dette eksemplet viser at en matrise kan være diagonaliserbar uten å være inverterbar. △

**Normale matriser**

En matrise er normal dersom $A * A = AA^*$.

**Om normale matriser**

En matrise er ortogonalt diagonaliserbar hvis og bare hvis den er normal.

**Eksempel 2.24.** En projeksjonsmatrise $P$ er definert ved likningen

$$P = P^2$$

Denne likningen sier at det ikke skjer noe nytt om man benytter projeksjonen for andre gang. La oss si at $P$ har en egenverdi $\lambda$, med egenvektor $\mathbf{x}$. I så fall må

$$\lambda\mathbf{x} = P\mathbf{x} = P^2\mathbf{x} = P(P\mathbf{x}) = P(\lambda\mathbf{x}) = \lambda P\mathbf{x} = \lambda^2\mathbf{x}.$$

Dersom $\mathbf{x} \neq \mathbf{0}$, må

$$\lambda = \lambda^2$$

eller

$$\lambda^2 - \lambda = \lambda(\lambda - 1) = 0.$$

Egenverdiene til en projeksjonsmatrise kan altså kun være 0 eller 1. $\triangle$

# Kapittel 3

# Funksjoner fra $\mathbb{N}$ til $\mathbb{R}$ og $\mathbb{C}$

Grunnene til å venne seg til følger og rekker er så mange at det nesten ikke er godt å vite hvor man skal begynne om man skal forklare det. Moderne signalbehandling er utenkelig uten muligheten for å skrive slike sprø ting som

$$x = 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nx \qquad x \in (-\pi, \pi)$$

Alternative kilder:

- Adams kap. 9
- Kreyszig kap. 15
- Lindstrøm I kap. 4 og 12

## Motiverende eksempel

Andregradslikningen

$$ax^2 + bx + c = 0$$

kan vi løse med formelen

$$x = \frac{b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Men i mange anvendelser dukker det opp likninger som ikke kan løses analytisk. Når man løser Schrödingers likning for en endelig kvantebrønn, må man for eksempel løse likningen

$$\tan\sqrt{x} = \frac{2\sqrt{x(1-x)}}{2x-1},$$

noe som ikke er praktisk gjennomførbart med penn og papir. Selv likninger som på papiret har en løsningsformel, slik som

$$x^4 + x^3 + x^2 + x + 1 = 0$$

kan være ganske vriene å finne ut av uten noen hjelpemidler.

Vi skal studere numeriske metoder for å løse slike likninger siden, men la oss begynne med et enkelt eksempel, nemlig likningen

$$x = \cos x.$$

Hvis du får til å løse denne med penn og papir, er det en sensasjon, for det går nemlig ikke an. Men løsningen til likningen eksisterer, selv om det ikke går an å regne den ut med noen enkel formel. Løsningen kalles Dottietallet, og er oppkalt etter en professor i fransk:
`https://www.maa.org/sites/default/files/Kaplan2007-131105.pdf`)
Professor Dottie oppdaget, som utallige skolebarn før henne, at dersom du trykker på cosinusknappen på kalkulatoren din igjen og igjen, vil du etter mange nok trykk alltid ende på tallet

$$r \approx 0.739085...$$

Det kalkulatoren gjør når du trykker på cosinusknappen mange ganger, er å utføre rekursjonen

$$x_{n+1} = \cos x_n$$

der den første verdien $x_1$ er det tallet som tilfeldigvis var lagret i kalkulatorens minne da man satte i gang og trykke på cosinusknappen.

Professor Dottie gjenoppdaget noe som kalles fikspunktiterasjonen. Hun oppdaget at rekursjonen sakte men sikkert konvergerer mot den korrekte løsningen til likningen $x = \cos x$. Dette er litt snodig, men virker av og til. Senere skal vi analysere når det virker, og når det ikke virker.

## Følger

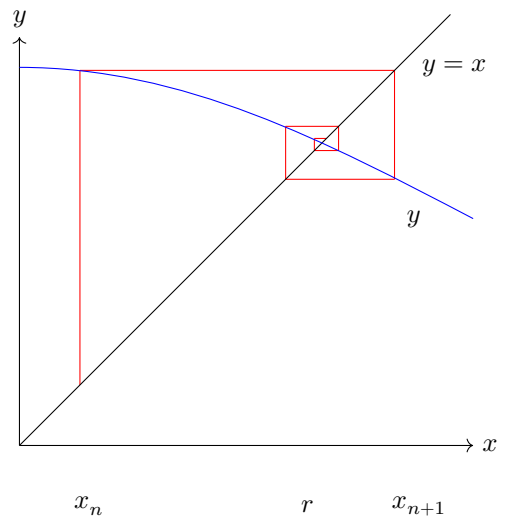
Rekursjonen

$$x_{n+1} = \cos x_n$$

produserer følgende tabell dersom vi setter $x_0 = \frac{3}{4}$:

| | |
|---|---|
| $x_1$ | 0.2000000000000000 |
| $x_2$ | 0.980066577841242 |
| $x_3$ | 0.556967252809642 |
| $x_4$ | 0.848862165658271 |
| $x_5$ | 0.660837551116615 |
| ⋮ | ⋮ |
| $x_{11}$ | 0.731977425258191 |
| $x_{21}$ | 0.738948732099227 |
| $x_{31}$ | 0.739082509617631 |
| $x_{79}$ | 0.739085133215145 |

Figuren under illustrerer omtrent hva som skjer.




En følge er kort og godt en haug med tall, lagt på rekke etter hverandre:

$$x_1, x_2, x_3, \cdots$$

Siden leddene i følgen er indeksert med de naturlige tallene kan vi like gjerne bruke funksjonsbegrepet, og definere:

**En følge**

En reell følge er en funksjon

$$x : \mathbb{N} \to \mathbb{R},$$

og en kompleks følge er en funksjon

$$z : \mathbb{N} \to \mathbb{C}.$$

Verdiene $x_n$ eller $z_n = a_n + ib_n$ kalles *leddene* i følgen. Du kan tenke på en følge som en en uendelig lang tabell:

| $n$ | 1 | 2 | 3 | ... |
|---|---|---|---|---|
| $x_n$ | $x_1$ | $x_2$ | $x_3$ | ... |

Vi skal hoppe litt mellom reelle og komplekse følger, men jeg skal være nøye på å bruke $x_n$ for reelle følger og $z_n = a_n + ib_n$ for komplekse følger. Hvis jeg fremsetter en påstand som er sann for begge typer, skriver jeg $z_n$.

**Eksempel 3.1.** Den enkleste følgen er kanskje de naturlige tallene:

$$\mathbb{N} = \{1, 2, 3, ...\}$$

Her er funksjonen $x : \mathbb{N} \to \mathbb{N}$ gitt ved

$$x_n = n. \qquad \triangle$$

**Eksempel 3.2.** En følge vi skal bli godt kjent med, er følgen gitt ved

$$x_n = \frac{1}{n}. \qquad \triangle$$

**Eksempel 3.3.** En annen følge vi skal bli godt kjent med, er følgen gitt ved

$$x_n = \frac{1}{n!} = \frac{1}{n \cdot (n-1) \cdots 3 \cdot 2}. \qquad \triangle$$

**Eksempel 3.4.** En klassiker i kompleks funksjonsteori er

$$z_n = \cos n + i \sin n. \qquad \triangle$$

Dersom leddene i følgen er definert som en funksjon av foregående ledd, slik som i Dotties eksperiment

$$x_{n+1} = \cos x_n$$

kalles en *rekursjon*.

**Eksempel 3.5.** Fibonaccitallene

$$\{1, 1, 2, 3, 5, 8, 13...\}$$

er gitt ved rekursjonen

$$x_n = x_{n-1} + x_{n-2},$$

og startverdiene $x_1 = x_2 = 1$. $\triangle$

Noen ganger har vi ikke en formel i det hele tatt.

**Eksempel 3.6.** Følgen av primtall

$$\{2, 3, 5, 7, 11, 13, 17, ...\}$$

har ingen kjent formel. Det er ikke så vanskelig å se at det finnes uendelig mange primtall. La oss anta at det kun finnes et endelig antall primtall, og kall dem $x_1$, $x_2$, $\cdots$, $x_n$. Tallet

$$x = x_1 \cdot x_2 \cdots x_n + 1$$

er helt klart ikke delelig med noen av primtallene i den endelige listen. Hvis du nå tror på at alle tall er delelig med et eller annet primtall (i tallteoribøker beviser man slikt, men mange mennesker tar det for gitt) må $x$ være delelig med et nytt primtall som ikke står i listen. Følgelig kan en endelig liste aldri inneholde alle primtall, og det må derfor være uendelig mange av dem.

Det finnes de som leter etter en formel for primtallene:
`https://en.wikipedia.org/wiki/Formula_for_primes` $\triangle$

Siden det er uendelig mange primtall, og de blir større og større utover i følgen, må de vokse over alle grenser når $n \to \infty$. Leddene i følgen $x_n = 1/n$ går derimot mot null. Vi skal nå definere hva vi mener med at en følge "går mot" noe. Husk at $|z| = a^2 + b^2$ og at dette blir vanlig absoluttverdi når $z$ er reell.

**Konvergent følge**

En følge $z$ sies å *konvergere* mot $L$ dersom det for en hver $\epsilon > 0$ finnes en $N$ slik at

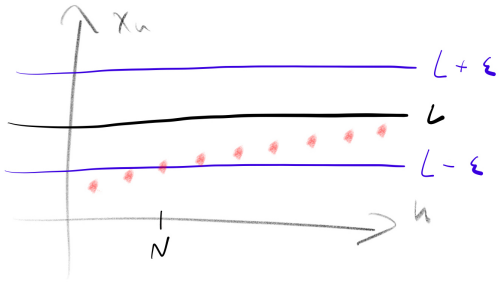
$$n \geq N \implies |z_n - L| < \epsilon,$$

og vi skriver i så fall

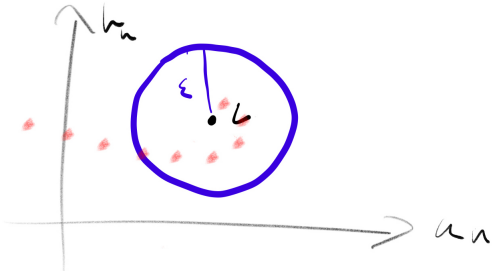
$$\lim_{n \to \infty} z_n = L.$$

En følge som konvergerer, sies å være *konvergent*. En følge som ikke er konvergent, kalles divergent.

Det er underforstått i definisjonen at vi først og fremst er interessert i små $\epsilon$. Uansett hvor liten $\epsilon$ man velger, er det bare å gå langt nok ut i følgen, og så vil alle etterfølgende ledd ligge og vake i en avstand fra $L$ som aldri blir større enn $\epsilon$. Her er en figur for reelle følger:




For komplekse følger blir figuren mer noe slikt:



**Eksempel 3.7.** Følgen

$$x_n = 1/n$$

går mot null. Hvordan ser man det? Hvis du velger $\epsilon = 0.1$, kan jeg velge $N \geq 11$, slik at $x_n < 0.1$. Hvis du velger $\epsilon = 0.05$, velger jeg $N \geq 21$, slik at $x_n < 0.05$. Poenget er nå at uansett hvor liten $\epsilon$ som velges, kan vi alltid finne $N$ slik at $x_n < \epsilon$ når $n > N$, og derfor kan vi si at

$$\lim_{n\to\infty} \frac{1}{n} = 0. \qquad \triangle$$

En følge kan ikke konvergere mot to forskjellige grenseverdier. Anta vi har to grenseverdier $L_1 \neq L_2$. Velg $\epsilon < |L_1 - L_2|$, og $N_1$ slik at

$$|z_n - L_1| < \epsilon/2$$

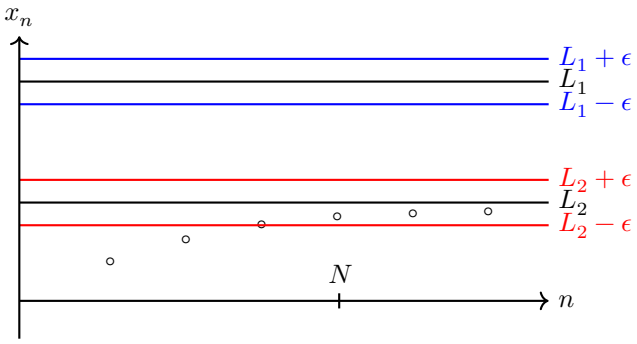
dersom $n > N_1$. Men vi kan også velge $N_2$ slik at

$$|z_n - L_2| < \epsilon/2,$$

og tar vi den største av $N_1$ og $N_2$, er begge ulikhetene oppfylt. Men nå er

$$\begin{aligned} |L_1 - L_2| = |L_1 - z_n - L_2 + z_n| \leq \\ |L_1 - z_n| + |L_2 - z_n| < \\ \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

Dette er en selvmotsigelse, for her står det at uansett hvor liten $\epsilon$ er, må avstanden mellom $L_1$ og $L_2$ være mindre. Siden $\epsilon$ er vilkårlig, kan det ikke stemme at $L_1 \neq L_2$, og vi må ha $L_1 = L_2$. En titt på figuren under kan være til hjelp.




La oss ofre en rød boks på dette resultatet:

**Konvergent følge**

En grenseverdi, dersom den eksisterer, er entydig.

Vi kan vise flere ting. For eksempel er

$$\lim_{n\to\infty} z_n + w_n = L_1 + L_2$$

dersom

$$\lim_{n\to\infty} z_n = L_1 \quad \text{og} \quad \lim_{n\to\infty} w_n = L_2$$

Velg $\epsilon$. Vi må vise at det går an å velge $N$ slik at $n > N$ impliserer

$$|z_n + w_n - L_1 - L_2| < \epsilon$$

Merk først at

$$\begin{aligned} |z_n + w_n - L_1 - L_2| = \\ |z_n - L_1 + w_n - L_2| \leq \\ |z_n - L_1| + |w_n - L_2| \end{aligned}$$

Siden $z$ konvergerer mot $L_1$, og $w$ mot $L_2$, kan vi velge $N$ slik at $n > N$ impliserer

$$|z_n - L_1| < \epsilon/2$$

og

$$|w_n - L_2| < \epsilon/2.$$

Men i så fall impliserer $n < N$ at

$$|z_n - L_1 + w_n - L_2| < \epsilon/2 + \epsilon/2 = \epsilon,$$

som var det vi skulle vise. Det går an å noen flere slike regler, men det er litt mer regning. Prøv selv om du er interessert, og slå opp i Rudin eller kom og spør om du står fast.

**Regneregler**

Anta vi har to følger $z$ og $w$, med

$$\lim_{n\to\infty} z_n = L_1 \quad \text{og} \quad \lim_{n\to\infty} w_n = L_2$$

og la $c$ være et tall. Da gjelder at

$$\begin{aligned}
\lim_{n\to\infty} z_n + w_n &= L_1 + L_2 \\
\lim_{n\to\infty} cz_n &= cL_1 \\
\lim_{n\to\infty} z_n w_n &= L_1 L_2 \\
\lim_{n\to\infty} z_n / w_n &= L_1 / L_2 \quad (L_2 \neq 0)
\end{aligned}$$

**Begrenset følge**

En følge $z$ sies å være *begrenset* dersom det finnes en konstant $C$ slik at

$$|z_n| \leq C$$

for alle $n$.

For reelle følger kan vi selvsagt løse opp absolutt-verditegnet snakke om følger som er begrenset enten ovenfra eller nedenfra.

**Eksempel 3.8.** En annen kjent rekursjon er

$$\begin{aligned}
z_1 &= 0 \\
z_n &= f_c(z_{n-1})
\end{aligned}$$

der

$$f_c(z) = z^2 + c$$

og $c$ er et komplekst tall. For noen verdier av $c$ vil følgen være begrenset, og for andre ikke. Mengden av alle $c$ slik at følgen er begrenset, kalles Mandelbrotmengden, og er en fraktal mengde. Dette var voldsomt hipt og skulle visst revolusjonere verden på 1980-tallet:
`https://en.wikipedia.org/wiki/Mandelbrot_set` △

# Cauchyølger

Her kommer litt teori.

**Monoton følge**

En reell følge sies å være monotont stigende dersom

$$x_{n+1} \geq x_n$$

og monotont synkende dersom

$$x_{n+1} \leq x_n$$

for alle $n$.

Nå skal vi bevise et lite teorem, for det er så fin illustrasjon av minste-øvre-skranke-egenskapen. Anta at en reell følge er monotont stigende, og at verdimengden til $x$ er begrenset ovenfra. Verdimengden har en minste øvre skranke $L$. Velg $\epsilon > 0$. For en eller annen $N$ er $x_n > L - \epsilon$, for ellers er ikke $L$ minste øvre skranke til verdimengden til $x$. Dersom $n > N$, er også

$$x_n > L - \epsilon$$

siden $x$ er monotont stigende. Siden $x_n > L - \epsilon$ for alle $n > N$, konvergerer følgen til $L$.

**Monoton + begrenset**

En monoton og begrenset reell følge er konvergent.

En kompleks følge kan ikke være monoton, for de komplekse tallene er ikke en ordnet kropp. Vi kan forsåvidt definere monotonisitet med hensyn på kompleks absoluttverdi, men en monoton og begrenset kompleks følge vil ikke nødvendigvis være konvergent.

Her kommer det noe litt mer teoretiske greier som er litt praktisk å ha tatt unna.

**Cauchyfølge**

Vi sier at en følge er en cauchyfølge dersom det for enhver $\epsilon > 0$ finnes $N$ slik at $m, n \geq N$ impliserer

$$|z_n - z_m| < \epsilon.$$

Vi skal nå se at

**Cauchys konvergenskriterium**

En følge er konvergent hvis og bare hvis den er en cauchyfølge.

Den ene veien er lett. Dersom $z$ er konvergent med grenseverdi $L$, og $\epsilon$ er gitt, kan vi velge $N$ med $n, m \geq N$ slik at

$$|z_n - L| \leq \epsilon/2$$

og

$$|z_m - L| \leq \epsilon/2$$

Følgelig er

$$\begin{aligned}
|z_n - x_m| =& |z_n - L - (z_m - L)| \\
\leq& |z_n - L| + |z_m - L| \leq \epsilon
\end{aligned}$$

Den andre veien er litt mer jobb. La $I$ være en delmengde av $\mathbb{N}$, ordnet i stigende rekkefølge. Dersom $z$ er en følge, sier vi at mengden av alle $z_n$ for $n \in I$ er en delfølge av $z$. Du kan tenke på dette som restriksjonen av $z$ til $I$.

Alle reelle følger har en monoton delfølge, av følgende grunn. La oss betrakte alle $x_k$ med den egenskap at

$$x_k \geq x_n$$

for alle $n > k$. Enten finnes det undelig mange slike, eller så finnes det endelig mange. I det første tilfellet danner de åpenbart en monotont synkende delfølge, og i det siste tilfellet går det an å lage en monotont stigende delfølge etter sistemann, for ellers må det være flere av dem.

Dette betyr igjen at den begrenset følge må ha en konvergent delfølge. En cauchyfølge må være begrenset, og la oss si at den har en delfølge som konvergerer til $p$. Dersom vi velger $\epsilon > 0$, er det bare å gå så langt ut følgen at både avstanden mellom $p$ og leddene i den konvergente delfølgen og avstanden mellom leddene i hele følgen er mindre enn $\epsilon/2$, og vips er alle ledd etter det nærmere $p$ enn $\epsilon$, slik at en reell følge er konvergent hvis og bare hvis den er en cauchyfølge.

Siden leddene i en kompleks rekke er på formen

$$z_n = x_n + iy_n$$

er både

$$|x_n - x_m| \leq |z_n - z_m|$$

og

$$|y_n - y_m| \leq |z_n - z_m|.$$

Dersom $z$ er en cauchyfølge er altså både $x$ og $y$ det, og dersom de konvergerer til henholdsvis $L_1$ og $L_2$, må $z$ konvergere til $L_1 + iL_2$. Derfor:

**Cauchys konvergenskriterium**

En følge er konvergent hvis og bare hvis den er en cauchyfølge.

Man kan bruke ekvivalensklasser av rasjonale cauchyfølger istedet for dedekindske snitt til å konstruere de reelle tallene. En rasjonal cauchyfølge er en cauchyfølge $x : \mathbb{N} \to \mathbb{Q}$. Vi sier at to rasjonale cauchyfølger $x$ og $y$ er ekvivalente dersom for en hver $\epsilon > 0$ finnes $N$ slik at

$$n > N \implies |x(n) - y(n)| < \epsilon$$

Man identifiserer for eksempel $\sqrt{2}$ med ekvivalensklassen av alle rasjonale cauchufølger som konvergerer mot $\sqrt{2}$, og så tar man det derfra. Man sjekker først alle aksiomene for kropp, og så beviser man minste-øvre-skranke-egenskapen ved intervallhalveringsmetoden.

**Eksempel 3.9.** Den rekursive følgen gitt ved $x_0 = 1$ og

$$x_{n+1} = \frac{x_n^2 + 2}{2x_n}$$

er rasjonal, altså at $x_n \in \mathbb{Q}$ for alle $n$. Den konvergerer mot $\sqrt{2}$. Vi skal se mer til denne i neste kapittel. △

# Rekker

Ingen lærebok i matematikk begynner et kapittel om rekker uten Xenos paradoks, som er rundt 2500 år gammelt. Anta at en langdistanseløper skal tilbake legge en distanse. Han tilbakelegger første halvdel på tiden $T$. Den neste fjerdedelen tibakelegger han på tiden $T/2$. Den neste åttendedelen tilbakelegger han på tiden $T/4$. Og slik fortsetter det.

Xeno trodde han hadde funnet et paradoks her, for han trodde at

$$T + \frac{T}{2} + \frac{T}{4} + \frac{T}{8} + \cdots = \infty$$

siden venstresiden er en sum av uendelig mange tall. Dette er ikke riktig. Det er åpenbart at løperen tilbakelegger distansen på tiden $2T$, og dersom man skal tillegge en den uendelige summen over noen verdi, er det naturlig å sette

$$T + \frac{T}{2} + \frac{T}{4} + \frac{T}{8} + \cdots = 2T.$$

**Rekke**

En rekke er summen av leddene i en følge:

$$\sum_{n=1}^{\infty} z_n.$$

Uttrykket

$$S_N = \sum_{n=1}^{N} z_n$$

kalles den $N$-te partialsummen. Vi sier at en rekke konvergerer dersom partialsummene danner en konvergent følge.

Dersom vi trenger å trykke inn en liten sum midt i en setning og indekseringen er innlysende, skriver vi bare $\sum x$. En reell følge er konvergent hvis og bare hvis den er en cauchyfølge. Brukt på partialsummer, blir dette

**Cauchys konvergenskriterium for rekker**

En rekke er konvergent hvis og bare hvis det for hver $\epsilon > 0$ finnes $N$ slik at $m, n \geq N$ impliserer

$$\left| \sum_{k=m}^{n} z_k \right| \leq \epsilon$$

**Eksempel 3.10.** Dersom Xeno hadde lest følgende eksempel, hadde han skjønt at han var på villspor. På skolen har du lært at den geometriske rekken

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n.$$

er konvergent så lenge $|x| < 1$. Det er ikke så vanskelig å se at dette er riktig, men vi kan gå et knepp lenger å gjøre alt for komplekse $z$:

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n.$$

Partialsummene er

$$S_N = \sum_{n=0}^{n} z^n = 1 + z + z^2 + \cdots + z^N,$$

og hvis vi tar

$$\begin{aligned}(1-z)S_N &= (1-z)\sum_{n=0}^{N} z^n \\ &= 1 + z + z^2 + \cdots + z^N \\ &\quad - (z + z^2 + z^3 + \cdots + z^{N+1}) = 1 - z^{N+1}\end{aligned}$$

får vi

$$S_N = \sum_{n=0}^{N} z^n = \frac{1 - z^{N+1}}{1-z}.$$

Dersom $|z| < 1$ og vi lar $N \to \infty$, får vi

$$\frac{1}{1-z} = \sum_{n=0}^{\infty} z^n.$$

og dersom $|z| > 1$ får vi

$$\infty = \sum_{n=0}^{\infty} z^n.$$

Dersom $|z| = 1$ får vi en eller annen form for divergens. Dersom $z = 1$ vokser partialsummene over alle grenser, mens for andre verdier vil de reise rundt på enhetssirkelen på ymse vis. △

**Eksempel 3.11.** Rekken

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{2} + \frac{1}{6} + \frac{1}{12} + \cdots$$

er veldig lett å analysere. Merk at

$$\frac{1}{n(n+1)} = \frac{1}{n} - \frac{1}{n+1},$$

slik at

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n(n+1)} &= \sum_{n=1}^{\infty} \frac{1}{n} - \frac{1}{n+1} \\ &= 1 - \frac{1}{2} + \frac{1}{2} - \frac{1}{3} + \frac{1}{3} + \cdots = 1.\end{aligned}$$

Dette kan minne litt om en teleskopfiskestang, og kalles derfor en *teleskoperende rekke*. △

Nå skal vi se på litt forskjellige tips og triks for å avgjøre om rekker konvergerer eller ikke. La $\sum z$ være en konvergent rekke. Dersom vi husker hva partialsummer er, kan vi utføre følgende beregning:

$$\begin{aligned}\lim_{n\to\infty} z_n &= \lim_{n\to\infty} (S_n - S_{n-1}) \\ &= \lim_{n\to\infty} S_n - \lim_{n\to\infty} S_{n-1} = 0\end{aligned}$$

som gir at

**Viktig!**

Dersom

$$\sum_{n=1}^{\infty} z_n$$

er konvergent, må

$$\lim_{n\to\infty} z_n = 0.$$

Den motsatte implikasjonen er ikke sann, for rekken

$$\sum_{n=1}^{\infty} \frac{1}{n}$$

er divergent.

**Eksempel 3.12.** La oss ta en titt på rekken

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \cdots$$

Vi kan lett vise at denne rekken divergerer ved å bruke et lite triks. Siden

$$\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

og

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$$

og så videre, må vi ha

$$\sum_{n=1}^{l} \frac{1}{n} \geq 1 + \sum_{n=1}^{k+1} \frac{1}{2} = 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots$$

der $l = 2 + \sum_{m=1}^{k} 2^k$ og $k > 0$. Siden partialsummene til $\sum \frac{1}{2}$ danner en ubegrenset følge, må også partialsummene til $\sum \frac{1}{n}$ gjøre det, og følgelig er rekken divergent. △

Hvis det kilte litt i magen når du leste eksemplet over, kan du ta en titt her: `https://en.wikipedia.org/wiki/Riemann_zeta_function`.

**Eksempel 3.13.** Vi kan bruke det samme trikset til å vise at rekken

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \frac{1}{49} + \frac{1}{64} + \cdots$$

er konvergent. Siden

$$\frac{1}{4} + \frac{1}{9} < \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

og

$$\frac{1}{16} + \frac{1}{25} + \frac{1}{36} + \frac{1}{49} < \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{1}{4}$$

og så videre, må vi ha

$$\sum_{n=1}^{l} \frac{1}{n^2} \leq 1 + \sum_{n=1}^{k} \frac{1}{2^n}$$

der $l = 1 + \sum_{m=1}^{k} 2^k$ og $k > 0$. Siden partialsummene til $\sum \frac{1}{2^k}$ danner en begrenset følge, må også partialsummene til $\sum \frac{1}{n^2}$ gjøre det, og følgelig er rekken konvergent. Det går faktisk an å regne ut at

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6},$$

men det må vi vente litt med. △

I de to foregående eksemplene sammenliknet vi partialsummer. La oss skrive opp et teorem bare for ordens skyld.

**Sammenlikningstesten**

La $x$ og $y$ være positive reelle følger, med $x_n \leq y_n$ for alle $n$.

- Dersom $\sum x$ divergerer, divergerer $\sum y$.
- Dersom $\sum y$ konvergerer, konvergerer $\sum x$.

Hvis du er av den typen som liker å bevise ting, er denne ikke så vanskelig. Bare husk definisjonen på hva det vil si at en rekke er konvergent, sammenlikne partialsummene til $x$ og $y$, og husk at en monoton og begrenset følge er konvergent.

**Eksempel 3.14.** Vi kan bruke sammenlikningstesten til å bygge videre på eksemplene over. Rekken

$$\sum_{n=1}^{\infty} \frac{1}{n^p}$$

konvergerer dersom $p > 1$ og divergerer dersom $p \leq 1$. Dette kalles en $p$-rekke. Vi har allerede sjekket at denne er konvergent for $p = 2$ og divergent for $p = 1$, så sammenlikningstesten gir konvergens for $p \geq 2$ og divergens for $p \leq 1$. For eksempel er

$$\sum_{n=1}^{\infty} \frac{1}{n^3}$$

konvergent, og

$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$$

divergent. Hva som skjer når $1 < s < 2$ er litt mer subtilt, så vi må utsette det litt. △

Her er et spesialtilfelle av sammenlikningstesten:

**Grensesammenlikningstesten**

La $x$ og $y$ være reelle følger, og la

$$\rho = \lim_{n\to\infty} \frac{x_n}{y_n}.$$

- Dersom $\rho > 0$ og $\sum y$ er divergent, er også $\sum x$ divergent.
- Dersom $\rho < \infty$ og $\sum y$ er konvergent, er også $\sum x$ konvergent.

**Eksempel 3.15.** Grensesammenligningstesten gir at

$$\sum_{n=0}^{\infty} \sin \frac{1}{n}$$

er divergent, siden

$$\rho = \lim_{n\to\infty} \frac{\sin \frac{1}{n}}{1/n} = \lim_{x\to 0} \frac{\sin x}{x} = 1. \qquad \triangle$$

Geometrisk rekke er også grei å sammenlikne med. Den er faktisk så grei å sammenlikne med at man kan gjøre det en gang for alle og skrive opp et teorem:

**Forholdstesten**

La $z$ være en følge, og la

$$\rho = \lim_{n\to\infty} \left|\frac{z_{n+1}}{z_n}\right|.$$

- Dersom $\rho > 1$ er $\sum z$ divergent.
- Dersom $\rho < 1$ er $\sum z$ konvergent.
- Dersom $\rho = 1$ kan alt skje.

**Eksempel 3.16.** Forholdstesten gir at

$$\sum_{n=0}^{\infty} \frac{1}{n!}$$

er konvergent, siden

$$\rho = \lim_{n\to\infty} \frac{1/(n+1)!}{1/n!} = \lim_{n\to\infty} \frac{1}{n} = 0. \qquad \triangle$$

Her er en rekketest som er veldig enkel i bruk.

**Alternerende rekketest**

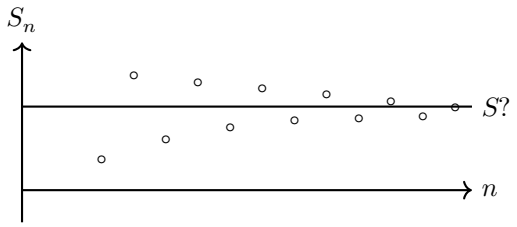
Dersom $x_n$ er en monotont synkende og positiv reell følge som konvergerer mot null, er rekken

$$\sum_{n=1}^{\infty} (-1)^n x_n$$

konvergent, og

$$|S_n| = \left|\sum_{n=1}^{N} (-1)^n x_n\right| < a_{n+1}$$

Beviset for denne er litt teknisk å skrive opp, men en figur av partialsummene forklarer omtrent hvordan det henger sammen:




Figuren forteller omtrent hvordan beviset går. La $m < n$ og $m$ være odde. Da er

$$\begin{aligned} |S_n - S_m| &= \sum_{k=m+1}^{n} (-1)^k x_k \\ &= x_{m+1} - (x_{m+2} - x_{m+3}) - \cdots \leq x_{m+1} \end{aligned}$$

og siden

$$\lim_{n\to\infty} x_n = 0$$

danner partialsummene en cauchyfølge. Lar du $n \to \infty$ får du feilestimatet.

**Eksempel 3.17.** Rekken

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \ln 2$$

er konvergent. Vi skal senere se hvorfor det blir $\ln 2$. △

**Absolutt konvergens**

Vi sier at rekken $\sum z$ konvergerer *absolutt* dersom

$$\sum_{n=1}^{\infty} |z_n|$$

konvergerer. En konvergent rekke som ikke konvergerer absolutt, sies å konvergere *betinget*.

**Eksempel 3.18.** Rekken

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

er betinget konvergent, siden

$$\sum_{n=1}^{\infty} \left| \frac{(-1)^{n+1}}{n} \right| = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

er divergent. △

Av ulikheten

$$\left| \sum_{k=m}^{n} z_k \right| \leq \sum_{k=m}^{n} |z_k|$$

og Cauchys konvergenskriterium, kan vi slutte at

**Absolutt konvergens og konvergens**

Dersom

$$\sum_{n=1}^{\infty} |z_n|$$

konvergerer, konvergerer også

$$\sum_{n=1}^{\infty} z_n.$$

**Eksempel 3.19.** Vi trenger altså ikke lure på om

$$\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n^2} = 1 - \frac{1}{4} + \frac{1}{9} - \frac{1}{16} + \cdots$$

er konvergent. △

Merk at sammenlikningstesten kan brukes til å sjekke absolutt konvergens både for reelle og komplekse rekker.

**Eksempel 3.20.** Forholdstesten gir at

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}$$

er absolutt konvergent for alle $z$, siden

$$\rho = \lim_{n\to\infty} \frac{|z^{n+1}/(n+1)!|}{|z^n/n!|} = |x| \lim_{n\to\infty} \frac{1}{n} = 0. \quad \triangle$$

Dersom du stokker om på rekkefølgen på leddene i en rekke, får du en annen rekke, la oss kalle det en omstokket rekke. Dersom en rekke er betinget konvergent, kan man ved omstokking få rekken til å konvergere til hva som helst. Dersom rekken er absolutt konvergent, konvergerer omstokkinger alltid til det samme. La $S_n$ og $T_n$ være partialsummene til en absolutt konvergent rekke $z$ og dens omstokking, og se på

$$|S_n - T_n|.$$

Velg $\epsilon$, og velg $N$ så stor at $n > m > N$ impliserer

$$\sum_{k=m}^{n} |z_k| \leq \epsilon.$$

Men i differansen $S_n - T_n$ er det en haug med kanselleringer, så vi kan velge både $p$, $m$ og $n$ store nok til at

$$|S_p - T_p| \leq \sum_{k=m}^{n} |z_k| \leq \epsilon.$$

slik at $S_n$ og $T_n$ konvergerer mot samme tall.

**Om stokking**

Dersom $\sum z$ konvergerer absolutt, konvergerer alle reorganiseringer til samme verdi.

# Kapittel 4

# Funksjoner fra $\mathbb{R}$ til $\mathbb{R}$

I dette kapitlet skal vi ta for oss funksjoner på $\mathbb{R}$. Dette er på noen måter mer komplisert enn funksjoner på $\mathbb{N}$.

Alternative kilder:

- Adams kap. P og 1-4
- Lindstrøms I kap. ??

## En sykt viktig funksjon

Mange vil hevde at eksponensialfunksjonen

$$\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

er den viktigste funksjonen av alle. Forholdstesten viser at rekken er absolutt konvergent for alle $x$:

$$\rho = \lim_{n\to\infty} \frac{|x|^{n+1}/(n+1)!}{|x|^n/n!} = \lim_{n\to\infty} \frac{|x|}{n+1} = 0,$$

og det går an å vise at

$$\exp(x+y) = \exp(x)\exp(y).$$

Tallet $e$ er definert ved

$$\exp(1) = \sum_{n=0}^{\infty} \frac{1}{n!} = 2.71182818284590452\ldots,$$

og det går an å vise at

$$\exp(x) = e^x.$$

Dette er relativt enkelt å vise for rasjonale $x$, man ser fra produktregelen at

$$\exp(2) = \exp(1)\exp(1) = e^2$$

og at

$$\exp(1) = \exp(1/2)\exp(1/2) = \sqrt{e}\sqrt{e}$$

og så videre.

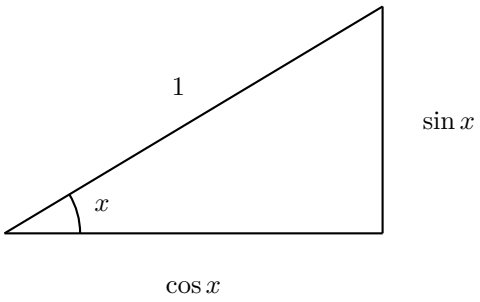
**Eksempel 4.1.** Det er vel på tide å introdusere sinusfunksjonen

$$\sin x = \frac{\exp(ix) - \exp(-ix)}{2i} = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

og cosinusfunksjonen

$$\cos x = \frac{\exp(ix) + \exp(-ix)}{2} = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}.$$

På skolen lærte du at disse var katetene i en rettvinklet trekant:




Det går an å vise at alt dette er det samme, men det er litt jobb. △

## Kontinuerlige funksjoner

Anta at man har en funksjon $f$. Grenseverdien

$$\lim_{x\to x_0} f(x)$$

forteller oss noe om hva som skjer med $f$ når den avhengige variabelen $x$ går mot verdien $x_0$. For de fleste funksjoner man støter på i dagliglivet, er

$$\lim_{x\to x_0} f(x) = f(x_0),$$

men dette er ikke alltid riktig.

**Definisjon.** En funksjon sies å ha grenseverdien $L$ i $x_0$ dersom det for hver $\epsilon > 0$ finnes en $\delta > 0$ slik at implikasjonen

$$0 < |x - x_0| < \delta \quad \Longrightarrow \quad |f(x) - L| < \epsilon$$

holder. Vi skriver i så fall

$$\lim_{x\to x_0} f(x) = L.$$

**Teorem 4.2.** *En grenseverdi, dersom den eksisterer, er entydig.*

*Bevis.* Dette er så og si identisk med beviset for tilsvarende teorem for følger. Prøv selv! □

**Eksempel 4.3.** For den konstante funksjonen

$$f(x) = 1$$

gjelder at

$$\lim_{x \to x_0} f(x) = 1$$

overalt. Dette er lett å se. Velg $\epsilon$. Vi har at

$$|f(x) - 1| = |1 - 1| = 0,$$

så her kan $\delta$ velges til hva som helst. △

**Eksempel 4.4.** Det første ordens polynomet

$$f(x) = ax + b$$

har grenseverdien

$$\lim_{x \to x_0} f(x) = f(x_0) = ax_0 + b.$$

Dette er også lett å se. Velg $\epsilon$. Vi har at

$$|f(x) - f(x_0)| = |a(x - x_0)|.$$

Dersom vi velger $\delta = \epsilon/|a|$, og krever $0 < |x - x_0| < \delta$, får vi

$$|f(x) - f(x_0)| = |a(x - x_0)| < |a|\frac{\epsilon}{|a|} = \epsilon.$$

Spesialtilfellet $a = 1$ og $b = 0$ forteller at

$$\lim_{x \to x_0} x = x_0,$$

som vi skal få bruk for lenger ned. △

**Eksempel 4.5.** Vi prøver å finne

$$\lim_{x \to 0} \exp(x).$$

Hva kan $L$ være? Vi ser først av definisjonen at

$$\exp(0) = 1,$$

Dette forteller ikke nødvendigvis at

$$\lim_{x \to x_0} \exp(x) = 1,$$

men vi kan ha det som arbeidshypotese, og undersøke saken nærmere. La $|x| < 1$. Det følger at

$$\begin{aligned} |\exp(x) - 1| &= \left|\sum_{n=1}^{\infty} \frac{x^n}{n!}\right| \leq \sum_{n=1}^{\infty} \left|\frac{x^n}{n!}\right| \\ &= \sum_{n=1}^{\infty} \frac{|x|^n}{n!} = |x| \sum_{n=1}^{\infty} \frac{|x|^{n-1}}{n!} \\ &\leq |x| \sum_{n=1}^{\infty} \frac{1}{n!} = |x|(e-1). \end{aligned}$$

Velg $\epsilon$. Dersom $|x| \leq \frac{\epsilon}{e-1}$ vil

$$\begin{aligned} |\exp(x) - 1| &\leq |x|(e-1) \\ &\leq \frac{\epsilon}{e-1}(e-1) = \epsilon, \end{aligned}$$

som er det vi måtte vise. △

Merk at i de tre foregående eksemplene er funksjonsverdier og grenseverdier identiske. Dette er et viktig poeng vi skal komme tilbake til om litt.

**Eksempel 4.6.** Vi kan på liknende vis som i forrige eksempel, se at

$$\lim_{x \to 0} \frac{\exp(x) - 1}{x} = 1.$$

La $|x| < 1$. Det følger at

$$\begin{aligned} \left|\frac{\exp(x) - 1}{x} - 1\right| &= \left|\sum_{n=1}^{\infty} \frac{x^n}{(n+1)!}\right| \leq \sum_{n=1}^{\infty} \left|\frac{x^n}{(n+1)!}\right| \\ &= \sum_{n=1}^{\infty} \frac{|x|^n}{(n+1)!} = |x| \sum_{n=1}^{\infty} \frac{|x|^{n-1}}{(n+1)!} \\ &\leq |x| \sum_{n=1}^{\infty} \frac{1}{(n+1)!} = |x|(e-2). \end{aligned}$$

Velg $\epsilon$. Dersom $|x| \leq \frac{\epsilon}{e-2}$ vil

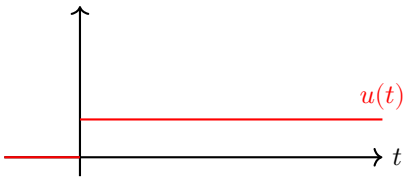
$$\begin{aligned} \left|\frac{\exp(x) - 1}{x} - 1\right| &\leq |x|(e-2) \\ &\leq \frac{\epsilon}{e-2}(e-2) = \epsilon, \end{aligned}$$

som er det vi måtte vise. Denne grenseverdien får vi bruk for når vi skal derivere eksponensialfunksjonen litt lenger ned. △

**Eksempel 4.7.** Heavisidefunksjonen er gitt ved

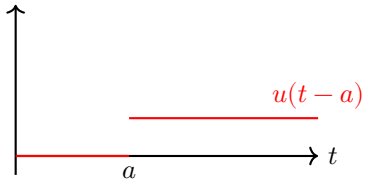
$$u(t) = \begin{cases} 0 & \text{for } t < 0 \\ 1 & \text{for } t \geq 0. \end{cases}$$

Man kan tenke på denne som en funksjon som slår noe på ved $t = 0$:



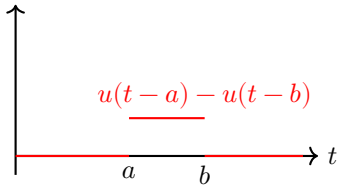

Vi kan slå på ved tiden $t = a$ istedet:

$$u(t-a) = \begin{cases} 0 & \text{for } t < a \\ 1 & \text{for } t \geq a, \end{cases}$$




Vi kan også slå på ved $t = a$ og av igjen ved $t = b$:

$$u(t-a) - u(t-b) = \begin{cases} 0 & \text{for } t < a \\ 1 & \text{for } a \leq t < b \\ 0 & \text{for } t \geq b. \end{cases}$$




Grenseverdien

$$\lim_{x\to 0} u(x)$$

eksisterer ikke. Hva skulle den i så fall vært? Fra venstre ser det ut til at $u$ går mot null, mens fra høyre ser det ut til at $u$ går mot en. △

**Eksempel 4.8.** Grenseverdien

$$\lim_{x\to 0} \sin\frac{1}{x}$$

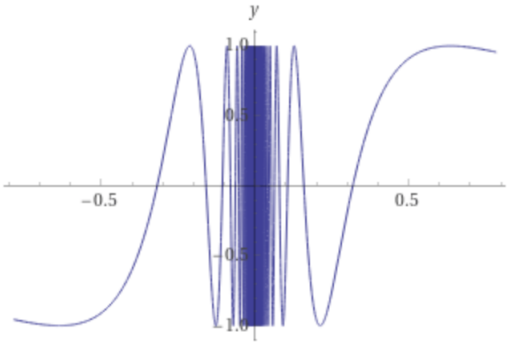
eksisterer ikke. Siden $\sin\frac{1}{x} = 1$ når

$$x = \frac{1}{2\pi n + \frac{\pi}{2}}$$

og $\sin\frac{1}{x} = -1$ når

$$x = \frac{1}{2\pi n + \frac{3\pi}{2}}$$

for alle $n \in \mathbb{Z}$, ser vi at dersom du står i $x = a$ og lar $x \to 0$ vil $\sin\frac{1}{x}$ ha uendelig mange oscillasjoner mellom $a$ og 0, uansett hvor liten $a$ er. Så konklusjonen er at dersom du har en kandidat for grenseverdi $L$, og velger $\epsilon > 0$ og $\delta > 0$ som du tror gjør jobben, vil $|\sin\frac{1}{x} - L| > 1 - |L|$ for en eller annen $|x| < \delta$, og $L$ er altså ikke en grenseverdi. Dette eksemplet er litt komplisert, men vi skal bruke det til å illustrere et viktig poeng litt senere. Ta en titt på figuren under. △




**Teorem 4.9.** *Anta vi har to funksjoner $f$ og $g$, med*

$$\lim_{x\to x_0} f(x) = L_1 \quad \text{og} \quad \lim_{x\to x_0} g(x) = L_2$$

*og la $c$ være et tall. Da gjelder at*

$$\lim_{x\to x_0} f(x) + g(x) = L_1 + L_2$$
$$\lim_{x\to x_0} cf(x) = cL_1$$
$$\lim_{x\to x_0} f(x)g(x) = L_1L_2$$
$$\lim_{x\to x_0} f(x)/g(x) = L_1/L_2 \quad (L_2 \neq 0)$$

*Bevis.* Velg $\epsilon$. Vi må vise at det går an å velge $\delta$ slik at $0 \le |x - x_0| \le \delta$ impliserer

$$|f(x) + g(x) - L_1 - L_2| < \epsilon$$

Merk først at

$$\begin{aligned}|f(x) + g(x) - L_1 - L_2| = \\ |f(x) - L_1 + g(x) - L_2| \le \\ |f(x) - L_1| + |g(x) - L_2|\end{aligned}$$

Siden $f$ og $g$ har respektive grenseverdier $L_1$ og $L_2$ i $x_0$, kan vi velge $\delta$ slik at $0 \le |x - x_0| \le \delta$ impliserer

$$|f(x) - L_1| < \epsilon/2$$

og

$$|g(x) - L_2| < \epsilon/2.$$

Men i så fall impliserer $0 \le |x - x_0| \le \delta$ at

$$|f(x) + g(x) - L_1 - L_2| < \epsilon/2 + \epsilon/2 = \epsilon,$$

som var det vi skulle vise. De andre reglene bevises på liknende måte, men det er litt mer regning. □

**Eksempel 4.10.** Nå kan vi vise at dersom

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

er

$$\lim_{x\to x_0} p(x) = p(x_0).$$

For det første vet vi at

$$\lim_{x\to x_0} 1 = 1$$

og at

$$\lim_{x\to x_0} x = x_0.$$

Teoremet om regneregler for grenseverdier sier at

$$\lim_{x\to x_0} f(x)g(x) = L_1L_2,$$

og dersom vi velger $f(x) = g(x) = x$, ser vi at

$$\lim_{x\to x_0} x^2 = x_0^2.$$

Det samme teoremet sier at

$$\lim_{x\to x_0} cf(x) = cL_1,$$

så derfor er

$$\lim_{x\to x_0} cx^2 = cx_0^2.$$

Til slutt kan vi bruke

$$\lim_{x\to x_0} f(x) + g(x) = L_1 + L_2,$$

og slutte at for eksempel

$$\lim_{x\to x_0} 2x^2 + x = 2x_0^2 + x_0.$$

Hvis vi fortsetter i samme stilen, ser vi at

$$\begin{aligned}\lim_{x\to x_0} p(x) &= \lim_{x\to x_0} a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \\ &= a_n x_0^n + a_{n-1}x_0^{n-1} + \cdots + a_1 x_0 + a_0 \\ &= p(x_0).\end{aligned}$$

△

**Eksempel 4.11.** Vi kan nå vise at

$$\lim_{x\to x_0} \exp x = \exp x_0.$$

Siden

$$\lim_{x\to x_0} x = x_0,$$

må

$$\lim_{x\to x_0} x - x_0 = 0.$$

Vi beregner så

$$\begin{aligned}\lim_{x\to x_0} \exp x &= \lim_{x\to x_0} \exp(x_0 + x - x_0)\\ &= \lim_{x\to x_0} \exp x_0 \exp(x - x_0)\\ &= \exp x_0 \lim_{x\to x_0} \exp(x - x_0)\\ &= \exp x_0 \exp(0) = \exp x_0.\end{aligned}$$ △

Det neste teoremet kalles gjerne skviseteoremet, for det handler om en funksjon som blir skvist mellom to andre funksjoner.

**Teorem 4.12.** *Anta at $f(x) \le g(x) \le h(x)$ på et intervall som inneholder $x_0$, og at*

$$\lim_{x\to x_0} f(x) = \lim_{x\to x_0} h(x) = L$$

*Da er*

$$\lim_{x\to x_0} g(x) = L.$$

*Bevis.* Denne tar vi i IF. □

**Eksempel 4.13.** Man kan bevise at

$$\lim_{x\to 0} \frac{\sin x}{x} = 1$$

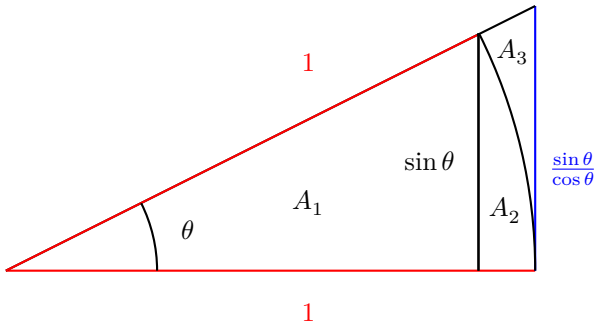
geometrisk ved å bruke skviseteoremet på ulikheten

$$\cos x \le \frac{\sin x}{x} \le \frac{1}{\cos x}.$$

Denne kan utledes ved å betrakte figuren under og se at

$$A_1 \le A_1 + A_2 \le A_1 + A_2 + A_3.$$

Husk at arealet av et kakestykke med radius en og vinkelutslag $\theta$ er $\theta/2$. △




**Eksempel 4.14.** Grenseverdien

$$\lim_{x\to 0} x \sin\frac{1}{x}$$

eksisterer. Siden $|\sin\frac{1}{x}| \le 1$ når $x \neq 0$, er

$$-|x| \le x \sin\frac{1}{x} \le |x|,$$

og siden

$$\lim_{x\to 0} |x| = \lim_{x\to 0} -|x| = 0,$$

må

$$\lim_{x\to 0} x \sin\frac{1}{x} = 0.$$ △

Til slutt kan nevnes at i noen situasjoner er vi nødt til å operere med ensidige grenser.

**Definisjon.** En funksjon sies å ha den venstresidige grenseverdien $L$ i $x_0$ dersom det for hver $\epsilon > 0$ finnes en $\delta > 0$ slik at implikasjonen

$$0 < x - x_0 < \delta \quad \Longrightarrow \quad |f(x) - L| < \epsilon$$

holder. Vi skriver i så fall

$$\lim_{x\downarrow x_0} f(x) = L$$

eller

$$\lim_{x\to x_0^+} f(x) = L.$$

Høyresidige grenser defineres tilsvarende.

En funksjon er kontinuerlig i et punkt dersom du kan tegne grafen gjennom punktet uten å løfte blyanten fra rutepapiret. For å gjøre dette presist, bruker vi grenseverdibegrepet.

**Definisjon.** En funksjon sies å være *kontinuerlig* i $x_0$ dersom:

- $f(x_0)$ er definert
- $\lim_{x\to x_0} f(x)$ eksisterer
- $f(x_0) = \lim_{x\to x_0} f(x)$

Som regel er det slik at man ikke spør om $f$ er kontinuerlig i et punkt der $f$ ikke er definert, for disse punktene ekskluderes per definisjon fra definisjonsmengden. Vi sier at en funksjon er kontinuerlig på et intervall dersom funksjonen er kontinuerlig for alle punkter i intervallet. Dersom intervallet er lukket, bruker man ensidige grenser i endepunktene.

**Eksempel 4.15.** Polynomer er kontinuerlige overalt. Vi har vist at for et polynom $p$, er

$$\lim_{x\to x_0} p(x) = p(x_0)$$

for alle $x_0$. △

**Eksempel 4.16.** En rasjonal funksjon er en funksjon på formen

$$\frac{p(x)}{q(x)}$$

der $p$ og $q$ er polynomer. Det samme resonnementet som for polynomer kan kjøres på disse funksjonene. Man bruker at

$$\lim_{x\to x_0}\frac{f(x)}{g(x)}=\frac{L_1}{L_2},$$

og slutter at siden $p$ og $q$ er polynomer, er

$$\lim_{x\to x_0}\frac{p(x)}{q(x)}=\frac{p(x_0)}{q(x_0)},$$

så lenge $q(x_0)\neq 0$. Derfor er rasjonale funksjoner kontinuerlige overalt der $q(x)\neq 0$. △

**Eksempel 4.17.** Heavisidefunksjonen er ikke kontinuerlig i $x=0$. △

**Eksempel 4.18.** Eksponensialfunksjonen er kontinuerlig overalt. Vi har jo vist at

$$\lim_{x\to x_0}\exp x=\exp x_0$$

for alle $x_0$. △

**Eksempel 4.19.** Funksjonen

$$f(x)=\begin{cases}\sin\frac{1}{x} & x\neq 0\\ L & x=0\end{cases}$$

er ikke kontinuerlig for noen $L$. △

**Eksempel 4.20.** Funksjonen

$$f(x)=\begin{cases}x\sin\frac{1}{x} & x\neq 0\\ 0 & x=0\end{cases}$$

er kontinuerlig overalt. △

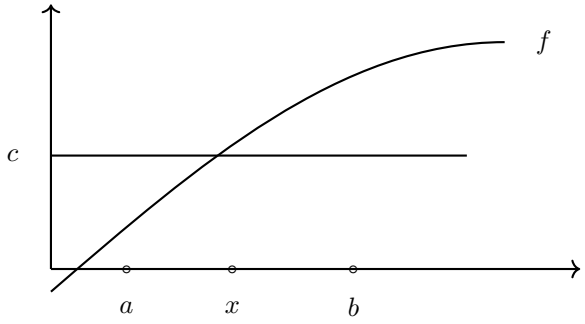
**Eksempel 4.21.** Funksjonen gitt ved $f(x)=1/x$ kan virke noe paradoksal, siden den er kontinuerlig overalt der den er definert, men allikevel ikke kan tegnes uten å løfte blyanten fra arket i $x=0$. Forklaringen er at $f$ er ikke definert i $x=0$, og dette deler definisjonsmengden i to adskilte deler. En kontinuerlig funksjon kan tegnes uten å løfte pennen fra papiret så lenge du ser på et sammenhengende intervall, men når definisjonsmengden er delt i to, bryter dette sammen. (Man kan spørre seg om det er naturlig i det hele tatt å stille spørsmålet hvorvidt en funksjon er kontinuerlig på en ikke sammenhengende definisjonsmengde.) Dette er lett å bli forvirret av i starten, og det finnes en strengere type kontinuitet som kalles uniform kontinuitet. En funksjon $f$ er uniformt kontinuerlig dersom det for hver $\epsilon>0$ finnes en $\delta>0$ slik at

$$0<|x-y|<\delta \quad\Longrightarrow\quad |f(x)-f(y)|<\epsilon$$

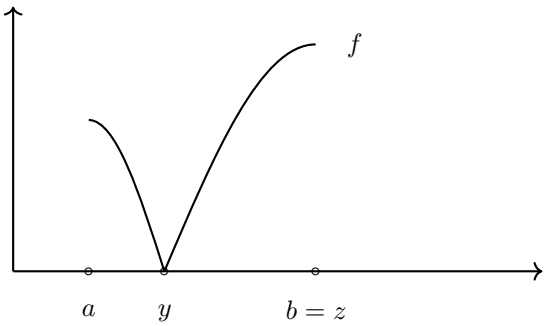
for alle $x$ og $y$. Funksjonen $f(x)=1/x$ er ikke uniformt kontinuerlig. △

De to neste teoremene kalles henholdsvis skjæringssetningen og ekstremalverdisetningen. Bevisene er litt for vanskelige for dette kurset, men figurene kan hinte om hvorfor teoremene er sanne. Merk også at teoremene ikke er sanne dersom definisjonsmengden er $\mathbb{Q}$ istedet for $\mathbb{R}$.

**Teorem 4.22.** *Anta at $f$ er en kontinuerlig funksjon på $[a,b]$, og at $f(a)<f(b)$. Dersom $f(a)<c<f(b)$, finnes en $x\in[a,b]$ slik at $f(x)=c$.*

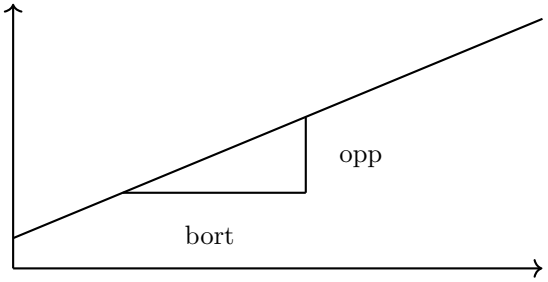



**Teorem 4.23.** *Anta at $f$ er en kontinuerlig funksjon på $[a,b]$. Det finnes punkter $y$ og $z$ slik at $f(y)\leq f(x)\leq f(z)$ for alle $x\in[a,b]$.*

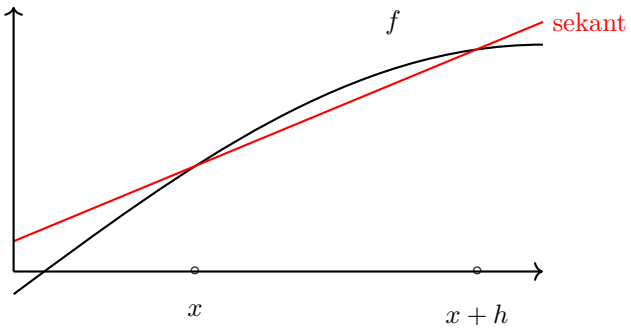



# Deriverbare funksjoner

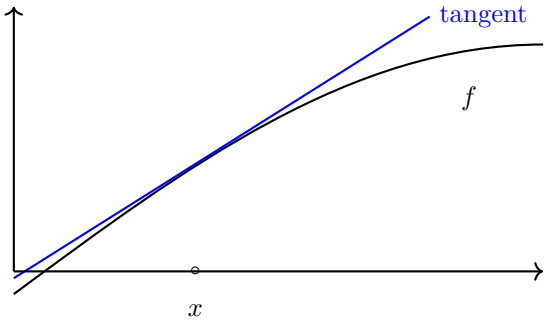
Det er et interessant faktum at Newton fant ut av differensialregning flere hundre år før noen skjønte hva de reelle tallene egentlig var. Stigningstallet til en rett linje er opp delt på bort.




La $f$ være en funksjon. En *sekant* er en rett linje som skjærer grafen til $f$ lokalt i to punkter:




En *tangent* er en rett linje som tangerer grafen til $f$ i et punkt:




Man kan tenke på dette som en sekant der de to skjæringspunktene er sammenfallende. Stigningstallet til tangenten i $x_0$ er det vi mener når vi snakker om stigningstallet til $f$ i $x_0$. Nå skal vi definere nøyaktig hva vi mener med dette.

**Definisjon.** Vi definerer den deriverte til $f$ i punktet $x$ som

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

dersom grenseverdien eksisterer. Vi sier i så fall at $f$ er *deriverbar* i $x$.

Vi skriver også

$$f'(x) = \frac{d}{dx} f(x).$$

Noen ganger er denne notasjonen mer praktisk. Siden grenseverdier er entydige, ser vi at den deriverte er entydig bestemt dersom den eksisterer. Likningen for tangenten til $f$ i punktet $x_0$ er gitt ved

$$y - f(x_0) = f'(x_0)(x - x_0).$$

**Eksempel 4.24.** La $f(x) = x$. Vi beregner

$$f'(x) = \lim_{h\to 0} \frac{x+h-x}{h} = \lim_{h\to 0} \frac{h}{h} = 1. \qquad \triangle$$

**Eksempel 4.25.** La $f(x) = \exp x$. Produktregelen for eksponensialfunksjonen gir

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\exp(x+h) - \exp x}{h} \\ &= \exp x \lim_{h\to 0} \frac{\exp h - 1}{h} = \exp x. \end{aligned} \qquad \triangle$$

**Eksempel 4.26.** Funksjonen

$$f(x) = \begin{cases} x \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

er kontinuerlig, men ikke deriverbar, i $x = 0$. Grenseverdien

$$\begin{aligned} f'(0) &= \lim_{h\to 0} \frac{f(h) - f(0)}{h} \\ &= \lim_{h\to 0} \frac{h \sin \frac{1}{h} - 0}{h} \\ &= \lim_{h\to 0} \sin \frac{1}{h} \end{aligned}$$

eksisterer ikke. $\triangle$

**Teorem 4.27.** *Dersom en funksjon er deriverbar i et punkt, er den også kontinuerlig i punktet.*

*Bevis.* Dersom

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

skal eksistere, må både $f(x)$ og

$$\lim_{h\to 0} f(x+h)$$

eksistere, og de må være like. Men dette er det samme som at $f$ kontinuerlig i $x$. $\square$

**Eksempel 4.28.** Hverken heavisidefunksjonen eller $\sin \frac{1}{x}$ er deriverbare i $x = 0$. Dette vet vi siden de ikke er kontinuerlige i $x = 0$. $\triangle$

**Eksempel 4.29.** Funksjonen

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

er deriverbar i $x = 0$, siden

$$\begin{aligned} f'(0) &= \lim_{h\to 0} \frac{f(h) - f(0)}{h} \\ &= \lim_{h\to 0} \frac{h^2 \sin \frac{1}{h} - 0}{h} \\ &= \lim_{h\to 0} h \sin \frac{1}{h} = 0. \end{aligned}$$

Men her er en artig tvist: $f'$ er ikke en kontinuerlig funksjon, for

$$f'(x) = 2x \sin \frac{1}{x} - \cos \frac{1}{x}$$

og $\lim_{x\to 0} f'(x)$ eksisterer ikke! $\triangle$

Eksemplet over viser at deriverbarhet ikke impliserer at den deriverte er en kontinuerlig funksjon. Den motsatte implikasjonen er derimot sann; dersom $f'$ er kontinuerlig i et punkt, må jo $f'$ være definert i dette punktet, så det er klart at $f$ er deriverbar om $f'$ er kontinuerlig. Denne lille skjebnens ironi inspirerer oss til følgende definisjon.

**Definisjon.** Dersom $f'$ er en kontinuerlig funksjon, sier vi at $f$ er kontinuerlig deriverbar. Dersom den $n$-te deriverte $f^n$ er kontinuerlig, sier vi at $f$ er $n$ ganger kontinuerlig deriverbar. Dersom $f^n$ er kontinuerlig for alle $n \in \mathbb{N}$, sier vi at $f$ er glatt.

**Eksempel 4.30.** Funksjonen

$$f(x) = \begin{cases} x^3 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

er kontinuerlig deriverbar i $x = 0$, siden

$$\begin{aligned} f'(0) &= \lim_{h \to 0} \frac{f(h) - f(0)}{h} \\ &= \lim_{h \to 0} \frac{h^3 \sin \frac{1}{h} - 0}{h} \\ &= \lim_{h \to 0} h^2 \sin \frac{1}{h} = 0. \end{aligned}$$

$$f'(x) = 3x^2 \sin \frac{1}{x} - x \cos \frac{1}{x}$$

som er kontinuerlig i $x = 0$. △

Her er noen regler for derivasjon.

**Teorem 4.31.** *La $f$ og $g$ være deriverbare funksjoner. Følgende regler gjelder. (Den tredje gjelder kun når $g(x) \neq 0$.)*

$$(f(x) + g(x))' = f'(x) + g'(x)$$

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$$

$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

*Bevis.* Vi vet at

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

og at

$$g'(x) = \lim_{h \to 0} \frac{g(x+h) - g(x)}{h}.$$

Addisjonsregelen er rett fram:

$$\begin{aligned} (f(x) + g(x))' &= \lim_{h \to 0} \frac{f(x+h) + g(x+h) - f(x) - g(x)}{h} \\ &= \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \\ &+ \lim_{h \to 0} \frac{g(x+h) - g(x)}{h} \\ &= f'(x) + g'(x) \end{aligned}$$

For produktregelen må vi sjonglere litt mer:

$$\begin{aligned} (f(x)g(x))' &= \lim_{h \to 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h \to 0} \frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} \\ &+ \lim_{h \to 0} \frac{f(x+h)g(x) - f(x)g(x)}{h} \\ &= \lim_{h \to 0} f(x+h) \frac{g(x+h) - g(x)}{h} \\ &+ \lim_{h \to 0} g(x) \frac{f(x+h) - f(x)}{h} \\ &= f'(x)g(x) + f(x)g'(x). \end{aligned}$$

Brøkregelen dropper vi. □

**Eksempel 4.32.** La $f(x) = x^2$. Vi beregner

$$f'(x) = (x \cdot x)' = 1 \cdot x + x \cdot 1 = 2x.$$ △

**Eksempel 4.33.** La $f(x) = x^n$, der $n \in \mathbb{N}$. Vi viser at

$$\frac{d}{dx} x^n = n x^{n-1}$$

ved induksjon. De to foregående eksemplene viser at regelen gjelder for $n = 1$ og $n = 2$. Induksjonssteget er å vise at likningen

$$\frac{d}{dx} x^n n = n x^{n-1}$$

impliserer

$$\frac{d}{dx} x^{n+1} = (n+1) x^n.$$

Vi bruker multiplikasjonsregelen på den første, og får

$$(x^{n+1})' = (x \cdot x^n)' = 1 \cdot x^n + x \cdot (n x^{n-1}) = (n+1) x^n.$$

Bruker vi derivasjonsregelen for rasjonale funksjoner, ser vi at regelen også gjelder for $n \in \mathbb{Z}$. Det går også an å vise at regelen gjelder dersom $n \in \mathbb{R}$, men dette er mye vanskeligere. △

**Teorem 4.34.** *Dersom $f$ er deriverbar i punktet $g(x)$, og $g$ er deriverbar i punktet $x$, er*

$$\frac{d}{dx} (f(g(x))) = f'(g(x))g'(x).$$

*Bevis.* Denne er litt teknisk, så vi nøyer oss med en skisse av hvordan beviset går. Da kan man ihvertfall skjønne hvorfor formelen ser ut som den gjør. Vi må beregne

$$\frac{d}{dx} (f(g(x))) = \lim_{h \to 0} \frac{f(g(x+h)) - f(g(x))}{h}.$$

Men dersom $h \neq 0$, er

$$\begin{aligned} &\frac{f(g(x+h)) - f(g(x))}{h} = \\ &\frac{f(g(x+h)) - f(g(x))}{g(x+h) - g(x)} \frac{g(x+h) - g(x)}{h}, \end{aligned}$$

slik at

$$\begin{aligned} \frac{d}{dx} (f(g(x))) &= \lim_{h \to 0} \frac{f(g(x+h)) - f(g(x))}{h} \\ &= f'(g(x))g'(x). \end{aligned}$$ □

**Eksempel 4.35.** La $f(x) = \cos x$. Vi bruker kjerneregelen, og får

$$\begin{aligned} f'(x) &= \frac{i\exp(ix) - i\exp(-ix)}{2} \\ &= -\frac{\exp(ix) - \exp(-ix)}{2i} = -\sin x. \end{aligned}$$ △

På skolen lærer man gjerne at maksimum og minimum er det samme som at tangenten er vannrett. Dette er en forenkling av virkeligheten, Det finnes også andre typer maksimums- og minimumspunkter.

**Eksempel 4.36.** La $f : [0,1] \to \mathbb{R}$ være gitt ved $f(x) = x$. Denne funksjonen har maksimum i $x = 1$. Dette er helt åpenbart det punktet der $f$ tar sin største verdi. △

**Eksempel 4.37.** Funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved $f(x) = |x|$ har et lokalt minimum i $x = 0$, siden

$$0 \leq |x|$$

for alle $x$. △

**Definisjon.** Dersom det finnes en $\delta$ slik at

$$|x - p| < \delta \implies f(p) \geq f(x),$$

sier vi at $p$ er et *lokalt maksimum* for $f$. Dersom

$$|x - p| < \delta \implies f(p) \leq f(x),$$

er $p$ et lokalt minimum.

**Eksempel 4.38.** Funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved $f(x) = 1 - (x-1)^2$ har et lokalt maksimum i $x = 1$. Dette er mulig å se ved å niglane litt på funksjonsuttrykket. Siden

$$(x-1)^2 \geq 0$$

oppnår funksjonen sin maksimale verdi når dette leddet er null, altså i $x = 1$. △

**Definisjon.** Et punkt $p$ der $f'(p) = 0$, kalles gjerne kritisk punkt. Punkter der $f'$ ikke eksisterer, kalles singulære punkter.

**Teorem 4.39.** *Dersom $f$ er en funksjon på $(a,b)$, har et lokalt maksimum i $p$, og $f'(p)$ er definert, er $f'(p) = 0$. Det samme gjelder om $p$ er et minimum.*

*Bevis.* La $p$ være et lokalt maksimum. Beviset for minimum er helt likt. Dersom $h > 0$ er liten, er

$$\frac{f(p+h) - f(p)}{h} \leq 0,$$

og dersom vi lar $h$ gå mot null, ser vi at

$$f'(p) = \lim_{x \to 0} \frac{f(p+h) - f(p)}{h} \leq 0.$$

Men vi kan gjøre det samme resonnementet med en liten $h < 0$. Da blir

$$\frac{f(p+h) - f(p)}{h} \geq 0,$$

og

$$f'(p) = \lim_{x \to 0} \frac{f(p+h) - f(p)}{h} \leq 0.$$

Med andre ord må både $f'(p) \leq 0$ og $f'(p) \geq 0$, så eneste mulighet er $f'(p) = 0$. □

**Eksempel 4.40.** Et andregradspolynom

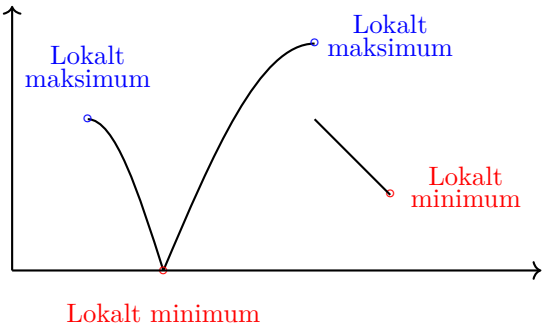
$$f(x) = ax^2 + bx + c$$

har ekstremaltpunkt i $x = -\frac{b}{2a}$, siden

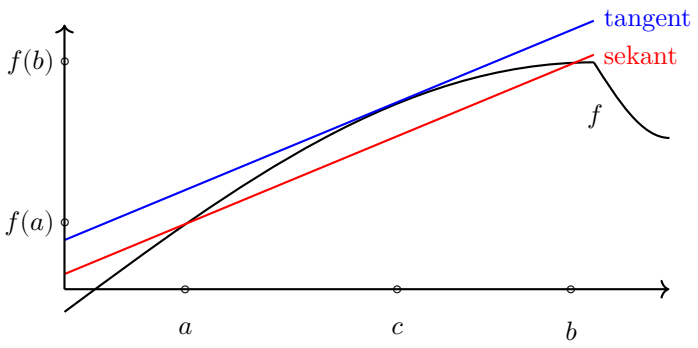
$$f'\left(-\frac{b}{2a}\right) = 2a \cdot \left(-\frac{b}{2a}\right) + b = -b + b = 0.$$ △

**Teorem 4.41.** *La $I$ være et intervall, og anta at $f : I \to \mathbb{R}$ har et ekstremalpunkt i $x_0 \in I$. Da er $x_0$ enten et endepunkt, eller et kritisk punkt, eller et singulært punkt.*

Under er en figur av en funksjon med verdimengde $[a,b]$.




Sekantsetningen kan brukes til å bevise noen andre interessante ting, og er motivert fra følgende figur. Dersom $f$ er deriverbar og du slår en sekant, må det mellom punktene der du slo sekanten finnes en tangent som er parallell med sekanten din.




**Teorem 4.42.** *Dersom $f$ er kontinuerlig på $[a,b]$ og deriverbar på $(a,b)$, finnes $c \in (a,b)$ slik at*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

*Bevis.* Vi beviser først spesialtilfellet der $f(a) = f(b)$. I dette tilfellet kalles gjerne sekantsetningen Rolles teorem, og teoremet sier nå at $f'(c) = 0$ for en eller annen $c \in (a, b)$. Dette er lett å bevise. Siden $f$ er kontinuerlig på $[a, b]$, må $f$ ha et maksimum og et minimum på $[a, b]$. Hvis begge disse sitter i endepunktene, må $f$ være en konstant funksjon, og da må minst ett av disse finnes på $(a, b)$. La oss kalle dette punktet $c$. Siden $f$ er deriverbar på $(a, b)$ må $f'(c) = 0$.

Dersom $f(a) \neq f(b)$, kan vi gjøre som følger. La $g$ være funksjonen gitt ved

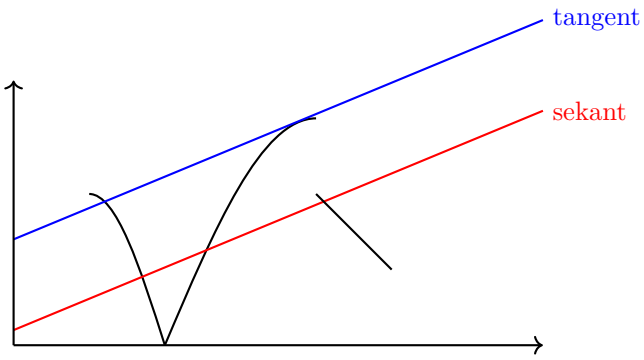
$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Siden $g(a) = g(b) = f(a)$, må Rolles teorem gjelde, og det må finnes et punkt $c \in (a, b)$ slik at $g'(c) = 0$. Men dette betyr at

$$0 = g'(c) = f'(c) - \frac{f(b) - f(a)}{b - a},$$

som er det sekantsetningen sier. □

Merk at $f$ må være deriverbar overalt på $(a, b)$, ellers trenger ikke utsagnet i teoremet være sant.



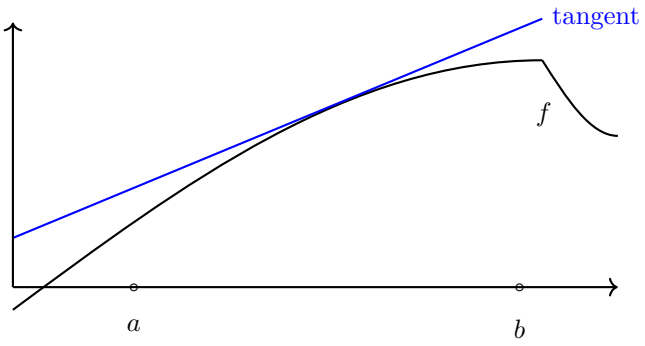

Sekantsetningen har en generalisering som kalles taylors teorem.

**Teorem 4.43.** *La $f$ være en $n + 1$ ganger kontinuerlig deriverbar funksjon på et intervall som inneholder $a$ og $x$. Det finnes en $s$ mellom $a$ og $x$ slik at*

$$f(x) = f(a) + f'(a)(x - a) + \cdots + \frac{f^n(a)}{n!}(x - a)^n + \frac{f^{n+1}(s)}{(n+1)!}(x - a)^{n+1}$$

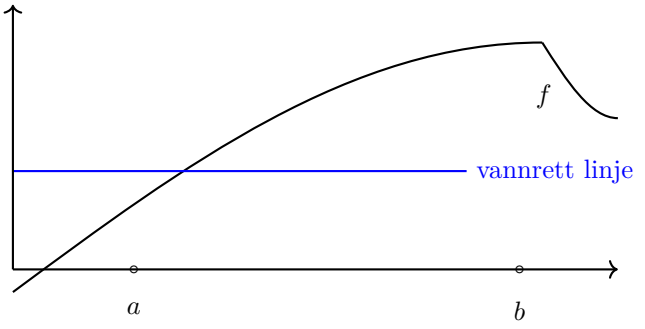
Vi avslutter med to teoremer.

**Teorem 4.44.** *Anta at $f$ er deriverbar på $(a, b)$. Dersom $f' > 0$ på $(a, b)$ er $f$ stigende på $(a, b)$. og dersom $f' < 0$ på $(a, b)$ er $f$ synkende på $(a, b)$.*




**Teorem 4.45.** *Anta at $f$ er deriverbar på $(a, b)$ og enten $f' > 0$ eller $f' < 0$ på $(a, b)$. Da eksisterer $f^{-1}$, og*

$$\frac{d}{dx}f^{-1}(x) = \frac{1}{f'(f^{-1}(x))}.$$




Dersom vi deriverer likningen

$$f(f^{-1}(x)) = x,$$

med kjerneregelen, får vi

$$f'(f^{-1}(x)) \cdot \frac{d}{dx}f^{-1}(x) = 1,$$

som gir at

$$\frac{d}{dx}f^{-1}(x) = \frac{1}{f'(f^{-1}(x))}.$$

# Analytiske funksjoner

Vi er vant med å skrive om funksjoner. For eksempel kan funksjonsuttrykket

$$f(x) = \cos^2 x - \sin^2 x$$

like gjerne skrives

$$f(x) = \cos 2x.$$

I dette kapitlet skal vi ta et steg videre, og skrive funksjoner som uendelige summer av tilsynelatende ikkerelaterte funksjoner.

Du kjenner allerede fire taylorrekker.

**Eksempel 4.46.**

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots \qquad \triangle$$

**Eksempel 4.47.**

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots \qquad \triangle$$

**Eksempel 4.48.**

$$\cos x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots \qquad \triangle$$

**Eksempel 4.49.**

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \cdots \quad (|x| < 1) \qquad \triangle$$

**Eksempel 4.50.** Funksjonen

$$f(x) = \begin{cases} e^{-\frac{1}{(x+1)(x-1)}} & |x| < 1 \\ 0 & |x| \geq 0 \end{cases}$$

er glatt, men ikke analytisk. Siden

$$\lim_{s \to 1} \frac{d}{dx^n} f(s) = \lim_{s \to -1} \frac{d}{dx^n} f(s) = 0$$

for alle $n$ er $f$ glatt, men det er klart at taylorrekken om et punkt $x = a$ ikke kan representere $f$ på hele $\mathbb{R}$. $\triangle$

Det er også mulig å skrive funksjoner som uendelige rekker av sinus- og cosinusfunksjoner, men dette skal vi vente med til senere.

**Eksempel 4.51.** Vi kan bruke rekken for eksponensialfunksjonen:

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

sinusfunksjonen:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

og cosinusfunksjonen:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!}$$

Dersom bruker den imaginære enheten $i$ til å skrive

$$\cos x = 1 + \frac{(ix)^2}{2!} + \frac{(ix)^4}{4!} + \cdots = \sum_{n=0}^{\infty} \frac{(ix)^{2n}}{(2n)!}$$

og

$$i \sin x = ix + \frac{(ix)^3}{3!} + \frac{(ix)^5}{5!} + \cdots = \sum_{n=0}^{\infty} \frac{(ix)^{2n+1}}{(2n+1)!},$$

og legger disse to sammen, får vi

$$\cos x + i \sin x = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} = e^{ix}. \qquad \triangle$$

# Numeriske likningsløsere

Det tilsynelatende mylderet av likninger man løser på skolen kan forlede en til å tro at man kan løse alle likninger analytisk, men i virkeligheten lærer man først og fremst standardteknikker for å løse noen veldig spesifikke likningstyper.

**Eksempel 4.52.** Likningen

$$ax^2 + bx + c = 0$$

kan løses ved å dele ut $a$, og skrive

$$x^2 + \frac{b}{a}x + \frac{c}{a} = \left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} - \frac{b^2}{4a^2} = 0$$

eller

$$\left(x + \frac{b}{2a}\right)^2 = \frac{b^2}{4a^2} - \frac{c}{a}.$$

Vi kvadrerer, og får

$$x = -\frac{b}{2a} \pm \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

som kalles *abc*-formelen. $\triangle$

Løsningen til en likning kalles gjerne en *rot*.

**Eksempel 4.53.** Likningen

$$x^3 + x^2 - 3x - 3 = 0$$

har en røtter $x = -1$ og $x = \pm\sqrt{3}$, men disse er ikke enkle å finne. Oppskriften for å løse en generell tredjegradslikning er lang og teknisk, og selv matematikkstudenter lærer det ikke. Løsningsteknikken til en generell fjerdegradslikning er enda mer komplisert, og for en generell femtegradslikning kan man ikke en gang utlede en løsningsformel. $\triangle$

**Eksempel 4.54.** Likningen

$$\cos(2x + 3) = \frac{1}{2}$$

løses ved å invertere cosinusfunksjonen

$$2x + 3 = \arccos \frac{1}{2},$$

huske at $\cos \frac{\pi}{3} = \frac{1}{2}$

$$2x + 3 = \frac{\pi}{3},$$

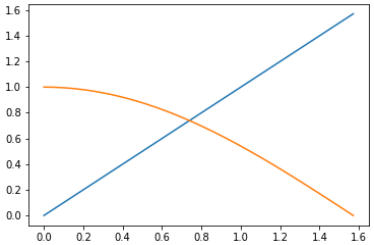
og løse for $x$

$$x = \frac{\pi}{6} - \frac{3}{2}. \qquad \triangle$$

**Eksempel 4.55.** Likningen

$$x = \cos x$$

har en rot på intervallet $[0, 1]$, se figur. Men denne likningen kan ikke løses med et endelig antall algebraiske operasjoner. $\triangle$




Noen likninger kan altså løses, mens andre ikke kan det. Andre likninger har løsningsteknikker som er for kompliserte til at det er praktisk å lære seg dem. Finnes det ingen løsningsteknikker som takler alle likninger? Svaret er nei, men det finnes teknikker for å beregne tilnærmede løsninger til likninger vi av en eller annen grunn ikke kan løse analytisk, og disse teknikkene kan ofte brukes på brede klasser av likninger. Disse teknikkene kalles *numeriske likningsløsere*.

En numerisk likningsløser produserer en følge av tilnærminger til løsningen. Dersom likningsløseren er tilpasset likningen vi prøver å løse, vil tilnærmingene bli bedre utover i følgen. Det neste eksemplet illustrerer tankegangen.

**Eksempel 4.56.** Siden

$$\cos\frac{1}{2} \approx 0.8776 \geq \frac{1}{2}$$

ser vi at roten ligger i intervallet $[\frac{1}{2}, 1]$. La oss sette $x_0 = \frac{3}{4}$, altså midt i dette invtervallet. Dette kan vi se på som en tilnærming til den analytiske løsningen. Siden

$$\cos\frac{3}{4} \approx 0.7317 \leq \frac{3}{4}$$

ser vi videre at roten må ligge i intervallet $[\frac{1}{2}, \frac{3}{4}]$, som er halvparten så bredt som det forrige, og vi setter $x_1 = \frac{5}{8}$, altså midt i det nye intervallet. Siden

$$\cos\frac{5}{8} \approx 0.8110 \geq \frac{5}{8},$$

må roten ligge i $[\frac{5}{8}, \frac{3}{4}]$, og vi setter $x_2 = \frac{11}{16}$. Slik kan vi fortsette så lenge vi ønsker. Denne metoden kalles gjerne *halveringsmetoden*. △

En løsning $r$ av en likning

$$f(x) = 0.$$

kalles gjerne en *rot*. Dersom $f$ bytter fortegn i $r$, kan vi benytte denne informasjonen til å finne en approksimasjon $x_n$ til $r$. La oss anta at vi kjenner $a$ og $b$ slik at $r \in [a, b]$ er den eneste roten til $f$ i $[a, b]$. Vi definerer nå $x_1 = a$, $x_2 = b$, og $x_3 = \frac{a+b}{2}$. Dersom $f$ ikke har noen sprang eller finner på noe annet tull på $[a, b]$, bil fortegnet til $f(x_3)$ fortelle oss om $x_3 < r$ eller $x_3 > r$, slik at vi med sikkerhet kan si om $r \in [x_1, x_3]$ eller $r \in [x_3, x_2]$.

Intervallene $[x_1, x_3]$ og $[x_3, x_2]$ er akkurat halvparten så lange som $[x_1, x_2]$, så etter å ha utført denne prosessen, har vi en mer nøyaktig ide om hvor $r$ befinner seg. Vi setter så $x_2 = x_3$ eller $x_1 = x_3$, alt etter fortegnet til $f(x_3)$, og repeterer prosedyren. Dersom vi gjør dette $n$ ganger, ender vi opp med et intervall med lengde

$$\frac{b-a}{2^n},$$

og hvis vi til slutt setter $r_n = \frac{x_1+x_2}{2}$, vet vi at

$$|r - x_n| \leq \frac{b-a}{2^{n+1}}.$$

**Halveringsmetoden**

Dersom $f$ er en kontinuerlig funksjon med en rot i intervallet $(a, b)$, produserer $n$ steg med halveringsmetoden et estimat som tilfredsstiller

$$|r - x_n| \leq \frac{b-a}{2^{n+1}}.$$

Professor Dottie gjenoppdaget med sine eksperimenter noe som kalles fikspunktiterasjonen. Hun oppdaget kort og godt at iterasjonen

$$x_{n+1} = \cos x_n$$

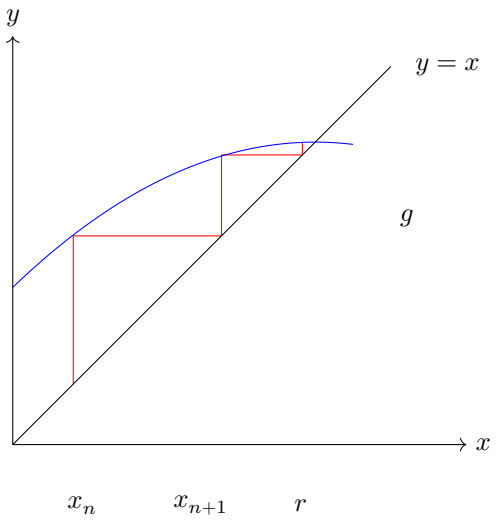
sakte men sikkert konvergerer mot den korrekte løsningen $r \approx 0.739085$. Dette ser sikkert rart ut, men i neste uke skal vi se på hvorfor dette noen ganger virker, og når det eventuelt virker. La oss anta at vi har en likning på formen

$$x = g(x).$$

En løsning $r$ av en slik likning, kalles et *fikspunkt*. Fikspunkmetoden er definert ved iterasjonen

$$x_{n+1} = g(x_n)$$

Det kan virke snodig at denne iterasjonen skal hjelpe oss til å finne $r$. Men det gjør den ofte. Figuren under illustrerer hvordan iterasjonen finner frem.




**Eksempel 4.57.** Likningen

$$x = \cos x$$

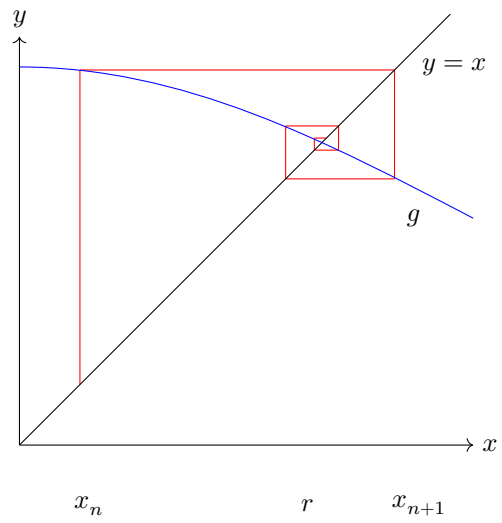
har som nevnt en foreløpig ukjent rot på intervallet $[0, 1]$. Iterasjonen

$$x_{n+1} = \cos x_n$$

produserer følgende tabell, dersom vi setter $x_0 = \frac{3}{4}$:

| | |
|---|---|
| $x_0$ | 0.7500000000000000 |
| $x_1$ | 0.731688868873821 |
| $x_2$ | 0.744047084788764 |
| $x_3$ | 0.735733618187236 |
| $x_4$ | 0.741338598887922 |
| $x_{78}$ | 0.739085133215161 |

Det går ikke så fort i svingene. Figurene under illustrerer hva som skjer. △



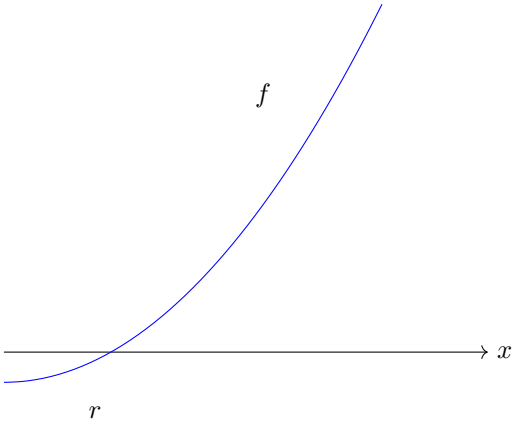

I neste bolk skal vi sette opp presise konvergenskriterier for fikspunktiterasjonen.

En av de metodene som er enklest å forstå, er *Newtons metode*. På samme måte som halveringsmetoden, produserer den en følge av tilnærminger til løsningen av likningen.

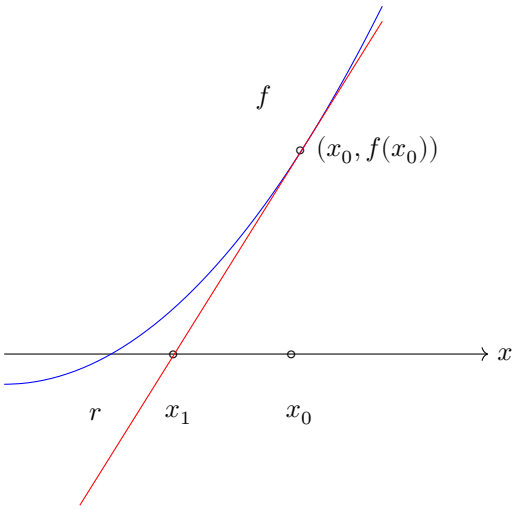
Newtons metode baserer seg på at likningen er skrevet på formen

$$f(x) = 0.$$

Den leter altså etter nullpunkter til funksjoner. La oss si at nullpunktet vi leter etter kalles $r$.




La oss anta at vi har en tilnærming $x_0$ til $r$. Vi slår tangenten til $f$ i $x_0$.




Punktet der tangenten skjærer $x$-aksen, kaller vi $x_1$. Dette punktet kan vi finne ved å sette opp likningen for tangenten til $f$ i $x_0$:

$$y - f(x_0) = f'(x_0)(x - x_0)$$

og så kreve at $y = 0$ i denne likningen:

$$-f(x_0) = f'(x_0)(x_1 - x_0)$$

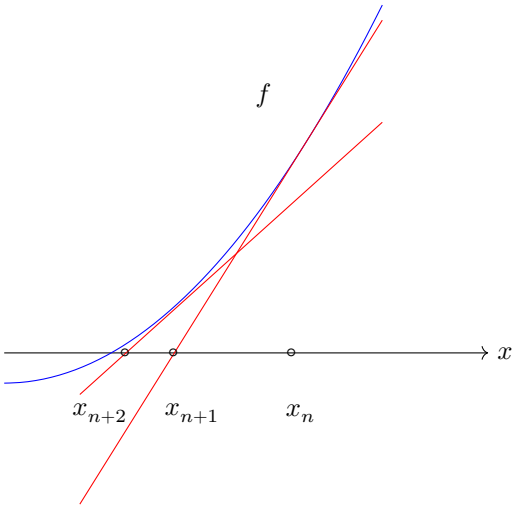
Løser vi denne likningen for $x_1$, får vi at

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Newtons metode er definert som den rekursive følgen

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}.$$

I mange situasjoner konvergerer denne følgen ganske fort mot $r$.




**Eksempel 4.58.** Likningen

$$x = \cos x$$

har som nevnt en foreløpig ukjent rot på intervallet $[0, 1]$. For å bruke Newtons metode, må vi skrive likningen

$$f(x) = x - \cos x = 0$$

og sette opp Newtons iterasjon

$$x_{n+1} = x_n - \frac{x_n - \cos x_n}{1 + \sin x_n}$$

Setter vi $x_0 = \frac{3}{4}$, produserer metoden følgende tabell:

| | |
|---|---|
| $x_0$ | 0.7500000000000000 |
| $x_1$ | 0.739111138752579 |
| $x_2$ | 0.739085133364485 |
| $x_3$ | 0.739085133215161 |
| $x_4$ | 0.739085133215161 |

Fra tredje til fjerde iterasjon er det ingen endring. Det betyr at vi mest sannsynlig har roten med seksten desimalers nøyaktighet. Senere i kurset skal vi sette opp presise kriterier for konvergens. △

**Teorem 4.59.** *La $g$ være en kontinuerlig deriverbar funksjon. Dersom både $a < g(x) < b$ og $|g'(x)| \leq L < 1$ på på $[a, b]$, finnes et entydig punkt $r$ slik at*

$$r = g(r).$$

*Fikspunktiterasjonen*

$$x_{n+1} = g(x_n)$$

*konvergerer mot $r$ dersom $x_0 \in [a, b]$.*

*Bevis.* Siden $a < g(x) < b$ vet vi at $x_1 \in (a, b)$ dersom $x_0 \in [a, b]$. Videre vet vi at dersom $x_n \in (a, b)$, er $x_{n+1} \in (a, b)$, så det er klart at $x_n \in (a, b)$ for alle $n \geq 1$.

Så la oss anta at $x_n \in (a, b)$. Vi bruker taylors teorem på $g$ i $r$, og skriver

$$g(x_n) = g(r) + g'(s)(x_n - r)$$

for en $s$ mellom $x_n$ og $r$. Hvis vi bruker at $r = g(r)$ og $x_{n+1} = g(x_n)$, kan vi skrive

$$x_{n+1} - r = g(x_n) - g(r) = g'(s)(x_n - r).$$

Vi tar absoluttverdi på begge sider, og bruker at $|g'(s)| \leq L < 1$, siden $s \in (a, b)$:

$$|x_{n+1} - r| = |g'(s)(x_n - r)| \leq L|x_n - r| \leq L^{n+1}|x_0 - r|$$

Lar vi så $n \to \infty$, får vi

$$\lim_{x\to\infty} |x_{n+1} - r| \leq |x_0 - r| \lim_{x\to\infty} L^{n+1} = 0,$$

siden $L < 1$.

For å vise at fikspunktet er entydig, kan vi anta at det finnes to fikspunkt, $r_1$ og $r_2$. Men da må

$$\begin{aligned} |r_2 - r_1| &= |x_{n+1} - r_1 - (x_{n+1} - r_2)| \\ &\leq |x_{n+1} - r_1| + |x_{n+1} - r_2| \\ &\leq L|x_n - r_1| + L|x_n - r_2| \\ &\leq L^{n+1}|x_0 - r_1| + L^{n+1}|x_0 - r_2|. \end{aligned}$$

Denne likningen gjelder for alle $n$. Dersom vi lar $n \to \infty$, går det siste uttrykket mot null, og vi ser at dette kun er mulig dersom $r_1 = r_2$. □

Sekantsetningen gir at det finnes et punkt $s$ slik at

$$x_{n+1} - r = g(x_n) - g(r) = g'(s)(x_n - r).$$

Dette er en likning som sier noe om størrelsen på $x_{n+1} - r$ som en funksjon av $x_n - r$, altså hvor mye feilen minker fra iterasjon til iterasjon, dersom fikspunktiterasjonen konvergerer. Feilen minker dersom $|g| \leq L < 1$.

**Eksempel 4.60.** Likningen

$$x = \frac{1}{2\cos x}$$

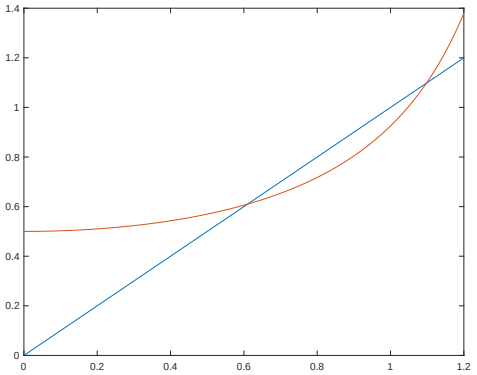
har en rot på intervallet $[1, 1.4]$. La oss prøve å finne den. Iterasjonen

$$x_{n+1} = \frac{1}{2\cos x_n}$$

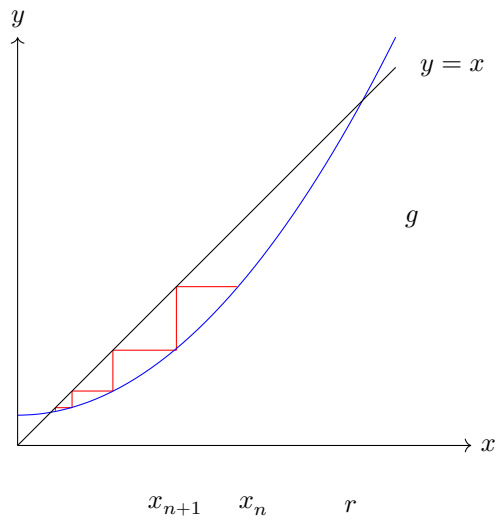
produserer følgende tabell, dersom vi setter $x_0 = 1$:

| | |
|---|---|
| $x_0$ | 1.0000000000000000 |
| $x_1$ | 0.925407858840463 |
| $x_2$ | 0.831243009591302 |
| $x_3$ | 0.741886011422627 |
| $x_4$ | 0.678246107990755 |
| $x_5$ | 0.642116945705959 |
| $x_6$ | 0.624352444102666 |
| $x_7$ | 0.616263059398898 |

Det gikk ikke i det hele tatt. △




I det siste eksemplet fant ikke fikspunktiterasjonen roten vi ville ha, men en annen. Figuren under illustrerer hvorfor. Fikspunktitrerasjonen finner aldri $r$ dersom $|g(r)| > 1$.




**Eksempel 4.61.** Vi løser polynomlikningen

$$x^3 + x^2 - 3x - 3 = 0.$$

Dette polynomet kan spaltes i

$$x^3 + x^2 - 3x - 3 = (x+1)(x-\sqrt{3})(x+\sqrt{3}),$$

og vi skal se hvordan fikspunkmetoden leter etter de forskjellige løsningene. Likningen kan skrives om til $x = g(x)$ på flere måter, men vi skal begynne med å skrive

$$x = \frac{1}{3}\left(x^3 + x^2 - 3\right),$$

slik at

$$g(x) = \frac{1}{3}\left(x^3 + x^2 - 3\right),$$

og

$$x_{n+1} = \frac{1}{3}\left(x_n^3 + x_n^2 - 3\right).$$

Vi prøver å finne løsningen $r = \sqrt{3} \approx 1.732050807568877$, og starter derfor en kjøring i $x_0 = 1.5$. Vi får:

| | |
|---|---|
| $x_0$ | 1.500000000000000 |
| $x_5$ | -0.995705356719772 |
| $x_{10}$ | -0.999982551541273 |
| $x_{15}$ | -0.999999928199386 |
| $x_{20}$ | -0.999999999704524 |
| $x_{25}$ | -0.999999999998784 |
| $x_{30}$ | -0.999999999999995 |
| $x_{35}$ | -1.000000000000000 |

Også nå ønsker metoden heller å finne $r = -1$. △

Fikspunktiterasjonen i forrige eksempel viste en sterk preferanse på hvilken rot den hadde lyst til å finne. Men man kan skrive om likningen til $x = g(x)$ på mange måter.

**Eksempel 4.62.** Vi skriver nå om likningen

$$x^3 + x^2 - 3x - 3 = 0.$$

til

$$x = \frac{3 + 3x - x^2}{x^2}.$$

Starter vi i $x_0 = -1.5$, får vi

| | |
|---|---|
| $x_0$ | -1.500000000000000 |
| $x_5$ | -1.732004423011461 |
| $x_{10}$ | -1.732050803458349 |
| $x_{15}$ | -1.732050807568513 |
| $x_{20}$ | -1.732050807568878 |
| $x_{25}$ | -1.732050807568877 |

Denne fikspunktiterasjonen klarte fint å finne roten $r = -\sqrt{3}$. △

Som vi ser av de to foregående eksemplene, kan fikspunktiterasjonen konvergere mot forskjellige røtter avhengig av hvordan vi skriver om likningen. Den kan også ikke konvergere i det hele tatt.

**Eksempel 4.63.** Vi prøver igjen

$$x = \frac{3 + 3x - x^2}{x^2}.$$

Starter vi i $x_0 = 1.5$, i håp om å finne $r = \sqrt{3}$, får vi

| | |
|---|---|
| $x_0$ | -1.500000000000000 |
| $x_1$ | 2.333333333333333 |
| $x_2$ | 0.836734693877551 |
| $x_3$ | 6.870315288518744 |
| $x_4$ | -0.499781154362809 |
| $x_5$ | 5.007884193672099 |
| $x_6$ | -0.281322161800267 |
| $x_7$ | 26.242541136990940 |
| $x_8$ | -0.881325585764740 |
| $x_9$ | -0.541641177723142 |
| $x_{10}$ | 3.687092259260734 |

Denne fikspunktiterasjonen klarte ikke å finne noe som helst når vi startet i $x_0 = 1.5$. △

**Eksempel 4.64.** I eksemplene over er

$$g(x) = \frac{3 + 3x - x^2}{x^2}.$$

og

$$g(x) = \frac{x^3 + x^2 - 3}{3}.$$

Hvis du deriverer disse og evaluerer i røttene til polynomet $x^3 + x^2 - 3x - 3$, vil du se et tydelig mønster. fikspunktiterasjonen greier ikke finne $r$ dersom $|g'(r)| > 1$. △

Vi kan også se på Newtons metode. Vi bruker taylors teorem (husk at $f(r) = 0$)

$$0 = f(x_n) + f'(x_n)(r - x_n) + \frac{f''(s)}{2}(r - x_n)^2$$

for $s$ mellom $x_n$ og $r$. Newtons metode er

$$-f(x_n) = f'(x_n)(x_{n+1} - x_n).$$

Trekker vi disse likningene fra hverandre, får vi

$$0 = f'(x_n)(r - x_{n+1}) + \frac{f''(s)}{2}(r - x_n)^2.$$

Her står det at

$$r - x_{n+1} = -\frac{f''(s)}{2f'(x_n)}(r - x_n)^2,$$

som sier at Newtons metode i mange situasjoner har kvadratisk konvergens. Det går an å sette opp presise kriterier for når dette skjer, men det skal vi ikke gjøre.

**Eksempel 4.65.** Vi søker løsningene til

$$x^4 - 10x^3 + 35x^2 - 50x + 24 = 0$$

Som Ingrid Espelid Hovig har vi jukset litt, og valgt et polynom som faktoriseres pent:

$$x^4-10x^3+35x^2-50x+24 = (x-1)(x-2)(x-3)(x-4)$$

Newtons metode blir:

$$x_{n+1} = x_n - \frac{x_n^4 - 10x_n^3 + 35x_n^2 - 50x_n + 24}{4x_n^3 - 30x_n^2 + 70x_n - 50}$$

La oss lete etter $r = 1$. Vi starter i $x_0 = 0.5$, og får:

| | |
|---|---|
| $x_0$ | 0.500000000000000 |
| $x_1$ | 0.798295454545455 |
| $x_2$ | 0.950817599863883 |
| $x_3$ | 0.996063283034122 |
| $x_4$ | 0.999971872651984 |
| $x_5$ | 0.999999998549667 |
| $x_6$ | 1.000000000000000 |

Konvergerer fort dette her. △

Som du ser i eksemplet over, dobles antall korrekte desimaler etter hver iterasjon.

**Eksempel 4.66.** Nå prøver vi å finne løsningene til

$$x^4 - 9x^3 + 27x^2 - 31x + 12 = 0.$$

Nok en gang er det et lettfaktorisert polynom:

$$x^4 - 9x^3 + 27x^2 - 31x + 12 = (x-1)^2(x-3)(x-4)$$

Newtons metode blir:

$$x_{n+1} = x_n - \frac{x_n^4 - 9x_n^3 + 27x_n^2 - 31x_n + 12}{4x_n^3 - 27x_n^2 + 54x_n - 31}$$

Nok en gang leter vi etter $r = 1$, ved å starte i $x_0 = 0.5$:

| | |
|---|---|
| $x_0$ | 0.500000000000000 |
| $x_1$ | 0.713414634146341 |
| $x_2$ | 0.842942878437971 |
| $x_3$ | 0.916937117337936 |
| $x_4$ | 0.957125910632705 |
| $x_5$ | 0.978193460613942 |
| $x_6$ | 0.988999465124112 |
| $x_7$ | 0.994474755305802 |
| $x_8$ | 0.997231047313292 |
| $x_9$ | 0.998613930094898 |
| $x_{10}$ | 0.999306565270834 |

Det ser ut til å konvergere, men mye saktere enn i sted. Hva skjedde? △

I eksemplet over er $f'(1) = 0$. Dette betyr at

$$g'(1) = -\frac{f(1)f''(1)}{(f'(1))^2}$$

ikke er definert. Grensen

$$\lim_{x \to 1} g'(x)$$

trenger ikke være null, og eksemplet demonstrerer tydelig at den kvadratiske konvergensen ikke kan garanteres dersom $f'(r) = 0$.

# Polynominterpolasjon

Dersom $x_i$ er $n+1$ er forskjellige punkter på $x$-aksen med korresponderende $y$-verdier $y_i$, finnes det et entydig polynom av maksimal grad $n$ som interpolerer punktene $(x_i, y_i)$. I dette kapitlet skal vi sette opp to forskjellige formler for dette polynomet.

## Lagranges interpolasjon

La $x_i$ være $n+1$ forskjellige punkter på intervallet $[a, b]$, med $x_0 = a$ og $x_n = b$. For hvert punkt $x_i$, definerer vi et polynom:

$$l_i(x) = \prod_{\substack{k=0 \\ k \neq i}}^{n} \frac{(x - x_k)}{(x_i - x_k)}.$$

Polynomet $l_i(x)$ har orden $n$, og tilfredsstiller

$$l_i(x_k) = \begin{cases} 1 & \text{for } i = k \\ 0 & \text{for } i \neq k \end{cases}.$$

La $f$ være en funksjon, med funksjonsverdier $f(x_i) = f_i$. Det er lett å se at

$$p_n(x) = \sum_{i=0}^{n} f_i l_i(x).$$

tilfredsstiller $p_n(x_i) = f(x_i)$ for alle $i$.

**Teorem 4.67.** *La $x_i$ være $n+1$ forskjellige punkter på intervallet $[a, b]$, og $f$ en funksjon med funksjonsverdier $f(x_i) = f_i$. Det finnes et entydig polynom som tilfredsstiller $p_n(x_i) = f(x_i)$ for alle $i$.*

*Bevis.* Konstruksjonen av Lagranges interpolasjon viser at det for en tabell med $n+1$ punkter eksisterer et interpolasjonspolynom av maksimal grad $n$; vi har jo nettopp konstruert det. Hvis vi antar at det finnes to forskjellige polynomer $p_n$ og $q_n$ av grad $n$ som interpolerer den samme tabellen, og evaluerer differansen $p_n - q_n$ i punktene $x_i$, ser vi at

$$p_n(x_i) - q_n(x_i) = 0 \qquad 0 \leq i \leq n.$$

Men polynomet $p - q$ har maksimal grad $n$, og kan følgelig ha maksimalt $n$ nullpunkter, så den eneste muligheten her er $p = q$, som betyr at interpolasjonspolynomet er entydig. □

**Eksempel 4.68.** En funksjon har følgende verdier:

| $i$ | 0 | 1 | 2 |
|---|---|---|---|
| $x_i$ | 1 | 2 | 3 |
| $f_i$ | 4 | 5 | 6 |

Vi setter opp

$$l_0(x) = \frac{(x-2)(x-3)}{(1-2)(1-3)} = \frac{1}{2}(x-2)(x-3)$$
$$l_1(x) = \frac{(x-1)(x-3)}{(2-1)(2-3)} = -(x-1)(x-3)$$
og
$$l_2(x) = \frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{1}{2}(x-1)(x-2).$$

Det andre ordens polynomet som interpolerer denne tabellen er:

$$\begin{aligned} p(x) &= 4l_0(x) + 5l_1(x) + 6l_2(x) \\ &= 2(x-2)(x-3) - 5(x-1)(x-3) \\ &\quad + 3(x-1)(x-2). \end{aligned}$$ △

I gamle dager sto det i numerikkbøkene at lagrangepolynomene ikke måtte brukes i numeriske beregninger, fordi det koster for mange flyttalsoperasjoner å evaluere dem. Dette er bare tull dersom man bruker de rette interpolasjonspunktene og evaluerer polynomene på rett måte, se Trefethens bok om approksimasjonsteori og approksimasjonspraksis.

## Newtons interpolasjon

For å konstruere Newtons interpolasjon trenger vi å beregne de *dividerte differansene*. De defineres rekursivt:

$$f[x_i] = f(x_i)$$
$$f[x_i, x_{i-1}] = \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}$$
$$f[x_i, ..., x_{i-k}] = \frac{f[x_i, ..., x_{i-k+1}] - f[x_{i-1}, ..., x_{i-k}]}{x_i - x_{i-k}}$$

Newtons interpolasjonspolynom er:

$$p_n(x) = f_0 + \sum_{i=1}^{n} f[x_i, ..., x_0] \prod_{k=0}^{i-1} (x - x_k).$$

Merk også at Lagranges og Newtons interpolasjon er bare to forskjellige formler for å sette opp det samme polynomet, siden interpolasjonspolynomet er entydig.

**Eksempel 4.69.** Polynomet

$$\begin{aligned} p_2(x) = \\ f_0 + \frac{f_1 - f_0}{x_1 - x_0}(x - x_0) \\ + \frac{\frac{f_2 - f_1}{x_2 - x_1} - \frac{f_1 - f_0}{x_1 - x_0}}{x_2 - x_0}(x - x_0)(x - x_1) \end{aligned}$$

interpolerer tabellen

| $x_0$ | $x_1$ | $x_2$ |
|---|---|---|
| $f_0$ | $f_1$ | $f_2$ |

så lenge $x_0 \neq x_1 \neq x_2$. △

Det er vanlig å sette opp følgende tableau for å illustrere de dividerte differansene:

$$\begin{array}{llll} x_0 & f[x_0] & & \\ & & f[x_0, x_1] & \\ x_1 & f[x_1] & & f[x_0, x_1, x_2] \\ & & f[x_1, x_2] & & f[x_0, x_1, x_2, x_3] \\ x_2 & f[x_2] & & f[x_1, x_2, x_3] \\ & & f[x_2, x_3] & \\ x_3 & f[x_3] & & \end{array}$$

**Eksempel 4.70.** La igjen

| $i$ | 0 | 1 | 2 |
|---|---|---|---|
| $x_i$ | 1 | 2 | 3 |
| $f_i$ | 4 | 5 | 6 |

Tableauet blir

$$\begin{array}{llll} 1 & 4 & & \\ & & 1 & \\ 2 & 5 & & 0 \\ & & 1 & \\ 3 & 6 & & \end{array}$$

og interpolasjonspolynomet blir

$$p(x) = 4 + x - 1 = 3 + x.$$

Dette er selvfølgelig det samme polynomet som i forrige eksempel. Merk at polynomet er av første grad, siden punktene tilfeldigvis ligger på en rett linje. △

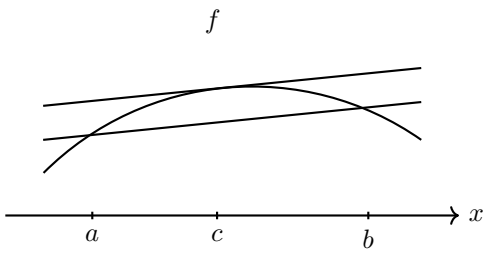
## Interpolasjonsfeilen

Det sier seg selv at et interpolasjonspolynom ikke kan være lik funksjonen som interpoleres med mindre denne funksjonen er et polynom av ikke høyere grad enn interpolanten. Nå er vi kommet til steg to i den generelle oppskriften for numeriske metoder, nemlig analyse av feilen. Da må vi begynne med en generalisering av middelverdisatsen.

Middelverdisatsen sier at

$$f'(c) = \frac{f(b) - f(a)}{b - a} = f[a, b]$$

for en funksjon som er deriverbar på $[a, b]$.




En generalisert variant for dividerte differenser går som følger.

**Teorem 4.71.** *Dersom $f$ er $n+1$ ganger deriverbar på $[a,b]$, og alle punktene $x_i$ er forskjellige, er*

$$f[x_n,...,x_0] = \frac{f^n(s)}{n!}$$

*for en eller annen $s$ i intervallet $(a,b)$.*

*Bevis.* Siden $p_n$ interpolerer $f$, må funksjonen $g = f - p_n$ ha minst $n+1$ nullpunkt på intervallet $(a,b)$. Gjentatt anvendelse av middelverdisatsen forteller oss at $g'$ har minst $n$ nullpunkt, at $g''$ har minst $n-1$ nullpunkt, og videre at $g^{n+1}$ har minst ett nullpunkt på $(a,b)$. Vi kaller dette $s$. Siden

$$\frac{d^n}{dx^n} p_n(x) = n! f[x_n,...,x_0],$$

må

$$f^n(s) = n! f[x_n,...,x_0]. \qquad \square$$

Vi kan nå utlede et uttrykk for interpolasjonsfeilen.

**Teorem 4.72.** *La $f$ være en $n+1$ ganger deriverbar funksjon på $[a,b]$, interpolert i punktene $x_i$. Det finnes en $s \in (a,b)$ slik at*

$$f(x) - p_n(x) = \frac{f^{n+1}(s)}{(n+1)!} \prod_{k=0}^{n} (x - x_k).$$

*Bevis.* Vi skriver opp Newtons interpolasjonspolynom

$$\begin{aligned} p_n(x) = \\ f_0 + \sum_{i=1}^{n-1} f[x_i,...,x_0] \prod_{k=0}^{i-1} (x - x_k) \\ + f[x_n,...,x_0] \prod_{k=0}^{n-1} (x - x_k), \end{aligned}$$

det siste leddet er skrevet ut kun av pedagogiske hensyn. Nå bytter vi ut $x_n$ med $x$ i uttrykket over (tenk på $x$ som et nytt interpolasjonspunkt), og får

$$\begin{aligned} f(x) = \\ f_0 + \sum_{i=1}^{n-1} f[x_i,...,x_0] \prod_{k=0}^{i-1} (x - x_k) \\ + f[x,...,x_0] \prod_{k=0}^{n-1} (x - x_k). \end{aligned}$$

Merk den snedige måten å skrive om $f$ på. Trekker vi de to foregående uttrykkene fra hverandre, får vi

$$\begin{aligned} f(x) - p_n(x) = \\ (f[x,x_{n-1}...,x_0] - f[x_n,x_{n-1},...,x_0]) \prod_{k=0}^{n-1} (x - x_k) \\ = f[x,x_n,...,x_0](x - x_n) \prod_{k=0}^{n-1} (x - x_k) \\ = f[x,x_n,...,x_0] \prod_{k=0}^{n} (x - x_k), \end{aligned}$$

og bruker vi den generaliserte middelverdisatsen over, får vi

$$f(x) - p_n(x) = \frac{f^{n+1}(s)}{(n+1)!} \prod_{k=0}^{n} (x - x_k),$$

for en eller annen $s \in [a,b]$. Merk at $s$ avhenger av $x$, akkurat som i Taylors teorem fra M1. $\square$

## Punktfordeling

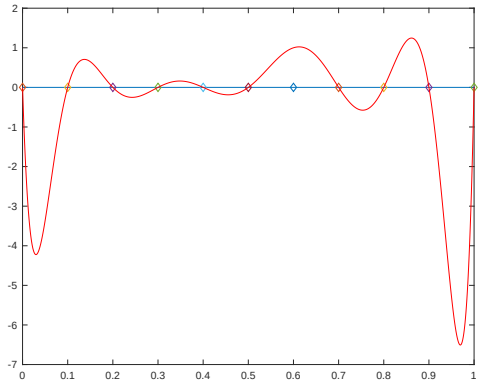
Steg tre i oppskriften på numeriskemetoder, er å finne ut om metoden kan ha noen gode egenskaper utover høy presisjon. I interpolasjonsfaget er det en relativt kjapp måte å avgjøre om en interpolasjonsmetode er god eller ikke: Det er avgjørende for kvaliteten på interpolasjonen at punktene står riktig fordelt på intervallet $[a,b]$.

Men hvordan skal vi finne gode punkter for polynominterpolasjon, og hva skiller de gode fra de dårlige punktene? Dette spørsmålet kan besvares på mange måter, og vi skal vende tilbake til spørsmålet i kapitlet om numerisk integrasjon. La oss begynne med å ta for oss en god og en dårlig punktmengde.

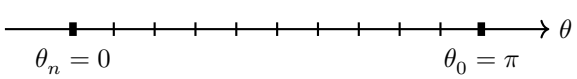
**Eksempel 4.73.** Den dårlige punktmengden er den kjente og kjære *ekvidistante* punktmengden. Et ekvidistant gitter med $n$ punkter på intervallet $[a,b]$ er gitt ved

$$x_i = a + hi$$

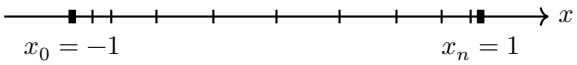
der $0 \le i \le n$ og $h = (b-a)/n$. I figuren under er et plot av en lagrangefunksjon på et ekvidistant gitter med elleve punkt. Merk oscillasjonene polynomet gjør mellom interpolasjonspunktene. $\triangle$



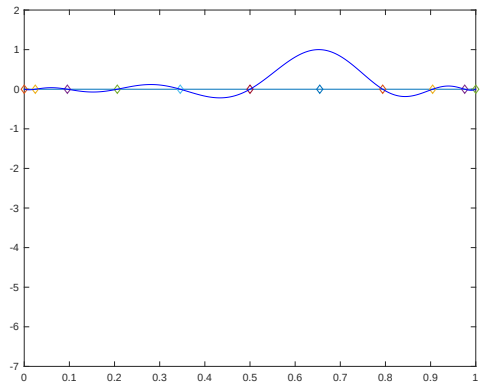

**Eksempel 4.74.** La $\theta_i$ være et ekvidistant gitter på $[0,\pi]$, med $\theta_n = 0$ og $\theta_0 = \pi$.




Nå definerer vi $x_i = \cos\theta_i$. Dette gitteret ligger på $[-1,1]$.



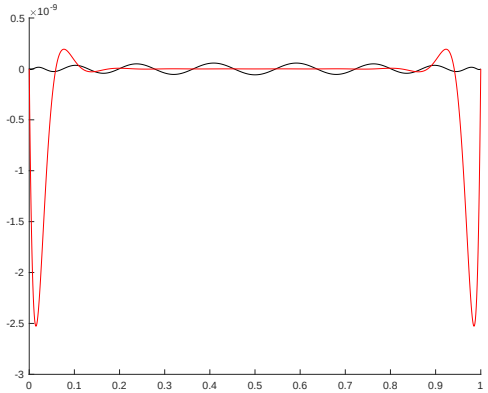

Det gitteret egner seg skikkelig godt for polynominterpolasjon. Under er et tilsvarende plot av en lagrangefunksjon på dette gitteret. Merk hvordan interpolasjonspolynomet ikke oscillerer særlig mellom interpolasjonspunktene. △



Man finner gode punkter for interpolasjon ved å plassere dem slik at de minimerer utslaget til polynomet

$$\prod_{k=0}^{n}(x - x_k)$$

i interpolasjonsfeilen, og alle gode punktmengder for polynominterpolasjon klumper seg i endene av intervallet. For ekvidistante gitre har feilpolynomet stort utslag i endepunktene, og det forklarer hvorfor lagrangefunksjonen i figuren over har størst utslag fra funksjonsverdiene der. Vi skal nå ta for oss noen forskjellige gode punktmengder.



Vi skal gjøre alt på intervallet $[-1, 1]$. Dersom man har en punktfordeling $x_i$ på dette intervallet, kan man flytte fordelingen til intervallet $[a, b]$ med formelen gitt i eksempel 4.77 under.

### Chebyshevpunkter

For å prøve å forklare hva som skjedde i figuren må vi introdusere noen polynomer.

**Teorem 4.75.** *Funksjonen*

$$T_n(x) = \cos(n \arccos x)$$

*er et polynom når $n$ er et naturlig tall.*

*Bevis.* La $T_n(\theta) = \cos(n\theta)$. Da har vi

$$\begin{aligned} T_0 &= 1 \\ T_1 &= \cos\theta \\ T_2 &= \cos 2\theta = 2\cos^2\theta - 1 \end{aligned}$$

Merk at alle disse er polynomer i $\cos\theta$. Dette er en sentral observasjon, for når man gjør variabelskiftet $x = \cos\theta$ vil man få polynomer i $x$.

Vi fortsetter med et induksjonsbevis for at $\cos n\theta$ er et polynom i $\cos\theta$ for alle naturlige tall $n$. Legg sammen

$$\cos(n+1)\theta = \cos n\theta \cos\theta - \sin n\theta \sin\theta$$

og

$$\cos(n-1)\theta = \cos n\theta \cos\theta + \sin n\theta \sin\theta$$

slik at

$$\cos(n+1)\theta = 2\cos n\theta \cos\theta - \cos(n-1)\theta.$$

Dersom vi antar at $\cos n\theta$ og $\cos(n-1)\theta$ er polynomer i $\cos\theta$, må

$$2\cos n\theta \cos\theta - \cos(n-1)\theta$$

være et polynom i $\cos\theta$, og følgelig må også

$$\cos(n+1)\theta$$

være et polynom i $\cos\theta$. Siden $T_0 = 1$ og $T_1 = \cos\theta$ (og $T_2 = 2\cos^2\theta - 1$) er polynomer i $\cos\theta$, må $\cos n\theta$ være det for alle naturlige $n$. Dersom $\cos n\theta$ er et polynom i $\cos\theta$, må $\cos(n \arccos x)$ være et polynom i $x$. □

Polynomene $T_n$ kalles *chebyshevpolynomer*. Disse kan beregnes ved rekursjonen

$$T_{n+1} = 2xT_n - T_{n-1},$$

som er likningen

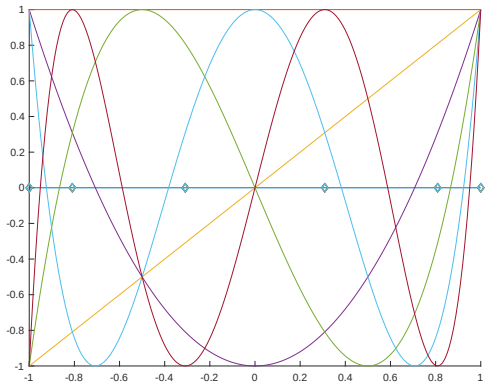
$$\cos(n+1)\theta + \cos(n-1)\theta = 2\cos n\theta \cos\theta.$$

fra forrige bevis, skrevet ut i form av $T_n$. De første er

$$\begin{aligned} T_0 &= 1 \\ T_1 &= x \\ T_2 &= 2x^2 - 1 \\ T_3 &= 4x^3 - 3x \\ T_4 &= 8x^4 - 8x^2 + 1 \\ &\vdots \end{aligned}$$

Gitteret i eksempel 4.74 kalles Chebyshevs ekstremalgitter, for punktene er ekstremalpunktene til et chebyshevpolynom. Det $n$-te ordens polynomet $T_n$ gir opphav til et gitter med $n+1$ punkter, der $n-1$ av dem er $T_n$s stasjonære punkter, og to er endepunktene i intervallet. De første seks polynomene er plottet under, sammen med 6-punktsgitteret som er ekstremalpunktene til $T_5$. En formel for gitterpunktene er:

$$x_i = \cos\frac{\pi i}{n} \quad 0 \le i \le n.$$



**Eksempel 4.76.** Gitteret i figuren over er gitt ved

| | |
|---|---|
| $x_0$ | -1.000000000000000 |
| $x_1$ | -0.809016994374947 |
| $x_2$ | -0.309016994374947 |
| $x_3$ | 0.309016994374947 |
| $x_4$ | 0.809016994374947 |
| $x_5$ | 1.000000000000000 |

Merk at endepunktene er en annen type ekstremalpunkter enn de andre; det er der $T_5$ er klippet av intervallgrensene. △

**Eksempel 4.77.** Hvis vi ønsker å sette opp gitteret fra $T_5$ på et intervall $[a,b]$, bruker vi bare formelen

$$y_i = a + (b-a)\frac{x_i+1}{2},$$

der $x_i$ er punktene på $[-1,1]$ og $y_i$ er punktene på $[a,b]$. På intervallet $[1,4]$ blir punktene

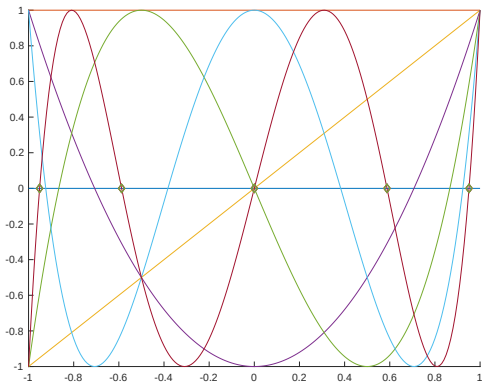
| | |
|---|---|
| $x_0$ | 1.000000000000000 |
| $x_1$ | 1.286474508437579 |
| $x_2$ | 2.036474508437579 |
| $x_3$ | 2.963525491562421 |
| $x_4$ | 3.713525491562421 |
| $x_5$ | 4.000000000000000 |

Det er viktig å forstå at det hvordan punktene er fordelt på intervallet som har noe å si for kvaliteten på interpolasjonen. △

Chebyshevpolynomenes nullpunkter gir opphav til en annen punktmengde som kalles *Chebyshevs nullpunktgitter*. Dette gitteret er på papiret enda bedre enn Chebyshevs ekstremalgitter, men noe mindre praktisk, siden det ikke inneholder endepunktene. Polynomet $T_{n+1}$ gir opphav til et gitter med $n+1$ punkter, gitt ved

$$x_i = \cos\frac{\pi(2i+1)}{2n+2} \quad 0 \le i \le n.$$

Under er et nok et plot av de første par chebyshevpolynomene, med nullpunktgitteret til $T_5$. Vi tar med



et teorem om interpolasjonsfeilen til Chebyshevs nullpunktgitter, men lar det stå ubevist.

**Teorem 4.78.** *La $f$ være en $n+1$ ganger deriverbar funksjon. Interpolasjonsfeil for interpolanten på chebyshevs nullpunktgitter på $[a,b]$ er:*

$$\max_{x\in[a,b]} |f(x) - p_n(x)| \le \frac{(b-a)^{n+1}}{2^{2n+1}} \max_{x\in[a,b]} |f^{n+1}(x)|$$

Under er et plot av feilpolynomet

$$\prod_{k=0}^{n}(x-x_k)$$

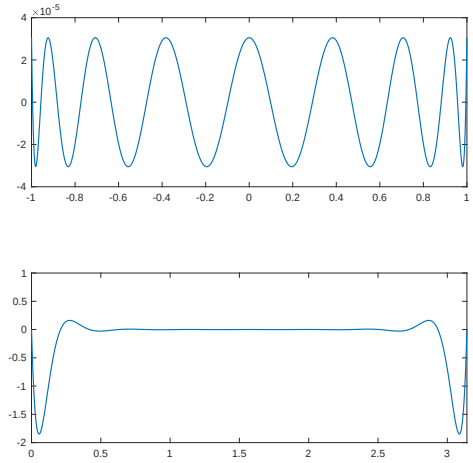
for chebyshevs nullpunktgitter og $n=16$. Dette polynomet har maksimalt utslag

$$\frac{(b-a)^{n+1}}{2^{2n+1}} \approx 1.525878906250000e-05.$$

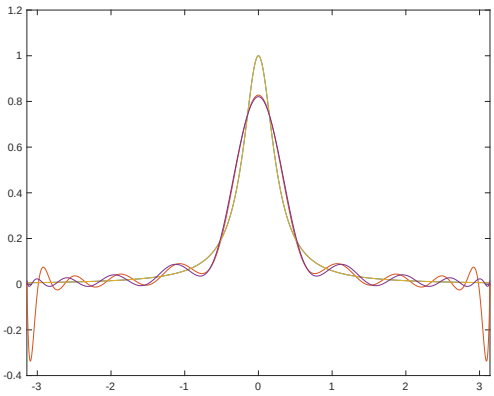
Vi tar med et plot av tilsvarende feilpolynom for ekvidistant gitter med $n=16$.

**Eksempel 4.79.** Nedenfor er en figur av 20. ordens interpolanter av Runges funksjon

$$f(x) = \frac{1}{1+16x^2}$$



på ekvidistant og chebyshevgitter. Denne funksjonen er kjent for sine patologisk store $n$-te deriverte. Dette er en størrelse vi ikke har kontroll på, merk nok en gang hvordan feilpolynomet illustrerer hvorfor ekvidistant tinærmer vesentlig dårligere enn chebyshev i endene. △



### Gauss-punkter

En punktmengde som likner på Chebyshev, og klumper seg i endene av intervallet, kalles Gauss-punkter. Akkurat som for Chebyshev, kommer disse i to varianter, som er henholdsvis null- og ekstremalpunkter til en følge av polynomer, som kalles Gauss-Legendre-polynomene. Legendre-polynomene er gitt ved rekursjonen

$$\begin{aligned} P_0 &= 1 \\ P_1 &= x \\ (n+1)P_{n+1}(x) &= (2n+1)xP_n(x) - nP_{n-1}(x). \end{aligned}$$

Vi skal ikke utlede disse, men i kapitlet om numerisk integrasjon skal vi vise en metode for å produsere nullpuntene.

Nullpunktgitrene til Legendre-polynomene kalles Gauss-Legendre-punkter. Her er en tabell med et par lavere ordens punktfordelinger på intervallet $[-1, 1]$

| $n$ | $x_i$ |
|---|---|
| 2 | $\pm\sqrt{\frac{1}{3}}$ |
| 3 | $0,\ \pm\sqrt{\frac{3}{5}}$ |
| 4 | $\pm\sqrt{\frac{3}{7} \pm \frac{2}{7}\sqrt{\frac{6}{5}}}$ |
| 5 | $0,\ \pm\frac{1}{3}\sqrt{5 \pm 2\sqrt{\frac{10}{7}}}$ |

Gauss-Legendre-punktene er på papiret enda bedre enn chebyshevpunktene, men mindre praktisk i bruk.

Ekstremalpunktene til Legendre-polynomene kalles Gauss-Lobatto-punkter. Vi hoster opp nok en tabell med et par lavere ordens punktfordelinger på intervallet $[-1, 1]$

| $n$ | $x_i$ |
|---|---|
| 3 | $0,\ \pm 1$ |
| 4 | $\pm\sqrt{\frac{1}{5}},\ \pm 1$ |
| 5 | $0,\ \pm\sqrt{\frac{3}{7}},\ \pm 1$ |
| 6 | $\pm\sqrt{\frac{1}{3} \pm \sqrt{\frac{2}{3\sqrt{7}}}},\ \pm 1$ |
| 7 | $0,\ \pm\sqrt{\frac{5}{11} \pm \frac{2}{11}\sqrt{\frac{5}{3}}},\ \pm 1$ |

Gauss-Lobatto er på papiret noe dårligere enn Gauss-Legendre, men er noe mer praktisk i bruk, for de inneholder intervallets endepunkter. Fremdeles mindre praktisk enn Chebyshev.

### Andre typer polynominterpolasjon

Til slutt kan nevnes at man trenger ikke nøye seg med å kreve at interpolanten $p$ skal ta $f$ sine verdier i interpolasjonspunktene. Man kan også skru opp graden på polynomet, og i tillegg kreve at $p$ skal ha samme stigningstall som $f$ i punktene. I dette tilfellet kalles det *Hermite-interpolasjon*.

Dersom man krever at $p$ skal ha samme verdi som $f$ sine $n+1$ første deriverte i et enkelt punkt, er vi tilkake i taylorpolynomene du kjenner fra M1.

Man kan også, istedet for å lage et interpolasjonspolynom som interpolerer $f$ i alle punktene, for eksempel sortere interpolasjonspunktene i grupper på fire og fire etterfølgende punkter, interpolere hver gruppe med et tredjeordens polynom, og så sette sammen en stykkvis kontinuerlig deriverbar polynominterpolant. Dette kalles *spline-interpolasjon*.

# Eulers formel

Med komplekse tall kan man belyse forholdet mellom eksponensialfunksjonen og de trigonometriske funksjonene. Dette forholdet er ikke mulig å få øye på dersom man kun har reelle tall til rådighet.

Brøken

$$\frac{\cos x + i \sin x}{e^{ix}}$$

må være konstant, siden

$$\frac{d}{dx}\left(\frac{\cos x + i\sin x}{e^{ix}}\right) =$$
$$-ie^{-ix}\left(\cos x + i\sin x\right) + e^{-ix}\left(-\sin x + i\cos x\right) = 0$$

og setter vi inn $x = 0$ får vi åpenbart 1, så

$$\frac{\cos x + i\sin x}{e^{ix}} = 1.$$

Richard Feynman kalte bare denne "vår juvel".

**Eulers formel**

$$e^{ix} = \cos x + i\sin x$$

Med Eulers formel kan vi skrive komplekse tall veldig kompakt:

**Polar form**

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta}$$

Hvis vi aksepterer Eulers formel, kan vi sette opp noen pene regneregler:

**Regneregler for polar form**

La $z = re^{i\theta}$ og $w = se^{i\alpha}$.
Da gjelder:

$$z \cdot w = rse^{i(\theta+\alpha)}$$

$$\frac{z}{w} = \frac{r}{s}e^{i(\theta-\alpha)}$$

**Eksempel 4.80.** Polar form er praktisk når man skal gange og dele komplekse tall for hånd. La $z = 1 + i$ og $w = 1 + \sqrt{3}i$, slik at

$$z = \sqrt{2}e^{i\frac{\pi}{4}}$$

og

$$w = 2e^{i\frac{\pi}{3}}.$$

Vi beregner

$$z \cdot w = 2\sqrt{2}e^{i\frac{7\pi}{12}}$$

og

$$\frac{z}{w} = \frac{1}{\sqrt{2}}e^{-i\frac{\pi}{12}}.$$ △

Nå er det imidlertid sjelden at noen ganger sammen komplekse tall for hånd etter at de har landet sin første jobb. Men de geometriske tolkningene som Eulers formel gir oss, og sammenhengen med de trigonometriske funksjonene, er viktig, spesielt om du skal drive på med signalbehandling eller liknende.

**Eksempel 4.81.** Eulers formel gir at $e^{\pi i/2} = i$, $e^{\pi i} = -1$, $e^{3\pi i/2} = -i$ og $e^{2\pi i} = 1$. Merk at det å gange med $i$ er det samme som å rotere et tall nitti grader (pi halve radianer) mot klokken. △

**Eksempel 4.82.** Bytter vi $x$ med $-x$ i Eulers formel, får vi

$$e^{-ix} = \cos x - i\sin x.$$

Hvis vi legger denne sammen med Eulers formel, ser vi at

$$\cos x = \frac{e^{ix} + e^{-ix}}{2}$$

og trekker vi dem fra hverandre, ser vi at

$$\sin x = \frac{e^{ix} - e^{-ix}}{2i}.$$ △

**Eksempel 4.83.** Dersom $z = re^{i\theta}$ gir Eulers formel $\overline{z} = re^{-i\theta}$. △

Husk ellers at

$$\cosh x = \frac{e^{x} + e^{-x}}{2}$$

og

$$\sinh x = \frac{e^{x} - e^{-x}}{2}.$$

Det er nå lett å se at

$$\cosh(ix) = \cos x$$

og

$$\sinh(ix) = i\sin x.$$

Dette kan vi bruke til å finne real- og imaginærdelen til sinus og cosinus:

**Trigonometriske funksjoner på kartesisk form**

$$\cos(a + bi) = \cos a\cosh b - i\sin a\sinh b$$
$$\sin(a + bi) = \sin a\cosh b + i\cos a\sinh b$$

Hvis du plukker opp en tilfeldig bok i algebra eller kompleks analyse, er det bevist følgende teorem et eller annet sted. Teoremet heter algebraens fundamentalteorem.

**Algebraens fundamentalteorem**

Et polynom

$$z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0$$

kan alltid faktoriseres

$$z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0 = \prod_{i=1}^{n}(z - z_i),$$

der $z_i \in \mathbb{C}$ er løsninger av likningen

$$z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0 = 0$$

**Merk.** I teoremet over er polynomet monisk, altså at $a_n = 1$. Det er for å slippe å luke ut tilfellet $a_n = 0$. Dersom $a_n \neq 0$, og noe annet enn 1, blir faktoriseringen

$$a_n z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0 = a_n\prod_{i=1}^{n}(z - z_i).$$

Dersom en faktor $(z - z_k)$ forekommer $m$ ganger i faktoriseringen, sier vi at $z_k$ har *multiplisitet* $m$.

**Eksempel 4.84.** Polynomet

$$z^3 - 3z^2 + 3z - 1 = (z-1)^3$$

har en rot $(z = 1)$ med multiplisitet 3. △

**Eksempel 4.85.** Polynomet

$$z^2 - 2z + 2$$

har to røtter

$$\lambda = \frac{2 \pm \sqrt{4-8}}{2} = 1 \pm i,$$

begge med multiplisitet 1, slik at

$$z^2 - 2z + 2 = (z - 1 - i)(z - 1 + i). \qquad \triangle$$

Vi skal ikke bevise algebraens fundamentalteorem, men et spesialtilfelle kan vi analysere med det vi kjenner til så langt, nemlig løsninger av polynomlikningen

$$z^n = w$$

for et vilkårlig komplekst tall $w$. Vi skal se med egne øyne at denne likningen alltid har $n$ løsninger. Vi begynner med å skrive $w$ på polar form med valgfritt antall omdreininger rundt origo

$$w = re^{i\theta} = re^{i(\theta + 2m\pi)}.$$

Dersom vi skriver

$$w^{1/n} = (re^{i(\theta+2m\pi)})^{1/n} = \sqrt[n]{r}e^{i(\theta/n + 2m\pi/n)},$$

ser vi at det nå finnes $n$ potensielle verdier for $\sqrt[n]{w}$, alle sammen gyldige løsninger av $z^n = w$. Hvis du velger $0 \leq m \leq n-1$ får du ut alle sammen. Vi definerer den prinsipale $n$-te roten av $w$ som

$$\sqrt[n]{w} = \sqrt[n]{r}e^{i\theta/n},$$

og så kan vi skrive de andre røttene som

$$\sqrt[n]{w} \cdot e^{2m\pi i/n}$$

for $1 \leq m \leq n-1$. Dette er analogt til hvordan man i det reelle tilfellet har to løsninger av ligningen

$$x^2 = 4,$$

definerer kvadratroten som den positive løsningen

$$\sqrt{4} = 2,$$

og skriver den andre løsningen som $-\sqrt{4}$.

**Eksempel 4.86.** Vi finner alle løsninger av ligningen

$$z^5 = -1.$$

Siden

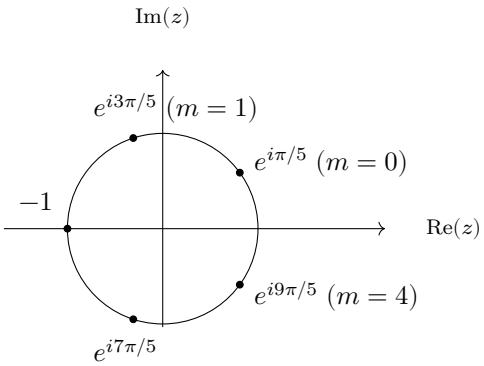
$$-1 = e^{i(\pi + 2m\pi)},$$

får vi

$$(-1)^{1/5} = e^{i(\pi/5 + 2m\pi/5)}.$$

Vi skriver opp løsningene for $0 \leq m \leq 4$:

$e^{i\pi/5} (= \sqrt[5]{-1})$, $e^{i3\pi/5}$, $e^{i5\pi/5} (= -1)$, $e^{i7\pi/5}$ og $e^{i9\pi/5}$




*Femterøttene til -1*

Merk hvordan røttene sprer seg jevnt ut på en sirkel om origo. Merk også at om vi lar $m > 4$ eller $m < 0$, får vi røtter som allerede er listet opp. △

Vi skal få bruk for algebraens fundamentalteorem senere i semesteret, når vi skal lære om egenverdier og egenvektorer.

# Integralet

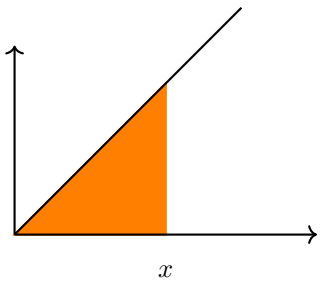
Integrasjon er som ketsjup. Det kan brukes til alt. Vi kan beregne areal, volum, arbeid, kraft, trykk, massesenter, væskeflyt og elektrisk ladning.

La $f$ være en begrenset funksjon på et intervall $[a, b]$. Det store spørsmålet er å finne arealet under grafen til $f$. Det går an å ta en lang diskusjon om hva areal egentlig er, sette opp aksiomer for areal, og så utlede integrasjonsteorien fra dem. Dette er imidlertid en litt langdryg prosess, så vi skal basere relasjonen mellom integralet og arealet under grafen på geometrisk intuisjon.

Alternative kilder:

- Adams kap. 5-7
- Lindstrøms I kap. ??

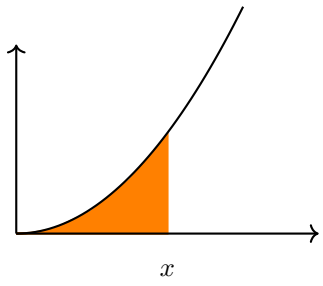
**Eksempel 4.87.** Noen integraler kan vi enkelt beregne geometrisk. La $f(x) = x$. En rask titt på grafen forteller at arealet under denne grafen og mellom 0 og $x$ er $\frac{1}{2}x^2$, siden dette er en rettvinklet trekant der høyde og bredde er $x$. △




**Eksempel 4.88.** La $(x) = x^2$. Arealet under grafen er $\frac{1}{3}x^3$. Dette ble oppdaget geometrisk av Arkimedes for over to tusen år siden. Han oppdaget det ved å dele inn området i bitte små trapeser, og bruke at

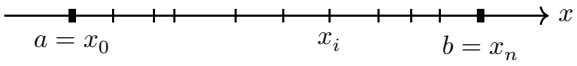
$$1 + 4 + 9 + \cdots = \sum_{i=1}^{n} i^2 = \frac{1}{6}n(n+1)(2n+1).$$

Vi skal gjøre en liknende beregning litt lenger ned. △

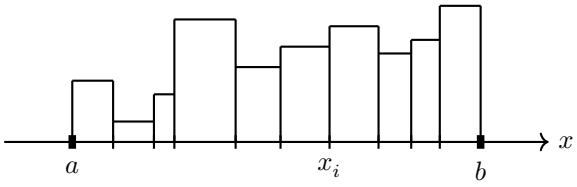



For å bygge opp integralteorien er det essensielt å starte med begensede funksjoner på lukkede intervaller. Det går an å utvide integrasjonsteorien til ubegrensede funksjoner og ubegrensede intervaller, og det skal vi se på så vidt til slutt i kapitlet.

**Definisjon.** En *partisjon* $P$ av intervallet $[a,b]$, er en endelig punktmengde som deler intervallet i mindre biter. Delingspunktene kaller vi $x_i$, der $x_0 = a < x_1 < \cdots < x_{n-1} < x_n = b$.




**Definisjon.** En *stegfunksjon* er en stykkvis konstant funksjon, som tar verdien $f_i$ på intervallet $[x_{i-1}, x_i)$.




Merk at vi kan skrive en stegfunksjon som en lineærkombinasjon av enhetssprangfunksjoner. Dette er som oftest ikke en hensiktsmessig notasjon, og det er enklere å bruke delt forskrift.

$$f(x) = \begin{cases} f_1 & x \in [x_0, x_1) \\ f_2 & x \in [x_1, x_2) \\ \vdots & \\ f_n & x \in [x_{n-1}, x_n] \end{cases}$$

Forbundet med en stegfunksjon er summen

$$s = \sum_{n=1}^{n} (x_i - x_{i-1}) f_i$$

som beskriver arealet under grafen til stegfunksjonen.

**Definisjon.** La $f$ være en begrenset funksjon, og $P$ en partisjon av intervallet $[a,b]$. La

$$m_i = \min_{x \in [x_{i-1}, x_i]} f(x)$$

og

$$M_i = \max_{x \in [x_{i-1}, x_i]} f(x).$$

En *øvre riemannsum* er

$$U(P) = \sum_{n=1}^{n} (x_i - x_{i-1}) M_i,$$

og en *nedre riemannsum* er

$$L(P) = \sum_{n=1}^{n} (x_i - x_{i-1}) m_i.$$

Det er ikke veldig vanskelig å vise at

$$U(P) - L(P) \geq 0$$

for alle partisjoner $P$, og at

$$U(P') \leq U(P) \quad \text{og} \quad L(P') \geq L(P)$$

dersom $P \subset P'$, altså at $P'$ inneholder alle punktene i $P$, og minst ett ekstra punkt. Vi sier da at $P'$ er en finere partisjon enn $P$.

**Definisjon.** La $f$ være en begrenset funksjon på $[a,b]$. Dersom det for hver $\epsilon > 0$ finnes en partisjon $P$ av intervallet $[a,b]$ slik at

$$U(P) - L(P) < \epsilon,$$

sier vi at $f$ er integrerbar.

Dersom $f$ er begrenset, er det klart at mengden av alle nedre riemannsummer må ha en minste øvre skranke, og at mengden av alle nedre riemannsummer må ha en største nedre skranke. Dersom $f$ er integrerbar, er det heller ikke så vanskelig å se at disse må være like. Dette tallet skriver vi

$$\int_a^b f(x)\, dx,$$

og vi har at

$$L(P) \leq \int_a^b f(x)\, dx \leq U(P)$$

for alle partisjoner $P$. Noen ganger ønsker man å uttrykke seg mer konsist, og da skriver man bare

$$\int_a^b f.$$

Det kan være ganske vanskelig å avgjøre hvilke funksjoner som er integrerbare direkte fra definisjonen, men noen eksempler er mulig å regne ut.

**Eksempel 4.89.** La $f$ være funksjonen

$$f(x) = \begin{cases} 1 & \text{rasjonale } x \\ 0 & \text{irrasjonale } x \end{cases}$$

Denne er diskontinuerlig overalt, og ikke integrerbar. Siden $m_i = 0$ og $M_i = 1$ for alle $i$ og alle partisjoner, er det klart at det ikke finnes partisjoner slik at

$$U(P) - L(P) < \epsilon$$

dersom $\epsilon < 1$. Dette eksemplet kan fremstå som noe patologisk, men vi skal se i et senere kapittel at denne funksjonen kan konstrueres ved hjelp av cosinusfunksjonen og to grenseverdiprosesser. △

Dette er ikke mulig å gjøre for mange funksjoner, og i neste avsnitt skal vi se at vi kan bruke antiderivasjon istedet. Men det er viktig å forstå riemannsummer når man skal igang med fysiske anvendelser av integrasjon, så vi tar med et eksempel.

**Eksempel 4.90.** Vi beregner

$$\int_0^b x\,dx.$$

La oss ta en jevn partisjon, der punktene er gitt ved:

$$x_k = \frac{bk}{n}.$$

Vi beregner en typisk nedre riemannsum:

$$\begin{aligned}\frac{b}{n}\sum_{k=0}^{n-1} x_k &= \frac{b}{n}\sum_{k=0}^{n-1}\frac{bk}{n} = \frac{b^2}{n^2}\sum_{k=0}^{n-1} k \\ &= \frac{b^2}{n^2}\frac{n(n-1)}{2} = \frac{b^2}{2}\frac{(n-1)}{n}\end{aligned}$$

Her er en øvre:

$$\begin{aligned}\frac{b}{n}\sum_{k=1}^{n} x_k &= \frac{b}{n}\sum_{k=1}^{n}\frac{bk}{n} = \frac{b^2}{n^2}\sum_{k=1}^{n} k \\ &= \frac{b^2}{n^2}\frac{n(n+1)}{2} = \frac{b^2}{2}\frac{(n+1)}{n}\end{aligned}$$

Vi kan nå sjekke at $f(x) = x$ er integrerbar. Velg $\epsilon > 0$. En rask kikk på

$$U(P) - L(P) = \frac{b^2}{2}\left(\frac{(n+1)}{n} - \frac{(n-1)}{n}\right) = \frac{b^2}{n}$$

forteller at dersom vi velger

$$n > \frac{b^2}{\epsilon}$$

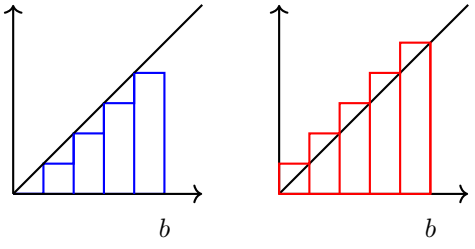
har vi

$$U(P) - L(P) < \epsilon,$$

og siden

$$\frac{b^2}{2}\lim_{n\to\infty}\frac{n+1}{n} = \frac{b^2}{2}\lim_{n\to\infty}\frac{n-1}{n} = \frac{b^2}{2}$$

ser det ut til at integralet blir $\frac{b^2}{2}$. △




**Eksempel 4.91.** Vi beregner

$$\int_0^b x^2\,dx.$$

på samme vis. La partisjonen være

$$x_k = \frac{bk}{n}.$$

Vi beregner en typisk nedre riemannsum:

$$\begin{aligned}\frac{b}{n}\sum_{k=0}^{n-1} x_k^2 &= \frac{b}{n}\sum_{k=0}^{n-1}\frac{b^2k^2}{n^2} = \frac{b^3}{n^3}\sum_{k=0}^{n-1} k^2 \\ &= \frac{b^3}{n^3}\frac{(n-1)n(2n-1)}{6} \\ &= \frac{b^3}{6}\frac{2n^3-3n^2+n}{n^3}\end{aligned}$$

Her er en øvre:

$$\begin{aligned}\frac{b}{n}\sum_{k=1}^{n} x_k^2 &= \frac{b}{n}\sum_{k=1}^{n}\frac{b^2k^2}{n^2} = \frac{b^3}{n^3}\sum_{k=1}^{n} k^2 \\ &= \frac{b^3}{n^3}\frac{n(n+1)(2n+1)}{6} \\ &= \frac{b^3}{6}\frac{2n^3+3n^2+n}{n^3}\end{aligned}$$

Vi kan nå sjekke at $f(x) = x$ er integrerbar. Velg $\epsilon > 0$. En rask kikk på

$$\begin{aligned}U(P) - L(P) &= \frac{b^3}{6}\left(\frac{2n^3+n^2+n}{n^3} - \frac{2n^3-n^2+n}{n^3}\right) \\ &= \frac{b^2}{6}\frac{2n^2}{n^3} = \frac{b^2}{3n}\end{aligned}$$

forteller at dersom vi velger

$$n > \frac{b^2}{3\epsilon}$$

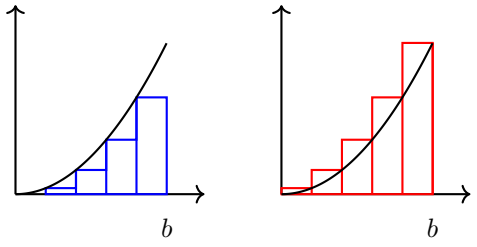
har vi

$$U(P) - L(P) < \epsilon,$$

og siden

$$\lim_{n\to\infty}\frac{b^3}{6}\frac{2n^3+n^2+n}{n^3} = \lim_{n\to\infty}\frac{b^3}{6}\frac{2n^3-n^2+n}{n^3} = \frac{b^3}{3}$$

ser det ut til at integralet blir $\frac{b^3}{3}$. △




Nå er det viktig å ikke fortvile. Det er ingen som beregner integraler på denne måten, men beregningen over er interessant, fordi den ble gjort av Arkimedes omtrent to tusen år før Newton og Leibniz oppfant integralregningen. Nå kommer en serie med små teoremer.

**Teorem 4.92.** *For integralet gjelder følgende:*

1 *Dersom $f$ er begrenset og stykkvis kontinuerlig på $[a,b]$ med et endelig antall diskontinuiteter, er $f$ integrerbar på $[a,b]$.*

2 *Dersom $f$ er kontinuerlig på $[a,b]$, finnes en $c \in [a,b]$ slik at*

$$f(c) = \frac{1}{b-a}\int_a^b f\,ds.$$

3 *Dersom $f_1$ og $f_2$ er integrerbare funksjoner, og $c_1$ og $c_2$ vilkårlige konstanter, er*

$$\int_a^b c_1 f_1(x) + c_2 f_2(x)\,dx = c_1\int_a^b f_1(x)\,dx + c_2\int_a^b f_2(x)\,dx.$$

4 *Dersom $f_1$ og $f_2$ er integrerbare funksjoner og $f_1 \leq f_2$, er*

$$\int_a^b f_1(x)\,dx \leq \int_a^b f_2(x)\,dx.$$

5 *Dersom $|f| \leq M$ på $[a,b]$, er*

$$\left|\int_a^b f\,dx\right| \leq M(b-a).$$

6 *Dersom $f$ er integrerbar, er $|f|$ integrerbar, og*

$$\left|\int_a^b f\,dx\right| \leq \int_a^b |f|\,dx.$$

7 *Dersom $f$ er integrerbar og $c \in (a,b)$, er*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

8 *Dersom $f$ er integrerbar, er*

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

Bevisene for disse teoremene er enten så lette at de ikke er spesielt interessante å lese, eller så vanskelige at vi ikke kan gjøre dem i dette kurset.

Egenskap 1 er ikke mulig å vise med det vi har gjort til nå i kurset. Man må bevise at en begrenset og kontinuerlig funksjon er såkalt uniformt kontinuerlig, og ta det derfra. En funksjon er uniformt kontinuerlig på $[a,b]$ dersom det for hver $\epsilon > 0$ finnes en $\delta$ slik at

$$0 < |x-y| < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \epsilon$$

for alle $x, y \in [a,b]$.

Egenskap 2 er en versjon av sekantsetningen, men for integraler. Den kalles gjerne *middelverdisatsen*. Navnet kommer av at uttrykket

$$\frac{1}{b-a}\int_a^b f\,ds$$

gir middelverdien $f$, i den forstand at dersom $f$ byttes ut med en konstant funksjon med denne verdien, får integralet den samme verdien. Middelverdisatsen sier at $f$ må innom denne verdien på vei fra $a$ til $b$. Denne er ikke så vanskelig å bevise, og følger av skjæringssetningen. Prøv selv!

Egenskap 3-7 følger av en serie litt kjedelige resonnementer basert på riemannsummer. Hvis du plukker opp en tilfeldig bok i envariabel funksjonsteori, vil du finne dem der.

Egenskap 8 handler egentlig om konvensjon. Dersom vi tillater $b < a$ i definisjonen av partisjoner, blir

$$x_i - x_{i-1} < 0$$

for alle $i$, slik at riemannsummene bytter fortegn i forhold til $f$. Dette kan vi bruke til å definere presist hva vi mener med

$$\int_a^b f(x)\,dx$$

når $b < a$. Det blir da relativt enkelt å bevise egenskap 8.

# Integrasjon og derivasjon

**Teorem 4.93.** *Dersom $f$ er integrerbar på $[a,b]$, er*

$$F(x) = \int_a^x f(s)\,ds$$

*en kontinuerlig funksjon på $[a,b]$.*

*Bevis.* Velg $\epsilon > 0$, og anta $|f| \leq M$. Da er

$$|F(x)| \leq M(b-a)$$

og følgelig er

$$|F(x) - F(y)| \leq \left|\int_x^y f\,ds\right| \leq M|x-y|.$$

Dersom vi velger

$$|x-y| < \frac{\epsilon}{M},$$

er

$$|F(x) - F(y)| \leq M|x-y| < \epsilon,$$

slik at $F$ er kontinuerlig i $x$. □

Integrasjon og derivasjon er på sett og vis inverse operasjoner, og det finnes en stor klasse av integrerbare funksjoner som har en antiderivert. Når man lærer integrasjon på skolen, blir man egentlig oppdratt til å tro at integrasjon er det samme som antiderivasjon. Dette er et forenklet bilde av virkeligheten, for integrasjon er mye mer enn bare antiderivasjon. Men det

er allikevel praktisk å kunne beregne integraler ved å antiderivere.

**Teorem 4.94.** *La $f$ være en kontinuerlig i $x = c \in [a, b]$. Da er*

$$F(x) = \int_a^x f(s)\, ds$$

*deriverbar i $x = c$, og*

$$F'(c) = f(c).$$

*Bevis.* Vi beregner

$$\begin{aligned} F'(x) &= \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} \int_a^{x+h} f(s)\, ds - \int_a^x f(s)\, ds \\ &= \lim_{h\to 0} \frac{1}{h} \int_x^{x+h} f(s)\, ds \\ &= \lim_{h\to 0} f(c) = f(x) \end{aligned}$$

siden det for hver $h$ må finnes en $c \in [x, x+h]$ slik at

$$\frac{1}{h} \int_x^{x+h} f(s)\, ds = f(c). \qquad \square$$

Det neste teoremet kalles gjerne *analysens fundamentalteorem*.

**Teorem 4.95.** *La $f$ være integrerbar på $[a, b]$. Dersom det finnes en funksjon $F$ slik at $F' = f$, er*

$$\int_a^b f(x)\, dx = F(b) - F(a).$$

*Bevis.* Velg en partisjon $P$.Hvis vi bruker sekantsetningen på hvert delintervall, kan vi for hvert intervall finne en $y_i$ slik at

$$F(x_i) - F(x_{i-1}) = f(y_i)((x_i - x_{i-1}),$$

og dersom vi legger sammen alle disse likningene, får vi

$$F(b) - F(a) = \sum_i f(y_i)((x_i - x_{i-1}).$$

Siden $m_i \le f(y_i) \le M_i$ for alle $i$, må

$$L(P) \le \sum_i f(y_i)((x_i - x_{i-1}) \le U(P)$$

og følgelig er også

$$L(P) \le F(b) - F(a) \le U(P).$$

Dette gjelder altså for alle partisjoner $P$. Nå er det kun ett tall som er større enn $L(P)$ og mindre enn $U(P)$ for alle $P$, nemlig

$$\int_a^b f(x)\, dx$$

så dersom $F(b)-F(a)$ også skal ha denne egenskapen, må

$$F(b) - F(a) = \int_a^b f(x)\, dx. \qquad \square$$

**Eksempel 4.96.** Siden polynomer er kontinuerlige, må åpenbart

$$\int_a^b x^n\, dx = \frac{b^{n+1} - a^{n+1}}{n+1},$$

siden

$$\frac{d}{dx} \frac{x^{n+1}}{n+1} = x^n.$$

Dette gjelder også for $n \in \mathbb{R}$, men dette er litt mer jobb å vise. △

**Eksempel 4.97.** Siden

$$\frac{d}{dx} e^x = e^x,$$

må

$$\int_a^b e^x\, dx = e^b - e^a. \qquad \triangle$$

**Eksempel 4.98.** Det finnes funksjoner som ikke har noen antiderivert som er enkel å skrive opp, for eksempel

$$f(x) = \exp(x^2),$$

$$g(x) = \frac{\sin x}{x},$$

og

$$h(x) = \sqrt{1 + x^4}.$$

Disse er integrerbare på alle lukkede intervaller. De antideriverte kan skrives som uendelige rekker, som vi skal se i kapitlet om rekkeutvikling. △

Vi tar med to regneregler for integrasjon som er umiddelbare konsekvenser av analysens fundamentalteorem.

**Teorem 4.99.** *La $F = f'$ og $G = g'$ på $[a, b]$. Dersom $f$ og $g$ er integrerbare, er*

$$\int_a^b F(x)g(x)\, dx = F(b)G(b) - F(a)G(a) - \int_a^b f(x)G(x)\, dx.$$

*Bevis.* Dette følger av analysens fundamentalteorem og produktregelen for derivasjon:

$$\begin{aligned} \frac{d}{dx} F(x)G(x) &= F(x)G'(x) + F'(x)G(x) \\ &= F(x)g(x) + f(x)G(x). \end{aligned} \qquad \square$$

**Teorem 4.100.** *Anta at $f$ er kontinuerlig på $[c,d]$, at $F'=f$ på $[c,d]$ og at $u$ er deriverbar på $[a,b]$, med $u(a)=c$ og $u(b)=d$. Da er*

$$\begin{aligned} F(u(b))-F(u(a)) &= \int_a^b f(u(x))u'(x)\,dx \\ &= \int_c^d f(u)\,du. \end{aligned}$$

*Bevis.* Dette følger av analysens fundamentalteorem og kjerneregelen for derivasjon:

$$\frac{d}{dx}F(u(x)) = f(u(x))u'(x). \qquad \square$$

Til slutt tar vi med en variant som kalles *taylors teorem*.

**Teorem 4.101.** *La $f$ være en $n+1$ ganger kontinuerlig deriverbar funksjon på et intervall som inneholder $a$ og $x$. Da er*

$$\begin{aligned} f(x) = {} & \\ & f(a)+f'(a)(x-a)+\cdots+\frac{f^n(a)}{n!}(x-a)^n \\ & +\frac{1}{n!}\int_a^x (x-s)^n f^{n+1}(s)\,ds \end{aligned}$$

**Eksempel 4.102.** En ellipse med halvakser $a$ og $b$ er mengden av alle punkter som tilfredsstiller likningen

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

Vi kan bruke integrasjon til å vise at arealet til ellipsen er gitt ved $\pi ab$. Den øvre delen av ellipsen er grafen til funksjonen $f:[-a,a]\to\mathbb{R}$

$$f(x)=b\sqrt{1-\frac{x^2}{a^2}}$$

og ved hjelp av substitusjonen $x=a\sin\theta$ kan vi beregne arealet:

$$\begin{aligned} A &= 4b\int_0^a \sqrt{1-\frac{x^2}{a^2}}\,dx \\ &= 4b\int_0^{\pi/2} i\sqrt{1-\frac{a^2\sin^2\theta}{a^2}}a\cos\theta\,d\theta \\ &= 4ab\int_0^{\pi/2}\sqrt{1-\sin^2\theta}\cos\theta\,d\theta \\ &= 4ab\int_0^{\pi/2}\cos^2\theta\,d\theta \\ &= 4ab\int_0^{\pi/2}\frac{1+\cos 2\theta}{2}\,d\theta = \pi ab \end{aligned}$$

Dersom $a=b=r$ er dette en sirkel med radius $r$, og arealet blir $\pi r^2$. $\triangle$

# Uegentlige integral

Et integral definerer en funksjon

$$F(x)=\int_a^x f$$

og det er ingenting i veien for å beregne

$$\lim_{x\to\infty}F(x)=\int_a^x f.$$

Dersom

$$\lim_{x\to c}f(x)=\infty,$$

kan vi også beregne

$$\lim_{x\to c}\int_a^x f.$$

Disse kalles *uegentlige integraler*. Dersom grenseverdiene eksisterer, sier vi at integralene konvergerer.

**Eksempel 4.103.**

$$\int_1^\infty \frac{1}{x}\,dx=\lim_{a\to\infty}\int_1^a\frac{1}{x}\,dx=\lim_{a\to\infty}\log a=\infty \ \triangle$$

**Eksempel 4.104.**

$$\int_1^\infty \frac{1}{x^2}\,dx=\lim_{a\to\infty}\int_1^a\frac{1}{x^2}\,dx=\lim_{a\to\infty}1-\frac{1}{a}=1\triangle$$

**Eksempel 4.105.**

$$\int_0^1 \frac{1}{x}\,dx=\lim_{a\to 0^+}\int_a^1\frac{1}{x}\,dx=\lim_{a\to 0^+}-\log a=\infty\triangle$$

**Eksempel 4.106.**

$$\int_0^1 \frac{1}{x^2}\,dx=\lim_{a\to 0^+}\int_a^1\frac{1}{x^2}\,dx=\lim_{a\to 0^+}\frac{1}{a}-1=\infty\triangle$$

**Eksempel 4.107.**

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx=\lim_{a\to 0^+}\int_a^1\frac{1}{\sqrt{x}}\,dx=\lim_{a\to 0^+}2-2\sqrt{a}=2\triangle$$

**Definisjon.** Vi sier at $\int_a^b f$ er absolutt konvergent dersom $\int_a^b |f|$ er konvergent.

**Eksempel 4.108.** Integralet

$$\int_1^\infty \frac{\sin x}{x^2}\,dx$$

er absolutt konvergent, siden

$$\int_1^\infty \left|\frac{\sin x}{x^2}\right|\,dx \le \int_1^\infty \frac{1}{x^2}\,dx = 1. \qquad \triangle$$

Uegentlige integraler åpner opp for noe som kalles integraltesten. Det er en test for å avgjøre om et integral og en korresponderende rekke konvergerer.

**Teorem 4.109.** *La $f$ være en begrenset funksjon. Integralet*

$$\int_a^\infty f(x)\,dx$$

*og rekken*

$$\sum_{n=a}^{\infty} f(n)$$

*er enten begge konvergente, eller begge divergente.*

**Eksempel 4.110.** Nå er det litt enklere å avgjøre om for eksempel

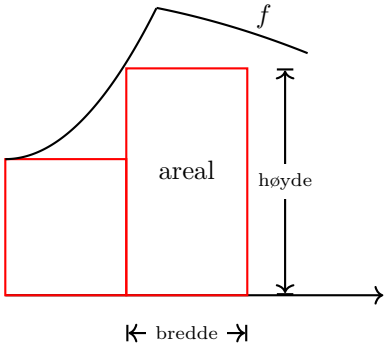
$$\sum \frac{1}{1+n^2}$$

konvergerer, for vi kan beregne

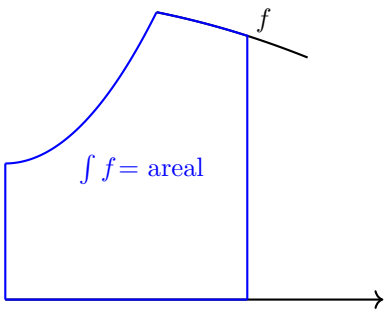
$$\int_0^\infty \frac{1}{1+x^2}\,dx = \lim_{x\to\infty} \arctan x - \arctan 0 = \frac{\pi}{2}. \quad \triangle$$

# Volum

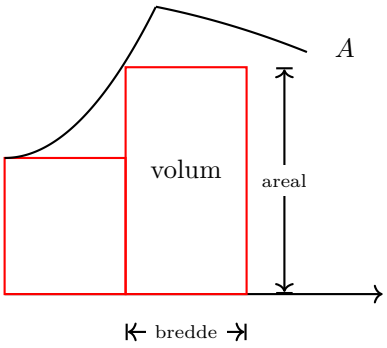
Til nå har vi brukt integralet til å beregne areal. I et bestemt rektangel i en riemannsum, representerer bredden bredden i et rektangel, og høyden høyden i et rektangel. Arealet til rektangel representerer da nettopp arealet til et rektangel:




Dermed vil $\int f$ representere et areal:




Men en av integralets mange styrker, er at de to lengdene som har representert høyde og bredde til nå, fint kan representere helt andre ting. I dette avsnittet skal vi tenke at man har en funksjon $A : [a, b] \to \mathbb{R}$ der funksjonsverdiene beskriver areal, og $x$-aksen beskriver lengde. Da vil både riemannsummene og $\int A$ representere volum:




**Eksempel 4.111.** En pyramide har et kvadratisk tverrsnitt med sidekant gitt ved

$$s(x) = b - x,$$

der $x$ er høyden til tverrsnittet, og $b$ er den totale høyden til pyramiden. Arealet til tversnittet er

$$A(x) = (b-x)^2,$$

og volumet blir

$$V = \int_0^b (b-x)^2\,dx = \left(-\frac{(b-x)^3}{3}\right)_0^b = \frac{b^3}{3}. \quad \triangle$$

Et av de enkleste eksemplene å visualisere, kalles omdreiningslegemer. Anta at du har en funksjon $f : [a, b] \to \mathbb{R}$, og dreier grafen en gang rundt $x$-aksen. Da vil du få et legeme med sirkulært tverrsnitt. Arealet av tverrsnittet for en gitt $x$-verdi er gitt ved

$$A(x) = \pi\left(f(x)\right)^2 = \pi f^2(x).$$

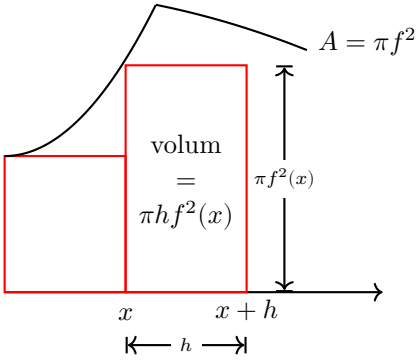
Man kan da tenke at dette er høyden i en riemannsum. Et rektangel i en riemannsum blir

$$\pi h f^2(x),$$

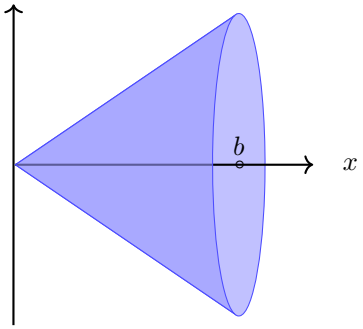
og dette representerer nå et volum. Totalvolumet til omdreiningslegemet er gitt ved

$$V = \pi \int_a^b f^2(x)\,dx.$$

Her er en figur som illustrerer riemannsummene:




For å illustrere hvordan volumet kan se ut i virkeligheten, er her omdreingingslegemet som fremkommer ved å dreie funksjonen $f(x) = x$ på intervallet $[0, b]$:




**Eksempel 4.112.** Omdreiningslegmet i figuren over har volum gitt ved

$$V = \pi \int_0^b x^2 \, dx = \frac{\pi b^3}{3}.$$

Dette kjenner vi igjen som volumet til en kjegle. Alle pyramider og kjegler har volum på formen $\frac{1}{3}Ah$, er $h$ er høyden, og $A$ er grunnflatens areal. △

**Eksempel 4.113.** Et artig eksempel kalles *Gabriels trompet*. Vi roterer funksjonen $f : [1, \infty] \to \mathbb{R}$ gitt ved

$$f(x) = \frac{1}{x}$$

om $x$-aksen. Volumet av dette legemet er

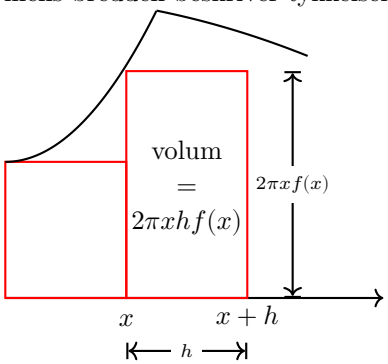
$$V = \pi \int_1^\infty \frac{1}{x^2} \, dx = \pi.$$

Eksemplet kalles Gabriels trompet fordi rotasjonslegemet ser ut som en uendelig lang trompet. △

Dersom vi roterer $f : [a, b] \to \mathbb{R}$ rundt $y$-aksen, får vi et annet omdreiningslegeme, hvis areal kan beregnes ved noe som kalles sylinderskallmetoden. Man tenker at høyden i riemannsummen beskriver arealet av et sylinderskall

$$A(x) = 2\pi x f(x)$$

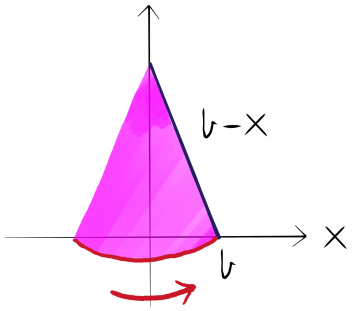
mens bredden beskriver tykkelsen av sylinderskallet:




Volumet blir

$$V = 2\pi \int_a^b x f(x) \, dx.$$

Her er en illustrasjon av hvordan dette ser ut. Den roterte funksjonen er $f(x) = b - x$ på intervallet $[0, b]$.



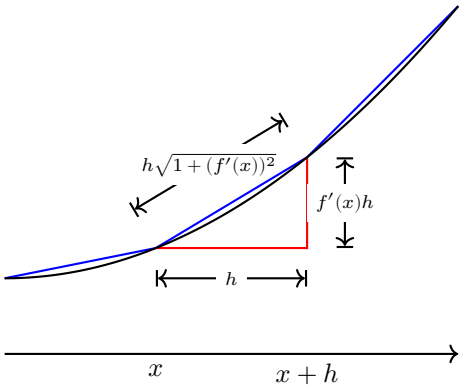

**Eksempel 4.114.** Volumet i figuren over blir

$$\begin{aligned} V &= 2\pi \int_0^b x f(x) \, dx \\ &= 2\pi \int_0^b x(b - x) \, dx \\ &= 2\pi \left( \frac{bx^2}{2} - \frac{x^3}{3} \right)_0^b \\ &= 2\pi \left( \frac{b^3}{2} - \frac{b^3}{3} \right) = \frac{\pi b^3}{3}. \end{aligned}$$

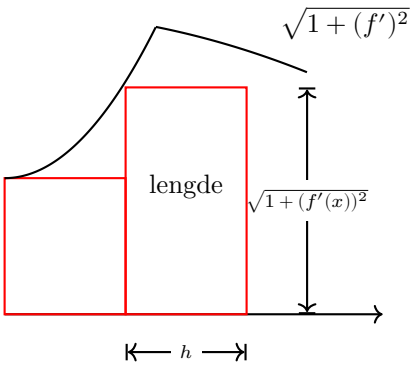
△

# Buelengde

Vi kan bruke integralet til å beregne lengden til kurven gitt ved $y = f(x)$ der $f : [a, b] \to \mathbb{R}$. Trikset er å se på en kurve bestående av sekantbiter til $f$:




Hvis du tenker riemannsummer, blir den korresponderende figuren slik:



Lengden av kurven fra $(a, f(a))$ til $(b, f(b))$ er

$$\int_a^b \sqrt{1 + (f'(x))^2}\, dx$$

**Eksempel 4.115.** La $b > 0$. Vi beregner lengden til kurven $y = x^2$, fra $(0, 0)$ til $(b, b^2)$:

$$\int_0^b \sqrt{1 + 4x^2}\, dx = 2\int_0^b \sqrt{\frac{1}{4} + x^2}\, dx = b\sqrt{\frac{1}{4} + b^2} + \frac{1}{4}\left(\ln\left(b + \sqrt{\frac{1}{4} + b^2}\right) + \ln 2\right).$$

Dette integralet måtte jeg slå opp i en tabell. Du trenger ikke huske det til eksamen. Å finne buelengden til selv enkle funksjoner, kan være ganske hardt. Slik er livet. △

# Numerisk integrasjon

I mange situasjoner er det enten umulig eller urealistisk å antiderivere for å finne integralet. Funksjonen $e^{-x^2}$ har ingen antiderivert som er enkel å evaluere, mens andre funksjoner har så kompliserte antiderivater at det ikke er vits i å prøve en gang:

$$\int \frac{\log(\cos(x))}{\sin(2x)}\, dx =$$

$$-\Big(\Big(\log(\cos(x))\Big(2\,\mathrm{Li}_2\Big(\frac{1}{2}\Big(1-\tan\Big(\frac{x}{2}\Big)\Big)\Big)-2\,\mathrm{Li}_2\Big(\frac{1}{2}\Big((1+i)-(1-i)\tan\Big(\frac{x}{2}\Big)\Big)\Big)+$$
$$2\,\mathrm{Li}_2\Big(-\tan\Big(\frac{x}{2}\Big)\Big)-2\,\mathrm{Li}_2\Big(-i\tan\Big(\frac{x}{2}\Big)\Big)-2\,\mathrm{Li}_2\Big(i\tan\Big(\frac{x}{2}\Big)\Big)+2\,\mathrm{Li}_2\Big(\tan\Big(\frac{x}{2}\Big)\Big)-$$
$$2\,\mathrm{Li}_2\Big(\Big(-\frac{1}{2}-\frac{i}{2}\Big)\Big(\tan\Big(\frac{x}{2}\Big)+i\Big)\Big)+2\,\mathrm{Li}_2\Big(\frac{1}{2}\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)\Big)-$$
$$2\,\mathrm{Li}_2\Big(\frac{1}{2}\Big((1-i)\tan\Big(\frac{x}{2}\Big)+(1+i)\Big)\Big)-2\,\mathrm{Li}_2\Big(\frac{1}{2}\Big((1+i)\tan\Big(\frac{x}{2}\Big)+(1-i)\Big)\Big)+$$
$$\log^2\Big(1-\tan\Big(\frac{x}{2}\Big)\Big)+\log^2\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)+$$
$$\log\Big(\tan\Big(\frac{x}{2}\Big)-i\Big)\log\Big(\tan^2\Big(\frac{x}{2}\Big)-1\Big)+\log\Big(\tan\Big(\frac{x}{2}\Big)+i\Big)\log\Big(\tan^2\Big(\frac{x}{2}\Big)-1\Big)+$$
$$4\log\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)\log\Big(1-\tan\Big(\frac{x}{2}\Big)\Big)-\log(4)\log\Big(1-\tan\Big(\frac{x}{2}\Big)\Big)-$$
$$2\log\Big(1-i\tan\Big(\frac{x}{2}\Big)\Big)\log\Big(\tan\Big(\frac{x}{2}\Big)\Big)-2\log\Big(1+i\tan\Big(\frac{x}{2}\Big)\Big)\log\Big(\tan\Big(\frac{x}{2}\Big)\Big)-$$
$$2\log\Big(\Big(-\frac{1}{2}-\frac{i}{2}\Big)\Big(\tan\Big(\frac{x}{2}\Big)-1\Big)\Big)\log\Big(\tan\Big(\frac{x}{2}\Big)-i\Big)-$$
$$2\log\Big(\Big(-\frac{1}{2}+\frac{i}{2}\Big)\Big(\tan\Big(\frac{x}{2}\Big)-1\Big)\Big)\log\Big(\tan\Big(\frac{x}{2}\Big)+i\Big)-$$
$$2\log\Big(\tan\Big(\frac{x}{2}\Big)-i\Big)\log\Big(\Big(\frac{1}{2}-\frac{i}{2}\Big)\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)\Big)-$$
$$2\log\Big(\tan\Big(\frac{x}{2}\Big)+i\Big)\log\Big(\Big(\frac{1}{2}+\frac{i}{2}\Big)\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)\Big)-$$
$$\log(4)\log\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)+\log\Big(\tan^2\Big(\frac{x}{2}\Big)-1\Big)\log\Big(\cos^2\Big(\frac{x}{2}\Big)\Big)-$$
$$2\log\Big(\tan\Big(\frac{x}{2}\Big)\Big)\log\Big(\cos^2\Big(\frac{x}{2}\Big)\Big)+\log\Big(\cos^2\Big(\frac{x}{2}\Big)\Big)\log\Big(\cos(x)\sec^2\Big(\frac{x}{2}\Big)\Big)+$$
$$\log\Big(\tan\Big(\frac{x}{2}\Big)-i\Big)\log\Big(\cos(x)\sec^2\Big(\frac{x}{2}\Big)\Big)+$$
$$\log\Big(\tan\Big(\frac{x}{2}\Big)+i\Big)\log\Big(\cos(x)\sec^2\Big(\frac{x}{2}\Big)\Big)\Big)\Big)\Big/$$
$$\Big(4\Big(\log\Big(1-\tan\Big(\frac{x}{2}\Big)\Big)+\log\Big(\tan\Big(\frac{x}{2}\Big)+1\Big)+\log\Big(\cos^2\Big(\frac{x}{2}\Big)\Big)\Big)\Big)+\text{constant}$$

**Eksempel 4.116.** Man kan bruke riemannsummer til å tilnærme integraler. Dersom en $f$ er strengt voksende på intervallet $[a, b]$, og punktene i en jevn partisjon $P$ gitt ved

$$a + \frac{b-a}{n}i$$

der $0 \le i \le n$, er øvre riemannsum gitt ved

$$U(P) = \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

og nedre riemannsum gitt ved

$$L(P) = \frac{1}{n}\sum_{i=0}^{n-1} f(x_i).$$

Dersom $f$ ikke er strengt voksende på $[a, b]$, er uttrykkene over ikke nødvendigvis riemannsummer, men de representerer fremdeles numeriske metoder for å tilnærme integralet:

$$L(P) \approx U(P) \approx \int_a^b f$$ △

**Eksempel 4.117.** Dersom man tar gjennomsnittet av uttrykkene i forrige eksempel, får man trapesregelen:

$$\int_a^b f \approx \frac{1}{2n}\left(\sum_{i=1}^{n} f(x_i) + \sum_{i=0}^{n-1} f(x_i)\right) = \frac{1}{n}\left(\frac{f(a)}{2} + \sum_{i=1}^{n-1} f(x_i) + \frac{f(b)}{2}\right)$$

Denne er noe mer nøyaktig enn riemannsummer, og vi skal senere i semesteret se hvorfor. △

**Eksempel 4.118.** Dersom man tar evaluerer $f$ midt mellom punktene i partisjonen, istedet for i endepunktene, får man midtpunktregelen:

$$\int_a^b f \approx \frac{1}{n}\sum_{i=1}^{n} f\left(\frac{x_{i-1} + x_i}{2}\right)$$

Denne er også noe mer nøyaktig enn riemannsummer. △

Det går selvfølgelig an å implementere disse metodene med ujevne partisjoner, men det er ikke så viktig for oss i dette semesteret.

Vi kan også finne mer avanserte tilnærminger til integralet

$$I[f] = \int_a^b f(x)\, dx$$

ved å interpolere $f$, og så integrere intepolasjonspolynomet analytisk. Dette kalles kvadratur, og en bestemt metode kalles gjerne *kvadraturregel*. La $x_i$ være interpolasjonspunkter på $[a, b]$, og $l_i$ de korresponderende lagrangefunksjonene, slik at

$$f(x) \approx \sum_{i=0}^{n} f(x_i) l_i(x).$$

Vi skriver

$$\int_a^b f(x)\, dx \approx \int_a^b \sum_{i=0}^{n} f(x_i) l_i(x)\, dx = \sum_{i=0}^{n} f(x_i) \int_a^b l_i(x)\, dx = \sum_{i=0}^{n} f(x_i) A_i = Q[f],$$

der vi har definert $A_i = \int_a^b l_i(x)\, dx$. Disse kalles *kvadraturvektene*, eller bare *vektene*.

**Eksempel 4.119.** En av de aller enkleste kvadraturreglene kjenner du fra M1. Den kalles trapesregelen, er gitt ved

$$\int_a^b f(x)\, dx \approx (b-a)\left(\frac{f(a) + f(b)}{2}\right)$$

og utledes enkelt ved å interpolere $f$ med et førsteordens polynom i $x_0 = a$ og $x_1 = b$ og integrere dette. Vektene er

$$A_0 = A_1 = \frac{b-a}{2}. \qquad \triangle$$

**Eksempel 4.120.** Gauss-Legendre-punktene for $n = 2$ på intervallet $[-1, 1]$ er $\pm\sqrt{\frac{1}{3}}$, og lagrange-funksjoner er

$$l_0(x) = \frac{\sqrt{3}}{2}\left(x + \sqrt{\frac{1}{3}}\right)$$

og

$$l_1(x) = -\frac{\sqrt{3}}{2}\left(x - \sqrt{\frac{1}{3}}\right).$$

Vi beregner

$$A_0 = \frac{\sqrt{3}}{2}\int_{-1}^{1} x + \sqrt{\frac{1}{3}}\, dx = 1$$

og

$$A_1 = -\frac{\sqrt{3}}{2}\int_{-1}^{1} x - \sqrt{\frac{1}{3}}\, dx = 1. \qquad \triangle$$

**Eksempel 4.121.** Simpsons regel, som du også kjenner fra M1,

$$Q[f] = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right),$$

har vekter

$$A_0 = A_2 = \frac{b-a}{6} \quad \text{og} \quad A_1 = \frac{2(b-a)}{3}.$$

Disse utledes ved å interpolere $f$ med andre ordens polynomer i $a$, $b$ og $\frac{a+b}{2}$, men dette er gjort i M1, så vi dropper det. $\triangle$

## Feilestimat

Det første vi kan gjøre, er å sette opp en generell regel for kvadraturfeil. Denne detter rett ut av feilestimatet for intepolasjon.

**Teorem 4.122.** *La $Q$ være en $n+1$-punkts kvadraturregel på $[a, b]$, og anta at $f$ er $n+1$ ganger kontinuerlig deriverbar på $[a, b]$. En øvre skranke for feilen er*

$$|I[f] - Q[f]| = \frac{M}{(n+1)!}\int_a^b \prod_{i=0}^{n} |x - x_i|\, dx$$

*der $M = \max_{s\in[a,b]} |f^{n+1}(s)|$.*

*Bevis.* Integrer feilestimatet

$$f(x) - p_n(x) = \frac{f^{n+1}(s)}{(n+1)!}\prod_{k=0}^{n}(x - x_k).$$

fra $a$ til $b$ og bruk at $|f^{n+1}| \le M$. $\square$

Fordelen med dette teoremet er at det er lett å bevise. Ulempen er at det stort sett er mulig å finne skarpere estimater, men disse er mer grisete å utlede. Vi tar derfor med noen bedre feilestimater for forskjellige kvadraturregler uten bevis.

**Teorem 4.123.** *For trapesregelen finnes det en $s \in (a, b)$ slik at*

$$I[f] - Q[f] = \frac{(b-a)^3}{12} f''(s).$$

**Teorem 4.124.** *For Simpsons regel finnes det en $s \in (a, b)$ slik at*

$$I[f] - Q[f] = -\frac{(b-a)^5}{2880} f^4(s).$$

Det går an å skrive opp feilestimater for kvadraturregler basert på Chebyshev-punkter, Gauss-Legendre-punkter og Gauss-Lobatto-punkter, men de er ganske grisete, og bevisen er altfor kompliserte for oss. Vi nøyer oss derfor med noen eksempler, der vi sammenlikner med den analytiske verdien til integralet.

**Eksempel 4.125.** Vi tilnærmer

$$\int_{-1}^{1} e^x\, dx = e - e^{-1} = 2.350402387287603$$

med trapesregelen:

$$Q[f] = e^{-1} + e \approx 3.086161269630488.$$

Feilen er i samme størrelsesorden som svaret. Triste greier. $\triangle$

**Eksempel 4.126.** Med Simpsons regel går det litt bedre:

$$Q[f] = \frac{1}{3}(e^{-1} + 4e^0 + e) \approx 2.362053756543496$$

Her er feilen $0.011651369255893 \approx 10^{-2}$. $\triangle$

**Eksempel 4.127.** Gauss-Legendre med $n = 2$ gir

$$Q[f] = e^{-\sqrt{1/3}} + e^{\sqrt{1/3}} \approx 2.342696087909730$$

Her er feilen på $-0.007706299377873 \approx 10^{-2}$, og faktisk mindre enn for Simpsons regel, til tross for at kvadraturregelen er basert på en lavere ordens interpolasjon. Dette eksemplet illustrerer at plasseringen av interpolasjonspunktene har mye å si. $\triangle$

## Presisjonsgrad

Har forskjellige kvadraturregler noen andre gode egenskaper enn høy presisjon? Vi sier at en integrasjonsformel er eksakt for funksjonen $f$ dersom

$$\int_a^b f\, dx = \sum_i f(x_i)A_i.$$

Dersom du interpolerer et polynom av grad $n$ elller lavere med et polynom av grad $n$, blir $f$ og $p$ identiske. Derfor må det være klart at

$$\int_a^b p(x)\, dx = \sum_{i=0}^{n} f(x_i)A_i,$$

for alle polynomer av grad $n$ eller lavere. Dersom kvadraturregelen er intelligent designet, kan man oppnå høyere presisjonsgrad enn som så.

**Eksempel 4.128.** Trapesregelen har presisjonsgrad 1, for

$$Q[1] = (b-a)\left(\frac{1+1}{2}\right) = b - a = \int_a^b dx$$

og

$$Q[x] = (b-a)\left(\frac{a+b}{2}\right) = \frac{b^2-a^2}{2} = \int_a^b x\, dx,$$

mens

$$\begin{aligned} Q[x^2] &= (b-a)\left(\frac{a^2+b^2}{2}\right) \\ &= \frac{b^3 - ab^2 + a^2b - a^3}{2} \neq \int_a^b x^2\, dx. \end{aligned}$$

Disse beregningene viser at trapesregelen integrerer alle første ordens polynomer riktig, siden

$$Q[cx+d] = cQ[x] + dQ[1].$$

Ingen andre ordens polynomer integreres riktig, siden

$$Q[cx^2+dx+e] = cQ[x^2] + dQ[x] + eQ[1],$$

og $Q[x^2]$ ikke integreres riktig. △

**Eksempel 4.129.** Gauss-Lobatto $n = 2$ har presisjonsgrad 3. Dette kan vises på samme måte som for trapesregelen, men vi venter litt med å se på det. Det er nemlig ikke så vanskelig å vise generelt at en $n+1$-punkts Gauss-Lobatto-regel har presisjonsgrad $2n+1$, og dette skal vi gjøre senere. △

**Eksempel 4.130.** Simpsons regel har presisjonsgrad 3. En rask titt på feilestimatet forteller at dersom $f$ er et tredjegradspolynom, er $f^4 = 0$. △

## Noen forskjellige kvadraturklasser

Kvadraturregler skilles av er punktfordelingen, og hvorvidt man interpolerer med et polynom av høy grad på alle punktene, eller deler opp intervallet og interpolerer med stykkvis kontinuerlige polynombiter, såkalt *sammensatte regler*. Vi skal ta for oss et par klassiske kvadraturregler, og avslutte med en diskusjon rundt sammensatte regler.

### Newton-Cotes

Både trapesregelen og Simpsons metode er eksempler på Newton-Cotes-regler. Dette er regler der interpolasjonen er gjort på ekvidistante gitre. En klassisk lærebok i numerisk analyse vil typisk inneholde en lengre utgreining om Newton-Cotes, men vi vet jo at man skal styre unna polynominterpolasjon på ekvidistante gitre, så derfor lar vi Newton-Cotes ligge.

**Eksempel 4.131.** Trapesregelen og Simpsons regel er Newton-Cotes-regler med henholdsvis $n = 2$ og $n = 3$. △

### Clenshaw-Curtis

Clenshaw-curtiskvadratur baserer seg på å integrere chebyshevinterpolanten. For chebyshevinterpolasjon er det lett å skrive opp formler for interpolasjonspunktene, og pene formler for kvadraturvektene, men disse formlene er vanskelige å utlede.

**Teorem 4.132.** *Dersom $n$ er et partall, og interpolasjonsgitteret er et $n$-punkts chebyshev ekstremalgitter, er vektene til kvadraturregelen gitt ved*

$$w_1 = w_n = \frac{1}{(n-1)^2}$$

*og*

$$w_i = \frac{2}{(n-1)}\left(1 - \sum_{j=1}^{(n-2)/2} \frac{2}{4j^2-1}\cos\frac{2j(i-1)}{n-1}\pi\right)$$

*for $2 \leq i \leq n-1$.*

### Gauss-Legendre

Vi skal beskrive prosessen som produserer Legendrepolynomene, og vise at denne prosessen produserer kvadraturformler med presisjonsgrad $2n+1$.

**Teorem 4.133.** *La $q$ være et polynom av grad $n+1$ slik at*

$$\int_a^b qp\, dx = 0 \tag{4.1}$$

*for alle polynomer $p$ av grad mindre enn eller lik $n$. En kvadraturregel med disse $n+1$ nullpunktene som noder, vil være eksakt for alle polynomer av grad mindre enn eller lik $2n+1$.*

*Bevis.* Vi begynner med å vise at polynomet $q$ har $n+1$ forskjellige nullpunkter på intervallet $[a,b]$. La $x_0, x_1 \ldots, x_{r-1}$ være de $r$ punktene der $q$ bytter fortegn på $[a,b]$. Polynomet

$$\prod_{i=0}^{r-1}(x - x_i)$$

har grad $r$, og bytter fortegn nøyaktig samtidig som $q$, slik at enten

$$q\prod_{i=0}^{r-1}(x - x_i) \leq 0$$

eller

$$q\prod_{i=0}^{r-1}(x - x_i) \geq 0$$

på $[a,b]$. Dette betyr at

$$\int_a^b q\prod_{i=0}^{r-1}(x - x_i)\, dx \neq 0,$$

og siden vi har

$$\int_a^b qp\, dx = 0 \tag{4.2}$$

for alle polynomer $p$ av grad mindre enn eller lik $n$, impliserer dette at $r \geq n+1$. Siden $r$ åpenbart ikke kan være større enn $n+1$, må $r = n+1$. Altså bytter

$q$ fortegn $n+1$ ganger på $[a,b]$, og må følgelig ha $n+1$ nullpunkter på $[a,b]$.

La nå $h$ være et polynom av grad $2n+1$ eller lavere, og del $h$ på $q$ med rest $r$:

$$h = qp + r.$$

der $p$ har grad $n$, og $r$ har maksimal grad $n$. Vi evaluerer $h$ i $x_i$, og ser at

$$h(x_i) = q(x_i)p(x_i) + r(x_i) = r(x_i),$$

siden $q(x_i) = 0$ for alle $i$. Siden kvadraturregelen er basert på integrasjon av et $n$-te ordens polyom, er den åpenbart er eksakt for polyonomer av orden $n$ eller lavere, og vi kan beregne

$$\begin{aligned}\int_a^b h(x)\,dx &= \int_a^b q(x)p(x) + r(x)\,dx = \int_a^b r(x)\,dx \\ &= \sum_{i=0}^{n} r(x_i)A_i = \sum_{i=0}^{n} h(x_i)A_i. \qquad \square\end{aligned}$$

**Eksempel 4.134.** Vi tar et eksempel der vi konstruerer $q$ fra grunnen av. La $n = 1$ og $[a,b] = [0,1]$. Vi må finne

$$q(x) = ax^2 + bx + c.$$

Vi begynner med å kreve at $q$ skal stå ortogonalt på alle polynomer av grad 1 eller lavere. Dersom

$$\int_0^1 q(x)\,dx = \int_0^1 ax^2 + bx + c\,dx = \frac{a}{3} + \frac{b}{2} + c = 0$$

og

$$\int_0^1 xq(x)\,dx = \int_0^1 ax^3 + bx^2 + cx\,dx = \frac{a}{4} + \frac{b}{3} + \frac{c}{2} = 0,$$

vil

$$\int_0^1 (ax+b)q(x)\,dx = 0.$$

Vi ganger den første likningen med seks, den andre med tolv, og får likningssystemet

$$\begin{pmatrix} 2 & 3 & 6 \\ 3 & 4 & 6 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Litt gausseliminasjon gir

$$\begin{pmatrix} 2 & 3 & 6 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Her står det at $a + b = 0$, og at $2a + 3b + 6c = 0$. Løsningsrommet er

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = c \begin{pmatrix} 6 \\ -6 \\ 1 \end{pmatrix},$$

slik at

$$q(x) = 6x^2 - 6x + 1,$$

er et mulig valg for $q$. Andre valg av konstanten $c$ vil gi andre polynomer, men alle vi ha de samme nullpunktene, og det er dem vi er ute etter. Nullpunktene til $q$ er

$$x_0 = \frac{1}{2} - \frac{1}{2\sqrt{3}} \qquad \text{og} \qquad x_1 = \frac{1}{2} + \frac{1}{2\sqrt{3}}.$$

Lagrangefunksjonene blir

$$L_0 = \frac{x - x_1}{x_0 - x_1} = \sqrt{3}\left(x - \left(\frac{1}{2} + \frac{1}{2\sqrt{3}}\right)\right)$$

og

$$L_1 = \frac{x - x_0}{x_1 - x_0} = \sqrt{3}\left(x - \left(\frac{1}{2} - \frac{1}{2\sqrt{3}}\right)\right)$$

slik at

$$A_1 = \sqrt{3}\int_0^1 x - \left(\frac{1}{2} - \frac{1}{2\sqrt{3}}\right)\,dx = \frac{1}{2} = A_0.$$

Integrasjonsrutinen vi har laget skal være eksakt for alle polynomer opp til grad 3. Vi tester:

$$\frac{1}{2}\left(\frac{1}{2} - \frac{1}{2\sqrt{3}}\right)^2 + \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2\sqrt{3}}\right)^2 = \frac{1}{3} = \int_0^1 x^2\,dx$$

$$\frac{1}{2}\left(\frac{1}{2} - \frac{1}{2\sqrt{3}}\right)^3 + \frac{1}{2}\left(\frac{1}{2} + \frac{1}{2\sqrt{3}}\right)^3 = \frac{1}{4} = \int_0^1 x^3\,dx.$$

Merk at punktene er bare Gauss-Legendre-punktene flyttet til intervallet $[0,1]$. Dette eksemplet illustrerer hvordan Gauss-Legendre-tabellen i interpolasjonskapitlet er konstruert. $\triangle$

**Kommentar.** Legendrepolynomene er definert på intervallet $[-1,1]$, så eksemplet over laget ikke et legendrepolynom, men derimot et polynom som har nullpunktene til legendrepolynomet

$$P_2 = \frac{1}{2}(3x^2 - 1)$$

flyttet til intervallet $[0,1]$.

Vi kan utvide tabellen fra forrige kapittel med vekter. Husk at tabellen gjelder for $[-1,1]$, så om du trenger integrasjonsrutine for andre intervaller, må punktene flyttes, og vektene beregnes på nytt.

| $n$ | $x_i$ | $A_i$ |
|---|---|---|
| 2 | $\pm\sqrt{\frac{1}{3}}$ | 1 |
| 3 | 0 | $\frac{8}{9}$ |
| | $\pm\sqrt{\frac{3}{5}}$ | $\frac{5}{9}$ |
| 4 | $\pm\sqrt{\frac{3}{7} - \frac{2}{7}\sqrt{\frac{6}{5}}}$ | $\frac{18+\sqrt{30}}{36}$ |
| | $\pm\sqrt{\frac{3}{7} + \frac{2}{7}\sqrt{\frac{6}{5}}}$ | $\frac{18-\sqrt{30}}{36}$ |
| 5 | 0 | $\frac{128}{225}$ |
| 5 | $0,\ \pm\frac{1}{3}\sqrt{5 - 2\sqrt{\frac{10}{7}}}$ | $\frac{322+13\sqrt{70}}{900}$ |
| 5 | $0,\ \pm\frac{1}{3}\sqrt{5 + 2\sqrt{\frac{10}{7}}}$ | $\frac{322-13\sqrt{70}}{900}$ |

Merk at alle vektene er positive. Kvadraturentusiaster regner dette som et kvalitetsstempel.

### Gauss-Lobatto

Vi koster på oss en tabell med vekter for Gauss-Lobatto også.

| $n$ | $x_i$ | $A_i$ |
|---|---|---|
| 3 | $0$ | $\frac{4}{3}$ |
| | $\pm 1$ | $\frac{1}{3}$ |
| 4 | $\pm\sqrt{\frac{1}{5}}$ | $\frac{5}{6}$ |
| | $\pm 1$ | $\frac{1}{6}$ |
| 5 | $0$ | $\frac{32}{45}$ |
| | $\pm\sqrt{\frac{3}{7}}$ | $\frac{49}{90}$ |
| | $\pm 1$ | $\frac{1}{10}$ |
| 6 | $\pm\sqrt{\frac{1}{3}-\sqrt{\frac{2}{3\sqrt{7}}}}$ | $\frac{14+\sqrt{7}}{30}$ |
| | $\pm\sqrt{\frac{1}{3}+\sqrt{\frac{2}{3\sqrt{7}}}}$ | $\frac{14-\sqrt{7}}{30}$ |
| | $\pm 1$ | $\frac{1}{15}$ |
| 7 | $0$ | $\frac{256}{525}$ |
| | $\pm\sqrt{\frac{5}{11}-\frac{2}{11}\sqrt{\frac{5}{3}}}$ | $\frac{124+7\sqrt{15}}{350}$ |
| | $\pm\sqrt{\frac{5}{11}+\frac{2}{11}\sqrt{\frac{5}{3}}}$ | $\frac{124-7\sqrt{15}}{350}$ |
| | $\pm 1$ | $\frac{1}{21}$ |

**Eksempel 4.135.** Simpsons regel er også en Gauss-Lobatto-regel med $n = 3$. △

# Kapittel 5

# Ordinære differensiallikninger

De aller fleste differensiallikninger kan ikke løses på en fornuftig måte med penn og papir. Vi skal begynne med å løse noen relativt enkle likninger (regningen blir allerede her ganske komplisert) og så går vi over på numeriske metoder. Til slutt kommer noen teoretiske greier.

Alternative kilder:

- Kreyszig kap. 1
- Lindstrøm I kap. 10
- Schaeffer/Cain kap. 1

## Analytiske løsningsteknikker

**Eksempel 5.1.** Det aller enkleste initialverdiproblemet er

$$\dot{y}(t) + ay(t) = 0 \quad y(0) = y_0,$$

og alle barn i barnehagen vet at løsningen er

$$y(t) = y_0 e^{-at}.$$

Siden $f(y) = y$ er kontinuerlig deriverbar overalt, må dette være den eneste løsningen, se avsnittet om vanskelig teori nederst. △

**Eksempel 5.2.** Hvis du ikke liker konseptet at læreren bare sier hva løsningen er, kan løsningen til likningen i forrige eksempel utledes ved å huske at likningen er separabel. En førsteordens differensiallikning kalles separabel om den kan skrives på formen

$$\dot{y}(t)f(y(t)) = g(t).$$

Disse løses ved å integrere med hensyn på $t$ på begge sider. Vi skriver først om

$$y(t)' + ay(t) = 0$$

til

$$\frac{\dot{y}(t)}{y(t)} = -a.$$

Vi integrerer

$$\int_0^t \frac{\dot{y}(u)}{y(u)}\, du = -\int_0^t a\, du$$

og får

$$\ln|y(t)| - \ln|y(0)| = -at.$$

Hvis man har lest den vanskelige teorien på slutten av kapitlet, vet man at ingen løsningstrajektorier kan krysse hverandre, og siden $y(t) = 0$ er den entydige løsningen dersom $y(0) = 0$, kan ingen løsningstrajektorier bytte fortegn. Derfor har $y(t)$ og $y(0)$ alltid samme fortegn, slik at

$$\ln|y(t)| - \ln|y(0)| = \ln\frac{|y(t)|}{|y(0)|} = \ln\frac{y(t)}{y(0)}$$

og

$$y(t) = e^{-at+\ln y(0)} = y_0 e^{-at}.$$ △

**Eksempel 5.3.** Man kan også bruke noe som kalles integrerende faktor. Vi ganger

$$\dot{y}(t) + ay(t) = 0$$

med $e^{at}$ og observerer at

$$\frac{d}{dt}\left(e^{at}y(t)\right) = e^{at}\dot{y}(t) + e^{at}ay(t) = 0$$

slik at

$$e^{at}y(t) = C$$

eller

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

og følgelig

$$y(t) = y_0 e^{-at}$$

△

**Eksempel 5.4.** I mange tilfeller vil luftmotstanden til en partikkel i bevegelse avhenge kvadratisk av hastigheten til partikkelen. En geværkules horisontale hastighetskomponent beskrives av den separable differensialliknen

$$-ky^2 = m\dot{y},$$

der $k$ er en konstant som avhenger av kulens form, og $m$ er kulens masse. La oss anta at kulens initialhastighet er $y(0) = v$. Vi skriver om til

$$-\frac{k}{m} = \frac{\dot{y}}{y^2}$$

integrerer opp

$$-\int_0^t \frac{k}{m} = \int_0^t \frac{\dot{y}}{y^2}$$

og får

$$-\frac{k}{m}t = \frac{1}{y(0)} - \frac{1}{y(t)} = \frac{1}{v} - \frac{1}{y(t)},$$

eller

$$y(t) = \frac{1}{\frac{k}{m}t + \frac{1}{v}} = \frac{mv}{kvt + m}. \qquad \triangle$$

**Eksempel 5.5.** Initialverdiproblemet

$$\dot{y}(t) = \sqrt{y(t)} \quad y(0) = 0,$$

har to løsninger

$$y_1(t) = 0$$

og

$$y_2(t) = \frac{t^2}{4}.$$

Problemet er at $f(y) = \sqrt{y}$ ikke er kontinuerlig deriverbar i $y = 0$. $\triangle$

**Eksempel 5.6.** Likningen

$$\dot{y}(t) + ay(t) = f(t)$$

dukker opp i ett bankende kjør i anvendelser, så vi skal bruke litt tid på å finne løsningen. Det første vi gjør, er å skrive opp løsningen for $f = 0$, altså løsningen av

$$\dot{y}(t) + ay(t) = 0.$$

Denne likningen kalles den homogene likningen. Løsningen kalles den homogene løsningen, og er som kjent gitt ved

$$y(t) = ce^{-at}.$$

Vi kan bruke denne til å utlede et uttrykk for løsningen til initialverdiproblemet

$$\dot{y}(t) + ay(t) = f(t) \quad y(0) = 0$$

En stor vitenskapsmann har en gang i tiden skjønt at dersom $f$ ikke er en grusom funksjon, kan løsningen fint skrives som

$$y(t) = c(t)e^{-at},$$

der $c(t)$ er en ukjent funksjon. Vi setter dette uttrykket inn i likningen, og deduserer at

$$\begin{aligned} f(t) &= \dot{y}(t) + ay(t) \\ &= c'(t)e^{-at} - ac(t)e^{-at} + ac(t)e^{-at} \\ &= c'(t)e^{-at}. \end{aligned}$$

Men dette betyr at

$$c'(t) = e^{at}f(t),$$

og følgelig at

$$c(t) = \int_0^t e^{as}f(s)\,ds.$$

Løsningen blir

$$y(t) = e^{-at}\int_0^t e^{as}f(s)\,ds = \int_0^t e^{a(s-t)}f(s)\,ds.$$

Det siste integralet kalles en konvolusjon mellom $f(s)$ og $e^{as}$, og dette blir veldig viktig i senere i studiet. Merk at løsningen tilfredsstiller $y(0) = 0$. Dersom man har initialbetingelse $y(0) = y_0$, kan løsningen enkelt justeres til

$$y(t) = y_0e^{-at} + \int_0^t e^{a(s-t)}f(s)\,ds. \qquad \triangle$$

**Eksempel 5.7.** Teknikken over kan utvides til å produsere en løsningsformel for

$$\dot{y}(t) + a(t)y(t) = f(t)$$

der $a(t)$ er en kontinuerlig funksjon. La oss igjen først finne løsningen til

$$\dot{y}(t) + a(t)y(t) = 0.$$

Vi ganger likningen med $\exp\int_0^t a(s)\,ds$, og får

$$\dot{y}(t)\exp\int_0^t a(s)\,ds + a(t)y(t)\exp\int_0^t a(s)\,ds = 0.$$

Men analysens fundamentalteorem gir oss at venstresiden kan skrives

$$\dot{y}(t)\exp\int_0^t a(s)\,ds + a(t)y(t)\exp\int_0^t a(s)\,ds = \frac{d}{dx}\left(y(t)\exp\int_0^t a(s)\,ds\right),$$

slik at

$$\frac{d}{dx}\left(y(t)\exp\int_0^t a(s)\,ds\right) = 0,$$

og følgelig

$$y(t)\exp\int_0^t a(s)\,ds = c$$

eller

$$y(t) = c\exp\left(-\int_0^t a(s)\,ds\right)$$

Vi bruker nå det samme trikset som isted, nemlig å anta at løsningen når $f \neq 0$ kan skrives

$$y(t) = c(t)\exp\left(-\int_0^t a(s)\,ds\right).$$

Derivasjon og innsetting i likningen gir

$$c'(t)\exp\left(-\int_0^t a(s)\,ds\right) = f(t),$$

slik at

$$c'(t) = f(t)\exp\left(\int_0^t a(s)\,ds\right),$$

eller

$$c(t) = \int_0^t f(u)\exp\left(\int_0^u a(s)\,ds\right).$$

Dette ser helt forferdelig ut. Løsningsmetoden kalles integrerende faktor. I praksis er det som regel ikke nødvendig å tenke så komplisert, se eksempel 5.3. $\triangle$

# Numeriske metoder

Vi skal lage numeriske metoder for å finne tilnærmede løsninger for initialverdiproblemet

$$\dot{y} = f(t, y) \quad y(a) = c$$

på intervallet $[a, b]$. La oss starte med en jevn partisjon med $n + 1$ punkter

$$t_i = a + hi$$

der

$$h = \frac{b - a}{n}$$

og $0 \leq i \leq n$. Vi er interessert i å finne tilnærminger til $y(t_i)$. Den første av disse kjenner vi, nemlig $y(t_0) = y(a) = c$, Tilnærmingene til de etterfølgende verdiene kaller vi $y_i$, og gjør vi alt rikig, vil vi ha

$$y_i \approx y(t_i)$$

for alle $i$.

Siden $\dot{y}$ en en kontinuerlig funksjon på $[a, b]$, går det knirkefint å bruke analysens fundamentalteorem, og skrive

$$y(t_{i+1}) - y(t_i) = \int_{t_i}^{t_{i+1}} \dot{y}(t)\, dt.$$

Siden $\dot{y} = f(t, y)$, kan vi skrive

$$y(t_{i+1}) - y(t_i) = \int_{t_i}^{t_{i+1}} f(t, y(t))\, dt.$$

Dette er arbeidshesten vår, og vi skal utlede numeriske metoder ved å sette inn forskjellige approksimasjoner for integralet på høyre side, samt erstatte $y(t_{i+1}) - y(t_i)$ med $y_{i+1} - y_i$ på venstre side.

**Eksempel 5.8.** Gjør man den særdeles enkle tilnærmingen

$$\int_{t_i}^{t_{i+1}} f(t, y(t))\, dt \approx \int_{t_i}^{t_{i+1}} f(t_i, y_i)\, dt = hf(t_i, y_i),$$

får man en metode som kalles Eulers eksplisitte metode:

$$y_{i+1} = y_i + hf(t_i, y_i). \qquad \triangle$$

For å komme i gang med beregningene, er det bare å definere $y_0 = c$, og så beregne

$$y_1 = y_0 + hf(t_0, y_0) = c + hf(a, c)$$

Når vi nå har $y_1$, kan vi tilsvarende beregne

$$y_2 = y_1 + hf(t_1, y_1)$$

og

$$y_3 = y_2 + hf(t_2, y_2)$$

og så videre, helt frem til

$$y_n = y_{n-1} + hf(t_{n-1}, y_{n-1}).$$

**Eksempel 5.9.** Velger man den tilsvarende enkle tilnærmingen

$$\int_{t_i}^{t_{i+1}} f(t, y(t))\, dt \approx hf(t_{i+1}, y_{i+1}),$$

får man Eulers implisitte metode

$$y_{i+1} = y_i + hf(t_{i+1}, y_{i+1}). \qquad \triangle$$

**Eksempel 5.10.** Trapesregelen

$$\int_{t_i}^{t_{i+1}} f(t, y(t))\, dt \approx \frac{h}{2}(f(t_i, y_i) + f(t_{i+1}, y_{i+1}))$$

gir trapesmetoden

$$y_{i+1} = y_i + \frac{h}{2}(f(t_i, y_i) + f(t_{i+1}, y_{i+1})). \qquad \triangle$$

**Eksempel 5.11.** Tilnærmer man trapesregelen med formelen

$$\int_{t_i}^{t_{i+1}} f(t, y(t))\, dt \approx \frac{h}{2}\left(f(t_i, y_i) + f(t_{i+1}, y^*)\right)$$

der $y^* = y_i + hf(t_i, y_i)$ er et eksplisitt eulersteg, får man Heuns metode:

$$\begin{aligned} y^* &= y_i + hf(t_i, y_i) \\ y_{i+1} &= y_i + \frac{h}{2}(f(t_i, y_i) + f(t_{i+1}, y^*)) \end{aligned} \qquad \triangle$$

**Eksempel 5.12.** Bruker man tilnærmingen

$$\int_{t_i}^{t_{i+1}} f(t, y(t))\, dt \approx hf\left(\frac{t_i + t_{i+1}}{2}, \frac{y_i + y_{i+1}}{2}\right)$$

får man *midtpunktmetoden*

$$y_{i+1} = y_i + hf\left(t_i + \frac{h}{2}, \frac{y_i + y_{i+1}}{2}\right). \qquad \triangle$$

Dette er alle metodene vi skal analysere, og alle er eksempler på noe som kalles *Runge-Kutta-metoder*. Metodene faller i to kategorier. I den ene kategorien (eksplisitt Euler og Heun) er $y_{i+1}$ alene på den ene siden av likningen for metoden, mens i den andre kategorien (implisitt Euler og trapesmetoden) finner du $y_{i+1}$ overalt i likningen, og det er ikke bare bare å løse for $y_{i+1}$.

Du lurer sikkert på hvorfor man har så mange forskjellige metoder, og det korte svaret er som ellers i anvendt matematikk: noen metoder eksisterer fordi de er lette å finne opp og forstå, mens andre metoder finnes fordi de er skikkelig bra. Dette skal vi se på nå.

**Eksempel 5.13.** Vi løser initialverdiproblemet

$$\dot{y} = -y \quad y(0) = 1$$

med Eulers eksplisitte metode på intervallet $[0, 1]$. Siden $f(t, y) = -y$, blir metoden

$$y_{i+1} = y_i - hy_i = (1 - h)y_i$$

med

$$y_0 = 1.$$

Løsning for $h = 0.25$ gir figuren under. De blå diamantene er $y_1$, $y_2$, $y_3$, $y_4$ og $y_5$, mens den røde kurven er den analytiske løsningen $y = e^{-x}$. Vi beregner $y(1) = 1/e \approx 0.367879441171442$, som kan sammenliknes med $y_5 = 0.31640625$:

$$y_5 - y(1) = -0.051473191171442. \qquad \triangle$$

**Eksempel 5.14.** Vi løser samme problem som i sted med Eulers implisitte metode. Metoden blir

$$y_{i+1} = y_i - hy_{i+1},$$

som vi løser for $y_{i+1}$, og får

$$y_{i+1} = \frac{y_i}{(1+h)}.$$

Figur under for $h = 0.5$. Vi får $y_5 = 0.4096$, og

$$y_5 - y(1) = 0.041720558828558. \qquad \triangle$$

**Eksempel 5.15.** Trapesmetoden:

$$y_{i+1} = y_i - \frac{1}{2}h(y_i + y_{i+1}).$$

Vi løser for $y_{i+1}$, og får

$$y_{i+1} = \frac{2-h}{2+h}y_i.$$

Denne treffer noe bedre:

$$y_5 - y(1) = -0.001929128719072. \qquad \triangle$$

**Eksempel 5.16.** Heuns metode:

$$y^* = y_i - hy_i$$
$$y_{i+1} = y_i - \frac{1}{2}h(y_i + y^*)$$

og

$$y_5 - y(1) = 0.004649588674749.$$

Bedre enn Euler, men ikke helt trapesmetoden. $\triangle$

Disse metodene oppfører seg litt forskjellig, og det ser ut som om noen er litt mer nøyaktige enn andre. Dette skal vi se en forklaring på i neste bolk, som handler om rekkeutvikling.

Metodene vi har utledet til nå, kalles *enstegsmetoder*, for kun $y_{i+1}$ og $y_i$ figurerer i likningene. Grunnen til at alle metodene er på denne formen, er at det kun er brukt en type endelig differansetilnærming på venstre side av

$$\dot{y}(x) = f(x, y(x)),$$

nemlig den første ordens differansen

$$\dot{y}(x_{i+1}) \approx \frac{y_{i+1} - y_i}{h}.$$

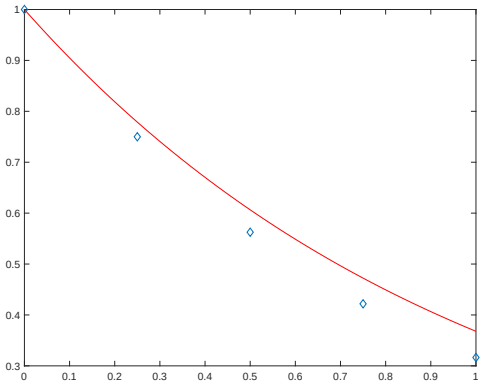
Nå er det ingenting i veien for å bruke en høyere ordens tilnærming, for eksempel sentraldifferansen

$$\dot{y}(x_i) \approx \frac{y_{i+1} - y_{i-1}}{2h}.$$

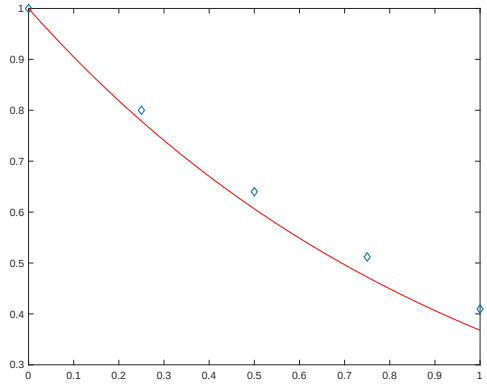
Setter man denne inn for $\dot{y}$, får man *leap-frog-metoden*

$$y_{i+1} = y_{i-1} + 2hf(x_i, y_i).$$

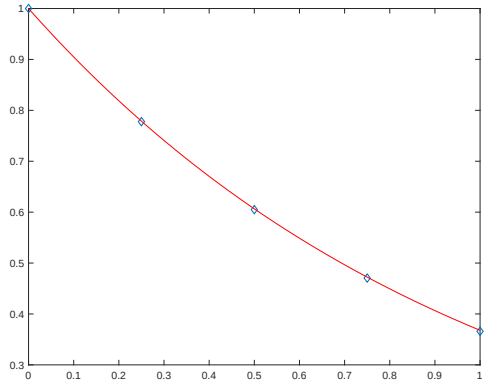
Her inngår både $y_{i+1}$, $y_i$ og $y_{i-1}$, og leap-frog er et eksempel på en *flerstegsmetode*. Flerstegsmetoder er ikke pensum i dette kurset.



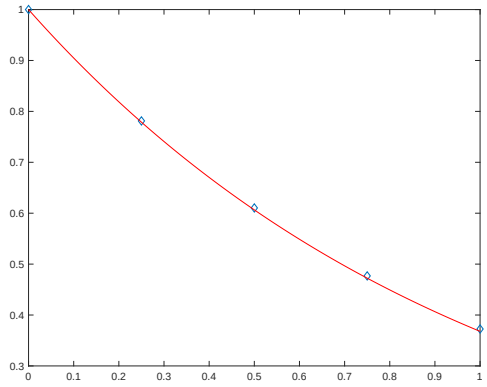
Figur 5.1: Eulers eksplisitte metode



Figur 5.2: Eulers implisitte metode



Figur 5.3: Trapesmetoden



Figur 5.4: Heuns metode

# Feilanalyse

I dette avsnittet skal vi ta en titt på hvorfor metodene treffer så foreskjellig. Vi skal indikere hvordan analysen får for eksplisitt Euler, og så skrive opp resultatet for de andre metodene.

Lineariseringen som gir det første eulersteget er

$$\begin{aligned} y(x_1) &= y(x_0 + h) \\ &\approx y(x_0) + hf(x_0, y(x_0)) \\ &= y(x_0) + h\dot{y}(x_0). \end{aligned}$$

Vi antar at $y$ er en analytisk funksjon, og taylorutvikler:

$$\begin{aligned} y(x_1) &= y(x_0 + h) \\ &= y(x_0) + h\dot{y}(x_0) + \frac{\dot{y}'(x_0)}{2}h^2 + \cdots \end{aligned}$$

Sammenlikner vi denne med

$$y_1 = y(x_0) + h\dot{y}(x_0),$$

ser vi at feilen i det første eulersteget er gitt ved

$$y(x_1) - y_1 = \frac{\dot{y}'(x_0)}{2}h^2 + \frac{\dot{y}''(x_0)}{6}h^3 + \cdots$$

altså taylorrekkehalen til $y$. Hvis vi antar at $h$ er liten, slik at $h^2$ er mye større enn $h^3$, og leddet

$$\frac{\dot{y}'(x_0)}{2}h^2$$

dominerer halen på taylorrekken, er det ikke urimelig å hevde at eksplisitt Euler har *lokal feil* av størrelsesorden $h^2$.

Feilen etter ett steg er altså av størrelsesorden $h^2$. Men hva er feilen etter $n$ steg? I eksemplene i forrige avsnitt, kjørte vi løserne på intervallet $[0, 1]$. La oss si at vi kjører på intervallet $[x_0, x_0 + a]$. Vi velger $h$ slik at

$$x_n = x_0 + hn = x_0 + a$$

og

$$n = \frac{a}{h}.$$

Hvis vi nå gjør $n$ steg med eksplisitt Euler, samler vi i hvert steg opp en lokal feil omtrent lik

$$\frac{\dot{y}'(x_i)}{2}h^2.$$

Feilen etter $n$ steg blir

$$\sum_{i=1}^{n} \frac{\dot{y}'(x_i)}{2}h^2,$$

og hvis vi antar at $\dot{y}' \leq M$ på $[x_0, x_0 + a]$, er det rimelig å hevde at

$$\sum_{i=1}^{n} \frac{\dot{y}'(x_i)}{2}h^2 \leq Mnh^2 = M\frac{a}{h}h^2 = Mah.$$

Vi sier derfor at eulers metode har *global feil* av størrelsesorden $h$.

**Teorem 5.17.** *Lokal og global feil for metodene:*

| *Metode* | *Lokal feil* | *Global feil* |
|---|---|---|
| *Eksplisitt Euler* | $h^2$ | $h$ |
| *Implisitt Euler* | $h^2$ | $h$ |
| *Trapesmetoden* | $h^3$ | $h^2$ |
| *Heuns metode* | $h^3$ | $h^2$ |
| *RK4* | $h^5$ | $h^4$ |

Vi skal ikke bevise dette teoremet, men nevner at bevisteknikken er den samme for alle metodene: taylorutvikle om $x_0$ for å finne lokal feil, og så se på hva som skjer etter $n$ steg. Dette teoremet forklarer langt på vei hva som skjedde i eksemplene i forrige avsnitt. Nå tar vi et par eksempler der vi lar $h \to 0$.

**Eksempel 5.18.** Vi kjører samme eksempel som i forrige avsnitt, men nå bruker vi Eulers eksplisitte metode for $h = 0.1$, $h = 0.01$ og så videre. Resultatene er oppsummert i følgende tabell:

| $h$ | $y_n - y(1)$ |
|---|---|
| $10^{-1}$ | $-1.920100107144218e-02$ |
| $10^{-2}$ | $-1.847099898213078e-03$ |
| $10^{-3}$ | $-1.840164004788258e-04$ |
| $10^{-4}$ | $-1.839473847314865e-05$ |
| $10^{-5}$ | $-1.839403194148215e-06$ |

Dette eksemplet demonstrerer tydelig at feilen etter $n$ steg er proporsjonal med $h$. På folkemunne sier man gjerne at man får en ekstra korrekt desimal hver gang man tideler $h$. Eulers implisitte metode oppfører seg omtrent likt, så den hopper vi over. △

**Eksempel 5.19.** Trapesmetoden for $h = 0.1$, $h = 0.01$ og så videre:

| $h$ | $y_n - y(1)$ |
|---|---|
| $10^{-1}$ | $-3.068987885734287e-04$ |
| $10^{-2}$ | $-3.065695217463471e-06$ |
| $10^{-3}$ | $-3.065658332745969e-08$ |
| $10^{-4}$ | $-3.069314802317535e-10$ |
| $10^{-5}$ | $-5.472844399889709e-12$ |

Her er feilen etter $n$ steg proporsjonal med $h^2$. På folkemunne sier man gjerne at man får to desimaler hver gang man tideler $h$. Heuns metode produserer omtrent den samme tabellen. △

**Eksempel 5.20.** RK4:

| $h$ | $y_n - y(1)$ |
|---|---|
| $10^{-1}$ | $3.332410560830112e-07$ |
| $10^{-2}$ | $3.091293887536040e-11$ |
| $10^{-3}$ | $3.996802888650564e-15$ |
| $10^{-4}$ | $2.664535259100376e-15$ |
| $10^{-5}$ | $-3.330669073875470e-15$ |

Hva skjedde her? Feilen etter $n$ steg proporsjonal med $h^4$, altså fire desimaler for hver tideling av $h$, men bare for de første tre tidelingene. Når $h = 10^{-3}$ har vi nådd såkalt *maskinpresisjon*. Matlab regner bare

med 16 desimaler, og dette setter en stopper for konvergensen. △

**Eksempel 5.21.** RK4, men nå har matlab fått beskjed om å regne med 32 desimaler:

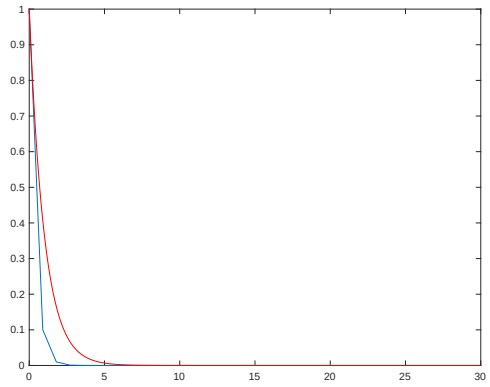
| $h$ | $y_n - y(1)$ |
|---|---|
| $10^{-1}$ | $3.332410561118064e-07$ |
| $10^{-2}$ | $3.091319001284922e-11$ |
| $10^{-3}$ | $3.068217823302200e-15$ |
| $10^{-4}$ | $3.065917492545278e-19$ |
| $10^{-5}$ | $3.065687557054922e-23$ |

Tabellene til nå har tatt en brøkdel av et sekund å produsere. Til sammenlikning tok denne her rundt ti minutter, pluss noen timer knoting for å finne ut av hvordan matlab skal regne riktig med 32 desimaler. Presisjon koster! △

# Stabilitet

Har metodene noen andre egenskaper? Eulers eksplisitte og implisitte metoder ser til forveksling like ut, og har akkurat samme orden. Men de oppfører seg ganske forskjellig.

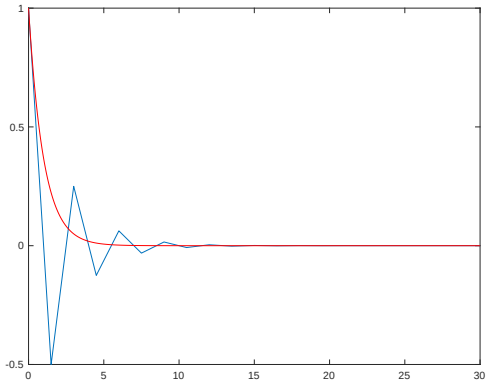
**Eksempel 5.22.** Vi kjører eksplisitt Euler på samme problem som over, men på intervallet $[0, 30]$, og $h = 0.9$. Det er trukket rette linjer mellom eulerstegene, så det skal bli litt enklere å se hva som skjer. △

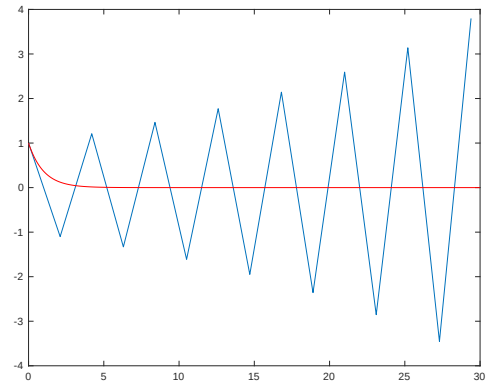


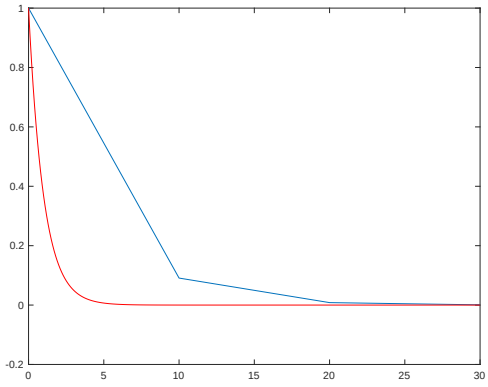
**Eksempel 5.23.** Vi kjører igjen på intervallet $[0, 30]$, men nå med $h = 1.5$. Den numeriske løsningen ser ut til å virre frem og tilbake en del før den sikter seg inn på rett spor. △

**Eksempel 5.24.** Intervallet $[0, 30]$, men nå med $h = 2.1$. Hva den numeriske løsningen tenker på, er ikke godt å si, men noe fornuftig er det ihvertfall ikke. △

**Eksempel 5.25.** Vi kan slå fast at eksplisitt Euler ikke fungerte, og det ser ut som om det går galt fordi $h$ er for stor. Vi prøver Euler implisitt på samme intervall, men med $h = 10$. Det går riktig bra. △







Forklaringen på hva som skjedde her, er ganske enkel. Eulers eksplisitte metode er, for eksemplet vi har studert,

$$\begin{aligned} y_{i+1} &= y_i - hy_i \\ &= (1-h)y_i \\ &= (1-h)^2 y_{i-1} \\ &= (1-h)^{i+1} y_0 = (1-h)^{i+1}, \end{aligned}$$

med andre ord en geometrisk følge. Siden gymnaset har du visst at denne følgen divergerer dersom

$$|1-h| \geq 1$$

og denne ulikheten blir innfridd akkurat i det $h$ bikker 2. Kjører vi samme resonnementet på Euler implisitt, får vi

$$y_{i+1} = \frac{1}{(1+h)^{i+1}}.$$

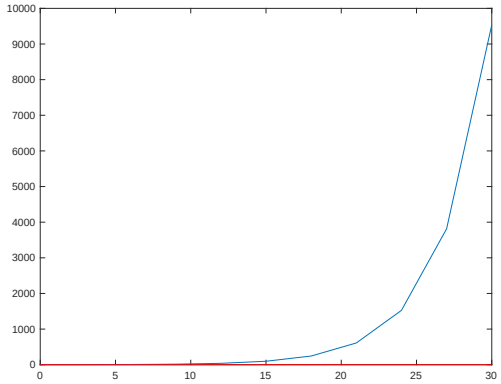
I vårt tilfelle er $h > 0$, så det må være klart at

$$0 < \frac{1}{(1+h)^{i+1}} < 1,$$

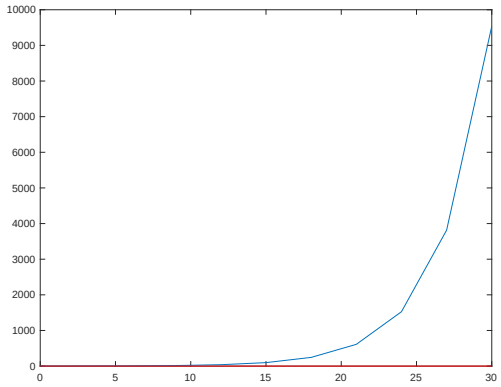
og følgelig konvergerer følgen mot 0 for alle valg av $h$.

Analysen vi har gjort, kalles *stabilitetsanalyse*, og $\dot{y} = -y$ er et såkalt *testproblem*. Vi får ikke eksakt informasjon om hvordan Eulers metode kommer til å oppføre seg for alle mulige differensiallikninger, men vi kan få en mageføelse allikevel. Vi skal ikke gå inn på en lengre diskusjon om stabilitetsanalyse, som er et forskningsfelt i seg selv, men nevne at stabilitetsproblemer er som regel betydelige for eksplisitte metoder, og ikke-eksisterende for implisitte metoder.

**Eksempel 5.26.** Vi prøver Heuns metode på $[0, 30]$, med $h = 3$. Som du ser, går det ganske dårlig. △




**Eksempel 5.27.** Vi prøver RK4 på $[0, 30]$, med $h = 2.1$. Det går ikke noe bedre. △




# Mer om implisitte metoder

Så hvorfor bør vi ikke alltid bruke implisitte metoder? Det er et komplisert spørsmål å svare på, men vi skal gjøre et forsøk i dette avsnittet. Vi begynner med et eksempel, der vi setter opp de forskjellige numeriske metodene.

**Eksempel 5.28.** Vi skriver opp de forskjellige metodene for

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

Eksplisitt Euler:

$$y_{i+1} = y_i + h\frac{y_i - 2x_i y_i^2}{1+x_i}$$

Imsplisitt Euler:

$$y_{i+1} = y_i + h\frac{y_{i+1} - 2x_{i+1} y_{i+1}^2}{1+x_{i+1}}$$

Trapesmetoden

$$y_{i+1} = y_i + \frac{h}{2}\left(\frac{y_i - 2x_i y_i^2}{1+x_i} + \frac{y_{i+1} - 2x_{i+1} y_{i+1}^2}{1+x_{i+1}}\right)$$

Heuns metode

$$y^*_{i+1} = y_i + h\frac{y_i - 2x_i y_i^2}{1+x_i}$$

$$y_{i+1} = y_i + \frac{h}{2}\left(\frac{y_i - 2x_i y_i^2}{1+x_i} + \frac{y^*_{i+1} - 2x_{i+1}(y^*_{i+1})^2}{1+x_{i+1}}\right)$$

RK4

$$\begin{aligned}
k_1 &= \frac{y_i - 2x_i y_i^2}{1+x_i} \\
k_2 &= \frac{\left(y_i + \frac{h}{2}k_1\right) - 2\left(x_i + \frac{h}{2}\right)\left(y_i + \frac{h}{2}k_1\right)^2}{1+\left(x_i + \frac{h}{2}\right)} \\
k_3 &= \frac{\left(y_i + \frac{h}{2}k_2\right) - 2\left(x_i + \frac{h}{2}\right)\left(y_i + \frac{h}{2}k_2\right)^2}{1+\left(x_i + \frac{h}{2}\right)} \\
k_4 &= \frac{(y_i + hk_3) - 2(x_i + h)(y_i + hk_3)^2}{1+(x_i + h)} \\
y_{i+1} &= y_i + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right).
\end{aligned}$$

Jeg har for RK4 beholdt parentesene for å prøve å beholde den visuelle likhetene med de generelle likningene som definerer metoden. Men det er strengt tatt ikke nødvendig. △

I dette eksemplet kan $y_{i+1}$ beregnes analytisk for de implisitte metodene, men merk at dette fort kan bli en smule håpløst om likningene ikke er kvadratiske i $y_{i+1}$, slik som her. Standardteknikken er da å slå til med en numerisk likningsløser.

Iterasjonen

$$y_{i+1} = y_i + h\frac{y_{i+1} - 2x_{i+1} y_{i+1}^2}{1+x_{i+1}}$$

allerede er på formen $y_{i+1} = g(y_{i+1})$, og dette gjør at fikspunktiterasjonen er et pedagogisk naturlig valg. Dersom $h$ er liten, blir gjerne $g'$ liten, og da husker du fra tidligere at fikspunktmetoden konvergerer ganske kjapt. Men hele poenget med implisitte metoder, er jo nettopp å slippe å måtte bruke små $h$, så Newtons metode er et bedre valg. Jeg skal allikevel bruke fikspunktmetoden for å spare litt kognitiv last på dere.

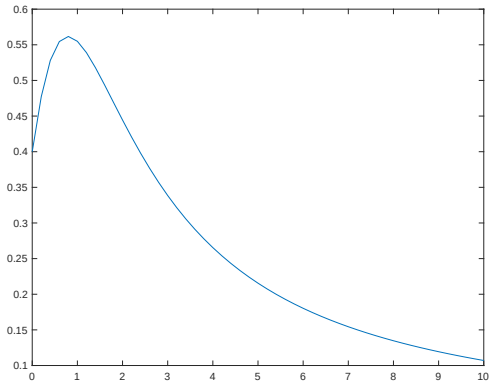
**Eksempel 5.29.** Vi løser

$$\dot{y} = \frac{y - 2xy^2}{1+x} \quad y(0) = \frac{2}{5}$$

med

$$y_{i+1} = y_i + h\frac{y_{i+1} - 2x_{i+1}y_{i+1}^2}{1 + x_{i+1}}.$$

Under er løsningskurve beregnet med $h = 0.2$ på intervallet $[0, 10]$. Fikspunktmetoden trengte med startgjetning $y_i$ et sted mellom 13 og 21 iterasjoner for å nå maskinpresisjon i hvert steg. Med lavere $h$ vil antall iterasjoner gå ned. △




Eksemplet over illustrerer et viktig moment. Koster det mange flyttallsoperasjoner å kjøre en metode til en gitt presisjon? Det hjelper ikke å ha en metode som beregner alt til maskinpresisjon om metoden tar ett år å kjøre. Implisitte metoder er ofte robuste og stabile, men de koster også mer å bruke.

# Vanskelig teori

Differensiallikningen vi skal løse er

$$\dot{y}(t) = f(y(t)) \quad t \in [a, b].$$

En *løsning* er en kontinuerlig deriverbar funksjon $y : [a, b] \to \mathbb{R}$ som passer i likningen. Dersom vi inkluderer en betingelse på formen

$$y(a) = c$$

har vi et initialverdiproblem. Om dette problemet kan løses, og hvorvidt det finnes flere løsninger, avhenger av $f$.

**Entydighet**

Dersom $f$ er kontinuerlig deriverbar på $\mathbb{R}$, og $y$ tilfredsstiller

$$\dot{y}(t) = f(y(t)) \quad y(a) = c,$$

er $y$ entydig bestemt.

*Bevis.* La oss anta at det finnes to kontinuerlig deriverbare funksjoner $y_1$ og $y_2$ slik at

$$y_1' = f(y_1) \quad y_1(a) = c$$

og

$$y_2' = f(y_2) \quad y_2(a) = c.$$

Vi ønsker å studere differansen $z = y_1 - y_2$. Hvis vi trekker de to initialverdiproblemene over fra hverandre, ser vi at funksjonen $z$ tilfredsstiller

$$z'(t) = \dot{y}_1(t) - \dot{y}_2(t) = f(y_1(t)) - f(y_2(t))$$

med initialverdibetingelse

$$z(a) = y_1(a) - y_2(a) = 0.$$

La oss først få rede på et par grunnleggende faktaopplysninger. Siden $z$ er kontinuerlig deriverbar på $[a, b]$, må $z$ ha maksimums- og minimumsverdier på $[a, b]$. La oss kalle disse $c$ og $d$.

Siden $f$ er kontinuerlig deriverbar, må $|f'|$ ha en maksimumsverdi $A$ på $[c, d]$. La nå $P$ være en jevn partisjon med punkter

$$t_i = a + i\frac{b-a}{n},$$

og velg $n$ slik at $A\frac{b-a}{n} < 1$. For det første gir sekantsetningen at

$$\left|\frac{f(y_1(t)) - f(y_2(t))}{y_1(t) - y_2(t)}\right| \leq A$$

på $[a, b]$, som gir at også

$$\begin{aligned} |z'(t)| &= |\dot{y}_1(t) - \dot{y}_2(t)| \\ &= |f(y_1(t)) - f(y_2(t))| \\ &< A|y_1(t) - y_2(t)| = A|z(t)| \end{aligned}$$

på $[a, b]$.

Nå ser vi på intervallet $[a, t_1]$. Siden $z$ er kontinuerlig deriverbar, må både $|z|$ og $|z'|$ ha maksimumsverdier på dette intervallet. Vi kaller disse $M_0$ og $M_1$, og ser at

$$|z(t)| \leq M_1(t_1 - a) = M_1\frac{b-a}{n}$$

på $[a, t_1]$ siden $z(a) = 0$. Men siden $|z'(t)| \leq A|z(t)|$, må $M_1 \leq M_0 A$, slik at

$$|z(t)| \leq M_1\frac{b-a}{n} \leq M_0 A\frac{b-a}{n}.$$

Men vi har valgt $n$ slik at

$$A\frac{b-a}{n} < 1.$$

Ulikheten

$$|z(t)| \leq M_0 A\frac{b-a}{n}$$

sier at på $[a, t_1]$ er $|z|$ mindre enn sin egen maksimumsverdi ganger et tall som er strengt mindre enn 1, og dette er kun mulig dersom $z = 0$ på $[a, t_1]$. Altså er $y_1$ og $y_2$ identiske på $[a, t_1]$.

Men dette betyr at $z(t_1) = 0$, og nå kan vi gjenta resonnementent på intervallet $[t_1, t_2]$, så på $[t_2, t_3]$ og så videre, og konkludere med at $y_1$ og $y_2$ er identiske på hele $[a, b]$. □

Det er relativt enkelt å utvide teoremet over til å gjelde for initialverdiproblemer av typen

$$\dot{y}(t) = f(t, y(t)) \quad y(a) = c.$$

Hvis du studerer beviset, vil du se at det vi trenger for teoremet skal være sant, er at det finnes en konstant $A$ slik at

$$\left| \frac{f(t,y) - f(t,z)}{y - z} \right| \leq A.$$

I dette tilfellet sier vi at $f$ er lipschitzkontinuerlig i det andre argumentet.

# Andre ordens lineære differensiallikninger

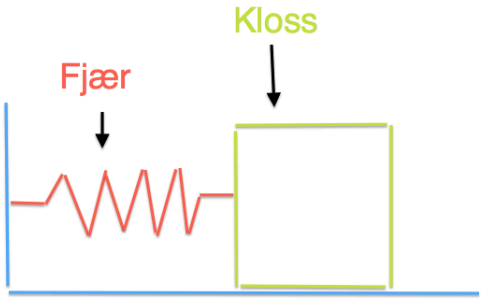
Nå skal vi behandle lineære andreordens differensiallikninger med konstante koeffisienter:

$$y'' + a_1 y' + a_0 y = f(t)$$

Det er vanlig å kreve at $y \in \mathcal{C}^2$, altså at $y$ har to kontinuerlige deriverte. På denne måten kan man sikre at likningen faktisk gir mening. Det finnes mange situasjoner der dette kravet kan slakkes noe, men det er pensum i M4.

### Hvor kommer andre ordens differensiallikninger fra?

En kloss sklir friksjonsfritt på underlaget, og er festet til veggen med en fjær. Hookes fjærlov sier at




$$F(y) = -ky,$$

der $y$ er hvor langt fjæren er strukket eller komprimert, $k$ er en konstant som avhenger av fjærens stivhet, og $F(y)$ er kraften fra fjæren på klossen. Dersom $y(t)$ er klossens posisjon, er klossens akselerasjon gitt ved $y''(t)$, og Newtons andre lov blir

$$-ky = my'',$$

der $m$ er klossens masse. Dette er en differensiallikning. Vi skriver vanligvis

$$my'' + ky = 0.$$

Vi kan komplisere det litt til. La oss innføre luftmotstand. Luftmotstand avhenger kvadratisk av farten:

$$F(y') = b(y')^2;$$

der $b$ er en proporsjonalitetskonstant som sier noe om luftmotstanden. Den totale kraften blir

$$F(y, y') = -ky + b(y')^2,$$

slik at Newtons andre lov gir

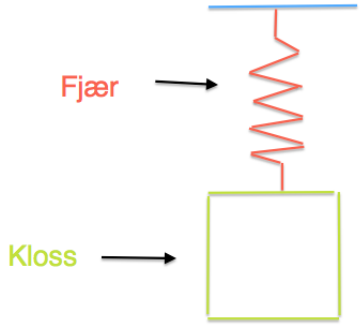
$$my'' - b(y')^2 + ky = 0.$$

Denne likningen har et problematisk ledd, $b(y')^2$. Men vi kan gjøre en forenkling. Dersom klossen ligger i en tyktflytende væske, blir motstanden proporsjonal med farten istedet for kvadratet av farten, og vi får likningen

$$my'' - cy' + ky = 0,$$

som er mye enklere å løse.

Nå skal vi komplisere det enda litt. La klossen henge fra taket. I tillegg til fjærkraften og luftmotstanden,




vil nå også gravitasjonen påvirke bevegelsen. Gravitasjonskraften er en konstant kraft $mg$ nedover. Den totale kraften er

$$F(y, y') = -ky + by' - mg,$$

og Newtons andre lov gir differensiallikningen

$$my'' - by' + ky = mg.$$

### Noen småting

Vi skal behandle *andre ordens differensiallikninger med konstante koeffisienter*:

$$a_2 y''(t) + a_1 y'(t) + a_0 y(t) = f(t)$$

Det er vanlig å sette $a_2 = 1$, for å forenkle analysen. Vi slipper å ha med $a_2$ i alle formler og utledninger, og vi slipper å luke ut $a_2 = 0$ hver gang vi skal sette opp et teorem. Dersom $a_2$ skulle slumpe til å være noe annet enn 1, kan du dele den ut av likningen før du setter igang.

Det er tradisjonelt og praktisk å sortere likninger i to kategorier, de homogene:

$$y''(t) + a_1 y'(t) + a_0 y(t) = 0$$

og de inhomogene:

$$y''(t) + a_1 y'(t) + a_0 y(t) = f(t)$$

### Løsningsteknikk for homogene likninger

Vi kaller gjerne løsningen av

$$y''(t) + a_1 y'(t) + a_0 y(t) = 0$$

for $y_h$, der $h$-en står for homogen. Det første man kan merke seg, er at vi har allerede lært å løse denne typen likning i forrige kapittel. Dersom vi innfører de nye variablene $v_1 = y$ og $v_2 = y'$, kan likningen skrives om til systemet

$$\begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}'$$

Vi vet derfor at vi kan forvente to lineært uavhengige løsninger. Det karakteristiske polynomet til matrisen er:

$$\lambda^2 + a_1\lambda + a_0.$$

Egenvektoren til $\lambda$ er:

$$\mathbf{x} = \begin{bmatrix} 1 \\ \lambda \end{bmatrix}.$$

Den karakteristiske likningen kjenner du forhåpentligvis igjen fra gymnaset, der du lærte å løse disse likningene. Den gang sa de noe sånt som at alle løsninger var på formen $Ce^{\lambda t}$ eller $A\cos t + B\sin t$, og så satte de dette uttrykket inn i differensiallikningen for å utlede den karakteristiske likningen.

Vi kan bruke analysen fra forrige oppgave til å liste opp løsningen til den homogene likningen for forskjellige typer egenverdier. Merk at vi er i utgangspunktet kun interessert i $v_1$, og derfor ikke har bruk for egenvektorene når vi skal skrive opp homogene løsninger.

Dersom $\lambda_1 \neq \lambda_2$ er reelle, kan vi plukke ut førstekomponenten av den generelle løsningen av systemet, og få

$$y_h(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}.$$

Dersom $\lambda = \alpha \pm \beta i$, slik at

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix}$$

og

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \beta \end{bmatrix},$$

får vi

$$y_h(t) = d_1 e^{\alpha t}\cos\beta t + d_2 e^{\alpha t}\sin\beta t.$$

Dersom $\lambda_1 = \lambda_2 = \lambda$, kan vi stokke litt om på verdiene til $c_1$ og $c_2$, og få

$$y_h(t) = c_1 e^{\lambda t} + c_2 t e^{\lambda t}.$$

**Eksempel 5.30.** Løsningen til

$$y'' - y = 0$$

er

$$y(t) = c_1 e^t + c_2 e^{-t}. \qquad \triangle$$

**Eksempel 5.31.** Løsningen til

$$y'' + y = 0$$

er

$$y(t) = c_1 \cos t + c_2 \sin t. \qquad \triangle$$

**Eksempel 5.32.** Løsningen til

$$y'' - 2y' + y = 0$$

er

$$y(t) = c_1 e^t + c_2 t e^t. \qquad \triangle$$

### Løsningsteknikk for inhomogene likninger

Tilsvarende kalles vi løsningen til

$$y''(t) + a_1 y'(t) + a_0 y(t) = f(t)$$

for $y_p$, der $p$ står for partikulær. La $y_h = y_1 + y_2$, der $y_1$ og $y_2$ er to lineært uavhengige homogene løsninger. På samme måte som i kapittelet om systemer av differensiallikninger, finnes det en formel for å finne løsningene til inhomogene systemer. Ved å sette inn de homogene løsningene i den, vil du få at partikulærløsningen kan beregnes på følgende måte:

$$\begin{aligned} y_p(t) = y_2 \int_0^t \frac{y_1(s)f(s)}{y_1(s)y_2'(s) - y_2(s)y_1'(s)}\, ds \\ - y_1 \int_0^t \frac{y_2(s)f(s)}{y_1(s)y_2'(s) - y_2(s)y_1'(s)}\, ds. \end{aligned}$$

Selve utledningen av formelen er ikke så interessant, så vi overlater det til spesielt interesserte.

**Eksempel 5.33.** Den homogene løsningen til

$$y'' - y = e^{2t}$$

er

$$y_h(t) = c_1 e^t + c_2 e^{-t} = c_1 y_1(t) + c_1 y_2(t),$$

slik at

$$\begin{aligned} y_p(t) &= y_2 \int \frac{y_1(t)f(t)}{y_1(t)y_2'(t) - y_2(t)y_1'(t)}\, dt \\ &\quad - y_1 \int \frac{y_2(t)f(t)}{y_1(t)y_2'(t) - y_2(t)y_1'(t)}\, dt \\ &= e^{-t} \int \frac{e^t e^{2t}}{-e^t e^{-t} - e^{-t}e^t}\, dt \\ &\quad - e^t \int \frac{e^{-t}e^{2t}}{-e^t e^{-t} - e^{-t}e^t}\, dt = \frac{1}{3}e^{2t}, \end{aligned}$$

og

$$y = y_h + y_p = c_1 e^t + c_2 e^{-t} + \frac{1}{3}e^{2t}. \qquad \triangle$$

**Eksempel 5.34.** Den homogene løsningen til

$$y'' - y = e^t$$

er

$$y_h(t) = c_1 e^t + c_2 e^{-t},$$

slik at

$$y_p(t) = e^{-t}\int \frac{e^t e^t}{-e^t e^{-t} - e^{-t}e^t}\,dt - e^t\int \frac{e^{-t}e^t}{-e^t e^{-t} - e^{-t}e^t}\,dt = \frac{1}{2}(t-1)e^t,$$

og

$$y = c_1 e^t + c_2 e^{-t} + \frac{1}{2}(t-1)e^t. \qquad \triangle$$

**Eksempel 5.35.** Den homogene løsningen til

$$y'' + y = \sin t$$

er

$$y_h(t) = c_1\cos t + c_2 \sin t,$$

slik at

$$\begin{aligned} y_p(t) &= \sin t\int \frac{\cos t\sin t}{\cos^2 t + \sin^2 t}\,dt \\ &\quad - \cos t\int \frac{\sin^2 t}{\cos^2 t + \sin^2 t}\,dt \\ &= \sin t\int \cos t\sin t\,dt - \cos t\int \sin^2 t\,dt \\ &= -\frac{1}{4}\sin t\cos 2t - \cos t\left(\frac{t}{2} - \frac{1}{4}\sin 2t\right) \\ &= -\frac{1}{2}t\cos t - \sin t. \end{aligned}$$

Merk at den homogene løsningen $y_1 = \sin x$ dukket opp i prosessen. Dette skjer av og til, men gjør ingenting. Vi har

$$y = c_1\cos t + c_2\sin t - \frac{1}{2}t\cos t. \qquad \triangle$$

Med litt trening vil du merke at du ikke alltid trenger å bruke den store formelen for å finne den partikulære løsningen. Med litt erfaring kan man gjette hva den er. En grundig diskusjon av dette er imidlertid kjedelig og langdryg, og det krever mye trening å se hva løsningen er i noen tilfeller. I andre tilfeller går det ikke an å gjette på partikulærløsningen, og også disse tilfellene krever mye erfaring å se med det blotte øye.

**Eksempel 5.36.** Det trengs ikke særlig rutine for å se at den inhomogene løsningen til

$$y'' + 2y = 1$$

må åpenbart være en konstant funksjon $f(x) = K$. Innsetting gir

$$0 + 2K = 1,$$

slik at $K = 1/2$. $\triangle$

**Eksempel 5.37.** Det trengs heller ikke særlig rutine for å se at den inhomogene løsningen til

$$y'' + y = e^t$$

er $Ke^t$. Innsetting gir $K$ på samme måte som i forrige eksempel. Det trengs derimot endel erfaring for å gjette at partikulærløsningen til

$$y'' + y = \sin t$$

er på formen $Kt\cos t$. Dette er fordi $\sin t$ er en homogen løsning. $\triangle$

### Initialverdiproblem

Til slutt kan vi registrere at den generelle løsningen til

$$a_2 y''(t) + a_1 y'(t) + a_0 y(t) = f(t)$$

er

$$y(t) = y_h(t) + y_p(t).$$

Merk at det finnes to ubestemte koeffisienter i $y_h$, så et initialverdiproblem trenger to betingelser - den vanligste formen er

$$y(t_0) = a, \quad y'(t_0) = b.$$

**Eksempel 5.38.** Løsningen til initialverdiproblemet

$$y'' - y = e^{2t}$$

med betingelser

$$y(0) = 1, \quad y'(0) = 0$$

er

$$y = \frac{2}{3}e^{-t} + \frac{1}{3}e^{2t}. \qquad \triangle$$

### Det ikkediagonaliserbare tilfellet

Dersom $a_1^2 = 4a_0$, slik at likningen

$$\lambda^2 + a_1\lambda + a_0 = 0$$

kun har en løsning, er vi i det ikkediagonaliserbare tilfellet fra forrige kapittel. Men i dette kapitlet finnes det et triks.

**Eksempel 5.39.** Vi ser på likningen

$$y'' + 2y + y = 0.$$

Den karakteristiske likningen er

$$\lambda^2 + 2\lambda + 1 = 0$$

som kun har løsningen $\lambda = -1$. Merk nå at

$$(ye^t)'' = e^t(y'' + 2y + y) = 0.$$

Dersom vi integrerer denne likningen to ganger, får vi

$$y = e^{-t}(c_1 t + c_0),$$

slik som i avsnittet om generaliserte egenvektorer i forrige kapittel. $\triangle$

Teknikken over fungerer generelt. La oss anta at den karakteristiske likningen

$$\lambda^2 + a_1\lambda + a_0 = 0$$

kun har en rot. I så fall er

$$a_1^2 - 4a_0 = 0$$

slik at den karakteristiske likningen kan skrives

$$\lambda^2 + a_1\lambda + \frac{a_1^2}{4} = 0.$$

Den doble roten er $\lambda = -\frac{a_1}{2}$. Prøv selv.

Numeriske metoder for andre ordens differensiallikninger venter vi litt med. Det er mer praktisk å ta når vi har lært om systemer av differensiallikninger.

# Kapittel 6

# Laplacetransform

Laplacetransform er en snedig teknikk for å løse ordinære differensiallikninger. La $x : [0, \infty) \to \mathbb{C}$ være en funksjon.

**Laplacetransformen til $x$ er**

$$X(s) = \mathcal{L}(x) = \int_0^\infty x(t)e^{-st}\,dt$$

der $s = \sigma + i\omega$.

Det er klart at laplacetransformen er en lineæroperator. Først kan man spørre seg hvilke funksjoner det er naturlig å finne laplacetransformen til. Vi noterer oss at $\mathcal{L}(x)$ er definert ved et uegentlig integral, og at dette bør konvergere.

**Hva kan du putte inn?**

La $x$ en stykkvis kontinuerlig funksjon, og la $a$ og $M > 0$ være reelle tall slik at

$$|x(t)| \leq Me^{at} \quad \text{for} \quad t \geq 0.$$

Integralet

$$\int_0^\infty x(t)e^{-st}\,dt.$$

konvergerer absolutt dersom $\sigma > a$.

Dette er ikke så vanskelig å tro på:

$$\left|\int_0^\infty x(t)e^{-st}\,dt\right| \leq \int_0^\infty |x(t)|e^{-st}\,dt \leq$$
$$M\int_0^\infty e^{at}e^{-st}\,dt \leq M\int_0^\infty e^{(a-s)t}\,dt = \frac{M}{s-a}.$$

**Eksempel 6.1.** Vi kan ikke beregne

$$\mathcal{L}(e^{t^2}) = \int_0^\infty e^{t^2-st}\,dt,$$

for $e^{t^2}$ vokser for fort. Siden

$$\lim_{t\to\infty} e^{t^2-st} = \infty$$

kan integralet ikke konvergere. △

**Eksempel 6.2.** Så lenge $s > a$, kan vi fint beregne

$$\mathcal{L}(e^{at}) = \int_0^\infty e^{(a-s)t}\,dt = \frac{1}{s-a}.$$ △

**Eksempel 6.3.** Vi kan også beregne

$$\mathcal{L}(0) = \int_0^\infty 0e^{-st}\,dt = 0.$$ △

Heretter skal vi alltid anta at $s$ er valgt slik at integralet konvergerer. Vi skal også anta $x$ er stykkvis kontinuerlig og tilfredsstiller $|x| \leq Me^{at}$. Stykkvis kontinurelig betyr at $x$ kan ha maksimalt et endelig antall sprang.

**Eksempel 6.4.** La $x(t) = \cos t$.

$$\begin{aligned}\mathcal{L}(x) &= \int_0^\infty e^{-st}\cos t\,dt \\ &= \frac{1}{2}\int_0^\infty \left(e^{it} + e^{-it}\right)e^{-st}\,dt \\ &= \frac{1}{2}\int_0^\infty e^{(i-s)t} + e^{-(i+s)t}\,dt \\ &= \frac{1}{2}\left(\frac{1}{s-i} + \frac{1}{s+i}\right) = \frac{s}{s^2+1}.\end{aligned}$$

Her har vi brukt en omskrivning av Eulers formel:

$$\cos t = \frac{e^{it} + e^{-it}}{2}.$$ △

**Eksempel 6.5.** Og nå $x(t) = \cosh t$.

$$\begin{aligned}\mathcal{L}(x) &= \mathcal{L}\left(\frac{e^t + e^{-t}}{2}\right) \\ &= \frac{1}{2}\left(\mathcal{L}(e^t) + \mathcal{L}(e^{-t})\right) \\ &= \frac{1}{2}\left(\frac{1}{s-1} + \frac{1}{s+1}\right) = \frac{s}{s^2-1}.\end{aligned}$$ △

La $\sigma > 0$. Vi beregner

$$\begin{aligned}\mathcal{L}(\dot{x}) &= \int_0^\infty \dot{x}(t)e^{-st}\,dt \\ &= x(t)e^{-st}\Big|_0^\infty + s\int_0^\infty x(t)e^{-st}\,dt \\ &= s\mathcal{L}(x) - f(0).\end{aligned}$$

**Derivasjonsregelen**

Dersom $s > 0$ og $|\dot{x}(t)| \leq Me^{at}$ er stykkvis kontinuerlig, er

$$\mathcal{L}(\dot{x}) = s\mathcal{L}(x) - f(0).$$

**Eksempel 6.6.**

$$\begin{aligned}\mathcal{L}(\sinh t) &= s\mathcal{L}(\cosh t) + \cosh(0)\\ &= -s\frac{s}{s^2-1} + 1 = \frac{1}{s^2+1}.\end{aligned}$$ △

**Eksempel 6.7.** Vi beregner først

$$\mathcal{L}(1) = \int_0^\infty e^{-st}\,dt = \frac{1}{s}.$$

La nå $x(t) = t^n$. Den $n$-te deriverte av $x$ er:

$$f^n(t) = n!$$

og siden $f^k(0) = 0$ uansett $k$, får vi

$$n!\mathcal{L}(1) = s^n\mathcal{L}(t^n),$$

slik at

$$\mathcal{L}(t^n) = \frac{n!}{s^{n+1}}.$$ △

Vi kan lage en regneregel for motsatt vei også. La $\sigma > 0$. Vi beregner

$$\begin{aligned}\mathcal{L}(g) &= \int_0^\infty g(t)e^{-st}\,dt\\ &= -\frac{1}{s}g(t)e^{-st}\Big|_0^\infty + \frac{1}{s}\int_0^\infty x(t)e^{-st}\,dt\\ &= \frac{1}{s}\mathcal{L}(x).\end{aligned}$$

**Integrasjonsregelen**

La $g(t) = \int_0^t f(u)\,du$. Dersom $g$ tilfredsstiller vekstkravet, har vi:

$$\mathcal{L}(g) = \frac{1}{s}\mathcal{L}(x)$$

**Eksempel 6.8.**

$$\mathcal{L}(\delta(t-a)) = \int_0^\infty \delta(t-a)e^{-st}\,dt = e^{-as}$$ △

**Eksempel 6.9.**

$$\begin{aligned}\mathcal{L}(\sin t) &= \frac{1}{s}\mathcal{L}(\cos t)\\ &= \frac{1}{s}\frac{s}{s^2+1} = \frac{1}{s^2+1}.\end{aligned}$$ △

**Eksempel 6.10.**

$$\begin{aligned}\mathcal{L}\left((t^2-1)e^{-t^2/2}\right) &= s^2\mathcal{L}\left(e^{-t^2/2}\right) - s\\ &= -s\frac{s}{s^2+1} + 1 = \frac{1}{s^2+1}.\end{aligned}$$ △

**En regel til**

La $g(t) = tx(t)$ og $\mathcal{L}(x) = F$. Anta at $x$ er deriverbar. Da har vi:

$$\mathcal{L}(g) = -\dot{x}$$

Dersom $x$ er stykkvis kontinuerlig og tilfredsstiller vekstkravet, gjelder dette også for $g$. Vi beregner

$$\begin{aligned}\mathcal{L}(g) &= \int_0^\infty tx(t)e^{-st}\,dt = \int_0^\infty -\frac{d}{ds}x(t)e^{-st}\,dt\\ &= -\frac{d}{ds}\int_0^\infty x(t)e^{-st}\,dt = -\dot{x}.\end{aligned}$$

**Eksempel 6.11.**

$$\mathcal{L}(t\sin t) = -\frac{d}{ds}\frac{1}{s^2+1} = \frac{2s}{(s^2+1)^2}.$$ △

**En regel til II**

La $g(t) = \frac{1}{t}x(t)$, la $\mathcal{L}(x) = F$, og $G(s) = \int_s^\infty F(u)\,du$. Da har vi:

$$\mathcal{L}(g) = G$$

Vi beregner

$$\begin{aligned}\mathcal{L}(g) &= \int_0^\infty \frac{1}{t}x(t)e^{-st}\,dt\\ &= -\int_s^\infty\int_0^\infty x(t)e^{-st}\,dt\,ds = G.\end{aligned}$$

Dersom $x(t)$ er stykkvis kontinuerlig og tilfredsstiller vekstkravet, gjelder dette også for $e^{at}x(t)$, slik at

$$\begin{aligned}\mathcal{L}(g) &= \int_0^\infty e^{at}x(t)e^{-st}\,dt\\ &= \int_0^\infty x(t)e^{at-st}\,dt\\ &= \int_0^\infty x(t)e^{-(s-a)t}\,dt = F(s-a).\end{aligned}$$

Dette kalles '$s$-skift'.

**En skifteregel**

La $g(t) = e^{at}x(t)$, $\mathcal{L}(x) = F$ og $\mathcal{L}(g) = G$. Da har vi

$$G(s) = F(s-a).$$

**Eksempel 6.12.**

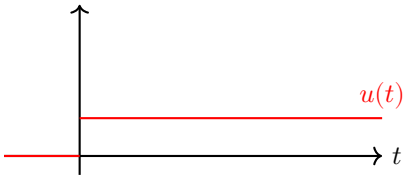
$$\begin{aligned}\mathcal{L}\left(e^{at}\sin(t)\right) &= \frac{1}{(s-a)^2+1}\\ \mathcal{L}\left(e^{at}\cos(t)\right) &= \frac{s-a}{(s-a)^2+1}\end{aligned}$$ △

## To viktige funksjoner

Husk at heavisidefunksjonen er gitt ved

$$u(t) = \begin{cases}0 & \text{for } t < 0\\ 1 & \text{for } t \geq 0.\end{cases}$$

Man kan tenke på denne som en funksjon som slår noe på ved $t = 0$.

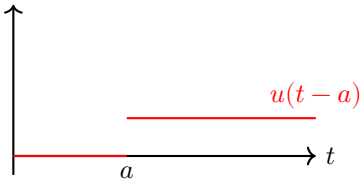



**Eksempel 6.13.**

$$u(t)e^t = \begin{cases} 0 & \text{for } t < 0 \\ e^t & \text{for } t \geq 0 \end{cases}$$ △

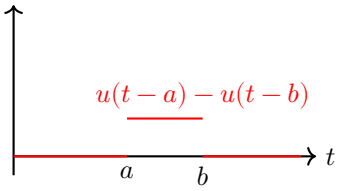
Vi kan slå på ved tiden $t = a$ istedet:

$$u(t-a) = \begin{cases} 0 & \text{for } t < a \\ 1 & \text{for } t \geq a, \end{cases}$$




Vi kan også slå på ved $t = a$ og av igjen ved $t = b$:

$$u(t-a) - u(t-b) = \begin{cases} 0 & \text{for } t < a \\ 1 & \text{for } a \leq t < b \\ 0 & \text{for } t \geq b. \end{cases}$$




**Eksempel 6.14.**

$$(u(t-a) - u(t-b))\, e^t = \begin{cases} 0 & \text{for } t < a \\ e^t & \text{for } a \leq t < b \\ 0 & \text{for } t \geq b \end{cases}$$ △

**Eksempel 6.15.** La $x(t) = u(t-a)$.

$$\mathcal{L}(x) = \int_a^\infty e^{-st}\, dt = \frac{e^{-sa}}{s}$$ △

Følgende teorem kalles $t$-skift.

**En skifterege til**

La $g(t) = f(t-a)u(t-a)$, $\mathcal{L}(x) = F$ og $\mathcal{L}(g) = G$. Vi har

$$G(s) = e^{-as}F(s).$$

Derfor:

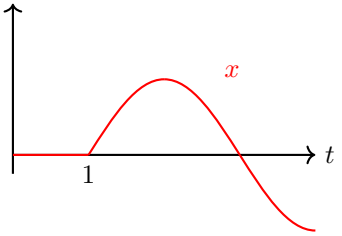
$$\begin{aligned} G(s) &= \int_0^\infty u(t-a)f(t-a)e^{-st}\, dt \\ &= \int_a^\infty f(t-a)e^{-st}\, dt \\ &= \int_0^\infty f(v)e^{-s(v+a)}\, dv \\ &= e^{-sa}\int_0^\infty f(v)e^{-sv}\, dv = e^{-as}F(s). \end{aligned}$$

**Eksempel 6.16.** La

$$\begin{aligned} x(t) &= u(t-1)\sin(t-1) \\ &= \begin{cases} 0 & \text{for } t < 1 \\ \sin(t-1) & \text{for } t \geq 1 \end{cases} \end{aligned}$$

Vi beregner

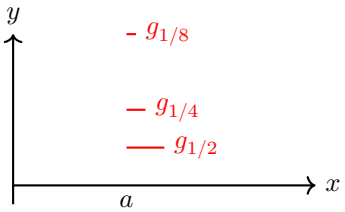
$$\mathcal{L}(x) = \frac{e^{-s}}{s^2+1}.$$ △




La

$$g_k(t-a) = \begin{cases} 1/k & \text{for } a < t < a+k \\ 0 & \text{ellers} \end{cases}$$

Slik ser de ut:




Vi definerer

$$\delta(t) = \lim_{k\to 0} g_k(t) = \begin{cases} \infty & \text{for } t = a \\ 0 & \text{ellers} \end{cases}$$

Deltafunksjonen brukes til å modellere impuls, altså energitilførsler der den påtrykte kraften er ekstremt høy og ekstremt kortvarig, for eksempel hammerslag. Man kan også tenke på den som noe som plukker ut funksjonsverdier:

$$\int_0^\infty f(t)\delta(t-a)\, dt = \lim_{k\to 0}\frac{1}{k}\int_a^{a+k} f(t)\, dt = f(a).$$

**Eksempel 6.17.**

$$\mathcal{L}(\delta(t-a)) = \int_0^\infty \delta(t-a)e^{-st}\, dt = e^{-as}$$ △

Deltafunksjonen er strengt tatt ikke noe funksjon i ordets rette forstand, og heavisidefunksjonen er ikke deriverbar. Det kan allikevel være fruktbart å tenke på deltafunksjonen som et forsøk på å sette opp den deriverte til heavisidefunksjonen.

### Konvolusjon og laplacetransform

En konvolusjon mellom $x$ og $g$, er et integral på formen

$$f * g = \int f(p)g(t-p)\,dp.$$

For laplacetransfrom bruker vi konvolusjonen

$$f * g = \int_0^t f(p)g(t-p)\,dp$$

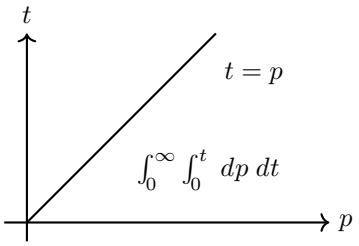
Når vi løser differensiallikninger med laplacetransform, kan vi støte på produkter av laplacetransformer. Følgende teorem, som kalles konvolusjonsteoremet, forteller oss hvordan vi skal nøste opp i et slikt produkt.

**Konvolusjonsteoremet**

Anta at både $\mathcal{L}(f*g)$, $\mathcal{L}(x)$ og $\mathcal{L}(g)$ eksisterer. Da er

$$\mathcal{L}(f * g) = \mathcal{L}(x)\,\mathcal{L}(g).$$

Vi laplacetransformerer $f * g$ og bytter integrasjonsvariable, slik som i M2. Hjelpefiguren under kan være grei å ha i bakhodet.




$$\begin{aligned}
\mathcal{L}(f*g) &= \int_0^\infty \int_0^t f(p)g(t-p)\,dp\,e^{-st}dt \\
&= \int_0^\infty \int_p^\infty f(p)g(t-p)e^{-st}\,dt\,dp \\
&= \int_0^\infty \int_0^\infty f(p)g(u)e^{-s(u+p)}\,du\,dp \\
&= \int_0^\infty f(p)e^{-sp}\,dp \int_0^\infty g(u)e^{-su}\,du \\
&= \mathcal{L}(x)\mathcal{L}(g)
\end{aligned}$$

### Regneeksempler

Oppskriften for å løse differensiallikninger med laplacetransform, er alltid å laplacetransformere hele differensiallikningen, bruke regnereglene, løse for laplacetransformen til den ukjente, og inverstransformere. Vi begynner med to elementære eksempler.

**Eksempel 6.18.**

$$y' + y = 0 \qquad y(0) = 1.$$

Vi laplacetransformerer likningen. Venstresiden blir

$$\begin{aligned}
\mathcal{L}(y' + y) &= \mathcal{L}(y') + \mathcal{L}(y) \\
&= s\mathcal{L}(y) - 1 + \mathcal{L}(y) \\
&= (s+1)\mathcal{L}(y) - 1.
\end{aligned}$$

Merk bruken av initialbetingelsen $y(0) = 1$. Høyresiden blir

$$\mathcal{L}(0) = 0.$$

Etter laplacetransformering står vi igjen med den algebraiske likningen

$$(s+1)\mathcal{L}(y) - 1 = 0,$$

slik at

$$\mathcal{L}(y) = \frac{1}{s+1}.$$

Dette er en laplacetransform vi har beregnet tidligere:

$$y(t) = e^{-t}. \qquad \triangle$$

**Eksempel 6.19.** Vi løser likningen

$$y'' + y = 0 \quad y(0) = 1\ y'(0) = 0.$$

Vi laplacetransformerer likningen

$$\mathcal{L}(y'') + \mathcal{L}(y) = s^2\mathcal{L}(y) - s + \mathcal{L}(y) = 0,$$

slik at

$$\mathcal{L}(y) = \frac{s}{s^2+1},$$

og

$$y(t) = \cos t. \qquad \triangle$$

Følgende eksempel illustrerer derivasjonsregelen 6.

**Eksempel 6.20.** Vi løser initialverdiproblemet

$$ty' - 2y = 0 \quad y(0) = 0$$

Vi laplacetransformerer likningen

$$\mathcal{L}(ty') - 2\mathcal{L}(y) = 0,$$

og lar $Y = \mathcal{L}(y)$. Først kan vi skrive

$$\mathcal{L}(ty') = -\frac{d}{ds}\left(\mathcal{L}(y')\right) = -\frac{d}{ds}\left(sY\right) = -Y - sY'.$$

Setter vi denne inn i likningen, får vi

$$-Y - sY' - Y = -sY' - 3Y = 0.$$

eller

$$sY' + 3Y = 0.$$

Dette er en differensialliking for $Y$. Integrerende faktor er $s^2$, slik at

$$(s^3Y)' = 0.$$

og

$$Y = \frac{C}{s^3}$$

Vi inverstransformerer og endrer litt på konstanten:

$$y = Ct^2$$

og bruker initialverdibetingelsen for andre gang:

$$y = t^2. \qquad \triangle$$

De tre foregående eksemplene løses enklere med metodene du kan fra før. Men disse teknikkene kommer til kort i eksemplene vi skal ta nå. Her er et par gamle eksamensoppgaver.

**Eksempel 6.21.** Vi løser

$$y'' + y = \delta(t-a) - \delta(t-b)$$

med initialbetingelser

$$y(0) = y'(0) = 0.$$

Vi får

$$(s^2+1)\mathcal{L}(y) = e^{-as} - e^{-bs},$$

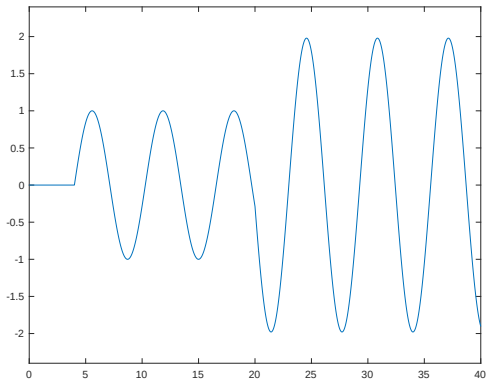
slik at

$$\mathcal{L}(y) = \frac{e^{-as}}{(s^2+1)} - \frac{e^{-bs}}{(s^2+1)}.$$

Vi bruker $t$-skift, og får

$$y(t) = \sin(t-a)u(t-a) - \sin(t-b)u(t-b).$$

Ligningens venstreside beskriver en kloss og en fjær på friksjonsfritt underlag. Ved tiden $t = 0$ er systemet i ro. Ved tiden $t = a$ blir klossen dengt av en hammer fra venstre, som gir opphav til svingningen $\sin(t-a)$. Ved tiden $t = b$ blir klossen dengt av en hammer fra høyre. Dette slaget gir en ny svingning $\sin(t-b)$ som legges oppå den gamle svingningen. △




Noen ganger kan det være komplisert å inverstransformere. I det neste eksemplet må man bruke både $t$-skift og $s$-skift.

**Eksempel 6.22.**

$$y'' + 2y' + y = \delta(t-1) \qquad y(0) = y'(0) = 0.$$

Vi transformerer ligningen til

$$s^2\mathcal{L}(y) + 2s\mathcal{L}(y) + \mathcal{L}(y) = \mathcal{L}(\delta(t-1))$$

slik at

$$\mathcal{L}(y) = \frac{e^{-s}}{(s+1)^2}.$$

Her lukter det $s$-skift på grunn av $(s+1)$, og $t$-skift på grunn av $e^{-s}$. Formelen for $s$-skift gir

$$\mathcal{L}(te^{-t}) = \frac{1}{(s+1)^2}.$$

Vi bruker i tillegg $t$-skift, slik at

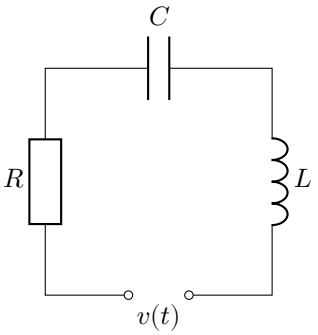
$$y = (t-1)e^{-(t-1)}u(t-1). \qquad \triangle$$

Elektrofolk er forresten veldig glade i laplacetransform. Det er nok fordi de trenger spenning som kan slås av og på. Heavisidefunksjonen gir dem en mulighet for dette, og laplacetransform takler heavisidefunksjonen med letthet.

**Eksempel 6.23.** Strømmen $i(t)$ i kretsen under tilfredsstiller integro-differensialligningen

$$Li'(t) + Ri(t) + \frac{1}{C}\int_0^t i(\tau)\,\mathrm{d}\tau = \delta(t-1),$$

der $R = 2$, $L = 1$, $C = 0.5$ og $\delta$ er Diracs deltafunksjon. Vi setter $i(0) = 0$ og finner strømmen $i(t)$.




Vi laplacetransformerer likningen, og får

$$s\mathcal{L}(i) + 2\mathcal{L}(i) + \frac{2}{s}\mathcal{L}(i) = e^{-s},$$

og løser vi for $\mathcal{L}(i)$, får vi

$$\mathcal{L}(i) = \frac{se^{-s}}{s^2+2s+2}.$$

Her er det enklest å fullføre kvadratet $s^2 + 2s + 2 = (s+1)^2 + 1$, og så skrive

$$\mathcal{L}(i) = \frac{(s+1)e^{-s}}{(s+1)^2+1} - \frac{e^{-s}}{(s+1)^2+1}.$$

Formelen for $s$-skift gir

$$\mathcal{L}(e^{-t}\cos t) = \frac{s+1}{(s+1)^2+1}$$

og

$$\mathcal{L}(e^{-t}\sin t) = \frac{1}{(s+1)^2+1}.$$

Inverstransformerer vi

$$\mathcal{L}(i) = \frac{(s+1)e^{-s}}{(s+1)^2+1} - \frac{e^{-s}}{(s+1)^2+1}.$$

med $t$-skift, får vi altså

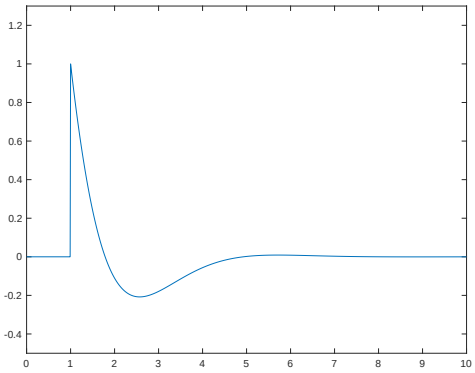
$$i(t) = e^{-(t-1)}u(t-1)\left(\cos(t-1) - \sin(t-1)\right).$$

Dette eksemplet modellerer en fugl som ved tiden $t = 1$ setter seg på en høyspentledning. Merk at strømmen er null frem til $t = 1$. I $t = 1$ opplever fuglen en voldsom påtrykt spenning i det den setter seg på høyspenten. Den faller umiddelbart av, slik at spenningsøkningen bare varer et kort øyeblikk. For $t > 1$ beskriver $i$ strømmen i fuglens kropp mens den daler ned mot bakken. Se figuren under.

I denne oppgaven kunne vi også brukt delbrøksoppspaltning:

$$\frac{s}{s^2+2s+2} = \frac{s}{(s-1-i)(s-1+i)}$$

men regningen blir noe svineri. △



Nå tar vi to eksempler der vi bruker konvolusjon.

**Eksempel 6.24.** Vi løser initialverdiproblemet

$$y + t * y' = t \quad y(0) = 0$$

Vi laplacetransformerer

$$\mathcal{L}(y) + \mathcal{L}(t * y') = \frac{1}{s^2}$$

og bruker konvolusjonsteoremet

$$\mathcal{L}(y) + \mathcal{L}(t)\mathcal{L}(y') = \mathcal{L}(y) + \frac{1}{s}\mathcal{L}(y) = \frac{1}{s^2}$$

eller

$$s\mathcal{L}(y) + \mathcal{L}(y) = \frac{1}{s}$$

Herfra er regningen standard, og vi får

$$\mathcal{L}(y) = \frac{1}{s(s+1)} = \frac{1}{s} - \frac{1}{s+1},$$

slik at

$$y = 1 - e^{-t}.$$ △

**Eksempel 6.25.** La oss løse initialverdiproblemet

$$y'' + y = \sin t \qquad y(0) = y'(0) = 0$$

Vi laplacetransformerer alt, og får

$$(s^2+1)\mathcal{L}(y) = \frac{1}{s^2+1}$$

eller

$$\mathcal{L}(y) = \frac{1}{(s^2+1)^2} = (\mathcal{L}(\sin t))^2$$

som gir at

$$y(t) = \int_0^t \sin(t-u)\sin u \, du$$

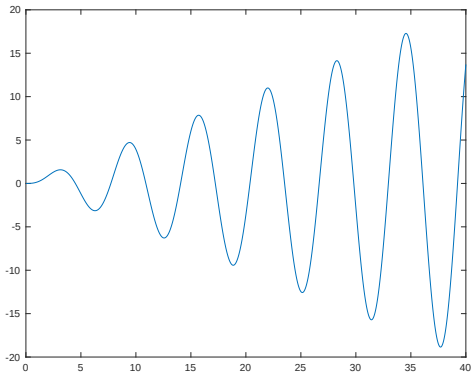
Dette siste integralet kan vi løse ved å bruke relasjonen

$$\sin u = \frac{e^{iu} - e^{-iu}}{2i}$$

og integrere i vei

$$\begin{aligned} y(t) &= \int_0^t \sin(t-u)\sin u \, du \\ &= -\frac{1}{4}\int_0^t (e^{i(t-u)} - e^{-i(t-u)})(e^{iu} - e^{-iu}) \, du \\ &= -\frac{1}{4}\int_0^t e^{it} - e^{i(t-2u)} - e^{-i(t-2u)} + e^{-it} \, du \\ &= -\frac{1}{2}\int_0^t \cos t - \cos(t-2u) \, du \\ &= \frac{1}{2}(\sin t - t\cos t) \end{aligned}$$

Dette eksemplet beskriver for eksempel en kloss og en fjær, der den påtrykte kraften $\sin t$ har samme frekvens som systemets naturlige svingefrekvens. Da oppstår det resonans. Resonansen kommer til uttrykk gjennom leddet $t\sin t$, som vokser mot uendelig når $t \to \infty$. Når dette skjer i et PA-anlegg, kalles det feedback, og alle må holde seg for ørene. △



Eksemplet over viser at selv om det er litt regning med laplacetransform, er det en penere løsningsmetode enn å dele opp i homogen og partikulær løsning, og så gjette i vei. Eksemplet illustrerer også en teknikk vi får bruk for når vi skal løse varmelikningen på hele $x$-aksen i kapittel 5.

Fouriertransform er egentlig et spesialtilfelle av laplacetransform, men vi kommer ikke til å komme så langt at vi ser det. Fourierrekker og fouriertransform er også to spesialtilfeller av en mer generell konstruksjon, men vi kommer ikke til å se det heller.

Vi skal senere bruke fouriertransform til å løse differensiallikninger på akkurat samme måte som vi gjorde med laplacetransform, men fouriertransform er teknisk vanskeligere å håndtere, og vi skal løse likninger som er mer kompliserte. I dette kapitlet skal vi gå gjennom fouriertransformens grunnleggende egenskaper.

# Kapittel 7

# Abstrakt lineæralgebra

I forrige semester lærte vi å løse lineære likningssystemer på en systematisk måte. Men det er ikke nødvendigvis prosessen med å løse likningssystemene som er viktigst. Dette kan en datamaskin gjøre på hundrevis av forskjellige måter. Det som er viktig for oss, er selve vektorregningen. Vektorrom er en algebraisk struktur som er oppfunnet med tanke på systematisering av vektorregning. Det er nemlig mange ting som ikke ser ut som vektorregning ved første øyekast, men som oppfører seg slik allikevel.

## Motivasjon

La oss begynne med

**Det første store spørsmålet**

Hva er en vektor?

Svaret på dette spørsmålet avhenger av hvem du spør. Noen vil si at det er en pil med lengde og retning som kan brukes til å beskrive for eksempel fart. Noen vil si at det er en liste med tall. En matematiker vil si at det er noe som tilfredsstiller aksiomene for vektorrom, mens en ingeniør vil kanskje si at det er noe du ikke trenger bry deg særlig med så lenge du forstår superposisjonsprinsippet. Alle har rett.

Vektorromskonseptet har som formål å binde sammen ting som ser forskjellig ut, men som oppfører seg veldig likt på noen måter. Her kommer et eksempel. La oss ta følgende vektor i $\mathbb{R}^2$:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

og skrive den slik:

$$\mathbf{x} = x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Du kan tenke på $\mathbb{R}^2$ som alle lineærkombinasjoner av disse to vektorene. Løsningene til

$$\ddot{x} + x = 0$$

er likeledes alle lineærkombinasjoner på formen

$$x(t) = c_1 \cos t + c_2 \sin t.$$

Disse to mengdene likner i den forstand at begge består av alle lineærkombinasjoner av to ting.

**Den folkelige definisjonen**

Et vektorrom er alle lineærkombinasjoner av en lineært uavhengig vektormengde.

La oss ta et spørsmål til.

**Det andre store spørsmålet**

Hva er en koordinat?

Det er ingenting spesielt med titallssystemet. Vi har valgt det fordi vi har ti fingre, men det er ingenting feil med andre tallsystemer. En moderne datamaskin er for eksempel utenkelig uten totallssystemet, siden all informasjon lagres med små brytere som står enten av eller på. Det vi tenker på som 7 heter 111 i totallssystemet, men det er bare navnet som er forskjellig. Syv epler er syv epler.

For vektorer finnes en tilsvarende problemstilling. La

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Koordinatene til $\mathbf{x}$ er $(1, 2)$, for

$$\mathbf{x} = 1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Men vi kunne like gjerne basert alt på

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

siden alle vektorer i $\mathbb{R}^2$ også kan skrives som en entydig lineærkombinasjon av disse to:

$$\mathbf{x} = 2 \cdot \begin{pmatrix} 2 \\ 3 \end{pmatrix} - 1 \cdot \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

Forskjellen er at det punktet vi prater om, altså det punktet du til vanlig tenker på som

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

har koordinater enten $(1, 2)$ eller $(2, -1)$, alt etter hvilke to vektorer vi baserer alt på. Men det er fortsatt det samme punktet, vi har bare skrevet det som to forskjellige lineærkombinasjoner. og vektene i denne lineærkombinasjonen er det vi tenker på som koordinatene til punktet.

Den samme situasjonen støter vi på for differensialligningen

$$\ddot{x} + x = 0.$$

Løsningene er alle lineærkombinasjoner på formen

$$x(t) = c_1 \cos t + c_2 \sin t,$$

men vi kan jo like gjerne skrive dem som lineærkombinasjoner på formen

$$x(t) = d_1 e^{it} + d_2 e^{-it}.$$

En av løsningene, for eksempel

$$x(t) = 2\cos t + 2\sin t,$$

er den samme som

$$x(t) = (1-i)e^{it} + (1+i)e^{-it},$$

de er bare skrevet på litt forskjellig måte. Hvis vi har lyst til å være skikkelig profesjonelle, kan vi bruke initialverdiene fra et eventuelt initialverdiproblem som koordinater, og skrive

$$x(t) = x_0 \cos t + v_0 \sin t,$$

der $x(0) = x_0$ og $\dot{x}(0) = v_0$, altså startposisjon og startfart.

På samme måte som totallssystemet i noen tilfeller er vesentlig bedre enn titallssystemet, er det i noen situasjoner koordinatsystemer som er bedre enn andre. Vi har allerede sett at det er lettere å regne med komplekse eksponensialfunksjoner en sinus og cosinus, og at man før eller siden må forholde seg til forskjellige koordinatsystemer, er ikke noe å lure på. Det er fremtiden, og dette vil du oppdage i løpet studiet. Så la oss si at vi løsriver oss fra koordinatsystemet vi er vant til. I så fall ser det ut som om koordinatbegrepet henger sammen med konseptet lineærkombinasjon, siden koordinater ser ut til å være det samme som vektene i en lineærkombinasjon.

**Det tredje store spørsmålet**

Finnes det et matematisk konsept som binder sammen disse greiene?

Svaret er ja. Det heter vektorrom. Noen elsker det og noen hater det. En funksjon kan tenkes på som en vektor. Følg med i neste bolk.

# De ti bud

Anta at vi har to vektorer $\mathbf{v}$ og $\mathbf{w}$, og to skalarer $a$ og $b$. Lineærkombinasjonen av $\mathbf{v}$ og $\mathbf{w}$ med vekter $a$ og $b$, er

$$a\mathbf{v} + b\mathbf{w}.$$

Denne er satt sammen av vektoraddisjon

$$\mathbf{v} + \mathbf{w}$$

og skalarmultiplikasjon

$$a\mathbf{v}.$$

Disse to operasjonene har du jobbet med siden videregående skole, og er sentrale for vektorrom.

**De ti bud**

La $\mathbf{V}$ være en (ikke tom) mengde med vektorer som kan adderes og skalarmultipliseres, og $\mathbb{K}$ en kropp. Vi sier at mengden $\mathbf{V}$ er et vektorrom over $\mathbb{K}$ dersom følgende er tilfredsstilt:

**Tre aksiomer om vektorrommets innhold:**

For alle $\mathbf{v}, \mathbf{w} \in V$ må

$$\mathbf{v} + \mathbf{w} \in V.$$

For alle $\mathbf{v} \in V$ må $a \in \mathbb{K}$ må

$$a\mathbf{v} \in V.$$

Det skal finnes en vektor, kalt $\mathbf{0}$, slik at

$$\mathbf{v} + \mathbf{0} = \mathbf{v}$$

for alle $\mathbf{v} \in V$.

**Syv aksiomatiske regneregler:**

For alle $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$, $a, b \in \mathbb{K}$: Addisjonen skal være assosiativ

$$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$$

og kommutativ

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}.$$

Skalarmultiplikasjonen skal være assosiativ

$$a(b\mathbf{v}) = (ab)\mathbf{v}$$

og distributiv både med hensyn på addisjon av skalarer

$$(a+b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$$

og vektorer

$$a(\mathbf{v} + \mathbf{w}) = a\mathbf{v} + a\mathbf{w}.$$

Vi må også kreve at

$$1\mathbf{v} = \mathbf{v}$$

og at

$$0\mathbf{v} = \mathbf{0}.$$

De tre første budene forteller oss noe om hva et vektorrom må inneholde, mens de resterende syv går på egenskapene til vektoraddisjon og skalarmultiplikasjon. Vi kommer bare til å trenge $\mathbb{K} = \mathbb{R}$ eller $\mathbb{K} = \mathbb{C}$.

**Eksempel 7.1.** Vi kan bruke aksiomene til å vise at $\mathbf{v}$ og $(-1)\mathbf{v}$ summerer til nullvektoren:

$$\mathbf{0} = 0\mathbf{v} = (1-1)\mathbf{v} = \mathbf{v} + (-1)\mathbf{v}$$

Det er vanlig å skrive $-\mathbf{v}$ istedet for $(-1)\mathbf{v}$. Det er kanskje mest vanlig å sette krav om existens av elementet $-\mathbf{v}$ som et aksiom, og så utlede at $\mathbf{0} = 0\mathbf{v}$, men da blir aksiomene litt vanskeligere å huske. Jeg liker best å gjøre det slik som vi gjør, for det gjør Hanche, og han er ganske smart. △

**Eksempel 7.2.** Det går an å vise at nullvektoren er entydig, i den forstand at det ikke finnes en annen vektor som tilfredsstiller

$$\mathbf{v} + \mathbf{w} = \mathbf{v}$$

for alle $\mathbf{v}$. La oss anta at en slik $\mathbf{w}$ finnes. Hvis vi velger $\mathbf{v} = \mathbf{0}$ i likningen over, ser vi at

$$\mathbf{0} + \mathbf{w} = \mathbf{0}.$$

Men nå kan vi bruke aksiomene, og se at

$$\mathbf{0} = \mathbf{0} + \mathbf{w} = 0\mathbf{w} + 1\mathbf{w} = (0+1)\mathbf{w} = \mathbf{w}.$$

Altså er $\mathbf{w} = \mathbf{0}$. △

**Eksempel 7.3.** Vi kan bevise at

$$a\mathbf{0} = \mathbf{0}.$$

Vi vet at

$$a\mathbf{v} = a(\mathbf{v} + \mathbf{0}) = a\mathbf{v} + a\mathbf{0},$$

men i forrige eksempel viste vi at nullvektoren er den eneste vektoren med denne egenskapen. Altså må $a\mathbf{0}$ være nettopp denne nullvektoren. △

**Eksempel 7.4.** Det går nå an å vise at dersom

$$a\mathbf{v} = \mathbf{0}$$

må enten $a = 0$ eller $\mathbf{v} = \mathbf{0}$. Dersom $a = 0$, har vi jo et aksiom som sier at $0\mathbf{v} = \mathbf{0}$, så det er greit. Dersom $a \neq 0$, er

$$\mathbf{v} = 1\mathbf{v} = \frac{1}{a}a\mathbf{v} = \frac{1}{a}\mathbf{0} = \mathbf{0}.$$ △

Nå går det an å fortsette på denne måten, og bevise alt mellom himmel og jord. Vel, faktisk ikke alt: Et av Gödels teoremer sier noe sånt som at i alle aksiomatiske systemer finnes det utsagn som er sanne, men som ikke lar seg bevise med aksiomene man har til rådighet. Gödel var veldig skarp, men sultet ihjel etter at han ble enkemann, fordi han nektet å spise annen mat enn den kona lagde til ham.

Arketypen på vektorrom er $\mathbb{R}^n$. Dette er mengden av alle vektorer på formen

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

der alle komponentene er reelle tall. Å vise at $\mathbb{R}^n$ er et vektorrom er trivielt, for aksiomene er avledet fra regnereglene for $\mathbb{R}^n$ du lærte allerede på gymnaset. Styrken i vektorromsabstraksjonen er at det er mange ting som ved første øyekast ikke ser ut som $\mathbb{R}^n$, men som oppfører seg som $\mathbb{R}^n$ allikevel.

Vi skriver $\mathbb{C}^n$ for vektorrommet der vektorene er søylevektorer med komplekse komponenter

$$\mathbf{z} = \begin{pmatrix} a_1 + b_1 i \\ a_2 + b_2 i \\ \vdots \\ a_n + b_n i \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$

Du kjenner til mange vektorrom. Vi har bare ikke kalt det vektorrom før. Her er noen enkle eksempler.

**Eksempel 7.5.** La $\mathbf{v}$ være en vektor, for eksempel

$$\mathbf{v} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Mengden av alle reelle eller komplekse skalarmultipler

$$a\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

utgjør et vektorrom. Dette er relativt enkelt å sjekke, for aksiomene for vektorrom er jo avledet fra de vanlige reglene for vektorregning. △

**Eksempel 7.6.** Alle lineærkombinasjoner av

$$\begin{pmatrix} 1 \\ 2 \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 3 \\ 4 \end{pmatrix}$$

utgjør også et vektorrom. Se øvingsoppegget. △

**Eksempel 7.7.** La $T : V \to V$ være en lineæravbildning. En egenvektor med egenverdi $\lambda$ er en vektor $\mathbf{v} \neq \mathbf{0}$ slik at

$$T\mathbf{v} = \lambda\mathbf{v}.$$

Som du så i Øysteins notat, er det vanlig å starte showet med å kreve at en egenvektor ikke kan være nullvektoren, slik at ikke alle skalarer skal passere nåløyet som egenverdier, og så reintrodusere nullvektoren som egenvektoren til en gitt egenverdi, slik at egenvektorene til denne egenverdien danner et rom. Dette er selvfølgelig et vektorrom, og kalles egenrommet til $\lambda$. △

**Eksempel 7.8.** Løsningene til differensiallikningen

$$\ddot{y}(t) + y(t) = 0$$

er

$$y(t) = a\cos t + b\sin t.$$

Disse utgjør et vektorrom. Det høres sikkert snodig ut, men dette er faktisk akkurat det samme vektorrommet som det i forrige eksempel. △

**Eksempel 7.9.** Alle polynomer på formen

$$p(x) = bx + c$$

utgjør et vektorrom. Dette er også det samme vektorrommet som i de to foregående eksemplene. △

**Eksempel 7.10.** La $\mathbf{V}$ være mengden av alle skalarmutlipler av $\mathbf{v}$, altså alle vektorer på formen

$$a\mathbf{v} = a\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Dette er et vektorrom. La oss sjekke alle aksiomene. Vektoraddisjonen $\mathbf{v}+\mathbf{w}$ tar vi som vanlig vektoraddisjon, slik vi lærte på skolen, og tilsvarende valg gjøres for skalarmultiplikasjon. Vi må nå sjekke alle aksiomene.

La oss begynne med nullvektoren. Vektoren

$$\begin{pmatrix}0\\0\end{pmatrix}$$

er med i $\mathbf{V}$, siden den er en skalarmultippel av $\mathbf{v}$:

$$\begin{pmatrix}0\\0\end{pmatrix}=0\cdot\begin{pmatrix}1\\2\end{pmatrix}$$

La oss definere

$$\mathbf{0}=\begin{pmatrix}0\\0\end{pmatrix}.$$

Vi har nå at

$$\mathbf{v}+\mathbf{0}=\mathbf{v}$$

for alle $\mathbf{v}\in\mathbf{V}$, og at

$$0\cdot\mathbf{v}=\mathbf{0}.$$

Siden

$$a\mathbf{v}+b\mathbf{v}=b\mathbf{v}+a\mathbf{v}$$

ser vi at kommutativitet for vektoraddisjon holder. Assosiativitet sjekkes på samme vis. Det samme gjelder regnereglene for skalarmultiplikasjon, altså assosiativitet, distributivitet, og at $1\mathbf{v}=\mathbf{v}$.

Vi må til slutt å sjekke at $\mathbf{V}$ er lukket under vektoraddisjon. Dette følger av at summen av to skalarmultipler av en og samme vektor, også er en skalarmultippel av den samme vektoren:

$$a\mathbf{v}+b\mathbf{v}=(a+b)\mathbf{v}.$$

Å sjekke at $\mathbf{V}$ er lukket under skalarmultiplikasjon, er sant per definisjon, siden vi har definert $\mathbf{V}$ som alle skalarmultipler av vektoren $\mathbf{v}$. △

**Eksempel 7.11.** La oss sjekke at alle lineærkombinasjoner av

$$\mathbf{v}=\begin{pmatrix}1\\2\end{pmatrix}\quad\text{og}\quad\mathbf{w}=\begin{pmatrix}3\\4\end{pmatrix}$$

utgjør et vektorrom. Vi må jo egentlig sjekke alle aksiomene her, Men jobben med å sjekke aksiomene som går på regneregler, er identisk med jobben som ble gjort i forrige oppgave, så vi må tenke over hva som er nytt.

Det som er annerledes, er egentlig bare spørsmålet om hvorvidt $\mathbf{V}$ er lukket under de to regneoperasjone. Anta $\mathbf{x}$ og $\mathbf{y}$ er vektorer i $\mathbf{V}$. Er $\mathbf{x}+\mathbf{y}$ også i $\mathbf{V}$? La oss skrive $\mathbf{x}=a\mathbf{v}+b\mathbf{w}$ og $\mathbf{y}=c\mathbf{v}+d\mathbf{w}$. Men da er

$$\mathbf{x}+\mathbf{y}=a\mathbf{v}+b\mathbf{w}+c\mathbf{v}+d\mathbf{w}=(a+c)\mathbf{v}+(b+d)\mathbf{w},$$

som er en lineærkombinasjon av $\mathbf{v}$ og $\mathbf{w}$, så $\mathbf{x}+\mathbf{y}$ er altså med i $\mathbf{V}$. Tilsvarende ser vi at

$$e\mathbf{x}=ea\mathbf{v}+eb\mathbf{w}$$

slik at $e\mathbf{x}$ er en lineærkombinasjon av $\mathbf{v}$ og $\mathbf{w}$, og følgelig med i $\mathbf{V}$. △

**Eksempel 7.12.** Mengden av alle vektorer på formen

$$\begin{pmatrix}a\\2\end{pmatrix}$$

utgjør ikke et vektorrom, ihvertfall ikke dersom vi bruker den vanlige vektoraddisjonen og skalarmultiplikasjonen som regneoperasjoner. Det første aksiomet som mangler, er det om nullvektoren. Det finnes ingen vektor $\mathbf{w}$ som har den egenskap at

$$\mathbf{v}+\mathbf{w}=\mathbf{v}.$$

For det andre er ikke denne mengden lukket under vanlig vektoraddisjon, siden

$$\begin{pmatrix}a\\2\end{pmatrix}+\begin{pmatrix}b\\2\end{pmatrix}=\begin{pmatrix}a+b\\4\end{pmatrix}\neq\begin{pmatrix}c\\2\end{pmatrix}.$$

For det tredje er den ikke lukket under den vanlige skalarmultiplikasjonen, siden

$$b\begin{pmatrix}a\\2\end{pmatrix}=\begin{pmatrix}ab\\4b\end{pmatrix}\neq\begin{pmatrix}c\\2\end{pmatrix}$$

så lenge $b\neq 1$. △

Det finnes også ting som ved første øyekast kan se ut som vektorrrom, men som ikke er det.

**Eksempel 7.13.** Mengden av alle vektorer på formen

$$\begin{pmatrix}a\\2\end{pmatrix}$$

utgjør ikke et vektorrom, ihvertfall ikke dersom vi bruker den vanlige vektoraddisjonen og skalarmultiplikasjonen som regneoperasjoner. Kan du se hvilke aksiomer som ikke er tilfredsstilt? △

**Underrom**

En delmengde av et vektorrom som i seg selv er et vektorrom, kalles et underrom.

Det er relativt enkelt å avgjøre om en delmengde av et vektorrom er et vektorrom. La $U$ være en delmengde av et vektorrom $V$. De syv aksiomene som går på regnereglene er ikke nødvendig å sjekke, for det ligger i sakens natur at vektorene i $U$ tilfredsstiller de samme regnereglene som de i $V$, siden vektorer i $U$ også er i $V$. Med andre ord er det bare nødvendig å sjekke de tre aksiomene som går på innhold. Men sjekker vi at

$$\mathbf{v},\mathbf{w}\in U\longrightarrow\mathbf{v}+\mathbf{w}\in U$$
$$\mathbf{v}\in U\longrightarrow c\mathbf{v}\in U$$

kommer nullvektoren med i $U$ av seg selv siden $0\mathbf{v}=\mathbf{0}$.

**Å sjekke om noe er et underrom**

En delmengde $U$ av vektorrommet $V$ er et underrom hvis og bare hvis

$$\mathbf{v},\mathbf{w}\in U\longrightarrow\mathbf{v}+\mathbf{w}\in U$$
$$\mathbf{v}\in U\longrightarrow c\mathbf{v}\in U$$

**Eksempel 7.14.** Alle skalarmultipler av

$$\begin{pmatrix}1\\2\end{pmatrix}$$

er et underrom av vektorrommet av alle lineærkombinasjoner av

$$\begin{pmatrix}1\\2\end{pmatrix} \quad \text{og} \quad \begin{pmatrix}3\\4\end{pmatrix}. \qquad \triangle$$

**Eksempel 7.15.** Alle konstante funksjoner er et underrom av vektorrommet av alle polynomer på formen

$$p(x) = bx + c. \qquad \triangle$$

# Dimensjon

Vi sier at vektorer $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er lineært uavhengige dersom

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

impliserer at

$$c_1 = c_2 = \cdots = c_n = 0.$$

Merk at en lineært uavhengig vektormengde ikke kan inneholde nullvektoren.

**Lineær uavhengighet er praktisk**

Dersom en vektor $\mathbf{w}$ kan skrives som en lineærkombinasjon av de lineært uavhengige vektorene $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$, kan dette kun gjøres på én måte.

*Bevis.* Anta at vi kan skrive både

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

og

$$\mathbf{w} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n.$$

I så fall er

$$\begin{aligned}\mathbf{0} &= \mathbf{w} - \mathbf{w}\\ &= (c_1 - d_1)\mathbf{v}_1 + (c_2 - d_2)\mathbf{v}_2 + \cdots + (c_n - d_n)\mathbf{v}_n.\end{aligned}$$

Men $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er lineært uavhengige, så likningen over impliserer at $c_k = d_k$ for alle $k$. □

Dersom alle $\mathbf{w} \in \mathbf{V}$ kan skrives som en lineærkombinasjon av en lineært uavhengig vektormengde $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$, sier vi at $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ spenner ut $\mathbf{V}$.

**Basis**

En vektormengde som spenner ut vektorrommet $\mathbf{V}$, kalles en basis for $\mathbf{V}$.

Basis er et sentralt verktøy.

**Et vektorrom har en basis**

Et vektorrom har enten en endelig basis, eller en uendelig følge av lineært uavhengige vektorer.

*Bevis.* Se her. □

Hanche kaller det neste teoremet for plassmangelteoremet, for det forteller noen om hvor mange lineært uavhengige vektorer det er plass til i et vektorrom.

**Plassmangelteoremet**

La $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ være en basis for $\mathbf{V}$, og la $\mathbf{w}_1$, $\mathbf{w}_2$ ... $\mathbf{w}_m$ være en lineært uavhengig vektormengde i $\mathbf{V}$. Da er $m \leq n$.

*Bevis.* Anta $m > n$. Siden $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er en basis for $\mathbf{V}$, kan vi skrive

$$\mathbf{w}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Siden $\mathbf{w}_1 \neq \mathbf{0}$, må minst en av koeffisientene $c_k$ være ulik 0. La oss anta at $c_1 \neq 0$. (Hvis ikke dette er tilfelle, bytter vi bare litt på rekkefølgen i $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$.) Men i så fall kan vi skrive

$$\mathbf{v}_1 = \mathbf{w}_1 - \frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_n}{c_1}\mathbf{v}_n,$$

og dette betyr at $\mathbf{w}_1$, $\mathbf{v}_2$ ... $\mathbf{v}_n$ er en basis for $\mathbf{V}$, for et vilkårlig element $\mathbf{u} \in \mathbf{V}$ vil alltid kunne skrives

$$\begin{aligned}\mathbf{u} =& d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n\\ =& d_1\left(\mathbf{w}_1 - \frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_n}{c_1}\mathbf{v}_n\right)\\ &+ d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n\\ =& d_1\mathbf{w}_1 + \left(d_2 - d_1\frac{c_2}{c_1}\right)\mathbf{v}_2 + \cdots + \left(d_n - d_1\frac{c_n}{c_1}\right)\mathbf{v}_n.\end{aligned}$$

Vi fortsetter på samme måte. Det er nå klart at vi kan skrive (med nye koeffisienter $d_1$ og $c_k$)

$$\mathbf{w}_2 = d_1\mathbf{w}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Her må minst en av koeffisientene $c_k$ være ulik null, for $\mathbf{w}_2$ er ikke en skalarmultippel av $\mathbf{w}_1$. La oss anta at $c_2 \neq 0$, hvis ikke, bytt rekkefølge på $\mathbf{v}_2$, $\mathbf{v}_3$ ... $\mathbf{v}_n$. Nå kan vi skrive

$$\mathbf{v}_2 = -\frac{d_1}{c_2}\mathbf{w}_1 + \mathbf{w}_2 - \frac{c_3}{c_2}\mathbf{v}_3 - \cdots - \frac{c_n}{c_2}\mathbf{v}_n,$$

og det samme argumentet som ovenfor forteller oss at $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{v}_3$ ... $\mathbf{v}_n$ er en basis for $\mathbf{V}$.

Vi tar en runde til. Vi kan skrive (nok en gang med nye koeffisienter)

$$\mathbf{w}_3 = d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n,$$

og her må minst en $c_k$ være ulik null, for $\mathbf{w}_1$, $\mathbf{w}_2$ og $\mathbf{w}_3$ er jo lineært uavhengige. Samme prosess gir nok en gang at $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$ $\mathbf{v}_4$ ... $\mathbf{v}_n$ er en basis for $\mathbf{V}$, og gjentar vi dette igjen og igjen, vil konklusjonen være at $\mathbf{w}_1$, $\mathbf{w}_2$ ... $\mathbf{w}_n$ er en basis for $\mathbf{V}$.

Men dersom dette er sant, kan vi skrive

$$\mathbf{w}_{n+1} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_n\mathbf{w}_n,$$

og dette går på tverke med den lineære uavhengigheten til $\mathbf{w}_1$, $\mathbf{w}_2$ ... $\mathbf{w}_m$. Altså kan ikke $m > n$. □

**Nå kan vi snart definere dimensjon dere!**

To endelige basiser for ett og samme vektorrom må bestå av samme antall vektorer.

*Bevis.* Anta at vi har to basiser $\mathbf{v}_1$, $\mathbf{v}_2$ ... $\mathbf{n}_n$ og $\mathbf{w}_1$, $\mathbf{w}_2$ ... $\mathbf{w}_m$. Siden både $n \leq m$ og $m \leq n$, må vi ha $n = m$. □

**NÅ kan vi definere dimensjon!**

Dersom et vektorrom har en endelig basis, sier vi at vektorrommet er endeligdimensjonalt. To forskjellige basiser for et og samme endeligdimensjonale vektorrom må inneholde det samme antall vektorer. Dette antallet kalles vektorrommets dimensjon.

**Eksempel 7.16.** $\mathbb{R}^n$ et $n$-dimensjonalt vektorrom over $\mathbb{R}$. En basis er de $n$ vektorene

$$\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} \ldots \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix} \qquad \triangle$$

**Eksempel 7.17.** Standardbasisen for $\mathbb{R}^n$ er også en basis for $\mathbb{C}^n$. Det er lett å se at alle elementer i $\mathbb{C}^n$ kan skrives som en entydig lineærkombinasjon av elementene i basisen:

$$\mathbf{z} = \begin{pmatrix}a_1 + b_1 i\\a_2 + b_2 i\\\vdots\\a_n + b_n i\end{pmatrix} = \begin{pmatrix}z_1\\z_2\\\vdots\\z_n\end{pmatrix}$$
$$= z_1 \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} + z_2 \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} + \ldots + z_n \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix} \qquad \triangle$$

**Eksempel 7.18.** Man kan også se på $\mathbb{C}^n$ som et vektorrom over $\mathbb{R}$, men da er ikke standardbasisen for $\mathbb{R}^n$ en basis for $\mathbb{C}^n$. En basis er

$$\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} + \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} + \ldots + \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}$$
$$+ \begin{pmatrix}i\\0\\\vdots\\0\end{pmatrix} + \begin{pmatrix}0\\i\\\vdots\\0\end{pmatrix} + \ldots + \begin{pmatrix}0\\0\\\vdots\\i\end{pmatrix},$$

og dimensjonen er $2n$ istedet for $n$. En tilfeldig vektor $\mathbf{z}$ i $\mathbb{C}^n$ kan skrives

$$\mathbf{z} = \begin{pmatrix}a_1 + b_1 i\\a_2 + b_2 i\\\vdots\\a_n + b_n i\end{pmatrix}$$
$$= a_1 \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} + a_2 \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} + \ldots + a_n \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}$$
$$+ b_1 \begin{pmatrix}i\\0\\\vdots\\0\end{pmatrix} + b_2 \begin{pmatrix}0\\i\\\vdots\\0\end{pmatrix} + \ldots + b_n \begin{pmatrix}0\\0\\\vdots\\i\end{pmatrix}. \qquad \triangle$$

**Eksempel 7.19.** $\mathbb{Q}^n$ et vektorrom over $\mathbb{Q}$. Selv om vi har sagt at vi har gjort lineær algebra på $\mathbb{R}^n$ til nå, har vi i praksis operert på $\mathbb{Q}^n$. Ikke et eneste matriseeksempel har involvert $\sqrt{2}$. $\triangle$

**Eksempel 7.20.** De hele tallene $\mathbb{Z}$ er ikke en kropp, så $\mathbb{Z}^n$ er ikke et vektorrom. $\triangle$

**Eksempel 7.21.** Alle polynomer på formen

$$p(x) = ax^2 + bx + c$$

utgjør et vektorrom med dimensjon 3. En basis er 1, $x$, $x^2$. Hvordan sr vi det? Det er klart at alle polynomer på formen

$$ax^2 + bx + c$$

kan skrives som en lineærkombinasjon av 1, $x$, $x^2$. (Du har jo brukt denne basisen til å definere hva du mener med et andregradspolynom siden første klasse på gymnaset.) De tre elementene er også lineært uavhengige, for om vi krever

$$ax^2 + bx + c = 0 \quad \text{for alle } t$$

må $a = b = c = 0$. En annen basis er de tre første Legendrepolynomene 1, $x$ og $\frac{1}{2}(3x^2 - 1)$. Det er litt mer regning å vise, men grunnstegene er de samme. $\triangle$

**Eksempel 7.22.** Løsningene til differensiallikningen

$$\ddot{y} + y = 0$$

utgjør et vektorrom over $\mathbb{C}$ med dimensjon 2. En basis er $\sin t$, $\cos t$, og en annen er $e^{it}$, $e^{-it}$. $\triangle$

**Eksempel 7.23.** Alle matriser på formen

$$\begin{pmatrix}a & b\\c & d\end{pmatrix}$$

utgjør et vektorrom med dimensjon 4. En pen basis er de fire matrisene

$$\begin{pmatrix}1 & 0\\0 & 0\end{pmatrix}, \begin{pmatrix}0 & 1\\0 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\1 & 0\end{pmatrix}, \begin{pmatrix}0 & 0\\0 & 1\end{pmatrix} \qquad \triangle$$

**Eksempel 7.24.** Alle kontinuerlige funksjoner, alle deriverbare funksjoner, alle analytiske funksjoner, og så videre, er eksempler på vektorrom med uendelig mange dimensjoner. Husk at en funksjon er noe som på en entydig måte tilordner et element i en mengde til et element i en annen mengde. Derfor kan du tenke på en funksjon som en uendelig lang søylevektor der komponentene er funksjonsverdiene til funksjonen på alle punktene i $\mathbb{R}$. $\triangle$

# Koordinater

I forrige avsnitt var rekkefølgen på basiselementene litt ad hoc. Fra nå av skal vi skrive basiselementene i en bestemt rekkefølge, for da kan basisen brukes til å definere koordinater.

**Nå kan vi definere koordinater!**

La $(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n)$ være en basis for et vektorrom, og la

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

Vi sier at skalarene $(c_1, c_2, ..., c_n)$ er koordinatene til $\mathbf{w}$ i basisen $(\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n)$.

**Eksempel 7.25.** Vi skriver jo at $z_1$, $z_2$,...$z_n$ er koordinatene til $\mathbf{z}$ fordi

$$\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = z_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + z_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + ... + z_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \qquad \triangle$$

**Eksempel 7.26.** Koordinatene til

$$p(x) = x^2 + 2x + 1$$

i basisen $(x^2, x, 1)$ er $(1, 2, 1)$. Koordinatene i basisen

$$\left(\frac{1}{2}(3x^2 - 1), x, 1\right)$$

er $(2/3, 2, 4/3)$, siden

$$p(x) = x^2 + 2x + 1 = \frac{2}{3}\left(\frac{1}{2}(3x^2 - 1)\right) + 2x + \frac{4}{3}. \quad \triangle$$

**Eksempel 7.27.** Koordinatene til vektoren

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

i vektorrommet av $2 \times 2$-matriser, er $(a, b, c, d)$, siden

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad \triangle$$

# Lineæravbildninger

Den vanligste funksjonen mellom vektorrom kalles lineæravbildning. Vanlige synonymer er lineærtransformasjon og lineæroperator.

**Superposisjonsprinsippet på matematikermåten**

La $V$ og $W$ være vektorrom. En lineæravbildning er en funksjon $T : V \to W$ slik at

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$$
$$T(c\mathbf{x}) = cT(\mathbf{x})$$

Arketypen på en lineæravbildning, er matrise-vektorproduktet.

**Eksempel 7.28.** La $A$ være en $n \times m$-matrise, og $\mathbf{x}$ og $\mathbf{y}$ $m \times 1$-vektorer. Vi vet jo fra forrige semester at

$$A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$$

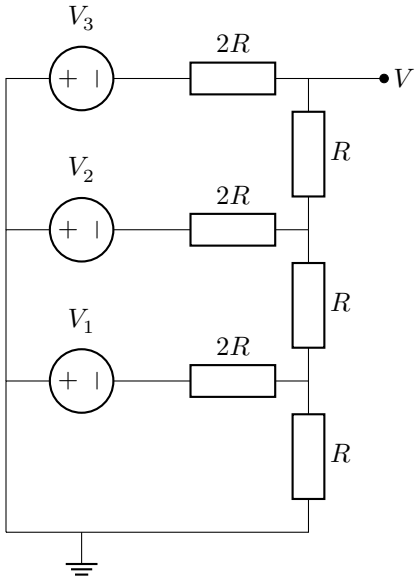
og at

$$A(c\mathbf{x}) = cA\mathbf{x}.$$

En lineæravbildning er bare en generalisering av denne operasjonen, og vi kan definere $T : \mathbb{R}^m \to \mathbb{R}^n$ ved

$$T(\mathbf{x}) = A\mathbf{x}. \qquad \triangle$$

**Eksempel 7.29.** En digital-til-analog-konverter består av en krets som ser slik ut:




Kretsen fungerer slik at spenningsnivåene $V_1$, $V_2$ og $V_3$ definerer en binærkode ved at de har enten høy (5V) eller lav (0V) spenning, og så får man ut en spenning $V$ på mellom 0 og 5 volt. La oss representere $V_1V_2V_3$ som en vektor med nullere og enere. Tabellen blir noe slikt:

| $V_1V_2V_3$ | $V$ |
|---|---|
| 0 0 0 | 0 |
| 0 0 1 | 0.625 |
| 0 1 0 | 1.25 |
| 0 1 1 | 1.875 |
| 1 0 0 | 2.5 |
| 1 0 1 | 3.125 |
| 1 1 0 | 3.75 |
| 1 1 1 | 4.375 |

Nå er det slik at siden denne typen resistiv krets modelleres av lineæralgebra, er superposisjonsprinsippet implisert av lineæralgebramodellen. Siden siden matrise-vektorproduktet er en lineæravbildning, vil også kretsen oppføre seg som en lineæravbildning. La for eksempel

$$\mathbf{x} = (0, 1, 0)$$

og

$$\mathbf{y} = (1, 0, 1).$$

Da er

$$\begin{aligned} T(\mathbf{x}+\mathbf{y}) &= T(\mathbf{x}) + T(\mathbf{y}) \\ &= T(0,1,0) + T(1,0,1) = \\ & T(1,1,1) = 4.375. \end{aligned}$$

Til slutt må vi være litt forsiktige her, for det er noen tilleggsregler som gjelder fordi dette er elektronikk i praksis. Man vil aldri legge sammen for eksempel vektorene $(0,1,0)$ og $(0,1,1)$, for dette korresponderer til å sette $V_2$ til 10V, og det gjør man ikke. △

**Eksempel 7.30.** Egenskapen

$$T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$$

er sentral for superposisjonsprinsippet, men en vanlig kilde til misforståelse. I noen fagfelt, for eksempel statistikk, bruker man ordet 'lineær' om en funksjon som beskriver en rett linje, altså funksjoner på formen $p(x) = ax + b$. For denne funksjonen gjelder jo ikke at

$$p(x+y) = p(x) + p(y)$$

med mindre $b = 0$, så ordet lineær har for dem en helt annen betydning. Dette er greit å vite om. △

**Eksempel 7.31.** Venstresiden i differensiallikningen

$$\ddot{y} + y = 0$$

er en lineæravbildning

$$T(y) = \ddot{y} + y.$$

Dette er matematikerens måte å si at superposisjonsprinsippet gjelder. Når du så likningen første gang, brukte læreren din egenskapene til lineæravbildninger, og sa antagelig noe slikt som at 'cos $t$ er jo selvfølgelig en løsning, men det er sin $t$ også, og så er det slik at da er faktisk $a\cos t + b\sin t$ en løsning'. Når det er snakk om differensiallikninger, bruker man ofte ordet lineæroperator. △

**Eksempel 7.32.** Setter man $c = 0$ i

$$T(c\mathbf{v}) = cT(\mathbf{v}),$$

ser man at for en vilkårlig lineærtransformasjon $T : V \to W$, må

$$T(\mathbf{0}) = \mathbf{0}.$$

Her er det underforstått at $\mathbf{0}$ på venstre side er nullvektoren i $V$, mens $\mathbf{0}$ på høyre side er nullvektoren i $W$. △

**Noe litt tekniske greier**

En lineæravbildning $T : V \to W$ er surjektiv dersom det for enhver $\mathbf{w} \in W$ finnes en $\mathbf{v} \in V$ slik at $T(v) = w$.

En lineæravbildning $T : V \to W$ som er både injektiv og surjektiv, kalles en isomorfi. Dersom det finnes en isomorfi mellom $V$ og $W$, sier vi at $V$ og $W$ er isomorfe. Dette betyr at det i bunn og grunn er snakk om samme vektorrom. To isomorfe vektorrom kan se veldig forskjellige ut. Vektorrom er isomorfe dersom de har samme dimensjon.

**Eksempel 7.33.** Vektorrommet av løsninger til

$$\ddot{y} + y = 0$$

er isomorft med både $\mathbb{R}^2$. En isomorfi er for eksempel

$$T(a\cos t + b\sin t) = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Disse vektorrommene er også isomorfe med $\mathbb{C}^2$, alle polynomer på formen $ax + b$, og $\mathbb{C}$, dersom du ser på det som et todimensjonalt vektorrom over $\mathbb{R}$. △

**To viktige vektorrom**

La $T : V \to W$ være en lineæravbildning. Mengden av alle $\mathbf{v} \in V$ slik at $T\mathbf{v} = \mathbf{0}$, kalles kjernen til $T$. Mengden av alle $\mathbf{w} \in W$ slik at $T\mathbf{v} = \mathbf{w}$, kalles bildet til $T$.

Kjernen og bildet gir opphav til to viktige vektorrom.

**De er faktisk vektorrom**

Kjernen til $T : V \to W$ er et underrom av $V$, og bildet er et underrom av $W$.

**Eksempel 7.34.** Dersom $T$ er gitt ved et matrise-vektor-produkt $A\mathbf{x} = \mathbf{y}$, er kjernen til $T$ noe som kalles nullrommet til $A$. Dette er alle vektorer slik at

$$A\mathbf{x} = \mathbf{0},$$

og man finner en basis for dette rommet ved å finne alle løsninger av dette lineære likningssystemet.

Bildet til $T$ kalles søylerommet, og er alle vektorer på formen

$$A\mathbf{x},$$

altså alle mulige lineærkombinasjoner av matrisens søyler. En basis for søylerommet kan man få ved radreduksjon. Hvis man plukker ut søyleindeksene til de søylene som ender opp med å ha pivotelementer etter radreduksjonen, vil de søylene med samme indeks i $A$ være en basis for søylerommet. Se øvingsopplegget. △

**Eksempel 7.35.** Dersom $T : V \to W$ er gitt ved en operasjon som gjør et eller annet, så finnes det alltid en matrise som gjør det samme med vektorer fra $\mathbb{R}^n$ til $\mathbb{R}^m$, der $n$ er dimensjonen til $V$, og $m$ er dimensjonen til $W$. La for eksempel

$$T(y) = \ddot{y} + y,$$

og la både $V$ og $W$ være vektorrommet av polynomer på formen $p(x) = ax^2 + bx + c$. Dette rommet er isomorft med $\mathbb{R}^3$, og derfor finnes det en $3\times 3$-matrise som gjør det samme med vektorer i $\mathbb{R}^3$ som $T$ gjør med andre ordens polynomer. Hvis vi anvender $T$ på $ax^2 + bx + c$, får vi

$$\begin{aligned} T(ax^2+bx+c) =& aT(x^2) + bT(x) + cT(1) \\ =& a(2+x^2) + bx + c \\ =& ax^2 + bx + 2a + c. \end{aligned}$$

En matrise som gjør det samme mellom vektorer i $\mathbb{R}^3$, er

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix},$$

siden

$$A \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a \\ b \\ 2a + c \end{pmatrix}.$$

Matrisen $A$ kalles lineæravbildningens standardmatrise, og kan enkelt finnes ved å se på hva lineæravbildningen gjør med vektorene i basisen man har valgt for $V$. I dette tilfellet er basisen $(x^2, x, 1)$. La oss begynne med $x^2$. Siden $T(x^2) = x^2 + 2$, ønsker vi en matrise $A$ slik at

$$A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix},$$

og dette impliserer at den første søylen i $A$ må være nettopp

$$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

Fortsetter man med $T(x)$ og $T(1)$, detter de andre søylene ut. △

**Eksempel 7.36.** Vi kan finne standardmatrisen til lineæravbildningen $T$ som speiler punkter om $x$-aksen i $\mathbb{R}^2$. Speilingen er definert av at

$$T \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

og

$$T \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

slik at matrisen blir

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$ △

**Eksempel 7.37.** Standardmatrisen til lineæravbildningen som roterer punkter vinkelen $\theta$ mot klokken, er gitt ved

$$A_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

siden

$$A_\theta \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$$

og

$$A_\theta \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$$ △

# Indreprodukt

Vektorrom kan, i tillegg til vektoraddisjon og skalarmultiplikasjon, ha forskjellige typer produkt mellom vektorer. Et *indreprodukt* er noe der du putter inn to vektorer og får ut et tall. Dersom dette tallet er reellt, sier vi at indreproduktet er reelt. Indreproduktet kan også være komplekst, men dette venter vi litt med. Indreproduktet er en generalisering av skalarproduktet du lærte om på videregående skole. Vi skriver gjerne

$$(\mathbf{v}, \mathbf{w})$$

og følgende aksiomer skal gjelde.

**Reelt indreprodukt**

La $V$ være et vektorrom over $\mathbb{R}$, la $\mathbf{u}$, $\mathbf{v}$ og $\mathbf{w}$ være vektorer, og $a$ og $b$ reelle tall. Indreproduktet skal være symmetrisk

$$(\mathbf{v}, \mathbf{w}) = (\mathbf{w}, \mathbf{v}),$$

lineært i andre argument

$$(\mathbf{u}, a\mathbf{v} + b\mathbf{w}) = a(\mathbf{u}, \mathbf{v}) + b(\mathbf{u}, \mathbf{w})$$

og positivt definitt

$$(\mathbf{v}, \mathbf{v}) > 0 \quad \text{dersom} \quad \mathbf{v} \neq \mathbf{0}.$$

Et vektorrom med et indreprodukt kalles et indreproduktrom.

**Eksempel 7.38.** Indreproduktet er lineært også i første argument:

$$\begin{aligned}(a\mathbf{v} + b\mathbf{w}, \mathbf{u}) &= (\mathbf{u}, a\mathbf{v} + b\mathbf{w}) \\ &= a(\mathbf{u}, \mathbf{v}) + b(\mathbf{u}, \mathbf{w}) \\ &= a(\mathbf{v}, \mathbf{u}) + b(\mathbf{w}, \mathbf{u})\end{aligned}$$

Siden det reelle indreproduktet er lineært i begge argumenter, sier vi gjerne at det er bilineært. Det går like greit å sette opp som aksiom at indreproduktet er lineært i første faktor, og så utlede at det er lineært i andre faktor. Vvilken vei man gjør det har ingenting å si for reelle indreprodukt, men det har litt å si for komplekse indreprodukt som vi skal se på lenger ned. △

**Eksempel 7.39.**

$$(\mathbf{v}, \mathbf{0}) = (\mathbf{v}, 0\mathbf{v}) = 0(\mathbf{v}, \mathbf{v}) = 0.$$ △

**Eksempel 7.40.** Fra videregående skole husker du forhåpentligvis skalarproduktet

$$\mathbf{v} \cdot \mathbf{w}$$

mellom vektorer i $\mathbb{R}^2$. Dette er et indreprodukt. Det finnes to formler for å beregne dette. Den ene ser slik ut:

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \cos\theta = \sqrt{v_1^2 + v_2^2}\sqrt{w_1^2 + w_2^2}\cos\theta$$

der $\theta$ er vinkelen mellom $\mathbf{v}$ og $\mathbf{w}$, og den andre slik:

$$\mathbf{v} \cdot \mathbf{w} = v_1 w_1 + v_2 w_2.$$

At disse to uttrykkene gir det samme tallet, følger av cosinussetningen, se øvingsopplegget. At uttrykket definerer et indreprodukt, er også lett å sjekke fra den siste formelen, for det er egenskapene til dette uttrykket som har gitt opphav til aksiomene for indreprodukt. △

**Eksempel 7.41.** I $\mathbb{R}^n$ kan vi bruke matriseregningsregler, og skrive skalarproduktet mellom to søylevektorer $\mathbf{v}$ og $\mathbf{w}$ som $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T\mathbf{w}$:

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$
$$= v_1w_1 + v_2w_2 + \cdots + v_nw_n.$$

Hvis du for eksempel er så gammel at du bruker kontanter, kan du bruke prikkprodukt til å regne ut hvor mange penger du har. Dersom du har tre hundrelapper, fire tohundrelapper og en femhundrelapp, har du

$$\begin{bmatrix} 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} 100 \\ 200 \\ 500 \end{bmatrix} = 1600$$

kroner. △

**Ortogonalitet mellom vektorer**

Vi sier at to vektorer $\mathbf{v}$ og $\mathbf{w}$ er ortogonale dersom

$$(\mathbf{v}, \mathbf{w}) = 0.$$

Merk at nullvektoren står ortogonalt på alle vektorer. Merk også at dersom $(\mathbf{v}, \mathbf{w}) = 0$, gir det første aksiomet at $(\mathbf{w}, \mathbf{v}) = 0$.

**Eksempel 7.42.** I $\mathbb{R}^2$ er $\mathbf{v}$ og $\mathbf{w}$ ortogonale dersom vinkelen mellom dem er $\pi/2$. △

**Eksempel 7.43.** Vektorene

$$\begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{og} \quad \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

er ortogonale. △

**Eksempel 7.44.** La $V$ være vektorrommet av kontinuerlige funksjoner $f : [-\pi, \pi] \to \mathbb{R}$. Dette rommet kalles gjerne $\mathcal{C}[-\pi, \pi]$. Vi definerer indreproduktet

$$(f, g) = \int_{-\pi}^{\pi} f(x)g(x)\, dx.$$

Vi sjekker at aksiomene holder. Dette indreproduktet er symmetrisk, siden

$$(f, g) = \int_{-\pi}^{\pi} f(x)g(x)\, dx = \int_{-\pi}^{\pi} g(x)f(x)\, dx = (g, f),$$

og lineært i andre faktor

$$\begin{aligned}(f, ag + bh) &= \int_{-\pi}^{\pi} f(x)\left(ag(x) + bh(x)\right)\, dx \\ &= a\int_{-\pi}^{\pi} f(x)g(x)\, dx + b\int_{-\pi}^{\pi} f(x)h(x)\, dx \\ &= a(f, g) + b(f, h).\end{aligned}$$

Siden $f^2 \geq 0$, ser vi også at

$$(f, f) = \int_{-\pi}^{\pi} f^2(x)\, dx > 0$$

dersom $f$ ikke er nullfunksjonen. Vi har tidligere sett at funksjonene $\sin nt$ og $\cos nt$ er ortogonale på dette vektorrommet. △

**Lengde**

Lengden til en vektor $\mathbf{v}$ er gitt ved $\|\mathbf{v}\| = \sqrt{(\mathbf{v}, \mathbf{v})}$.

**Eksempel 7.45.** Vektoren $\mathbf{v} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ har lengde 1 for alle $\mathbf{w} \neq 0$. For eksempel er

$$\left\|\frac{\cos nt}{\pi}\right\| = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2 nt\, dt = 1$$

for alle $n$. △

Pytagoras' teorem om rettvinklede trekanter i $\mathbb{R}^2$ gjelder også for indreproduktrom.

**Pytagoras**

Vektorene $\mathbf{v}$ og $\mathbf{w}$ er ortogonale hvis og bare hvis

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.$$

*Bevis.* Vi beregner

$$\begin{aligned}\|\mathbf{v} + \mathbf{w}\|^2 &= (\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}) \\ &= (\mathbf{v}, \mathbf{v}) + (\mathbf{v}, \mathbf{w}) + (\mathbf{w}, \mathbf{v}) + (\mathbf{w}, \mathbf{w}) \\ &= \|\mathbf{v}\|^2 + (\mathbf{v}, \mathbf{w}) + (\mathbf{w}, \mathbf{v}) + \|\mathbf{w}\|^2 \\ &= \|\mathbf{v}\|^2 + 2(\mathbf{v}, \mathbf{w}) + \|\mathbf{w}\|^2.\end{aligned}$$

Dersom $(\mathbf{v}, \mathbf{w}) = 0$, er det klart at

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.$$

Dersom

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2,$$

kan vi trekke denne fra likningen over, og se at

$$2(\mathbf{v}, \mathbf{w}) = 0,$$

som betyr at $\mathbf{v}$ og $\mathbf{w}$ er ortogonale. □

Vi sier at vektorene $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ er innbyrdes ortogonale dersom

$$(\mathbf{v}_i, \mathbf{v}_j) = 0$$

for alle $i$ og $j$. Dersom i tillegg $\|\mathbf{v}_j\| = 1$ for alle $j$, sier vi at vektorene er ortonormale. Det er lett å se at en innbyrdes ortogonal vektormengde må være lineært uavhengig. Hvis vi tar indreproduktet mellom $\mathbf{v}_k$ og begge sider av likningen

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

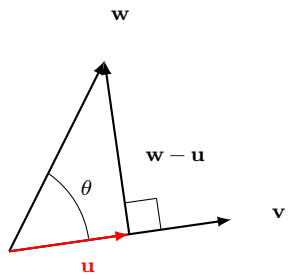
ser vi at $c_k = 0$ for alle $k$.

# Projeksjon

I dette avsnittet skal vi ta for oss en viktig lineæravbildning, nemlig projeksjon. La oss begynne med å se på dette i $\mathbb{R}^2$.

**Et viktig spørsmål**

Hvordan kan vi skrive vektoren $\mathbf{u}$ i figuren under?




*Hva er projeksjon?*

**Svaret:**

Vi kan begynne med å utlede en formel for lengden:

$$\|\mathbf{u}\| = \|\mathbf{w}\|\cos\theta = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|}\|\mathbf{w}\|\cos\theta = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|}.$$

Lengden $\|\mathbf{u}\|$ kalles $\mathbf{w}$ sin skalarprojeksjon på $\mathbf{v}$. Vi kan så bruke denne skalarprojeksjonen til å skrive

$$\mathbf{u} = \|\mathbf{u}\|\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\mathbf{v}\cdot\mathbf{w}}{\|\mathbf{v}\|^2}\mathbf{v} = \frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}.$$

Vektoren $\mathbf{u}$ kalles gjerne $\mathbf{w}$ sin komponent i retningen til $\mathbf{v}$, eller vektorprojeksjonen av $\mathbf{w}$ på $\mathbf{v}$. Komponenten til $\mathbf{w}$ ortogonalt på $\mathbf{v}$ er og

$$\mathbf{w} - \mathbf{u}.$$

**Eksempel 7.46.** Vektoren

$$\mathbf{w} = \begin{bmatrix}2\\1\end{bmatrix}$$

sin komponent i retningen gitt av

$$\mathbf{v} = \begin{bmatrix}1\\2\end{bmatrix}$$

er:

$$\mathbf{u} = \frac{\mathbf{v}\cdot\mathbf{w}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v} = \frac{4}{5}\begin{bmatrix}1\\2\end{bmatrix}$$

Komponenten ortogonalt på $\mathbf{v}$ er

$$\mathbf{w} - \mathbf{u} = \begin{bmatrix}2\\1\end{bmatrix} - \frac{4}{5}\begin{bmatrix}1\\2\end{bmatrix} = \frac{3}{5}\begin{bmatrix}2\\-1\end{bmatrix}. \qquad \triangle$$

**Ortogonal projeksjon**

La $V$ være et vektorrom med et indreprodukt. Inspirert av de geometriske betraktningene over, definerer vi for $\mathbf{v} \neq \mathbf{0}$ en lineæravbildning

$$P_{\mathbf{v}}(\mathbf{w}) = \frac{(\mathbf{v},\mathbf{w})}{(\mathbf{v},\mathbf{v})}\mathbf{v},$$

kalt ortogonal projeksjon på $\mathbf{v}$.

Å se at dette er en lineæravbildning, er ikke så vanskelig, se øvingsopplegg. Man kan tenke at $P_{\mathbf{v}}(\mathbf{w})$ er skyggen $\mathbf{w}$ kaster på $\mathbf{v}$ dersom man lyser på $\mathbf{w}$ med en lommelykt, og lommelykten står uendelig langt borte, rett over $\mathbf{w}$, og vinkelrett på $\mathbf{v}$.

# Noen vanskelige teoremer

La oss ta en ny titt på Pytagoras. Dersom $\mathbf{v} \neq \mathbf{0}$ og $\mathbf{w}$ er vektorer i et indreproduktrom, er $\mathbf{v}$ og $\mathbf{w} - P_{\mathbf{v}}(\mathbf{w})$ ortogonale:

$$\begin{aligned}(\mathbf{v}, \mathbf{w} - P_{\mathbf{v}}(\mathbf{w})) &= \left(\mathbf{v}, \mathbf{w} - \frac{(\mathbf{v},\mathbf{w})}{(\mathbf{v},\mathbf{v})}\mathbf{v}\right)\\ &= (\mathbf{v},\mathbf{w}) - \frac{(\mathbf{v},\mathbf{w})}{(\mathbf{v},\mathbf{v})}(\mathbf{v},\mathbf{v})\\ &= (\mathbf{v},\mathbf{v}) - (\mathbf{v},\mathbf{v}) = 0\end{aligned}$$

Siden $P_{\mathbf{v}}(\mathbf{w})$ er parallell med $\mathbf{v}$, og

$$\mathbf{w} - P_{\mathbf{v}}(\mathbf{w}) + P_{\mathbf{v}}(\mathbf{w}) = \mathbf{w},$$

gir pytagoras at

$$\|\mathbf{w} - P_{\mathbf{v}}(\mathbf{w})\|^2 + \|P_{\mathbf{v}}(\mathbf{w})\|^2 = \|w\|^2.$$

Siden $\|\mathbf{w} - P_{\mathbf{v}}(\mathbf{w})\| \geq 0$, må

$$\|P_{\mathbf{v}}(\mathbf{w})\|^2 \leq \|w\|^2.$$

Men

$$\begin{aligned}\|P_{\mathbf{v}}(\mathbf{w})\|^2 &= (P_{\mathbf{v}}(\mathbf{w}), P_{\mathbf{v}}(\mathbf{w}))\\ &= \frac{(\mathbf{v},\mathbf{w})^2}{(\mathbf{v},\mathbf{v})^2}(\mathbf{v},\mathbf{v}) = \frac{(\mathbf{v},\mathbf{w})^2}{(\mathbf{v},\mathbf{v})} = \frac{(\mathbf{v},\mathbf{w})^2}{\|\mathbf{v}\|^2}\end{aligned}$$

slik at

$$\frac{(\mathbf{v},\mathbf{w})^2}{\|\mathbf{v}\|^2} \leq \|w\|^2.$$

Dersom vi ganger opp med $\|\mathbf{v}\|^2$ og tar det positive rotutdraget, får vi

**Cauchy-Schwarz' ulikhet**

La $\mathbf{v}$ og $\mathbf{w}$ være vektorer i et indreproduktrom. Da gjelder at

$$|(\mathbf{v},\mathbf{w})| \leq \|v\|\|w\|.$$

Det ble riktignok antatt at $\mathbf{v} \neq \mathbf{0}$ i resonnementet over, men ulikheten er trivielt sann dersom $\mathbf{v} = \mathbf{0}$.

Vi kan spinne litt videre, og beregne at

$$\begin{aligned}\|v+w\|^2 &= (v+w, v+w)\\ &= \|v\|^2 + (v,w) + (w,v) + \|w\|^2\\ &\leq \|v\|^2 + |(v,w)| + |(w,v)| + \|w\|^2\\ &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 =\\ &\left(\|v\|^2 + \|w\|^2\right)^2,\end{aligned}$$

som gir

**Trekantulikheten**

La $\mathbf{v}$ og $\mathbf{w}$ være vektorer i et indreproduktrom. Da gjelder at

$$\|v+w\| \leq \|v\| + \|w\|$$

Dersom man har et vektorrom med indreprodukt, er det noen basiser som er bedre enn andre.

**Ortogonal basis**

Et endeligdimensjonalt vektorrom har en ortogonal basis.

*Bevis.* La $V$ være vektorrommet, og la $\mathbf{v}_1$, $\mathbf{v}_2$, ...,$\mathbf{v}_n$ være en basis for $V$. Vi skal lage oss en ortogonal basis $\mathbf{u}_1,\ \mathbf{u}_2, \cdots, \mathbf{u}_n$ for $V$. Prosedyren kalles Gram-Schmidts metode, og går som følger.

Husk at ingen av vektorene $\mathbf{v}_1,\ \mathbf{v}_2, \cdots, \mathbf{v}_n$ kan være null. Vi begynner med å definere

$$\mathbf{u}_1 = \mathbf{v}_1.$$

Vektoren $\mathbf{u}_1$ ikke nødvendigvis ortogonal på $\mathbf{v}_2$, men det er lett å sjekke at vektoren

$$\mathbf{u}_2 = \mathbf{v}_2 - P_{\mathbf{u}_1}\mathbf{v}_2 = \mathbf{v}_2 - \frac{(\mathbf{u}_1, \mathbf{v}_2)}{(\mathbf{u}_1, \mathbf{u}_1)}\mathbf{u}_1$$

er det:

$$\begin{aligned}(\mathbf{u}_1, \mathbf{u}_2) &= \left(\mathbf{u}_1, \mathbf{v}_2 - \frac{(\mathbf{u}_1, \mathbf{v}_2)}{(\mathbf{u}_1, \mathbf{u}_1)}\mathbf{u}_1\right)\\ &= (\mathbf{u}_1, \mathbf{v}_2) - \frac{(\mathbf{u}_1, \mathbf{v}_2)}{(\mathbf{u}_1, \mathbf{u}_1)}(\mathbf{u}_1, \mathbf{u}_1)\\ &= (\mathbf{u}_1, \mathbf{v}_2) - (\mathbf{u}_1, \mathbf{v}_2) = 0\end{aligned}$$

Det er også lett å se at $\mathbf{u}_2$ ikke er nullvektoren. Dersom så var tilfelle, impliserer

$$\mathbf{v}_2 - \frac{(\mathbf{u}_1, \mathbf{v}_2)}{(\mathbf{u}_1, \mathbf{u}_1)}\mathbf{u}_1 = \mathbf{0}$$

at $\mathbf{u}_1$ og $\mathbf{v}_2$ er lineært avhengige, noe som ikke er sant. Siden $\mathbf{u}_1$ og $\mathbf{u}_2$ er ortognonale, og ingen av dem nullvektoren, er de lineært uavhengige. De spenner ut det samme rommet som $\mathbf{v}_1$ og $\mathbf{v}_2$, for dersom $\mathbf{w}$ kan skrives som en lineærkombinasjon av $\mathbf{v}_1$ og $\mathbf{v}_2$, kan $\mathbf{w}$ også skrives som en lineærkombinasjon av $\mathbf{u}_1$ og $\mathbf{u}_2$:

$$\begin{aligned}\mathbf{w} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2\\ &= c_1\mathbf{u}_1 + c_2\left(\mathbf{u}_2 + \frac{(\mathbf{u}_1, \mathbf{v}_2)}{(\mathbf{u}_1, \mathbf{u}_1)}\mathbf{u}_1\right)\end{aligned}$$

På samme vis kan vi sjekke at vektoren

$$\begin{aligned}\mathbf{u}_3 &= \mathbf{v}_3 - P_{\mathbf{u}_1}\mathbf{v}_3 - P_{\mathbf{u}_2}\mathbf{v}_3\\ &= \mathbf{v}_3 - \frac{(\mathbf{u}_1, \mathbf{v}_3)}{(\mathbf{u}_1, \mathbf{u}_1)}\mathbf{u}_1 - \frac{(\mathbf{u}_2, \mathbf{v}_3)}{(\mathbf{u}_2, \mathbf{u}_2)}\mathbf{u}_2\end{aligned}$$

ikke er nullvektoren, står ortogonalt på både $\mathbf{u}_1$ og $\mathbf{u}_2$, og sammen med $\mathbf{u}_1$ og $\mathbf{u}_2$ spenner ut det samme rommet som $\mathbf{v}_1$, $\mathbf{v}_2$ og $\mathbf{v}_3$. Vi kan nå fortsette slik, og definere rekursivt

$$\begin{aligned}\mathbf{u}_k &= \mathbf{v}_k - \sum_{j=1}^{k-1} P_{\mathbf{u}_j}\mathbf{v}_k\\ &= \mathbf{v}_k - \sum_{j=1}^{k-1} \frac{(\mathbf{u}_j, \mathbf{v}_k)}{(\mathbf{u}_j, \mathbf{u}_j)}\mathbf{u}_j,\end{aligned}$$

og så sjekke ved induksjon at vi får en ortogonal basis for $V$. □

Dersom en ortogonal mengde $\mathbf{u}_1$, $\mathbf{u}_2$, ...,$\mathbf{u}_n$ spenner ut et rom $V$, sier vi at mengden er en *ortogonal basis* for $V$. Hvis vi har en ortogonal basis for et rom, er det veldig lett å finne en vektors kordinater i dette rommet. La oss si at vi ønsker å finne vektoren $\mathbf{v}$ sine komponenter i basisen $(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n)$. Vektoren skrives

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + ... + c_n\mathbf{u}_n$$

og komponentene er $(c_1, c_2, \cdots, c_n)$. Tar vi indreproduktet av begge sidene av likningen med $\mathbf{u}_k$, får vi en haug med kanselleringer, og det eneste som står igjen, er

$$(\mathbf{u}_k, \mathbf{v}) = c_k(\mathbf{u}_k, \mathbf{u}_k),$$

eller

$$c_k = \frac{(\mathbf{u}_k, \mathbf{v})}{(\mathbf{u}_k, \mathbf{u}_k)}.$$

**Det er så lett!**

La $\mathbf{u}_1$, $\mathbf{u}_2$, ...,$\mathbf{u}_n$ være en ortogonal basis for $V$, og la $\mathbf{v} \in V$. Da er

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$$

der

$$c_k = \frac{(\mathbf{u}_k, \mathbf{v})}{(\mathbf{u}_k, \mathbf{u}_k)}.$$

Dersom basisen er ortonormal, blir teoremet enda enklere, og pytagoras gir en ekstra fun fact.

**Pytagoras II**

La $\mathbf{u}_1$, $\mathbf{u}_2$, ...,$\mathbf{u}_n$ være en ortonormal basis for $V$, og la $\mathbf{v} \in V$. Da er

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$$

der

$$c_k = (\mathbf{u}_k, \mathbf{v}),$$

og

$$\|\mathbf{v}\|^2 = \sum_{k=1}^{n} |c_k|^2.$$

*Bevis.* Dersom basisen er ortonormal, blir

$$(\mathbf{u}_k, \mathbf{u}_k) = 1$$

for alle $k$. Tar man indreproduktet av

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n$$

med seg selv og gjør alle kanselleringer, detter

$$\|\mathbf{v}\|^2 = \sum_{k=1}^{n} |c_k|^2.$$

ut. Jeg har skrevet absoluttverditegn for at det skal være tydelig hvordan teoremet ser ut for komplekse indreprodukt, se lenger ned. □

Det er ikke uvanlig å kalle

$$c_k = \frac{(\mathbf{u}_k, \mathbf{v})}{(\mathbf{u}_k, \mathbf{u}_k)}$$

for fourierkoeffisientene. Dette vil bli klart litt lenger ned. Det faktum at

$$\|\mathbf{v}\|^2 = \sum_{k=1}^{n} |c_k|^2.$$

kalles gjerne Parsevals teorem, ihvertfall dersom $n = \infty$.

**Eksempel 7.47.** Dersom du ønsker å løse et lineært likningssystem

$$U\mathbf{x} = \mathbf{b},$$

der søylene i matrisen er reelle og ortogonale vektorer $\mathbf{u}_1$, $\mathbf{u}_2$, ...,$\mathbf{u}_n$, er det bare å gange på begge sider med $U^T$:

$$U^TU\mathbf{x} = U^T\mathbf{b}$$

For siden kolonnene til $U$ er ortogonale, blir den kvadratiske matrisen $U^TU$ diagonal:

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{u}_2^T\mathbf{u}_2 & 0 & \dots & 0 \\ 0 & 0 & \mathbf{u}_3^T\mathbf{u}_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{u}_n^T\mathbf{u}_n \end{bmatrix}$$

Løsningen av systemet $U^TU\mathbf{x} = U^T\mathbf{b}$ bare å skrive rett opp uten gausselimiasjon:

$$\mathbf{v} = \frac{\mathbf{u}_1^T\mathbf{b}}{\mathbf{u}_1^T\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{u}_2^T\mathbf{b}}{\mathbf{u}_2^T\mathbf{u}_2}\mathbf{u}_2 + \dots + \frac{\mathbf{u}_n^T\mathbf{b}}{\mathbf{u}_n^T\mathbf{u}_n}\mathbf{u}_n$$

Vi sier at $U$ er en *ortogonal matrise* dersom søylene i $U$ er ortogonale. △

Dersom $\mathbf{v} \notin V$, har uttrykket

$$\frac{(\mathbf{u}_1, \mathbf{v})}{(\mathbf{u}_1, \mathbf{u}_1)}\mathbf{u}_1 + \frac{(\mathbf{u}_2, \mathbf{v})}{(\mathbf{u}_2, \mathbf{u}_2)}\mathbf{u}_2 + \dots + \frac{(\mathbf{u}_n, \mathbf{v})}{(\mathbf{u}_n, \mathbf{u}_n)}\mathbf{u}_n$$

allikevel en artig egenskap: Dette er vektoren i $V$ som er nærmest $\mathbf{v}$, målt i lengden gitt av indreproduktet.

**Kortest avstand**

La $\mathbf{u}_1$, $\mathbf{u}_2$, ...,$\mathbf{u}_n$ være en ortogonal basis for $V$, og la $\mathbf{v} \notin V$. Punktet

$$\mathbf{v}' = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_n\mathbf{u}_n$$

der

$$c_k = \frac{(\mathbf{u}_k, \mathbf{v})}{(\mathbf{u}_k, \mathbf{u}_k)}$$

er det punktet i $V$ som har kortest avstand til $\mathbf{v}$:

$$\|\mathbf{v} - \mathbf{v}'\| = \min_{\mathbf{w}\in V} \|\mathbf{v} - \mathbf{w}\|$$

*Bevis.* Vi må først bevise at $\mathbf{v}-\mathbf{v}'$ står ortogonalt på $V$. Rommet $V$ er utspent av $\mathbf{u}_1$, $\mathbf{u}_2$, ...,$\mathbf{u}_n$. Vi sjekker at $\mathbf{v} - \mathbf{v}'$ står ortogonalt på hver $\mathbf{u}_k$:

$$\begin{aligned}(\mathbf{v} - \mathbf{v}', \mathbf{u}_k) &= (\mathbf{v}, \mathbf{u}_k) - (\mathbf{v}', \mathbf{u}_k) \\ &= (\mathbf{v}, \mathbf{u}_k) - (\mathbf{v}, \mathbf{u}_k) = 0\end{aligned}$$

Dersom $\mathbf{w} \in V$, ligger også $\mathbf{w} - \mathbf{v}'$ i $V$, og da står $\mathbf{w} - \mathbf{v}'$ og $\mathbf{v} - \mathbf{v}'$ ortogonalt på hverande. Pytagoras' teorem gir

$$\begin{aligned}\|\mathbf{v} - \mathbf{w}\|^2 &= \|\mathbf{v} - \mathbf{v}' - (\mathbf{w} - \mathbf{v}')\|^2 \\ &= \|\mathbf{v} - \mathbf{v}'\|^2 + \|\mathbf{w} - \mathbf{v}'\|^2 \geq \|\mathbf{v} - \mathbf{v}'\|^2,\end{aligned}$$

for alle $\mathbf{w} \in V$, slik at

$$\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - \mathbf{v}'\|,$$

og

$$\|\mathbf{v} - \mathbf{v}'\| = \min_{\mathbf{w}\in V} \|\mathbf{v} - \mathbf{w}\|.$$ □

# Komplekse indreprodukt

Dersom $V$ er et vektorrom over $\mathbb{C}$, er det mest naturlig å kreve at indreproduktet blir et komplekst tall. Reglene blir litt annerledes, og følgende aksiomer skal gjelde.

**Komplekse indreprodukt**

La $V$ være et vektorrom over $\mathbb{C}$, la $\mathbf{u}$, $\mathbf{v}$ og $\mathbf{w}$ være vektorer, og $a$ og $b$ komplekse tall. Indreproduktet skal være konjugert symmetrisk

$$(\mathbf{v}, \mathbf{w}) = \overline{(\mathbf{w}, \mathbf{v})},$$

og positivt definitt

$$(\mathbf{v}, \mathbf{v}) > 0 \quad \text{dersom} \quad \mathbf{v} \neq \mathbf{0},$$

og lineært enten første eller andre faktor.

Linearitet i den ene faktoren, impliserer antilinearitet i den andre faktoren. Dersom det komplekse indreproduktet er lineært i andre faktor,

$$(\mathbf{u}, a\mathbf{v} + b\mathbf{w}) = a(\mathbf{u}, \mathbf{v}) + b(\mathbf{u}, \mathbf{w})$$

blir det antilineært i første faktor:

$$(a\mathbf{u} + b\mathbf{v}, \mathbf{w}) = \overline{a}(\mathbf{u}, \mathbf{w}) + \overline{b}(\mathbf{v}, \mathbf{w})$$

Dersom det er lineært i første faktor,

$$(a\mathbf{u}+b\mathbf{v},\mathbf{w}) = a(\mathbf{u},\mathbf{w}) + b(\mathbf{v},\mathbf{w})$$

blir det antilineært i andre faktor:

$$(\mathbf{u}, a\mathbf{v}+b\mathbf{w}) = \overline{a}(\mathbf{u},\mathbf{v}) + \overline{b}(\mathbf{u},\mathbf{w})$$

Forskjellige fagfelt foretrekker forskjellig versjon, så det er greit å vite om at dette kan gjøres på to måter. Vi skal ha mest bruk for den varianten som er lineær i første faktor, for det er denne som er vanlig i fourieranalyse. Relle indreprodukt er lineære i begge faktorer, så da faller problemstillingen bort.

**Eksempel 7.48.** La $V$ være et vektorrom av integrerbare funksjoner $f : [-\pi, \pi] \to \mathbb{C}$. Uttrykket

$$(f,g) = \int_{-\pi}^{\pi} f(x)\overline{g(x)}\,dx.$$

definerer et komplekst indreprodukt på $V$, og funksjonene $e^{int}$ er ortogonale med hensyn på dette indreproduktet. Dette indreproduktet er lineært i første faktor. △

**Eksempel 7.49.** La

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

være en kompleks $m\times n$-matrise. Hvis vi transponerer $A$ og komplekskonjugerer komponentene, får vi den *adjungerte* matrisen

$$A^* = \begin{bmatrix} \overline{a}_{11} & \overline{a}_{21} & \cdots & \overline{a}_{m1} \\ \overline{a}_{12} & \overline{a}_{22} & \cdots & \overline{a}_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a}_{1n} & \overline{a}_{2n} & \cdots & \overline{a}_{mn} \end{bmatrix}.$$

La $\mathbf{v}$ og $\mathbf{w}$ være søylevektorer i $\mathbb{C}^n$. *Indreproduktet* mellom dem er som regel definert som:

$$(\mathbf{v},\mathbf{w}) = \mathbf{v}^*\mathbf{w} = \overline{v}_1 w_1 + \overline{v}_2 w_2 + \cdots + \overline{v}_n w_n$$

Merk at $(\mathbf{v},\mathbf{v})$ består av de kvadrerte absoluttverdiene til komponentene til $\mathbf{v}$, slik at at vi får et fornuftig mål på lengden til $\mathbf{v}$. (Dersom $n = 1$, klapper dette sammen til å bli modulus til et komplekst tall.)

Dette indreproduktet er lineært i andre faktor, som antagelig springer ut fra tradisjonen med å fokusere på søylevektorer. Skulle vi hatt et indreprodukt som var lineært i første faktor, hadde det vært mer naturlig å si at $\mathbf{v}$ og $\mathbf{w}$ var radvektorer i $\mathbb{C}^n$, og definere

$$(\mathbf{v},\mathbf{w}) = \mathbf{v}\mathbf{w}^* = v_1\overline{w}_1 + v_2\overline{w}_2 + \cdots + v_n\overline{w}_n,$$

men dette er altså ikke denne varianten som er vanligst i bruk. △

Akkurat som for reelle indreprodukt, har vi at $(\mathbf{w},\mathbf{v}) = 0$ dersom $(\mathbf{v},\mathbf{w}) = 0$. De fleste teoremene vi beviste for reelle indreproduktrom, gjelder for komplekse indreproduktrom, men unntaket er Pytagoras' teorem, som blir litt annerledes; implikasjonen går bare en vei når indreproduktet er komplekst.

**Pytagoras III**

Dersom vektorene $\mathbf{v}$ og $\mathbf{w}$ er ortogonale, er

$$\|\mathbf{v}+\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.$$

*Bevis.* Vi vet at

$$\begin{aligned} \|\mathbf{v}+\mathbf{w}\|^2 &= (\mathbf{v}+\mathbf{w},\mathbf{v}+\mathbf{w}) \\ &= (\mathbf{v},\mathbf{v}) + (\mathbf{v},\mathbf{w}) + (\mathbf{w},\mathbf{v}) + (\mathbf{w},\mathbf{w}) \\ &= \|\mathbf{v}\|^2 + (\mathbf{v},\mathbf{w}) + (\mathbf{w},\mathbf{v}) + \|\mathbf{w}\|^2 \\ &= \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 \end{aligned}$$

siden $(\mathbf{v},\mathbf{w}) = (\mathbf{w},\mathbf{v}) = 0$. □

**Eksempel 7.50.** I den komplekse varianten av Pytagoras' teorem er det kun enveis implikasjon. Det er nemlig ikke nødvendigvis sant at $(\mathbf{v},\mathbf{w}) = 0$ dersom

$$\|\mathbf{v}-\mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2,$$

for uttrykket

$$(\mathbf{v},\mathbf{w}) + (\mathbf{w},\mathbf{v})$$

kan kansellere på andre måter enn at begge indreprodukt er null. Se øvingsopplegg. △

**Eksempel 7.51.** Vektorene

$$\begin{bmatrix} 1 \\ i \end{bmatrix} \quad \text{og} \quad \begin{bmatrix} i \\ 1 \end{bmatrix}$$

er ortogonale. △

Den ortogonale projeksjonen må defineres som

$$P_{\mathbf{v}}(\mathbf{w}) = \frac{(\mathbf{w},\mathbf{v})}{(\mathbf{v},\mathbf{v})}\mathbf{v},$$

eller

$$P_{\mathbf{v}}(\mathbf{w}) = \frac{(\mathbf{v},\mathbf{w})}{(\mathbf{v},\mathbf{v})}\mathbf{v},$$

alt etter om man opererer med et indreprodukt som er lineært i første eller andre faktor. En lineæravbildning skal jo være lineær, ikke antilineær!

**Eksempel 7.52.** La oss projisere vektoren

$$\mathbf{w} = \begin{bmatrix} -5i \\ 0 \\ 2i \end{bmatrix}$$

både på og normalt på

$$\mathbf{v} = \begin{bmatrix} 3 \\ -i \\ 4 \end{bmatrix}.$$

Vi beregner:

$$(\mathbf{v},\mathbf{v}) = 3\cdot 3 + i\cdot(-i) + 4\cdot 4 = 26$$

og

$$(\mathbf{v},\mathbf{w}) = 3\cdot(-5i) + i\cdot 0 + 4\cdot 2i = -7i$$

slik at

$$P_{\mathbf{v}}(\mathbf{w}) = \frac{(\mathbf{v},\mathbf{w})}{(\mathbf{v},\mathbf{v})}\mathbf{v} = \frac{-7i}{26}\begin{bmatrix} 3 \\ -i \\ 4 \end{bmatrix}$$

og

$$\mathbf{w} - P_{\mathbf{v}}(\mathbf{w}) = \mathbf{w} - \frac{(\mathbf{v},\mathbf{w})}{(\mathbf{v},\mathbf{v})}\mathbf{v}$$
$$= \begin{bmatrix} -5i \\ 0 \\ 2i \end{bmatrix} - \frac{-7i}{26}\begin{bmatrix} 3 \\ -i \\ 4 \end{bmatrix} = \frac{1}{26}\begin{bmatrix} -109i \\ 7 \\ 80i \end{bmatrix} \triangle$$

**Eksempel 7.53.** Partialsummen til en fourierrekke

$$S_N(t) = \sum_{n=-N}^{N} c_n e^{int}$$

der

$$c_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt,$$

er en ortogonal projeksjon av $f$ ned i vektorrommet utspent av funksjonene $e^{int}$ for $-N \leq n \leq N$. Dersom $N \to \infty$, utgjør disse funksjonene en basis for et uendeligdimensjonalt vektorrom som heter $L^2(-\pi,\pi)$, og for alle $f$ i dette rommet er det riktig å skrive at

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}.$$

Dette kan ikke vi vise, for det er ganske kompliert, og man må først gjennom et komplisert kurs i noe som heter funksjonalanalyse. Men det er greit å vite at fourierrekker egentlig bare er lineæralgebra. Matematikkens styrke er at mange ting som ser helt forskjellige ut, oppfører seg temmelig likt. Man må bare studere oppførselen nøye, litt som en etolog som studerer sjimpanser, gorillaer, orangutanger eller fjellbavianer. △

# Minste kvadraters metode

Dette er en teknikk for å finne tilnærmede løsninger til lineære systemer med flere likninger enn ukjente. Teoremet om kortest avstand forteller hvordan vi kan gjøre dette på en fornuftig måte om vi har en ortogonal basis for søylerommet til matrisen. Minste kvadraters metode er en teknikk som finner den samme løsningen, men som ikke avhenger av at vi har en ortogonal basis for søylerommet.

La oss si at $A$ er en $m \times n$-matrise, at $\mathbf{x}$ og $\mathbf{b}$ er kolonnevektorer i $\mathbb{C}^n$, og at vi ønsker å betrakte systemet

$$A\mathbf{x} = \mathbf{b}$$

for $m > n$. Dette systemet vil ikke ha noen løsning med mindre $\mathbf{b}$ tilfeldigvis ligger i kolonnerommet til $A$, så vi ønsker istedet å finne den $\mathbf{x}$ som minimerer avstanden fra $A\mathbf{x}$ til $\mathbf{b}$. Hvis vi krever at vektoren $A\mathbf{x}-\mathbf{b}$ står ortogonalt på kolonnerommet til $A$, oppnår vi dette. Altså må vi ha

$$A^*(A\mathbf{x} - \mathbf{b}) = \mathbf{0}$$

eller

$$A^*A\mathbf{x} = A^*\mathbf{b}.$$

Dette er et $n \times n$-system som kalles normallikningene. Løsningen av systemet gir den $\mathbf{x}$ som minimerer avstanden fra $A\mathbf{x}$ til $\mathbf{b}$.

**Eksempel 7.54.** Vi ønsker å bruke minste kvadrats metode på systemet med totalmatrise

$$\left(\begin{array}{cc|c} 0 & 1 & 1-i \\ i & i & 1+i \\ 0 & i & i \end{array}\right)$$

Vi ganger matrisen på venstre side av likningssystemet med sin adjungerte

$$\begin{bmatrix} 0 & -i & 0 \\ 1 & -i & -i \end{bmatrix}$$

og får

$$\begin{bmatrix} 0 & -i & 0 \\ 1 & -i & -i \end{bmatrix}\begin{bmatrix} 0 & 1 \\ i & i \\ 0 & i \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}.$$

Vi ganger høyresiden med den adjungerte av venstresiden, og får

$$\begin{bmatrix} 0 & -i & 0 \\ 1 & -i & -i \end{bmatrix}\begin{bmatrix} 1-i \\ 1+i \\ i \end{bmatrix} = \begin{bmatrix} 1-i \\ 3-2i \end{bmatrix}$$

Løsningen av systemet med totalmatrise

$$\left(\begin{array}{cc|c} 1 & 1 & 1-i \\ 1 & 3 & 3-2i \end{array}\right)$$

er

$$\begin{bmatrix} -i/2 \\ 1-i/2 \end{bmatrix}.$$

Dette betyr at vektoren

$$\begin{bmatrix} 0 & 1 \\ i & i \\ 0 & i \end{bmatrix}\begin{bmatrix} -i/2 \\ 1-i/2 \end{bmatrix} = \begin{bmatrix} 1-i/2 \\ 1+i \\ 1/2+i \end{bmatrix}$$

er det punktet i kolonnerommet til matrisen

$$\begin{bmatrix} 0 & 1 \\ i & i \\ 0 & i \end{bmatrix}$$

som minimerer avstanden til punktet

$$\begin{bmatrix} 1-i \\ 1+i \\ i \end{bmatrix} \qquad \triangle$$

# Regresjon

Hvis du har $n+1$ punkter $(x_i, y_i)$ i $\mathbb{R}^2$, der $x_i$ er forskjellig for alle punktene, vil det alltid være mulig å finne et reelt polynom

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

hvis graf går gjennom alle disse punktene, altså at

$$p(x_i) = y_i$$

for alle $1 \leq i \leq n+1$. Dette kalles *interpolasjon*. Likningene over utgjør et $(n+1)\times(n+1)$-likningssystem for koeffisientene $a_i$ med totalmatrise

$$\left(\begin{array}{ccccc|c} x_1^n & x_1^{n-1} & \ldots & x_1 & 1 & y_1 \\ x_2^n & x_2^{n-1} & \ldots & x_2 & 1 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{n+1}^n & x_{n+1}^{n-1} & \ldots & x_{n+1} & 1 & y_{n+1} \end{array}\right)$$

Vi vet jo fra Lagrangeinterpolasjon at dette likningssystemet må ha en entydig løsning siden det finnes et entydig interpolasjonspolynom så lenge $x_j \neq x_k$ for $j \neq k$.

**Eksempel 7.55.** Vi prøver å finne et annengradspolyom som går gjennom punktene

$$\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix} \quad \text{og} \quad \begin{bmatrix}2\\1\end{bmatrix}.$$

Et annengradspolynom skrives $p(x) = ax^2 + bx + c$, så likningssystemet blir

$$\begin{aligned} c &= 1 \\ a + b + c &= 0 \\ 4a + 2b + c &= 1 \end{aligned}$$

Løsningen er $a = 1$ , $b = -2$ og $c = 1$, slik at polynomet blir $p(x) = x^2 - 2x + 1 = (x-1)^2$. Det er lett å sjekke at polynomet tar de rette verdiene i $x = 0$, $x = 1$ og $x = 2$. △

Dersom man prøver å gjøre den samme prosessen med et polynom som har orden $m < n$ , vil man få det overbestemte $(n+1) \times (m+1)$systemet

$$\left(\begin{array}{ccccc|c} x_1^m & x_1^{n-1} & \dots & x_1 & 1 & y_1 \\ x_2^m & x_2^{n-1} & \dots & x_2 & 1 & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{n+1}^m & x_{n+1}^{n-1} & \dots & x_{n+1} & 1 & y_{n+1} \end{array}\right)$$

Bruker man så minste kvadrats metode på dette systemet, får man et polynom som passer ganske bra til punktene uten at grafen går gjennom hvert enkelt punkt - dette kalles *regresjon*.

**Eksempel 7.56.** Vi prøver å finne et annengradspolyom som går gjennom punktene

$$\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix} \quad \text{og} \quad \begin{bmatrix}3\\2\end{bmatrix}.$$

Likningssystemet blir nå

$$\begin{aligned} c &= 1 \\ a + b + c &= 0 \\ 4a + 2b + c &= 1 \\ 9a + 3b + c &= 2 \end{aligned}$$

Dette systemet har ingen løsning, men vi kan bruke minste kvadrats metode til å finne et polynom som passer ganske bra. Matrisen er:

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix},$$

mens høyresiden $\mathbf{b}$ er:

$$\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

Den adjungerte $A^*$ er:

$$A^* = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Vi ganger $A^*$ med $A$ og $\mathbf{b}$, og får

$$A^*A = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 98 & 36 & 14 \\ 36 & 14 & 6 \\ 14 & 6 & 4 \end{bmatrix}$$

og

$$A^*\mathbf{b} = \begin{bmatrix} 0 & 1 & 4 & 9 \\ 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 22 \\ 8 \\ 4 \end{bmatrix}$$

Vi må løse systemet $A^*A = A^*\mathbf{b}$, altså systemet med totalmatrise

$$\left(\begin{array}{ccc|c} 98 & 36 & 14 & 22 \\ 36 & 14 & 6 & 8 \\ 14 & 6 & 4 & 4 \end{array}\right).$$

Løsningen er

$$\begin{bmatrix} 1/2 \\ -11/10 \\ 9/10 \end{bmatrix}$$

slik at polynomet blir

$$p(x) = \frac{1}{2}x^2 - \frac{11}{10}x + \frac{9}{10}.$$ △

**Eksempel 7.57.** Når statistikere snakker om regresjonslinje, snakker de om et første ordens polynom som prøver å reise gjennom en punktmengde med litt for mange punkter, der minste kvadraters metode er brukt for å finne noen koeffisienter som får linjen til å passe ganske bra til punktene. La oss gjøre dette for punktmengden i eksemplet over:

$$\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\1\end{bmatrix} \quad \text{og} \quad \begin{bmatrix}3\\2\end{bmatrix}.$$

Regresjonslinjen er gitt ved $p(x) = ax + b$, så likningssystemet blir

$$\begin{aligned} b &= 1 \\ a + b &= 0 \\ 2a + b &= 1 \\ 3a + b &= 2 \end{aligned}$$

Dette systemet har ingen løsning. Vi bruker minste kvadrats metode, og beregner

$$A^*A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}$$

og

$$A^*\mathbf{b} = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

Løsningen til $A^*A = A^*\mathbf{b}$ blir

$$\frac{1}{5}\begin{bmatrix}2\\2\end{bmatrix}$$

slik at regresjonslinjen blir

$$p(x) = \frac{2}{5}(x+1)\,. \qquad \triangle$$

# Kapittel 8

# Fourieranalyse

Tradisjonelt sett har fourierrekker vært pensum i TMA4120/25/30/35 på Gløshaugen. Se på gamle eksamener i 4K her for et rikholdig utvalg av regneeksempler. Se også

- Adams kap. 9.9
- Kreyszig kap. 11
- Rudin kap. 8

Joseph Fourier ble satt i fengsel under den franske revolusjon, dro til Egypt med Napoleon Bonaparte, og regnes som oppdageren av drivhuseffekten. Han oppfant også fourierrekker. De satte opp en statue av ham i hjembyen Auxerre i 1849, men den ble visst rekvirert av staten og smeltet om under andre verdenskrig.

## Ka e problemet?

Du har lært at mange glatte funksjoner kan skrives som en potensrekke. For eksempel er

$$\sin x = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

for alle $x$. Joseph Fourier støtte på et annet problem da han prøvde å løse varmelikningen tidlig på 1800-tallet. Han måtte

**PROBLEMET**

skrive en gitt funksjon $f : [0, \pi] \to \mathbb{R}$ som en uendelig rekke av sinusfunksjoner:

$$f(t) = \sum_{n=1}^{\infty} b_n \sin nt$$

Hvis du har fiklet litt med indreproduktrom, går det an å gjette på hvordan det skal gjøres.. Det er nemlig slik at sinusfunksjoner med forskjellige heltallsmultipler av en grunnfrekvens er ortogonale:

$$\int_0^{\pi} \sin nt \sin mt \; dt = \begin{cases} \pi/2 & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

Fourier benyttet seg av dette omtrent som følger. Han ganget likningen

$$f(t) = \sum_{n=1}^{\infty} b_n \sin nt,$$

med $\sin mt$ og integrerte:

$$\int_0^{\pi} f(t) \sin mt \; dt = \int_0^{\pi} \sum_{n=1}^{\infty} b_n \sin nt \sin mt \; dt$$

Så byttet han plass på integraltegnet og summen, og fikk

$$\int_0^{\pi} f(t) \sin mt \; dt = \sum_{n=1}^{\infty} b_n \int_0^{\pi} \sin nt \sin mt \; dt$$

Siden sinusfunksjonene er ortogonale, blir det nå en haug med kanselleringer, slik at

$$\int_0^{\pi} f(t) \sin mt \; dt = \frac{\pi b_m}{2}$$

eller

$$b_n = \frac{2}{\pi} \int_0^{\pi} f(t) \sin mt \; dt.$$

Dette kjenner du forhåpentligvis igjen som projeksjonen av $f$ på $\sin mt$, og vi gjorde bunn og grunn den samme beregningen i kapitlet om indreproduktrom, men da i et endelig antall dimensjoner. Det at vi nå har fått uendelig mange dimensjoner, gjør alt mer komplisert, men det viser seg at veldig mange funksjoner kan skrives som uendelige summer av sinus- eller cosinusfunksjoner.

Joseph Fouriers ideer ble til dels avskrevet av samtidige matematikere, men eksemplene han oppdrev på at en funksjon med knekkpunkt kan skrives som en uendelig rekke av glatte sinusfunksjoner var banebrytende.

## Periodiske funksjoner

Når man skal forstå fourierrekker er det ikke mulig å komme utenom periodiske funksjoner.

**Periode**

En funksjon sies å ha *periode* $p > 0$ dersom

$$f(t + p) = f(t)$$

for alle $t$ i definisjonsmengden til $f$. Den minste $p$ slik at likningen holder for alle $t$, kalles *fundamentalperioden* til $f$.

**Eksempel 8.1.** Funksjonen

$$f(t) = \sin t$$

har perioder $2n\pi$ for alle $n \in \mathbb{N}$. Fundamentalperioden er $2\pi$. △

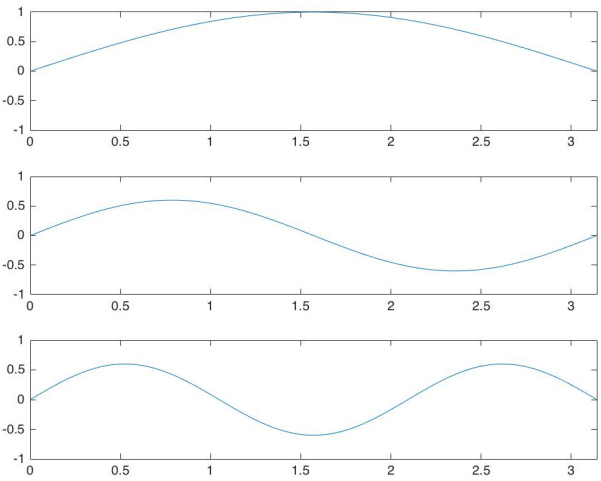
Dersom $f$ har periode $p$, og $g(t) = f(kt)$, har $g$ periode $p/k$, for

$$\begin{aligned} g(t) &= f(kt) = f(kt + p) \\ &= f\left(k\left(t + p/k\right)\right) = g(t + p/k). \end{aligned}$$

**Eksempel 8.2.** Funksjonen

$$f(t) = \sin(3t)$$

har perioder $2n\pi/3$ for alle $n \in \mathbb{N}$. Fundamentalperioden er $2\pi/3$. Under er plot av funksjonene $\sin t$, $\sin 2t$ og $\sin 3t$. △



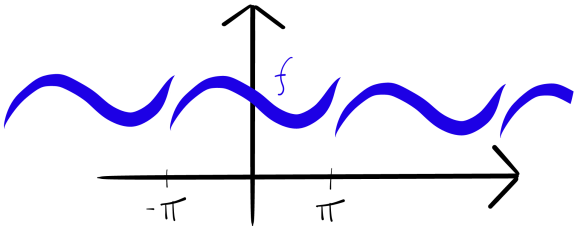

**Eksempel 8.3.** Den komplekse eksponensialfunksjonen $e^{it} = \cos t + i \sin t$ har fundamentalperiode $2\pi$. △

**Eksempel 8.4.** La vi $k = 2\pi/T$ og sammenlikner med forrige eksempel, ser vi at den komplekse eksponensialfunksjonen $e^{2\pi it/T}$ har fundamentalperiode $T$. Det er vanlig å skrive $\omega = 2\pi/T$. △

Hvis man tar en funksjon

$$f : [-\pi, \pi) \to \mathbb{R}$$

og kopierer funksjonen slik at den ser identisk ut på $[\pi, 3\pi)$, $[3\pi, 5\pi)$ og så videre, får man noe som kalles den periodiske utvidelsen til $f$:




**Eksempel 8.5.** Den $2\pi$-periodiske utvidelsen av funksjonen $f : [-\pi, \pi) \to \mathbb{R}$ gitt ved

$$f(t) = \begin{cases} 0 & t < 0 \\ 1 & t \geq 0 \end{cases}$$

kalles en firkantbølge. △

**Eksempel 8.6.** Den $2\pi$-periodiske utvidelsen av funksjonen $f : [-\pi, \pi) \to \mathbb{R}$ gitt ved

$$f(t) = \begin{cases} \pi + t & t < 0 \\ \pi - t & t \geq 0 \end{cases}$$

kalles en trekantbølge. △

**Eksempel 8.7.** Den $2\pi$-periodiske utvidelsen av funksjonen $f : [-\pi, \pi) \to \mathbb{R}$ gitt ved

$$f(t) = t$$

kalles en sagtannbølge. △

Av og til slurver vi litt og omtaler den periodiske utvidelsen til en funksjon $f : [-\pi, \pi] \to \mathbb{R}$. Det er da underforstått at funksjonsverdiene i endepunktene er irrelevanwte.

# Fourierrekker på kompleks form

**Det vakreste**

En *fourierrekke* er en rekke

$$h(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

der

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt$$

kalt fourierkoeffisientene.

Dette er en geometrisk rekke med multiplikasjonsfaktor $e^{it}$, og den er en periodisk funksjon med fundamentalperiode $2\pi$. [1] Av og til er det mer praktisk å skrive

$$h(t) = a_0 + \sum_{n=1}^{\infty} a_n \cos nt + b_n \sin nt.$$

eller

$$h(t) = d_0 + \sum_{n=1}^{\infty} d_n \cos(nt + \phi_n).$$

Følgende teorem skal vi bevise nederst i kapitlet:

[1] Vi skal sette opp fourierrekker med andre perioder enn $2\pi$, men formlene blir mer grisete, så det er lurt å lære seg $2\pi$ først.

**Det pene konvergensteoremet**

La $f : \mathbb{R} \to \mathbb{R}$ være en kontinuerlig deriverbar $2\pi$-periodisk funksjon, og la

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt.$$

Da er

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

for alle $t$.

Nå skal vi gjøre noe som er ganske vanlig i ingeniørmatematikk. Vi skriver opp et presist teorem med strenge betingelser på $f$ fordi dette er greit å bevise, men når vi skal regne eksempler, går vi friskt på med funksjoner ikke tilfredsstiller disse betingelsene, og dette går irriterende nok stort sett bra. Det går selvfølgelig an å si noe om hva som skjer under svakere forutsetninger på $f$, men da blir bevisene mye vanskeligere, og ingeniører har andre ting å gjøre enn å lese lange bevis. Rapporter skal jo skrives.

Dersom $f$ er stykkvis kontinuerlig, og den venstre- og høyrederiverte eksisterer i eventuelle bruddpunkter, er

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

dersom $f$ er kontinuerlig i $t$, og

$$\lim_{h \to 0^+} \frac{f(t-h) + f(t+h)}{2} = \sum_{n=-\infty}^{\infty} c_n e^{int}$$

dersom $f$ ikke er kontinuerlig i $t$. $f$ ikke er det. Vi skriver

$$f(t) \sim \sum_{n=-\infty}^{\infty} c_n e^{int}$$

dersom fourierrekken ikke konvergerer til $f$ for alle $t$. (Det går fint å beregne fourierrekken så lenge $f$ er riemannintegrerbar.)

Dersom $f : [-\pi, \pi] \to \mathbb{R}$ er kontinuerlig deriverbar, blir fourierrekken omtrent lik den $2\pi$-periodiske utvidelsen til $f$. Det er ofte denne ingeniører er interesserte i, siden de bruker fourierrekker til å analysere signaler som er periodiske i tid, for eksempel pipetone eller vekselspenning.

**Det store spørsmålet**

Hvorfor er fourierkoeffisientene gitt ved

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt \quad ?$$

Fourierkoeffisientene $c_n$ er ortogonale projeksjoner, og de komplekse eksponensialfunksjonene

$$e^{int} \quad n \in \mathbb{Z}$$

er innbyrdes ortogonale på $[-\pi, \pi]$:

$$\int_{-\pi}^{\pi} e^{int} \overline{e^{imt}}\, dt = \int_{-\pi}^{\pi} e^{int} e^{-imt}\, dt =$$

$$\int_{-\pi}^{\pi} e^{i(n-m)t}\, dt = \begin{cases} 2\pi & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

Ved første øyekast kan man bli forledet til å tro at alt dette kun funker for harmoniske svingebevegelser, og dette trodde de fleste matematikere helt frem til 1700-tallet eller så. [2] Det viser seg imidlertid at veldig mange funksjoner kan skrives som trigonometriske rekker. Fourier klarte ikke bevise dette, men Dirichlet fant ut av det i 1829, og oppfant en konvergentstest i samme slengen. Det finnes noen berømte eksempler på kontinuerlige funksjoner der fourierrekken ikke konvergerer til funksjonen, men deriverbarhet gjør susen.

**Eksempel 8.8.** Vi finner fourierrekken til $f : \mathbb{R} \to \mathbb{R}$ gitt ved $f(t) = t$. Fourierkoeffisientene blir:

$$c_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} t\, dt = 0$$

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} t e^{-int}\, dt = -\frac{1}{2in}(e^{-in\pi} + e^{in\pi})$$

$$= -\frac{1}{in} \cos n\pi = \frac{(-1)^{n+1}}{in}.$$

Siden $f$ er glatt på intervallet $(-\pi, \pi)$, kan vi skrive

$$t = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{(-1)^{n+1}}{in} e^{int}$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{in} e^{int} - \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{in} e^{-int}$$

$$= 2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin nt$$

på $(-\pi, \pi)$. I $t = \pi$ og $t = -\pi$ konvergerer rekken til 0. Fourierrekken er $2\pi$-periodisk, utenfor intervallet $(-\pi, \pi)$ konvergerer den ikke til $f$. △

La oss merke oss et par ting fra eksemplet over.

- Fourierrekken er reell siden $f$ er det, selv om den er en sum av komplekse eksponensialfunksjoner
- Fourierrekken er lik den $2\pi$-periodiske utvidelsen til funksjonen $g : [-\pi, \pi)$ gitt ved

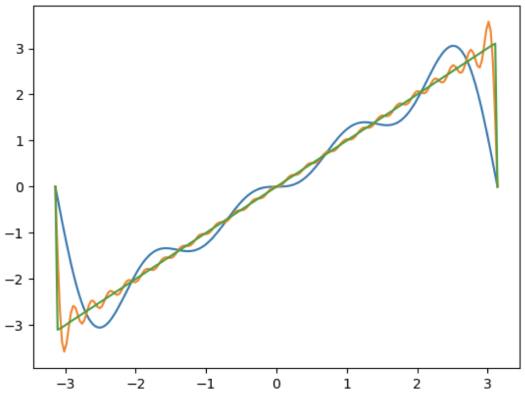
$$g(t) = \begin{cases} t & \text{for } t \in (-\pi, \pi) \\ 0 & \text{for } t = -\pi \end{cases}$$

Uttrykket

$$S_N(t) = \sum_{n=-N}^{N} c_n e^{int}$$

[2] https://en.wikipedia.org/wiki/Fourier_series

kalles *den N-te partialsummen* til fourierrekken. Under er et plot av tre av partialsummene til sagtannbølgen. Den blå er $S_5$, den gule er $S_{25}$ og grønne er$S_{5000}$. Merk hvordan det er umulig å se forskjell på $S_{5000}$ og $f$ unntatt i endepunktene.




# Fourierrekker på reell form

Dersom $f$ er reell, blir fourierrekken også reell. Dette følger av formelen for koeffisientene $c_n$. Dersom $f$ er reell, kan vi skrive:

$$\begin{aligned} c_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{e^{int}}\,dt \\ &= \overline{\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{int}\,dt} = \overline{c_{-n}} \end{aligned}$$

Derfor blir

$$c_n e^{int} + c_{-n}e^{-int} = c_n e^{int} + \overline{c_n e^{int}}$$

blir en reell funksjon, og følgelig er hele fourierrekken reell. Vi kan fint sette opp fourierrekken direkte med sinus- og cosinusfunksjoner. I noen tilfeller gir dette enklere regning, og noen tilfeller vanskeligere.

**Fourierrekker på reell form**

En annen variant er

$$f \sim a_0 + \sum_{n=1}^{\infty} a_n \cos nt + b_n \sin nt$$

med koeffisienter:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\,dt \quad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt$$

Det er vanlig å kalle alt dette 'reell fourierrekke', men det er litt misvisende, siden en fourierrekke kan være reell selv om den er skrevet ned med komplekse eksponensialfunksjoner.

**Eksempel 8.9.** Vi finner enhetssprangfunksjonen

$$u(t) = \begin{cases} 1 & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases}$$

sin reelle fourierrekke på $(-\pi, \pi)$. Vi beregner

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} u(t)\,dt = \frac{1}{2\pi}\int_{0}^{\pi} dt = \frac{1}{2}$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} u(t)\cos nt\,dt = \frac{1}{\pi}\int_{0}^{\pi} \cos nt\,dt = 0$$

$$\begin{aligned} b_n &= \frac{1}{\pi}\int_{-\pi}^{\pi} u(t)\sin nt\,dt = \frac{1}{\pi}\int_{0}^{\pi} \sin nt\,dt \\ &= \begin{cases} \frac{2}{n\pi} & \text{for } n \text{ oddetall} \\ 0 & \text{for } n \text{ partall.} \end{cases} \end{aligned}$$

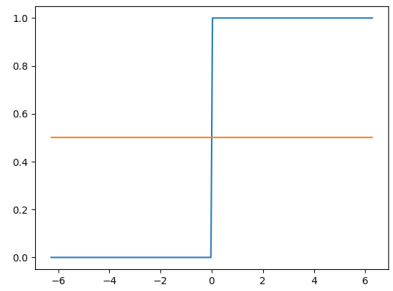
Her kan vi bare skrive

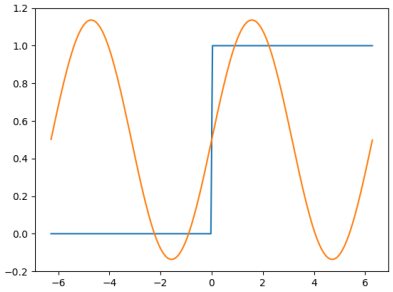
$$u(t) \sim \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{2n-1}\sin(2n-1)t,$$

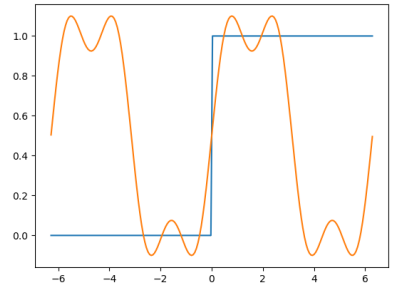
for fourierrekken konvergerer til $u$ på intervallene $(-\pi, 0)$ og $(0, \pi)$, men til $1/2$ i $t = 0$ og $t = \pm\pi$. Partialsummene er gitt ved

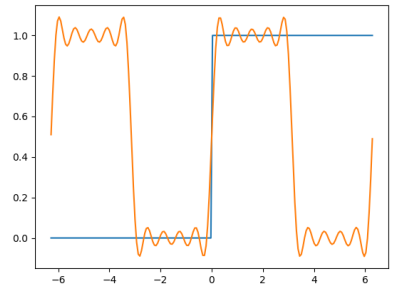
$$S_N = \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{N} \frac{1}{2n-1}\sin(2n-1)t.$$

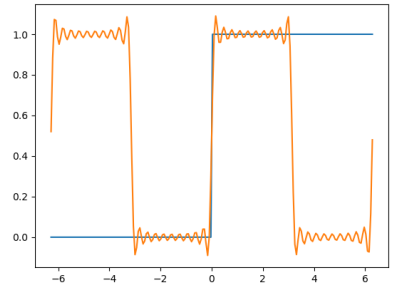
Under er plot av partialsummer for $n = 1$, $n = 2$, $n = 5$, og $n = 10$. Jeg har plottet på intervallet $[-2\pi, 2\pi]$ for å illustrere hvordan fourierrekken oppfører seg utenfor intervallet $[-\pi, \pi]$. Merk den lille over- og underskytningen partialsummene gjør på hver side av spranget. Denne er alltid på rundt 9% av sprangets høyde, og oppførelen kalles Gibbs fenomen, til tross for at det først ble oppdaget av en som het Wilbraham. Spranget forsvinner på mystisk vis når $n \to \infty$. △

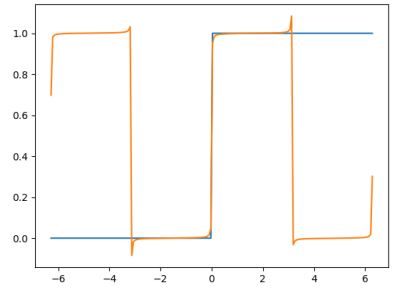












Vi kan utlede formler for overgangen mellom fourierrekker på reell og kompleks form:

$$\begin{aligned} c_n + c_{-n} &= \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} f(t)e^{-int}\,dt + \int_{-\pi}^{\pi} f(t)e^{int}\,dt\right) \\ &= \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} f(t)\left(e^{-int} + e^{int}\right)\,dt\right) \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt = a_n \end{aligned}$$

$$\begin{aligned} c_n - c_{-n} &= \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} f(t)e^{-int}\,dt - \int_{-\pi}^{\pi} f(t)e^{int}\,dt\right) \\ &= \frac{1}{2\pi}\left(\int_{-\pi}^{\pi} f(t)\left(e^{-int} - e^{int}\right)\,dt\right) \\ &= \frac{-i}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt = -ib_n \end{aligned}$$

**Overgang mellom reell og kompleks**

$$c_n = \frac{a_n - ib_n}{2} \qquad c_{-n} = \frac{a_n + ib_n}{2}$$

$$a_n = c_n + c_{-n} \qquad b_n = i(c_n - c_{-n})$$

# Andre intervaller enn $[-\pi, \pi]$

Å utlede formler for fourierrekker på andre intervaller enn $[-\pi, \pi]$ er ikke vanskelig, men formlene blir mer grisete og vanskeligere å huske.

**Fourierrekker på generelt intervall**

For intervallet $[-L, L]$ skriver man

$$f(t) \sim \sum_{n=-\infty}^{\infty} c_n e^{i\frac{n\pi t}{L}}. \tag{8.1}$$

eller

$$f \sim a_0 + \sum_{n=1}^{\infty} a_n \cos\frac{n\pi t}{L} + b_n \sin\frac{n\pi t}{L}.$$

Koeffisientene er gitt ved

$$c_n = \frac{1}{2L}\int_{-L}^{L} f(t)e^{-i\frac{n\pi t}{L}}\,dt,$$

og

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f(t)\,dt$$

$$a_n = \frac{1}{L}\int_{-L}^{L} f(t)\cos\frac{n\pi t}{L}\,dt$$

$$b_n = \frac{1}{L}\int_{-L}^{L} f(t)\sin\frac{n\pi t}{L}\,dt$$

henholdsvis.

Utledning er identisk med utledning på intervallet $[-\pi, \pi]$. På samme måte kan man sette opp fourierrekker på intervallet $[a, b]$, men det dropper vi.

**Eksempel 8.10.** Vi beregner den komplekse fourierrekken til heavisidefunksjonen på $[-1, 1]$. Koeffisientene blir

$$\begin{aligned} c_n &= \frac{1}{2}\int_{-1}^{1} f(t)e^{-i\frac{n\pi t}{L}}\,dt \\ &= \frac{1}{2}\int_{0}^{1} e^{-i\frac{n\pi t}{L}}\,dt = \begin{cases} \frac{1}{2} & \text{for } n = 0 \\ \frac{1}{n\pi i} & \text{for odde } n \\ 0 & \text{for jevne } n, \end{cases} \end{aligned}$$

slik at

$$u(t) \sim \frac{1}{2} + \frac{1}{\pi i}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{(2n-1)} e^{(2n-1)\pi i t}.$$

Merk nok en gang at

$$c_n e^{n\pi i t} + c_{-n} e^{-n\pi i t} = \frac{2}{n\pi}\sin n\pi t,$$

slik at

$$\begin{aligned} u(t) &\sim \frac{1}{2} + \frac{1}{\pi i}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1}{(2n-1)} e^{(2n-1)\pi i t} \\ &= \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{2n-1}\sin(2n-1)\pi t. \end{aligned}$$ △

# Odde og jevne funksjoner

Vi sier at en funksjon er odde dersom

$$f(-t) = -f(t)$$

og jevn dersom

$$f(-t) = f(t).$$

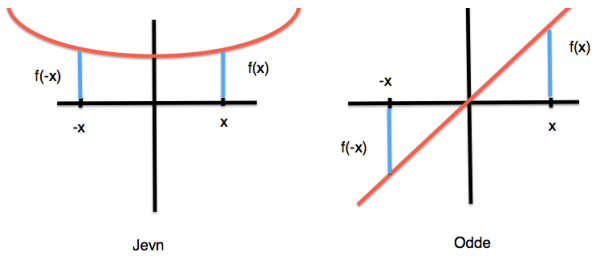
for alle $t$ i definisjonsmengden til $f$. Grafen til en odde funksjon blir identisk dersom du dreier den $\pi$ radianer om origo, mens grafen til en jevn funksjon blir identisk dersom du speiler den om $y$-aksen. En rask kikk på figur viser at

$$\int_{-L}^{L} f(t)\,dt = 0$$

for odde funksjoner, og

$$\int_{-L}^{L} f(t)\,dt = 2\int_{0}^{L} f(t)\,dt$$

for jevne funksjoner.




Dersom $f(-t) = -f(t)$, og $g(-t) = g(t)$, ser vi at

$$f(-t)g(-t) = -f(t)g(t),$$

altså er $fg$ en odde funksjon dersom $f$ er odde og $g$ er jevn. Dersom både $f$ og $g$ er enten jevne eller odde blir $fg$ jevn.

Dersom $f$ er odde, blir $a_n = 0$ for alle $n$, og dersom $f$ er jevn, blir $b_n = 0$ for alle $n$. Fourierrekken til en odde funksjon inneholder derfor kun sinusfunksjoner, mens fourierrekken til jevne funksjoner inneholder kun cosinusfunksjoner.

For $f : [0, L] \to \mathbb{R}$ kan vi definere den odde utvidelsen

$$f_o = \begin{cases} f(t) & \text{for } 0 \le t \le L \\ -f(-t) & \text{for } -L \le t < 0 \end{cases}$$

og den jevne utvidelsen

$$f_j = \begin{cases} f(t) & \text{for } 0 \le t \le L \\ f(-t) & \text{for } -L \le t < 0 \end{cases}$$

Siden både $f_o$ og $f_j$ er identiske med $f$ på $[0, L]$, vil fourierrekkene deres konvergere til $f$ på $[0, L]$. Man kan således velge mellom sinus eller cosinus når man skal fourierutvikle $f$. Disse kalles henholdsvis sinus- og cosinusrekkene til $f$ på $[0, L]$. For fourierutviklingen til $f_o$ er

$$\begin{aligned} b_n &= \frac{1}{L}\int_{-L}^{L} f_o(t)\sin\frac{n\pi t}{L}\,dt \\ &= \frac{2}{L}\int_{0}^{L} f(t)\sin\frac{n\pi t}{L}\,dt. \end{aligned}$$

og for fourierutviklingen til $f_j$ er

$$\begin{aligned} a_n &= \frac{1}{L}\int_{-L}^{L} f_j(t)\cos\frac{n\pi t}{L}\,dt \\ &= \frac{2}{L}\int_{0}^{L} f(t)\cos\frac{n\pi t}{L}\,dt \end{aligned}$$

og

$$a_0 = \frac{1}{2L}\int_{-L}^{L} f_j(t)\,dt = \frac{1}{L}\int_{0}^{L} f(t)\,dt.$$

**Eksempel 8.11.** Vi beregner cosinusrekken til $f(t) = t$ på $(0, \pi)$. Koeffisientene blir

$$a_0 = \frac{1}{\pi}\int_{0}^{\pi} t\,dt = \pi/2$$

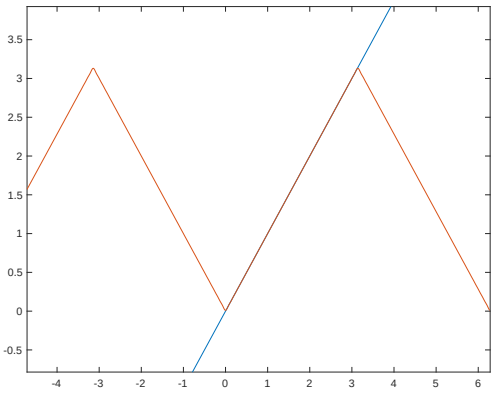
og

$$a_n = \frac{2}{\pi}\int_{0}^{\pi} t\cos nt\,dt = \begin{cases} \frac{-4}{\pi n^2} & \text{for odde } n \\ 0 & \text{for for jevne } n. \end{cases}$$

slik at

$$t = \frac{\pi}{2} - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2}\cos(2n-1)t$$

på intervallet $(0, \pi)$. Under er et plot av $f(t) = t$ og cosinusrekken på et litt større intervall enn $[0, \pi]$. △




# Parsevals teorem

En uendeligdimensjonal variant av pytagoras, kalles Parsevals teorem.

**Parsevals teorem**

Anta $f$ er riemannintegrerbar, og at

$$f(t) \sim \sum_{n=-\infty}^{\infty} c_n e^{i\frac{n\pi t}{L}}.$$

Da gjelder Parsevals identitet:

$$\begin{aligned}\frac{1}{2L}\int_{-L}^{L} f^2(t)\,dt &= \sum_{n=-\infty}^{\infty} |c_n|^2 \\ &= a_0^2 + \frac{1}{2}\sum_{n=1}^{\infty} a_n^2 + b_n^2\end{aligned}$$

**Eksempel 8.12.** Vi kan bruke fourierrekken til heavisidefunksjonen til å finne summen til rekken

$$1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1}$$

Siden

$$u(t) \sim \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{2n-1}\sin(2n-1)t$$

og $u$ er glatt i $t = \pi/2$, ser vi at

$$\begin{aligned}1 = u(\pi/2) &= \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{1}{2n-1}\sin(2n-1)\frac{\pi}{2} \\ &= \frac{1}{2} + \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1}.\end{aligned}$$

Dette betyr at

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1} = \left(1 - \frac{1}{2}\right)\frac{\pi}{2} = \frac{\pi}{4}. \qquad \triangle$$

**Eksempel 8.13.** Vi kan bruke Parsevals identitet til å finne summen til den kjente og kjære rekken

$$1 + \frac{1}{4} + \frac{1}{9} + \cdots = \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Siden

$$t \sim 2\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}\sin nt,$$

gir Parsevals identitet at

$$\frac{2\pi^2}{3} = \frac{1}{\pi}\int_{-\pi}^{\pi} t^2\,dt = \sum_{n=1}^{\infty} b_n^2 = 4\sum_{n=1}^{\infty} \frac{1}{n^2},$$

eller

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2}. \qquad \triangle$$

# Vanskelig teori

Hvis du slår opp i et kapittel om fourierrekker i en vanlig lærebok i matematikk, begynner de gjerne med noe slikt:

Anta at vi har en funksjon $f : [-\pi, \pi] \to \mathbb{R}$ som vi har lyst til å skrive

$$f(t) = \sum_{n=1}^{\infty} b_n \sin nt.$$

Ok, så hva må koeffisientene være? La oss gange hele greia med $\sin mt$ og integrere alt sammen:

$$\int_{-\pi}^{\pi} f(t)\sin mt\,dt = \int_{-\pi}^{\pi}\sum_{n=1}^{\infty} b_n \sin nt \sin mt\,dt$$

Det neste steget er å bytte plass på integraltegnet og summen, slik at vi får

$$\int_{-\pi}^{\pi} f(t)\sin mt\,dt = \sum_{n=1}^{\infty} b_n \int_{-\pi}^{\pi} \sin nt \sin mt\,dt$$

og så bruke at

$$\int_{-\pi}^{\pi} \sin nt \sin mt\,dt = \begin{cases}\pi & \text{for } n = m \\ 0 & \text{for } n \neq m\end{cases}$$

for å se at det blir en haug med kanselleringer på høyresiden, og at vi derfor får

$$\int_{-\pi}^{\pi} f(t)\sin mt\,dt = \pi b_n.$$

Men denne 'utledningen' gir en litt falsk følelse av forståelse. Det er et problem her, som er illustrert av et eksempel fra kapittel 7 i Rudin. Problemet ligger i likningen

$$\begin{aligned}\int_{-\pi}^{\pi}\sum_{n=1}^{\infty} b_n \sin nt \sin mt\,dt = \\ \sum_{n=1}^{\infty} b_n \int_{-\pi}^{\pi} \sin nt \sin mt\,dt\end{aligned}$$

Følgende eksempel viser at det å bytte om på integrasjon og uendelig sum kan være farlig. La

$$f_n(t) = nt\left(1 - t^2\right)^n.$$

Fra rekketeorien ser vi enkelt at

$$\lim_{n\to\infty} f_n(t) = \lim_{n\to\infty} nt\left(1 - t^2\right)^n = 0$$

dersom $0 \leq t \leq 1$, slik at

$$\int_0^1 \lim_{n\to\infty} f_n(t)\, dt = 0.$$

Men

$$\lim_{n\to\infty} \int_0^1 f_n(t)\, dt = \lim_{n\to\infty} \frac{n}{2n+2} = \frac{1}{2}.$$

Det ligger altså ingen nødvendighet i at det går greit å bytte om på uendelig sum og integral:

$$\int_0^1 \lim_{n\to\infty} f_n(t)\, dt \neq \lim_{n\to\infty} \int_0^1 f_n(t)\, dt$$

Følgende strategi er derfor å foretrekke. Strategien er litt teknisk, men leder frem mot riktig resultat uten alt for mange vanskeligheter. En *generell fourierrekke* er et uttrykk på formen

$$\sum_{n=1}^{\infty} c_n \phi_n(t).$$

Vi skal nå utlede litt teori om disse. La $f$ og $g$ være komplekse funksjoner av en reell variabel på intervallet $[a, b]$. Husk at

$$\int_a^b f\overline{g}$$

er et indreprodukt og at $f$ og $g$ er *ortogonale* på $[a, b]$ dersom

$$\int_a^b f\overline{g} = 0.$$

Alle lineærkombinasjoner av en endelig mengde med innbyrdes ortogonale funksjoner $\phi_n$ er et vektorrom, og $\{\phi_n\}$ er en basis for dette vektorrommet. Dersom

$$\int_a^b \phi_n \overline{\phi_m} = \begin{cases} 1 & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

sier vi at familien er *ortonormal*.

**Eksempel 8.14.** Funksjonene

$$\frac{1}{2\pi} e^{int} \quad n \in \mathbb{Z}$$

er et ortonormalt system på intervallet $[-\pi, \pi]$. △

**Eksempel 8.15.** La $m, n \geq 1$. Siden

$$\int_{-\pi}^{\pi} \cos nt \cos mt\, dt = \begin{cases} \pi & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

$$\int_{-\pi}^{\pi} \sin nt \sin mt\, dt = \begin{cases} \pi & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

og

$$\int_{-\pi}^{\pi} \cos nt \sin mt\, dt = 0$$

utgjør $\{\sin nt, \cos mt\}$ også et ortogonalt system på $[-\pi, \pi]$. Vi tar cosinusfunksjonene. Først skriver vi om litt:

$$\int_{-\pi}^{\pi} \cos nt \cos mt; dt =$$
$$\frac{1}{2} \int_{-\pi}^{\pi} \cos(n+m)t + \cos(n-m)t\, dt$$

Det siste integralet er lett å beregne. Det forsvinner for alle verdier av $m$ og $n$, unntatt når $m = n$, for da er

$$\frac{1}{2} \int_{-\pi}^{\pi} \cos(n-m)t\, dt = \int_{-\pi}^{\pi} dt = \pi,$$

slik at

$$\int_{-\pi}^{\pi} \cos nt \cos mt; dt = \begin{cases} \pi & \text{for } n = m \\ 0 & \text{for } n \neq m \end{cases}$$

De to andre formlene bevises på samme måte. △

**Eksempel 8.16.** Legendrepolynomene er en familie av polynomer gitt ved rekursjonen

$$\begin{aligned} P_0(t) &= 1 \\ P_1(t) &= t \\ (n+1)P_{n+1}(t) &= (2n+1)tP_n(t) - nP_{n-1}(t). \end{aligned}$$

Disse er et ortogonalt system på intervallet $[-1, 1]$. De første fem polynomene er:

$$\begin{aligned} P_0(t) &= 1 \\ P_1(t) &= t \\ P_2(t) &= \frac{1}{2}(3t^2 - 1) \\ P_3(t) &= \frac{1}{2}(5t^3 - 3t) \\ P_4(t) &= \frac{1}{8}(35t^4 - 30t^2 + 3) \end{aligned}$$

Disse dukker opp mange steder i anvendelser, for eksempel i interpolasjon, numerisk integrasjon, samt Schrødingers likning for hydrogenatomet. △

Dersom en funksjon kan skrives som en lineærkombinasjon av ortonormale funksjoner:

$$f(t) = \sum_{n=1}^{N} c_n \phi_n(t),$$

vet vi fra kapitlet om indreproduktrom at fourierkoeffisientene kan skrives

$$c_n = \int_a^b f(t)\overline{\phi_n(t)}\, dt$$

og at

$$\sum_{n=1}^{N} |c_n|^2 = \int_a^b f^2(t)\, dt.$$

Dersom $f$ ikke kan skrives som en lineærkombinasjon av funksjonene $\phi_n$, har allikevel uttrykket

$$\sum_{n=1}^{N} c_n \phi_n,$$

noen gunstige egenskaper. Lineærkombinasjonen med fourierkoeffisientene som vekter er en god approksimasjonen til $f$.

**Teorem 8.17.** *Dersom*

$$h(t) = \sum_{n=1}^{N} c_n \phi_n(t),$$

*med*

$$c_n = \int_a^b f(t)\overline{\phi_n(t)}\, dt$$

*og*

$$g(t) = \sum_{n=1}^{N} d_n \phi_n(t),$$

*er*

$$\int_a^b |f(t) - h(t)|^2\, dt \leq \int_a^b |f(t) - g(t)|^2\, dt$$

*Bevis.* Denne beregningen er litt hårete, men du finner nok ut av det med penn og papir:

$$\begin{aligned}
\int_a^b |f(t) - g(t)|^2\, dt = & \\
\int_a^b |f(t)|^2\, dt - \int_a^b f(t)\overline{g(t)}\, dt & \\
- \int_a^b \overline{f(t)} g(t)\, dt + \int_a^b |g(t)|^2\, dt = & \\
\int_a^b |f(t)|^2\, dt - \sum_{n=1}^{N} c_n \overline{d_n} - \sum_{n=1}^{N} \overline{c_n} d_n + \sum_{n=1}^{N} d_n \overline{d_n} = & \\
\int_a^b |f(t)|^2\, dt - \sum_{n=1}^{N} c_n \overline{c_n} + \sum_{n=1}^{N} |c_n - d_n|^2. &
\end{aligned}$$

Det siste uttrykket er helt klart minimert dersom man velger $d_n = c_n$. □

Det neste teoremet kalles gjerne *Bessels ulikhet*, ihvertfall dersom $n \to \infty$.

**Teorem 8.18.** *Dersom*

$$h(t) = \sum_{n=1}^{N} c_n \phi_n(t),$$

*med*

$$c_n = \int_a^b f(t)\overline{\phi_n(t)}\, dt$$

*er*

$$\sum_{n=1}^{N} |c_n|^2 \leq \int_a^b |f(t)|^2\, dt$$

*Bevis.* Fra forrige bevis vet vi at

$$\int_a^b |f(t) - h(t)|^2\, dt = \int_a^b |f(t)|^2\, dt - \sum_{n=1}^{N} c_n \overline{c_n}.$$

Siden

$$\int_a^b |f(t) - h(t)|^2\, dt \geq 0,$$

må

$$\int_a^b |f(t)|^2\, dt \geq \sum_{n=1}^{N} c_n \overline{c_n}.$$ □

Merk til slutt at dersom vi har uendelig mange funksjoner $\phi_n$, impliserer Bessels ulikhet at

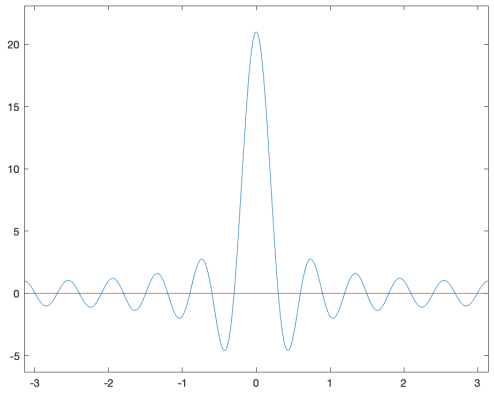
$$\sum_{n=1}^{\infty} |c_n|^2$$

er en konvergent rekke, og at $\lim_{n\to\infty} |c_n| = 0$.

Uttrykket

$$D_N(t) = \sum_{n=-N}^{N} e^{int} = 1 + 2\sum_{n=1}^{N} \cos nt$$

kalles *dirichletkjernen*. Under er et plot av dirichletkjernen for $N = 10$.




Vi kan også lage en enda penere formel. Først utnytter vi at en trigonometrisk rekke er en geometrisk rekke med faktor $e^{it}$, og skriver

$$e^{iNt} \sum_{n=-N}^{N} e^{int} = \sum_{n=0}^{2N} e^{int} = \frac{e^{i(2N+1)t} - 1}{e^{it} - 1},$$

så lenge $e^{it} \neq 1$. Siden

$$\frac{e^{-i(N+1/2)t}}{e^{-it/2}} e^{iNt} = 1$$

kan vi beregne

$$\begin{aligned}
\sum_{n=-N}^{N} e^{int} &= \frac{e^{-i(N+1/2)t}}{e^{-it/2}} e^{iNt} \sum_{n=-N}^{N} e^{int} \\
&= \begin{cases} \frac{e^{i(N+1/2)t} - e^{-i(N+1/2)t}}{e^{it/2} - e^{-it/2}} & t \neq 0 \\ 2N+1 & t = 2k\pi \end{cases} \\
&= \begin{cases} \frac{\sin(N+\frac{1}{2})t}{\sin \frac{1}{2}t} & t \neq 0 \\ 2N+1 & t = 2k\pi \end{cases}
\end{aligned}$$

Vi kan bruke dirichletkjernen til å skrive partialsummen til en fourierrekke som noe som kalles en konvolusjon.

$$\begin{aligned} 2\pi \sum_{n=-N}^{N} c_n e^{int} &= \sum_{n=-N}^{N} \int_{-\pi}^{\pi} f(y) e^{-iny}\, dy\, e^{int} \\ &= \int_{-\pi}^{\pi} f(y) \sum_{n=-N}^{N} e^{in(t-y)}\, dy \\ &= \int_{-\pi}^{\pi} f(y) D_n(t-y)\, dy = f * D_n. \end{aligned}$$

Konvolusjonsbegrepet er ekstremt viktig i signalbehandling, men vi skal vente til neste semesteret med å se nøye på dette.

**Teorem 8.19.** *La $f$ være en $2\pi$-periodisk funksjon slik at $\int_{-\pi}^{\pi} f^2$ konvergerer. Dersom $f$ er deriverbar i $t$, er*

$$\lim_{N\to\infty} \sum_{n=-N}^{N} c_n e^{int} = f(t).$$

*Bevis.* Dette er et pent bevis laget av en kar som heter Paul Chernoff. Formlene i beviset blir litt enklere å skrive opp dersom vi antar $t = 0$ og at $f(0) = 0$. Hvis ikke er dette er tilfelle, kan vi flytte hele problemet slik at $t$ ligger på origo, noe som er helt greit siden $f$ er $2\pi$-periodisk, og så subtrahere $f(0)$ fra $f$, som også er greit siden $f(x) - f(0)$ også tilfredsstiller betingelsene i teoremet. Vi må vise at

$$\lim_{N\to\infty} S_N(0) = \lim_{N\to\infty} \sum_{n=-N}^{N} c_n = 0$$

$$\begin{aligned} \sum_{n=-N}^{N} c_n &= \sum_{n=-N}^{N} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \sum_{n=-N}^{N} e^{-int}\, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{e^{i(N+1/2)t} - e^{-i(N+1/2)t}}{e^{it/2} - e^{-it/2}}\, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{e^{i(N+1)t} - e^{-iNt}}{e^{it} - 1}\, dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(t)}{e^{it} - 1} e^{-i(N+1)t}\, dt \\ &+ \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{f(t)}{e^{it} - 1} e^{iNt}\, dt \end{aligned}$$

De to siste integralene er fourierkoeffisienter til funksjonen

$$g(t) = \frac{f(t)}{e^{it} - 1}.$$

og må derfor gå mot null når $N \to \infty$ dersom $\int_{-\pi}^{\pi} g^2$ konvergerer. Siden $f$ er deriverbar og $f(0) = 0$, er $g$ begrenset i området rundt 0, og følgelig må $\int_{-\pi}^{\pi} g^2$ konvergere siden $\int_{-\pi}^{\pi} f^2$ gjør det. □

# Fouriertransform

Dette er noe som har en spesiell plass i elektroingeniørens hjerte. I dette kapitlet skal vi gå gjennom fouriertransformens grunnleggende egenskaper.

**Fouriertransform**

Fouriertransformen til $f$ er

$$X(\omega) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t) e^{-i\omega t}\, dt.$$

der $\omega$ er en reell variabel.

Siden

$$X(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, dt,$$

ser vi at det gir ingen mening å stappe inn en funksjon som ikke lar seg integrere på hele $t$-aksen. Det er vanlig å kreve at $x$ er absolutt integrerbar, altså at

$$\int_{-\infty}^{\infty} |x(t)|\, dt < \infty.$$

Det er mulig å slakke noe på dette kravet, med det skal ikke vi gjøre. Funksjonen $e^{-i\omega t} = \cos wx - i \sin wx$ er en parametrisering av enhetssirkelen, med $|e^{-i\omega t}| = 1$ for alle $\omega$ og $x$.

**Inngangsbillett til fouriertansform**

Dersom $x$ er absolutt integrerbar, konvergerer integralet

$$\int_{-\infty}^{\infty} x(t) e^{-i\omega t}\, dt.$$

absolutt.

*Bevis.* Dersom $\omega$ og $x$ er reelle, er $|e^{-i\omega t}| = 1$, slik at

$$\begin{aligned} \int_{-\infty}^{\infty} |x(t) e^{-i\omega t}|\, dt &= \int_{-\infty}^{\infty} |x(t)| \cdot |e^{-i\omega t}|\, dt \\ &= \int_{-\infty}^{\infty} |x(t)|\, dt < \infty. \end{aligned}$$ □

**Eksempel 8.20.** La $a > 0$. Vi beregner

$$\begin{aligned} \mathcal{F}(e^{-a|x|}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-a|x|} e^{-iwx}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \left( \int_{0}^{\infty} e^{-ax} e^{-iwx}\, dx + \int_{-\infty}^{0} e^{ax} e^{-iwx}\, dx \right) \\ &= \frac{1}{\sqrt{2\pi}} \left( \int_{0}^{\infty} e^{-x(a+iw)}\, dx + \int_{-\infty}^{0} e^{x(a-iw)}\, dx \right) \\ &= \frac{1}{\sqrt{2\pi}} \left( \frac{1}{a+iw} + \frac{1}{a-iw} \right) \\ &= \frac{1}{\sqrt{2\pi}} \frac{2a}{a^2+w^2} = \sqrt{\frac{2}{\pi}} \frac{a}{a^2+w^2} \end{aligned}$$ △

Vi har tilgang på en formel for den inverse transformen, men det er for vanskelig å bevise for oss.

**Invers fouriertransform**

Dersom $X(\omega)$ er absolutt integrerbar, har vi

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}\,d\omega.$$

For å få en mageføelse for hvorfor inversformelen er som den er, kan vi ta en titt på fourierrekker. Man kan tenke at fouriertransform er en slags fourierrekke der $[-L, L]$ strekkes til å bli hele $x$-aksen. La $f$ være en kontinuerlig og absolutt integrerbar funksjon. Vi setter opp fourierrekken til $f$ på intervallet $(-L, L)$, og gjør en omskrivning:

$$\begin{aligned} x(t) &= \sum_{n=-\infty}^{\infty} c_n e^{i\frac{n\pi t}{L}} \\ &= \sum_{n=-\infty}^{\infty}\left(\frac{1}{2L}\int_{-L}^{L} x(s)e^{-i\frac{n\pi}{L}s}\,ds\right) e^{i\frac{n\pi t}{L}} \\ &= \frac{1}{2\pi}\sum_{n=-\infty}^{\infty}\left(\frac{\pi}{L}\int_{-L}^{L} x(s)e^{-i\frac{n\pi}{L}s}\,ds\right) e^{i\frac{n\pi t}{L}}. \end{aligned}$$

Det siste uttrykket kan tolkes som en riemannsum på aksen der $n$ telles fra $-\infty$ til $\infty$. Gitteravstanden er $\frac{\pi}{L}$, og punktene er gitt ved $\frac{n\pi}{L}$. Hvis vi lar $L \to \infty$, vifter vi det vi kan med armer og bein og får

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x(s)e^{-i\omega s}\,ds\right) e^{iwt}\,d\omega.$$

Vi kjenner igjen det innerste integralet som fouriertransformen til $x$:

$$X(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{-i\omega t}dt.$$

Det ytterste integralet kalles den inverse fouriertransformen, og med noen strenge krav på signalet $x$, er det riktig å skrive at

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega$$

**Teorem 8.21.** *Hvis $f$ er glatt, og alle deriverte synker fryktelig raskt når $|x| \to \infty$, er det riktig at*

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{i\omega t}d\omega$$

Hva betyr det at alle deriverte synker raskt nok når $t \to \infty$? Det er vanlig å kreve at

$$\sup_{t\in\mathbb{R}} |t|^k \left|\frac{d^n}{dt^n}x(t)\right| < \infty$$

for alle $k$ og $n$, for da er det ikke så vanskelig å bevise at inversformelen gjelder (det er fremdeles litt for vanskelig for oss). Funksjoner som tilfredsstiller dette kravet, utgjør et vektorrom som kalles Schwartzrommet. Det går an å slakke på dette kravet, men da må alt baseres på en mye mer komplisert integrasjonsteori oppfunnet av Henri Lebesgue. Ingeniører må jo stadig vekk fouriertransformere funksjoner som ikke tilhører Schwartzrommet, så det er bare å knipe igjen øynene og håpe på det beste som vi pleier.

**Eksempel 8.22.** La $a > 0$. Vi beregner

$$\begin{aligned} \mathcal{F}(e^{-a|t|}) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-a|t|}e^{-i\omega t}\,dt \\ &= \frac{1}{\sqrt{2\pi}}\left(\int_0^{\infty} e^{-at}e^{-i\omega t}\,dt + \int_{-\infty}^{0} e^{at}e^{-i\omega t}\,dt\right) \\ &= \frac{1}{\sqrt{2\pi}}\left(\int_0^{\infty} e^{-x(a+i\omega)}\,dt + \int_{-\infty}^{0} e^{x(a-i\omega)}\,dt\right) \\ &= \frac{1}{\sqrt{2\pi}}\left(\frac{1}{a+i\omega} + \frac{1}{a-i\omega}\right) \\ &= \frac{1}{\sqrt{2\pi}}\frac{2a}{a^2+\omega^2} = \sqrt{\frac{2}{\pi}}\frac{a}{a^2+\omega^2} \end{aligned}$$ △

# Regneregler

Den første er grei. Siden integralet er en en lineæroperator, er også fouriertransformen det.

**Fouriertransform er en lineæroperator**

Dersom $a$ og $b$ er tall og $f$ og $g$ funksjoner, er

$$\mathcal{F}\{ax + by\} = a\mathcal{T}\{x\} + b\mathcal{T}\{y\}.$$

Husk at $\lim_{x\to\pm\infty} x(t) = 0$ siden $x$ er absolutt konvergent.

$$\begin{aligned} \mathcal{F}(\dot{x}) =& \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \dot{x}(t)e^{-i\omega t}\,dt \\ =& \frac{1}{\sqrt{2\pi}}x(t)e^{-i\omega t}\Big|_{-\infty}^{\infty} + \frac{i\omega}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(t)e^{-i\omega x}\,dx \\ =& i\omega\mathcal{F}(x) \end{aligned}$$

slik at

**Derivasjonsregelen**

La $x'$ være absolutt integrerbar på $x$-aksen, og anta at $x(t) \to 0$ når $x \to \pm\infty$. Da er

$$\mathcal{F}\{\dot{x}\} = i\omega\mathcal{F}\{x\}$$

**Eksempel 8.23.** Vi beregner

$$\mathcal{F}(e^{-t^2}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2}e^{-i\omega t}\,dt$$

Dette er litt jobb. Derivasjonsregelen over gir

$$\mathcal{F}(-2te^{-t^2}) = i\omega\mathcal{F}(e^{-t^2}).$$

Men vi kan også observere at

$$\begin{aligned}\mathcal{F}(-2te^{-t^2}) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -2te^{-t^2}e^{-i\omega t}\,dt \\ &= \frac{-2i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -ite^{-t^2}e^{-i\omega t}\,dt \\ &= \frac{-2i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2}\frac{d}{d\omega}e^{-i\omega t}\,dt \\ &= -2i\frac{d}{d\omega}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2}e^{-i\omega t}\,dt \\ &= -2i\frac{d}{d\omega}\mathcal{F}(e^{-t^2}).\end{aligned}$$

Vi mangler litt vanskelig teori for å være helt sikker på den siste beregningen, men vi lar den passere allikevel. Hvis vi setter disse uttrykkene lik hverandre, får vi differensiallikningen

$$i\omega\mathcal{F}(e^{-t^2}) = -2i\frac{d}{d\omega}\mathcal{F}(e^{-t^2})$$

eller

$$\frac{d}{d\omega}\mathcal{F}(e^{-t^2}) + \frac{\omega}{2}\mathcal{F}(e^{-t^2}) = 0$$

for $\mathcal{F}(e^{-t^2})$. Integrerende faktor er

$$e^{\omega^2/4},$$

slik at

$$\frac{d}{d\omega}\left(e^{\omega^2/4}\mathcal{F}(e^{-t^2})\right) = 0$$

eller

$$\mathcal{F}(e^{-t^2}) = Ce^{-\omega^2/4}.$$

Senere i semesteret skal vi se at

$$\hat{f}(0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-t^2}\,dt = \frac{1}{\sqrt{2\pi}}\sqrt{\pi} = \frac{1}{\sqrt{2}},$$

som gir

$$\mathcal{F}(e^{-t^2}) = \frac{1}{\sqrt{2}}e^{-\omega^2/4}.$$

Dersom $a > 0$, kan vi gjøre den samme beregningen og få

$$\mathcal{F}(e^{-at^2}) = \frac{1}{\sqrt{2a}}e^{-\omega^2/4a}. \qquad \triangle$$

**Flere regneregler**

**Tidsskift**

$$\mathcal{F}\left\{x(t-\theta)\right\} = e^{-i\omega\theta}X(\omega)$$

**Frekvensskift**

$$\mathcal{F}\left\{e^{i\theta t}x(t)\right\} = X\left(\omega-\theta\right)$$

**Tidsskalering**

$$\mathcal{F}\left\{x(at)\right\} = \frac{1}{|a|}X\left(\frac{\omega}{a}\right)$$

# Konvolusjon

Konvolusjon er et litt merkelig produkt som dukker opp her og der, og er definert ved

$$x * y = \int_{-\infty}^{\infty} x(\theta)y(t-\theta)\,d\theta. \tag{8.2}$$

Det finnes forskjellige typer konvolusjoner, og det er stort sett forskjellige integrasjonsgrenser som skiller dem. Hvilken type integrasjonsgrenser som er mest relevant, kommer litt an på anvendelsen, men vi begynner med den her.

Konvolusjon kan fremstå som noe umotivert, men har fryktelig mange bruksområder. Mange artige ting er basert på konvolusjon. Noen eksempler er reverbknappen på gitarforsterkeren din, bakgrunnsuskarpheten i det vakre konfirmasjonsbildet ditt, eller autofokusfunksjonen på speilreflekskameraet du fikk til ovennevnte konfirmasjon. Sannsynlighetstettheten til summen av to stokastiske variable er konvolusjonen mellom tetthetene til hver variabel.

**Eksempel 8.24.** Anta at du har to uavhengige stokastiske variable $X$ og $Y$, med sannsynlighetstettheter $f(x)$ og $g(y)$. Hva er sannsynlighetstettheten til $X+Y$?

Dersom $X$ tar verdien $v$ og $Y$ tar verdien $z-v$, tar $X+Y$ verdien $z$. For å finne sannsynlighetstettheten til $X+Y$ må vi derfor summere opp bidraget fra alle sannsynligheter på formen $f(v)g(z-v)$, slik at sannsynlighetstettheten $h(z)$ til $X+Y$ blir

$$h(z) = f * g = \int_{-\infty}^{\infty} f(v)g(z-v)\,dv. \qquad \triangle$$

**Viktig regel**

Det er lett å vise at konvolusjonsoperatoren er kommutativ:

$$x * y = y * x$$

**Konvolusjonsteoremet**

$$x(t) * y(t) \overset{\mathcal{F}}{\longleftrightarrow} X(\omega)Y(\omega),$$

eller på norsk: "Konvolusjon i tidsdomenet tilsvarer multiplikajon i frekvensdomenet."

*Bevis.* Vi bruker varablelskiftet $u = x - v$, $v = v$, og

beregner

$$\begin{aligned}
\mathcal{F}\left\{x(t) * y(t)\right\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(v)y(x-v)\ dv\ e^{-i\omega t}\ dt \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x(v)y(x-v)\ e^{-i\omega t}\ dv\ dt \\
&= \int_{-\infty}^{\infty} x(v)y(u)\ e^{-i\omega(u+v)}\ dv\ du \\
&= \int_{-\infty}^{\infty} y(u)e^{-i\omega u}\int_{-\infty}^{\infty} x(v)\ e^{-i\omega v}\ dv\ du \\
&= \int_{-\infty}^{\infty} y(u)e^{-i\omega u}\ \mathcal{F}(x)\ du \\
&= \mathcal{F}\left\{x\right\}\int_{-\infty}^{\infty} y(u)e^{-i\omega u}\ du \\
&= \mathcal{F}\left\{x\right\}\mathcal{F}\left\{y\right\} = X(\omega)Y(\omega) \qquad \square
\end{aligned}$$

Vi skal løse differensiallikninger med fouriertransform, og da vil vi få bruk for å inverstransformere produkter av fouriertransformer. Konvolusjonsteoremet forteller oss nøyaktig hvordan vi inverstransformerer et slikt produkt.

# Kapittel 9

# Funksjoner fra $\mathbb{R}$ til $\mathbb{R}^n$

En vektorfunksjon er en vektor der komponentene er funksjoner. Vi skal begynne med å studere vektorfunksjoner der definisjonsmengden $\mathbb{R}$ eller en bit av denne, og verdimengden er $\mathbb{R}^2$ eller $\mathbb{C}$. Grafen til en vektorfunksjon kan man tenke på som trajektorien til en flue som surrer rundt i rommet en varm sommerdag. Mengden av alle vektorfunksjoner utgjør et vektorrom med uendelig mange dimensjoner.

## Parametriseringer

Du kjenner allerede mange vektorfunksjoner.

**Eksempel 9.1.** Funksjonen $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$ gitt ved

$$\mathbf{x}(t) = \begin{pmatrix} 1+t \\ 2t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

beskriver en rett linje. Disse var i min tid pensum på videregående skole. △

**Eksempel 9.2.** Funksjonen $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$ gitt ved

$$\mathbf{x}(t) = \begin{pmatrix} 1+2t \\ 4t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ 4 \end{pmatrix}$$

beskriver den samme rette linjen som i forrige eksempel, men fluen flyr her med dobbel fart. △

**Eksempel 9.3.** Funksjonen $\mathbf{y} : [0, 2\pi] \to \mathbb{R}^2$ gitt ved

$$\mathbf{y}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

beskriver enhetssirkelen. Jeg vet ikke om denne vektorfunksjonen er pensum på videregående skole, men det er definitivt pensum å vite at $(\cos t, \sin t)$ er et punkt på enhetssirkelen for alle $t$. △

**Eksempel 9.4.** Funksjonen $z : [0, \pi] \to \mathbb{C}$ gitt ved

$$z(t) = e^{it} = \cos t + i \sin t$$

beskriver også enhetssirkelen, men nå i det komplekse planet. △

**Eksempel 9.5.** Dersom $f : \mathbb{R} \to \mathbb{R}$ er en vanlig envariabel funksjon, vil

$$\mathbf{x}(t) = \begin{pmatrix} t \\ f(t) \end{pmatrix}$$

tegne den samme grafen i $\mathbb{R}^2$. △

Eksemplene over illustrerer at mange forskjellige vektorfunksjoner kan representere det samme grafiske objektet, i den forstand at to forskjellige vektorfunksjoner kan ha nøyaktig den samme grafen, selv om funksjonsuttrykkene er forskjellige. Funksjonsuttrykket $\mathbf{x}(t)$ kalles gjerne en parametrisering for kurven det er snakk om.

## Tangentvektoren

For envariable funksjoner definerer man stigningstall, og så bruker man dette til å skrive opp en likning for tangenten. For vektorfunksjoner definerer vi bare tangent på direkten.

**Fartsvektoren**

Tangenten til vektorfunksjonen $\mathbf{x}(t)$ er

$$\dot{\mathbf{x}}(t) = \lim_{h \to 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h}.$$

Dersom tangenten eksisterer, sier vi at $\mathbf{x}$ er deriverbar. Definisjonen impliserer at dersom

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$$

er

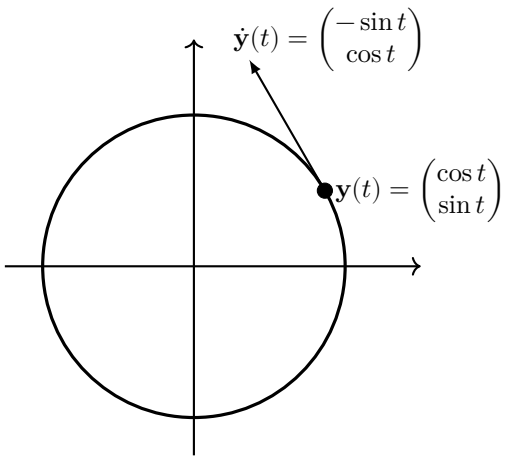
$$\dot{\mathbf{x}}(t) = \begin{pmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{pmatrix}.$$

**Eksempel 9.6.** Enhetssirkelfunksjonen

$$\mathbf{y}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

har tangentvektor gitt ved

$$\mathbf{y}(t) = \begin{pmatrix} -\sin t \\ \cos t \end{pmatrix}.$$




Her er det viktig å holde styr på den geometriske betydningen av $\mathbf{y}(t)$ og $\dot{\mathbf{y}}(t)$. Den første kan du tenke på som et punkt i planet. Den andre kan du tenke på som kursen til partikkelen. △

Her er en haug med derivasjonsregler.

**Noen regneregler**

Dersom $\mathbf{x}$, $\mathbf{y}$ og $\lambda$ er deriverbare, er

$$\frac{d}{dt}(\mathbf{x}(t) + \mathbf{y}(t)) = \dot{\mathbf{x}}(t) + \dot{\mathbf{y}}(t)$$
$$\frac{d}{dt}(\lambda(t)\mathbf{x}(t)) = \dot{\lambda}(t)\mathbf{x}(t) + \lambda(t)\dot{\mathbf{x}}(t)$$
$$\frac{d}{dt}(\mathbf{x}(t) \cdot \mathbf{y}(t)) = \dot{\mathbf{x}}(t) \cdot \mathbf{y}(t) + \mathbf{x}(t) \cdot \dot{\mathbf{y}}(t)$$
$$\frac{d}{dt}(\mathbf{x}(t) \times \mathbf{y}(t)) = \dot{\mathbf{x}}(t) \times \mathbf{y}(t) + \mathbf{x}(t) \times \dot{\mathbf{y}}(t)$$
$$\frac{d}{dt}\mathbf{x}(\lambda(t)) = \dot{\lambda}(t)\dot{\mathbf{x}}(t)$$

Hvis man tenker at fluen har speedometer, er $\|\dot{\mathbf{x}}(t)\|$ farten som vises på speedometeret ved tiden $t$.

**Speedometerets hemmelighet**

Banefarten til vektorfunksjonen $\mathbf{x}(t)$ er

$$v(t) = \|\dot{\mathbf{x}}(t)\|$$

**Eksempel 9.7.** Enhetssirkelfunksjonen

$$\mathbf{y}(t) = \begin{pmatrix} \cos t \\ \sin t \end{pmatrix}$$

konstant fart:

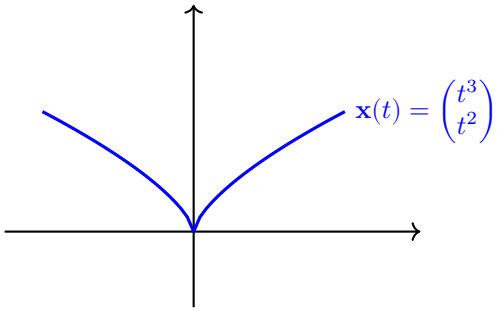
$$\begin{aligned} \|\mathbf{y}(t)\| &= \sqrt{(-\sin t)^2 + (\cos t)^2} \\ &= \sqrt{\sin^2 t + \cos^2 t} = 1 \end{aligned}$$ △

En av de store forskjellene mellom vektorfunksjoner og vanlige funksjoner, er dette med glattheten.

**Eksempel 9.8.** Funksjonen

$$\mathbf{x}(t) = \begin{pmatrix} t^3 \\ t^2 \end{pmatrix}$$

ser slik ut:




Komponentfunksjonene $t^3$ og $t^2$ er deriverbare funksjoner, men grafen til $\mathbf{x}$ er allikevel ikke en glatt kurve. △

Det vi trenger for å luke ut slik patologisk oppførsel, er enhetstangentvektoren.

**Enhetstangentvektoren**

Enhetstangentvektoren til vektorfunksjonen $\mathbf{x}(t)$ er

$$\mathbf{T}(t) = \frac{\dot{\mathbf{x}}(t)}{\|\dot{\mathbf{x}}(t)\|} = \frac{\|\dot{\mathbf{x}}(t)\|}{v(t)}.$$

**Eksempel 9.9.** La oss beregne enhetstangentvektoren til

$$\mathbf{y}(t) = \begin{pmatrix} t^3 \\ t^2 \end{pmatrix}.$$

Den er

$$\mathbf{T}(t) = \frac{\dot{\mathbf{y}}(t)}{\|\dot{\mathbf{y}}(t)\|} = \begin{pmatrix} \frac{3t^2}{\sqrt{9t^4+4t^2}} \\ \frac{2t}{\sqrt{9t^4+4t^2}} \end{pmatrix}.$$

Hva skjer i $t = 0$? La oss regne ut noen grenseverdier. Førstekomponenten er grei nok, siden

$$\lim_{x\to 0^+} \frac{3t^2}{\sqrt{9t^4+4t^2}} = \lim_{x\to 0^-} \frac{3t^2}{\sqrt{9t^4+4t^2}} = 0.$$

Men i andrekomponenten får vi

$$\lim_{x\to 0^+} \frac{2t}{\sqrt{9t^4+4t^2}} = \lim_{x\to 0^+} \frac{t}{|t|}\frac{2}{\sqrt{9t^2+4}} = 1$$

og

$$\lim_{x\to 0^-} \frac{2t}{\sqrt{9t^4+4t^2}} = \lim_{x\to 0^-} \frac{t}{|t|}\frac{2}{\sqrt{9t^2+4}} = -1.$$

Dette betyr at

$$\lim_{x\to 0^+} \mathbf{T}(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

mens

$$\lim_{x\to 0^-} \mathbf{T}(t) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

slik at $\lim_{x\to 0} \mathbf{T}(t)$ ikke eksisterer. △

Enhetstangentvektoren har alltid lengde en. Dette betyr at endringen til enhetsvektoren kun sier noe om hvordan kurven dreier, og dersom $\lim_{t\to a} \mathbf{T}(t)$ ikke eksisterer, vil kurven ha en knekk eller noe slikt i $t = a$.

**Glatt som et olja lyn**

Vi sier at kurven til $\mathbf{x}(t)$ er glatt dersom $\dot{\mathbf{x}}$ er kontinuerlig og $v(t) \neq 0$.

# Normalvektoren

La oss kikke litt på enhetstangentvektoren. Merk at

$$\mathbf{T}(t) \cdot \mathbf{T}(t) = 1,$$

siden $\mathbf{T}$ er en enhetsvektor. Hvis vi deriverer hver side av likningen, får vi

$$2\dot{\mathbf{T}}(t) \cdot \mathbf{T}(t) = 0$$

som sier at den deriverte av enhetstangentvektoren står normalt på enhetstangentvektoren.

**Enhetsnormalvektoren**

Dersom $\mathbf{T}(t) \neq \mathbf{0}$, er enhetsnormalvektoren til kurven gitt ved

$$\mathbf{N}(t) = \frac{\dot{\mathbf{T}}(t)}{\|\dot{\mathbf{T}}(t)\|}.$$

**Eksempel 9.10.** En rett linje har ikke noen definert enhetsnormalvektor. △

La oss anta $\mathbf{T}(t) \neq \mathbf{0}$. Vi kan nå utlede at

$$\dot{\mathbf{y}}(t) = v(t)\mathbf{T}(t),$$

og at

$$\begin{aligned}\ddot{\mathbf{y}}(t) &= \dot{v}(t)\mathbf{T}(t) + v(t)\dot{\mathbf{T}}(t) \\ &= \dot{v}(t)\mathbf{T}(t) + v(t)\|\dot{\mathbf{T}}(t)\|\mathbf{N}(t).\end{aligned}$$

Av dette ser vi at akselerasjonen til en partikkel som dreier har to komponenter, Den ene komponenten tangentiell til banen med størrelse $v(t)$, og den andre står normalt på banen med størrelse $v(t)\|\dot{\mathbf{T}}(t)\|$. Den siste kalles sentripetalakselerasjonen.

En sving på en bilvei trenger ikke være en sirkulær bane, men ingeniører i Vegvesenet snakker allikevel om svingens krumningsradius. Dette er fornuftig mål på hvor krapp svingen er. Vi tenker at det ligger en sirkel som tangerer kurven i $\mathbf{x}(t)$, med sentrum i et sted på linjen spent ut av normalvektoren $\mathbf{N}(t)$ og radius $R(t)$. Dersom en partikkel reiser gjennom $\mathbf{x}(t)$, kan vi stille oss spørsmålet hvilken radius som ville gitt den samme sentripetalakselerasjonen dersom partikkelen hadde gått i en sirkulær bane med samme banefart $v(t)$. For en sirkulær bane er sentripetalakselerasjonen er gitt ved $v^2/R$, og setter vi denne lik den faktiske sentripetalakselerasjonen til partikkelen,

$$v(t)\|\dot{\mathbf{T}}(t)\| = \frac{v^2(t)}{R(t)}$$

kan vi regne ut

$$R(t) = \frac{v(t)}{\|\dot{\mathbf{T}}(t)\|}.$$

Sirkelen kalles oskulasjonssirkelen, og $R(t)$ kalles kurvens krumningsradius. Akkurat som at stigningstall er stigningstallet til en rett linje vi assosierer med hvert punkt på kurven, er krumningsradien radien i en sirkel assosiert med hvert punkt på kurven. Krumningsradien kan også skrives

$$R(t) = \frac{v^3(t)}{|x_1(t)x_2(t) - x_1(t)x_2(t)|},$$

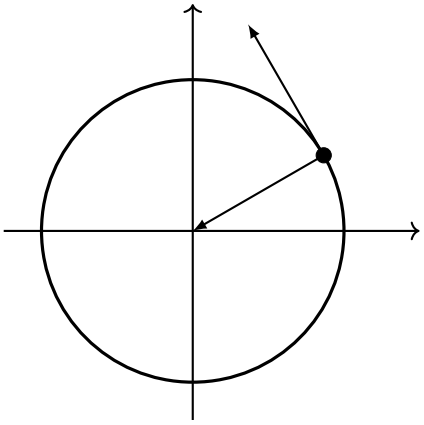
men å vise dette er ganske hårete, se øvingsopplegg neste semester.

**Eksempel 9.11.** Enhetssirkelfunksjonen

$$\mathbf{y}(t) = \begin{pmatrix}\cos t \\ \sin t\end{pmatrix}$$

har enhetsnormalvektor gitt ved

$$\mathbf{N}(t) = \begin{pmatrix}-\cos t \\ -\sin t\end{pmatrix}.$$



Merk at dersom kurven faktisk er en sirkel, vil enhetsnormalen peke inn mot sirkelens sentrum. △

# Kapittel 10

# Funksjoner fra $\mathbb{R}^n$ til $\mathbb{R}$

Nå skal vi studere funksjoner fra $\mathbb{R}^2$ til $\mathbb{R}$. Slike funksjoner kalles skalarfelt, og er ofte gitt ved et funksjonsuttrykk i to variable:

$$z = f(x, y) \quad \text{eller} \quad x_3 = f(x_1, x_2) = f(\mathbf{x})$$

## Nivåkurver

Du bør til å begynne med tenke på $x$ og $y$ (eller $x_1$ og $x_2$) som koordinater, og $z$ (eller $x_3$) som den korresponderende høyden over havet. Funksjonen $f$ angir terrenget, og den positive $y$-aksen (eller $x_2$-aksen) peker mot nord. Siden vi nå må visualisere i tre dimensjoner ($x$, $y$ og $z$ eller $x_1$, $x_2$ og $x_3$) blir livet mer komplisert, og dette tar som regel litt tid å venne seg til.

På et kart forteller ekvidistanselinjene noe om høyden over havet; dersom du følger en ekvidistanselinje, går du hverken opp eller ned. Den matematiske ekvivalenten til ekvidistanselinjene kalles nivåkurve. Disse er gitt ved

$$c = f(x, y),$$

og forskjellige $c$ gir forskjellige høyder over $xy$-planet.

**Eksempel 10.1.** Nivåkurvene til

$$f(x, y) = x + 2y$$

blir rette linjer på formen $c = x + 2y$. △

Akkurat som på kart, kan man bruke avstanden mellom ekvidistanselinjene til å indikere hvor bratt funksjonen stiger.

**Eksempel 10.2.** Nivåkurvene til

$$g(x, y) = (x + 2y)^2$$

er de samme rette linjene som i forrige eksempel. △

Et polynom i to variable er gitt ved

$$p(x, y) = \sum_{k,m} a_{km} x^k y^m$$

Dersom $k + m \leq n$ og $k + m = n$ for minst en kombinasjon av $k$ og $m$, sier vi at polynomet har orden $n$. For eksempel er et generelt førsteordens polynom gitt ved

$$p(x, y) = ax + by + c.$$

Grafen til denne blir alltid et plan i $\mathbb{R}^3$. Et generelt andreordens polynom er gitt ved

$$p(x, y) = ax^2 + bxy + cy^2 + dx + ey + f$$

Nivåkurvene til andreordens polynomer er de berømte kjeglesnittene:
`https://en.wikipedia.org/wiki/Conic_section`
Parabelen og sirkelen kjenner du fra før. For en generasjon siden måtte sivingstudenter kunne alle disse på fingrene, men idag fokuserer vi mer på andre ting. Kjeglesnittene kan være nyttige å kjenne til en gang i blant.

**Eksempel 10.3.** Finn og skisser nivåkurvene til flaten gitt ved $z = 4x^2 + 8x + 5y^2 - 10y + 9$. △

## Derivasjon

Vi har nå to uavhengige variable, og det er interessant å vite hvordan $f$ endres med hensyn på begge. Derfor finnes det to deriverte. De skrives

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

$$\frac{\partial f}{\partial y} = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h}$$

eller

$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1 + h, x_2) - f(x_1, x_2)}{h}$$

$$\frac{\partial f}{\partial x_2} = \lim_{h \to 0} \frac{f(x_1, x_2 + h) - f(x_1, x_2)}{h}$$

og uttales henholdsvis "$f$ derivert med hensyn på $x$ og $y$ (eller $x_1$ og $x_2$)".

Å partiellderivere er enkelt: Man finner $\frac{\partial f}{\partial x}$ ved å betrakte $y$ som en konstant, og så derivere i vei med hensyn på $x$. Samme for $y$.

**Eksempel 10.4.** 5 Finn de partiellderiverte til $f(x, y) = x^2 + xy + y^2 + x + y$. △

Det er vanlig å sette opp der partiellderiverte i en vektor, kalt gradientvektoren:

$$\nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right]$$

Vi kan fortsette å derivere. Det finnes fire andrederiverte, hvorav to er rene:

$$\frac{\partial f^2}{\partial x^2} \quad \text{og} \quad \frac{\partial f^2}{\partial y^2},$$

og to er blandede:

$$\frac{\partial f^2}{\partial x \partial y} \quad \text{og} \quad \frac{\partial f^2}{\partial y \partial x}.$$

De to siste er identiske dersom de er kontinuerlige, men det finnes patologiske eksempler der de ikke er like. Dette skal vi se på senere, nå skal vi bare få til å derivere.

**Eksempel 10.5.** Finn gradientvektoren til $f(x, y) = x^2 + xy + y^2 + x + y$. △

**Eksempel 10.6.** 7 Finn gradientvektoren til $g(x, y) = \sin x \sin y$. △

Tangentplanet til $f$ i punktet $(x_0, y_0)$ er gitt ved

$$z = (x - x_0) + b(y - y_0) + f(x_0, y_0)$$

der

$$a = \frac{\partial f}{\partial x}(x_0, y_0) \quad \text{og} \quad b = \frac{\partial f}{\partial y}(x_0, y_0)$$

**Eksempel 10.7.** 8 Finn tangentplanet til $f(x, y) = x^2 + xy + y^2 + x + y$ i $(1, 2)$. △

**Eksempel 10.8.** 9 Finn tangentplanet til $g(x, y) = \sin x \sin y$ i $(\pi/2, \pi/2)$. △

De partiellderiverte forteller noe om stigningen til funksjonen. Tenk at $f$ er et fjell, og at du går på ski. Dersom du peker skiene i enhetsretningen $\mathbf{v}$, er stigningen i denne retningen gitt ved

$$\mathbf{v} \cdot \nabla f$$

der prikken $\cdot$ betegner skalarproduktet du lærte på videregående.

**Eksempel 10.9.** 10 Finn stigningen på fjellsiden $f(x, y) = x^2 + xy + y^2 + x + y$ når du står i punktet $(1, 2)$ og skiene peker rett nordvest. △

**Eksempel 10.10.** 11 Hvilken vei må du peke skiene dersom du vil gå langs med en ekvidistanselinje på fjellet $g(x, y) = \sin x \sin y$, og står i punktet $(\pi/4, \pi/4)$? △

**Eksempel 10.11.** 12 Hva om du vil kjøre rett utfor så bratt som mulig?

△

# Dobbeltintegral

Tenk på $f$ som taket i et hus eller ladningstettheten på en plate. Integralet

$$\iint_D f(\mathbf{x})\, d\mathbf{x}$$

kan du tenke på som volumet av huset eller den totale ladningen på platen. Grunnflaten i huset er gitt ved $D \in \mathbb{R}^2$.

Dersom $D$ er et rektangel blir dette ikke noe vanskeligere enn i envariabel kalkulus. La oss finne volumet under flaten gitt ved $x_3 = f(x_1, x_2) = x_1 x_2$ på kvadratet avgrenset av $0 \leq x_1 \leq 1$ og $0 \leq x_2 \leq 1$. Vi uttrykker våre følelser for dette kvadratet ved å skrive

$$D = [0, 1] \times [0, 1]$$

som kalles et kartesisk produkt. Integralet blir (vi bytter til $x$ og $y$ for lesbarhetens skyld):

$$\iint_D f(\mathbf{x})\, d\mathbf{x} = \int_0^1 \int_0^1 xy\, dxdy = \int_0^1 \left( \int_0^1 xy\, dx \right) dy$$

Dette kalles et iterert integral, for det er et integral inni et annet integral. Først tar vi det innerste integralet. Det er med hensyn på $x$, fordi $dx$ står innerst. Vi går fremad med samme taktikk som ved partiellderivasjon; man integrerer med hensyn på $x$, og later som om $y$ er en konstant:

$$\begin{aligned} \int_0^1 \left( \int_0^1 xy\, dx \right) dy &= \int_0^1 \left[ \frac{1}{2} x^2 y \right]_{x=1}^{x=1} dy \\ &= \int_0^1 \left( \frac{1}{2} \cdot 1^2 \cdot y - \frac{1}{2} \cdot 0^2 \cdot y \right) dy \\ &= \frac{1}{2} \int_0^1 y\, dy \end{aligned}$$

Funksjonen $g(y) = \frac{1}{2} y$ gir arealet av tverrsnittet av volumet vi beregner. (Tenk at du for hver $y$-verdi har saget volumet i to med en motorsag, parallelt med $x$-aksen). Hvis du skjønte riemannsumtankegangen i forrige semester, skjønte du forhåpentligvis at dersom man har en funksjon som beskriver arealet av tverrsnittet av et volum, finner man volumet ved å integrere tverrsnittsfunksjonen:

$$\frac{1}{2} \int_0^1 y\, dy = \left[ \frac{1}{4} y^2 \right]_0^1 = \frac{1}{4}$$

Det som kan gjøre dobbeltintegraler litt knotete, er at integrasjonsområdet $D$ kan være noe mer komplisert enn et rektangel. La oss prøve trekanten avgrenset av $x = 0$, $y = 0$, og linjen $x = 1 - y$. Men nå blir det litt mer komplisert, for vi får et funksjonsuttrykk i en integrasjonsgrense i det innerste integralet:

$$\iint_D f(x, y)\, dA = \int_0^1 \left( \int_0^{1-y} xy\, dx \right) dy$$

Akkurat som isted, angir det innerste integralet arealet av tverrsnittet til volumet vi beregner, men nå skal vi bare integrere ut til linjen $x = 1 - y$, ikke helt ut til $x = 1$. Vi integrerer:

$$\int_0^1 \left( \int_0^{1-y} xy \, dx \right) dy = \int_0^1 \left( \frac{1}{2}x^2 y \right)_{x=0}^{x=1-y} dy$$
$$= \frac{1}{2} \int_0^1 (1-y)^2 y \, dy$$

Vi integrerer nå tverrsnittsfunksjonen:

$$\frac{1}{2} \int_0^1 (1-y)^2 y \, dy = \frac{1}{2} \int_0^1 y - 2y^2 + y^3 \, dy$$
$$= \frac{1}{2} \left( \frac{1}{2}y^2 - \frac{2}{3}y^3 + \frac{1}{4}y^4 \right)_{y=0}^{y=1} = \frac{1}{24}$$

# Linjeintegraler

En av de tingene som gjør funksjoner av flere variable (inn og ut) komplisert, er at det blir et forvirrende antall forskjellige integraltyper å holde styr på. Nå skal vi se på en annen måte å integrere funksjoner fra $\mathbb{R}^2$ til $\mathbb{R}$. La oss nå si at en brugde svømmer langs med en trajektorie gitt ved $\mathbf{z}$, og at planktontettheten i vannet langs kurven er gitt ved $f(x, y)$. Hvis du tenker riemannsummer, er det kanskje ikke så vanskelig å se at det brugden spiser på en infinitesimal del av kurven er gitt ved

$$f(x(t), y(t)) \sqrt{\dot{x}^2(t) + \dot{y}^2(t)} dt,$$

og derfor er brugdens totale måltid gitt ved

$$\int_\Gamma f ds = \int_a^b f(x(t), y(t)) \sqrt{\dot{x}^2(t) + \dot{y}^2(t)} \, dt.$$

Dette kalles linjeintegralet til $f$ over $\Gamma$. Du kan også tenke på linjeintegralet som arealet til en vegg i et hus. Veggen følger kurven $\Gamma$ sett ovenfra, og husets tak er gitt av $f$.

By Greg Skomal / NOAA Fisheries Service

# Newtons metode

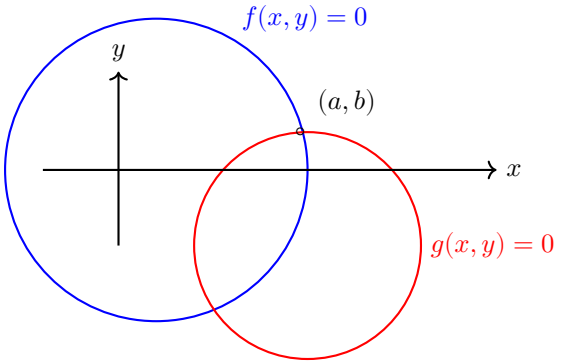
Anta at vi har to funksjoner $f(x, y)$ og $g(x, y)$. Vi leter etter et punkt $(a, b)$ slik at både

$$f(a, b) = 0 \quad \text{og} \quad g(a, b) = 0.$$

likningene

$$f(x, y) = 0 \quad \text{og} \quad g(x, y) = 0$$

angir nivåkurver for funksjonene $f$ og $g$. Punktet $(a, b)$ må ligge på skjæringspunktet mellom disse.




Anta at du har en iterasjon $(x_n, y_n)$. Vi setter opp tangentplanene til $f$ og $g$ i $(x_n, y_n)$

$$z - f(x_n, y_n) = f_x(x_n, y_n)(x - x_n) + f_y(x_n, y_n)(y - y_n)$$

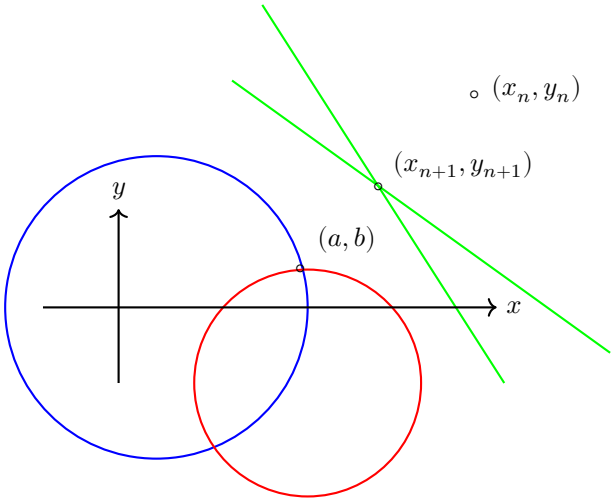
$$z - g(x_n, y_n) = g_x(x_n, y_n)(x - x_n) + g_y(x_n, y_n)(y - y_n).$$

Hvis vi krever at $z = 0$, får vi likninger for skjæringslinjene mellom disse tangentplanene og $(x, y)$-planet

$$-f(x_n, y_n) = f_x(x_n, y_n)(x - x_n) + f_y(x_n, y_n)(y - y_n)$$

$$-g(x_n, y_n) = g_x(x_n, y_n)(x - x_n) + g_y(x_n, y_n)(y - y_n).$$

Iterasjonen $(x_{n+1}, y_{n+1})$ defineres som skjæringen mellom disse linjene.




Altså er

$$-f(x_n, y_n) = f_x(x_n, y_n)(x_{n+1} - x_n) + f_y(x_n, y_n)(y_{n+1} - y_n)$$

$$-g(x_n, y_n) = g_x(x_n, y_n)(x_{n+1} - x_n) + g_y(x_n, y_n)(y_{n+1} - y_n),$$

et lineært likningssystem som definerer $(x_{n+1}, y_{n+1})$. Matrisen er

$$\begin{pmatrix} f_x(x_n, y_n) & f_y(x_n, y_n) \\ g_x(x_n, y_n) & g_y(x_n, y_n) \end{pmatrix}$$

og vi skriver

$$-\begin{pmatrix} f(x_n,y_n) \\ g(x_n,y_n) \end{pmatrix} = \begin{pmatrix} f_x(x_n,y_n) & f_y(x_n,y_n) \\ g_x(x_n,y_n) & g_y(x_n,y_n) \end{pmatrix} \begin{pmatrix} x_{n+1}-x_n \\ y_{n+1}-y_n \end{pmatrix}.$$

Nå kan vi gange med

$$\begin{pmatrix} f_x(x_n,y_n) & f_y(x_n,y_n) \\ g_x(x_n,y_n) & g_y(x_n,y_n) \end{pmatrix}^{-1}$$

fra venstre, legge til $(x_n, y_n)$ på begge sider, og få Newtons metode for systemer

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} - \begin{pmatrix} f_x(x_n,y_n) & f_y(x_n,y_n) \\ g_x(x_n,y_n) & g_y(x_n,y_n) \end{pmatrix}^{-1} \begin{pmatrix} f(x_n,y_n) \\ g(x_n,y_n) \end{pmatrix}.$$

Jeg gadd ikke skrive ut hva

$$\begin{pmatrix} f_x(x_n,y_n) & f_y(x_n,y_n) \\ \\ g_x(x_n,y_n) & g_y(x_n,y_n) \end{pmatrix}^{-1}$$

er, men du kan regne det ut ved å huske fra lineæralgebraen at

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

**Eksempel 10.12.** Vi leter etter løsninger til systemet

$$\begin{aligned} x^2 + y^2 &= 1 \\ \frac{x^2}{4} + 2y^2 &= 1 \end{aligned}$$

Som du husker fra M2 er den første likningen for enhetssirkelen, mens den andre er likningen for en ellipse med halvakser 2 og $\frac{1}{2}$. Trekker man to ganger den første fra den andre, kan man regne ut at skjæringspunktet mellom disse kurvene i første kvadrant er

$$\left(\frac{2}{\sqrt{7}}, \sqrt{\frac{3}{7}}\right) \approx (0.755928946018454, 0.654653670707977).$$

La oss se om Newtons metode finner dette punktet. Siden

$$f(x,y) = x^2 + y^2 - 1$$

og

$$g(x,y) = \frac{x^2}{4} + 2y^2 - 1$$

blir

$$\begin{pmatrix} f_x(x_n,y_n) & f_y(x_n,y_n) \\ g_x(x_n,y_n) & g_y(x_n,y_n) \end{pmatrix} = \begin{pmatrix} 2x & 2y \\ \frac{x}{2} & 4y \end{pmatrix}$$

som har invers

$$\frac{1}{7xy}\begin{pmatrix} 4y & -2y \\ -\frac{x}{2} & 2x \end{pmatrix}.$$

Metoden blir

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} - \frac{1}{7xy}\begin{pmatrix} 4y & -2y \\ -\frac{x}{2} & 2x \end{pmatrix}\begin{pmatrix} f(x_n,y_n) \\ g(x_n,y_n) \end{pmatrix},$$

og starter vi i $(1,1)$, får vi:

| $n$ | $x_n$ | $y_n$ |
|---|---|---|
| 1 | 1.000000000000000 | 1.000000000000000 |
| 2 | 0.785714285714286 | 0.714285714285714 |
| 3 | 0.756493506493507 | 0.657142857142857 |
| 4 | 0.755929156680230 | 0.654658385093168 |
| 5 | 0.755928946018484 | 0.654653670724952 |
| 6 | 0.755928946018454 | 0.654653670707977 |
| 7 | 0.755928946018455 | 0.654653670707977 |

Maskinpresisjon etter syv iterasjoner. Bra greier det her. △

# Kapittel 11

# Funksjoner fra $\mathbb{R}^m$ til $\mathbb{R}^n$

Nå skal vi studere funksjoner fra $\mathbb{R}^m$ til $\mathbb{R}^n$. En slik funksjon kalles et vektorfelt, og er en kolennevektor i med $n$ komponenter der komponentene er funksjonsuttrykk i $m$ variable:

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_n(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} F_1(x_1, 2_x, \cdots, x_m) \\ F_2(x_1, 2_x, \cdots, x_m) \\ \vdots \\ F_n(x_1, 2_x, \cdots, x_m) \end{pmatrix}$$

I forrige kapittel skrev jeg gradientvektoren $\nabla f$ som radvektor. Dette var ikke tilfeldig. Når man skal lære funksjoner fra $\mathbb{R}^m$ til $\mathbb{R}^n$ er det lurt å være litt systematisk med rader og kolonner, for da kan man utnytte at man kan matrisemultiplikasjon.

**Eksempel 11.1.** Det eksemplet du er godt kjent med fra før, er funksjonen

$$\mathbf{F}(\mathbf{x}) = A\mathbf{x}$$

der $A$ er en $m \times n$-matrise og $\mathbf{x}$ er en en vektor i $\mathbb{R}^m$. △

Det er antagelig lurt å tenke på $\mathbf{x} \in \mathbb{R}^m$ som en søylevektor, samt skrive $\mathbf{F}$ som en kolonnevektor, for det er dette vi er vant med fra lineæralgebraen, og $A\mathbf{x}$ er en av de mest elementære funksjonene vi har fra $\mathbb{R}^m$ til $\mathbb{R}^n$.

**Eksempel 11.2.** Hvis vi regner ut de partiellderiverte til $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$, ser vi at de $m$ partiellderiverte til hver komponent er det samme som radene i $A$. Det er derfor ikke helt tilfeldig at vi skriver gradientvektoren som en radvektor når $\mathbf{F}$ skrives som kolonnevektor. △

De partiellderiverte til $\mathbf{F}$ organiseres i en *jacobimatrisen* til $\mathbf{F}$:

$$D_{\mathbf{x}}\mathbf{F} = \begin{pmatrix} \nabla F_1 \\ \nabla F_2 \\ \vdots \\ \nabla F_n \end{pmatrix} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_m} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_n}{\partial x_1} & \frac{\partial F_n}{\partial x_2} & \cdots & \frac{\partial F_n}{\partial x_m} \end{pmatrix}$$

Til syvende og sist har nok denne konvensjonen å gjøre med at vi skriver fra venstre mot høyre. Vi skriver jo likningsssystemer slik:

$$\begin{aligned} 2x_1 + 3x_2 + 4x_3 &= 4 \\ 3x_1 + 4x_2 + 5x_3 &= 5 \\ 4x_1 + 5x_2 + 6x_3 &= 6 \end{aligned}$$

og derfor skriver vi matriser på en slik måte at radene er koeffisientene i hver likning, mens høyresiden er en kolonnevektor. Det er kanskje ikke fullt så innlysende hvorfor vi også skriver variabelen $\mathbf{x}$ som en kolonnevektor.

Uansett er det nå praktisk å gjøre det slik vi gjør, for kjerneregelen blir lett å skrive opp:

**Kjerneregelen i flere variable**

La $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^p$ og $\mathbf{G} : \mathbb{R}^m \to \mathbb{R}^n$, og la $H : \mathbb{R}^m \to \mathbb{R}^p$ være gitt ved $\mathbf{H} = \mathbf{F}(\mathbf{G})$. Da er $D\mathbf{H}$ gitt ved matriseproduktet mellom $D\mathbf{F}$ og $D\mathbf{G}$:

$$D\mathbf{H} = D\mathbf{F}D\mathbf{G}$$

# Kapittel 12

# Systemer av differensiallikninger

I dette kapitlet skal vi bruke det vi har lært om lineær algebra til å studere systemer av differensiallikninger.

## Fundamentalsystemer

I dette kapiitlet må vi være litt nøye på forskjellen mellom vektorfunksjonen $\mathbf{x}$, som er et element i et uendeligdimensjonalt vektorrom av vektorfunksjoner, og funksjonsverdien $\mathbf{x}(t)$, som er en søylevektor i $\mathbb{R}^n$ eller $\mathbb{C}^n$.

**Eksempel 12.1.** La

$$\mathbf{x}(t) = \begin{pmatrix} t \\ t \end{pmatrix} \quad \text{og} \quad \mathbf{y}(t) = \begin{pmatrix} t \\ e^{t-1} \end{pmatrix}.$$

Merk at

$$\mathbf{x}(1) = \mathbf{y}(1) = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

og at

$$\mathbf{x}(0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

For alle andre verdier av $t$ er $\mathbf{x}(t)$ og $\mathbf{y}(t)$ ikke parallelle. Med andre ord: $\mathbf{x}(t)$ og $\mathbf{y}(t)$ er lineært avhengige for $t = 0$ og $t = 1$, men ikke for andre verdier av $t$. Som vektorfunksjoner er $\mathbf{x}$ og $\mathbf{y}$ lineært uavhengige, siden

$$c_1\mathbf{x} + c_2\mathbf{y} = \mathbf{0}$$

impliserer $c_1 = c_2 = 0$. △

**Eksempel 12.2.** La

$$\mathbf{x}(t) = t\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad \mathbf{y}(t) = t\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \quad \text{og} \quad \mathbf{z}(t) = t\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

I et tidligere kapittel fant vi at

$$\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} - 2\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} + \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

og i dette eksemplet er det lett å se at

$$\mathbf{x} - 2\mathbf{y} + \mathbf{z} = \mathbf{0}.$$

Dette betyr at $\mathbf{x}$, $\mathbf{y}$ og $\mathbf{z}$ er lineært avhengige som vektorfunksjoner. △

**Eksempel 12.3.** La nå

$$\mathbf{x}(t) = t\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad \mathbf{y}(t) = t\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \quad \text{og} \quad \mathbf{z}(t) = e^t\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

De tre vektorene $\mathbf{x}(t)$, $\mathbf{y}(t)$ og $\mathbf{z}(t)$ er lineært avhengige for hver $t$, for vi kan alltid finne $c_1$, $c_2$ og $c_3$, ikke alle lik null, slik at

$$c_1\mathbf{x}(t) + c_2\mathbf{y}(t) + c_3\mathbf{z}(t) = \mathbf{0}.$$

Men $c_1$, $c_2$ og $c_3$ må endres for hver $t$, for det er ikke mulig å finne ett enkelt valg slik at likningen

$$c_1\mathbf{x} + c_2\mathbf{y} + c_3\mathbf{z} = \mathbf{0}$$

alltid er sann. Derfor er $\mathbf{x}$, $\mathbf{y}$ og $\mathbf{z}$ lineært uavhengige som vektorfunksjoner. △

La oss anta at vi har en mengde

$$\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n$$

med vektorfunksjoner. Det er vanlig å sette funksjonene opp som kolonner i en $n \times n$-matrise $\mathbf{Y}$, kalt *fundamentalmatrisen*. Dersom $\det \mathbf{Y}(t) \neq 0$ for alle $t$, er funksjonsverdiene til mengden en basis for $\mathbb{R}^n$ for hver $t$, og vi kaller mengden et *fundamentalsystem*.

**Eksempel 12.4.** Vi kan sette sammen

$$\mathbf{x}(t) = e^t\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad \mathbf{y}(t) = e^t\begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \quad \text{og} \quad \mathbf{z}(t) = e^t\begin{pmatrix} 4 \\ 5 \\ 7 \end{pmatrix}$$

til

$$\mathbf{Y}(t) = e^t\begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 7 \end{pmatrix},$$

og siden $\det \mathbf{Y}(t) = -e^t \neq 0$ for alle $t$, utgjør $\mathbf{x}$, $\mathbf{y}$ og $\mathbf{z}$ et fundamentalsystem. △

**Eksempel 12.5.** La

$$\mathbf{x}(t) = t\begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \text{og} \quad \mathbf{y}(t) = e^t\begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

Funksjonene $\mathbf{x}$ og $\mathbf{y}$ er lineært uavhengige siden det ikke finnes konstanter $c_1$ og $c_2$ slik at

$$c_1\mathbf{x} + c_2\mathbf{y} = 0.$$

Men funksjonsverdiene $\mathbf{x}(t)$ og $\mathbf{y}(t)$ er parallelle vektorer for alle $t$, og derfor utgjør disse funksjonene ikke et fundamentalsystem. △

# Systemer av differensiallikninger

I dette kapitlet skal

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{m2} & \cdots & a_{nn} \end{pmatrix}$$

alltid være en reell matrise. Det blir mer enn komplisert nok. Et *førsteordens lineært og homogent system av differensiallikninger med konstante koeffisienter* er et sett med $n$ likninger og $n$ ukjente

$$\begin{pmatrix} \dot{y}_1(t) \\ \dot{y}_2(t) \\ \vdots \\ \dot{y}_n(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{m2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$

På kortform skriver vi enkelt og greit

$$\dot{\mathbf{y}} = A\mathbf{y}.$$

og forkorter den lange tittelen til *system*.

Før man i det hele tatt begynner å løse systemet over, kan man få en ide om hvordan løsningskurvene kommer til å oppføre seg, ved å skissere systemets *vektorfelt*. Dette får man ved å evaluere høyresiden i likningssystemet for forskjellige verdier av $\mathbf{y}$, slik at man får ut stigningstallet til en eventuell løsningskurve i dette punktet, og så tegne disse i et koordinatsystem.

En konstant funksjon som løser systemet $\dot{\mathbf{y}} = A\mathbf{y}$, kalles en *likevektsløsning*. Dette er en løsning som står stille i rommet. Merk at dette er alle vektorer i nullrommet til $A$. Denne klassen av løsninger sitter inni en mer interessant klasse av løsninger, som er beskrevet i neste teorem.

**Det er enkelt å finne løsninger**

La $\mathbf{x}$ være en vektor i $\mathbb{R}^n$, og la

$$\mathbf{y}(t) = \mathbf{v}e^{\lambda t}.$$

Funksjonen $\mathbf{y}$ er en løsning av systemet

$$\dot{\mathbf{y}} = A\mathbf{y}$$

hvis og bare hvis $\lambda$ er en egenverdi, og $\mathbf{v}$ den korresponderende egenvektor, til matrisen A.

*Bevis.* Siden (husk at $e^{\lambda t}$ er en skalar)

$$A\mathbf{y} = A\mathbf{v}e^{\lambda t} = \lambda \mathbf{v}e^{\lambda t}$$

og

$$\dot{\mathbf{y}} = \lambda \mathbf{v}e^{\lambda t}$$

er $\mathbf{y}$ en løsning av systemet. Omvendt kan vi se at dersom $\mathbf{v}e^{\lambda t}$ skal være en løsning av systemet, må

$$A\mathbf{v}e^{\lambda t} = \frac{d}{dt}\left(\mathbf{v}e^{\lambda t}\right) = \lambda \mathbf{v}e^{\lambda t},$$

og hvis vi deler ut $e^{\lambda t} \neq 0$ får vi

$$A\mathbf{v} = \lambda \mathbf{v},$$

som sier at $\mathbf{v}$ er en egenvektor med egenverdi $\lambda$. □

**Eksempel 12.6.** Vi løser systemet

$$\dot{\mathbf{y}} = A\mathbf{y}$$

der

$$A = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}.$$

Egenvektorer er som kjent

$$c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{og} \quad c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

med egenverdier $-1$ og 3, henholdsvis. Derfor er to løsninger av likningssystemet

$$\mathbf{y}_1 = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-t} \quad \text{og} \quad \mathbf{y}_2 = c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{3t}. \qquad \triangle$$

**Eksempel 12.7.** Matrisen

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 6 \end{pmatrix}$$

har egenvektorer

$$\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} -4 \\ 1 \\ 1 \end{pmatrix}.$$

med egenverdier henholdsvis 4, 9 og 0. Løsningene av $\dot{\mathbf{y}} = A\mathbf{y}$ er

$$c_1 \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} e^{4t}, \quad c_2 \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} e^{9t} \quad \text{og} \quad c_3 \begin{pmatrix} -4 \\ 1 \\ 1 \end{pmatrix}.$$

Merk at den siste er en likevektsløsning. △

**Et vanskelig teorem eller ikke**

For systemet

$$\dot{\mathbf{y}} = A\mathbf{y}$$

kan vi alltid finne $n$ kontinuerlig deriverbare løsninger som utgjør et fundamentalsystem.

Det generelle beviset er litt for hardt for oss, men det er lett å se at det må være sant dersom $A$ er diagonaliserbar. Finnes det $n$ lineært uavhengige egenvektorer, finnes det også $n$ løsninger på formen

$$\mathbf{y}(t) = e^{\lambda t}\mathbf{v},$$

og siden eksponensialfunksjonen aldri er null, er det lett å se at disse løsningene utgjør et fundamentalsystem.

Teoremet er imidlertid sant også for ikkediagonaliserbare matriser, men dette er mer jobb å vise, se Xavier Raynauds notat. Trikset er å sette opp et fundamentalsystem av løsninger på formen

$$\mathbf{y}(t) = e^{\lambda t}\mathbf{v}(t),$$

der $\mathbf{v}(t)$ er en polynomisk vektorfunksjon basert på noe som kalles *generaliserte egenvektorer*. Dette er viktig i automatisering og regulering.

Siden både venstre- og høyresiden av systemet $\dot{\mathbf{y}} = A\mathbf{y}$ følger superposisjonsprinsippet (begge sider av likningen er lineæroperatorer) er det lett å se at mengden av løsninger danner et vektorrom. Vi skal nå se at dette rommet har nøyaktig $n$ dimensjoner.

**Det blir et vektorrom!**

Alle løsninger av

$$\dot{\mathbf{y}} = A\mathbf{y}$$

utgjør et vektorrom av dimensjon $n$.

*Bevis.* Vi lar $\mathbf{Y}$ være et fundamentalsystem av løsninger. Vi at dette finnes. (Dersom $A$ er diagonaliserbar, er det lett å finne et fundamentalsystem. Dersom $A$ ikke er diagonaliserbar, er det vanskeligere, men alltid mulig.) Husk at kolonnene i $\mathbf{Y}(t)$ alltid er lineært uavhengige, slik at $\mathbf{Y}(t)$ kan inverteres for alle $t$.

La $\mathbf{z}$ være en løsning. Siden kolonnene i $\mathbf{Y}(t)$ utgjør en basis for $\mathbb{R}^n$ for alle $t$, kan vi for hver verdi av $t$ skrive

$$\mathbf{z}(t) = \mathbf{Y}(t)\mathbf{c}(t).$$

og siden $\mathbf{Y}(t)$ er inverterbar for alle $t$, kan vi skrive

$$\mathbf{Y}^{-1}(t)\mathbf{z}(t) = \mathbf{c}(t),$$

som viser at også $\mathbf{c}$ er kontinuerlig deriverbar. Vi kan derfor trygt derivere

$$\dot{\mathbf{z}}(t) = \dot{\mathbf{Y}}(t)\mathbf{c}(t) + \mathbf{Y}(t)\dot{\mathbf{c}}(t),$$

og siden både

$$\dot{\mathbf{z}}(t) = A\mathbf{z}(t)$$

og

$$\dot{\mathbf{Y}}(t)\mathbf{c}(t) = A\mathbf{Y}(t)\mathbf{c}(t) = A\mathbf{z}(t),$$

kansellerer disse mot hverandre, og vi står igjen med

$$0 = \mathbf{Y}(t)\dot{\mathbf{c}}(t).$$

Siden kolonnene i $\mathbf{Y}(t)$ er lineært uavhengige, impliserer denne likningen at $\dot{\mathbf{c}}(t) = \mathbf{0}$, og følgelig er $\mathbf{c}(t)$ en konstant vektor. Men dette betyr at

$$\mathbf{z}(t) = \mathbf{Y}(t)\mathbf{c},$$

og denne likningen sier at funksjonen $\mathbf{z}$ er en (konstant) lineærkombinasjon av løsningene i fundamentalsystemet. Disse løsningene utgjør dermed en basis for løsningsrommet, og følgelig er løsningsrommet $n$-dimensjonalt. □

Fra nå av skal vi studere systemer der matrisen $A$ er diagonaliserbar, slik at løsningen kan skrives

$$\mathbf{y}(t) = c_1\mathbf{v}_1 e^{\lambda_1 t} + c_2\mathbf{v}_2 e^{\lambda_2 t} + \ldots + c_n\mathbf{v}_n e^{\lambda_n t}$$

der $\mathbf{v}_k$ er egenvektorene til $A$, med korresponderende egenverdier $\lambda_k$. Dette kalles *den generelle løsningen.*

**Eksempel 12.8.** Den generelle løsningen til systemet

$$\dot{\mathbf{y}} = A\mathbf{y}$$

der

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

er

$$\mathbf{y} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-t}.$$

△

**Eksempel 12.9.** Den generelle løsningen til systemet med matrise

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 6 \end{pmatrix}$$

er

$$\mathbf{y}(t) = c_1 \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} e^{4t} + c_2 \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} e^{9t} + c_3 \begin{pmatrix} -4 \\ 1 \\ 1 \end{pmatrix}. \quad \triangle$$

I noen tilfeller er det naturlig å spesifisere et punkt $\mathbb{R}^n$ der løsningskurven skal starte.

**Initialverdiproblem**

Et *initialverdiproblem* er et likningssystem

$$\dot{\mathbf{y}} = A\mathbf{y}$$

med initialbetingelse

$$\mathbf{y}(t_0) = \mathbf{y}_0,$$

der $\mathbf{y}_0 \in \mathbb{R}^n$. En kontinuerlig deriverbar løsning som tilfredsstiller dette kravet, kalles en *spesiell løsning.*

Det er ikke så vanskelig å se at et initialverdiproblem har entydig løsning. Løsningen

$$\mathbf{y}(t) = \mathbf{Y}(t)\mathbf{Y}^{-1}(t_0)\mathbf{y}_0$$

tilfredsstiller helt klart initialbetingelsen $\mathbf{y}(t_0) = \mathbf{y}_0$. Anta at det finnes en annen løsning $\mathbf{z}$. Vi vet at $\mathbf{z}$ tilfredsstiller

$$\mathbf{z}(t) = \mathbf{Y}(t)\mathbf{c},$$

der $\mathbf{c}$ er en konstant. Men siden $\mathbf{z}(t_0) = \mathbf{y}_0$, må $\mathbf{c} = \mathbf{Y}^{-1}(t_0)\mathbf{y}_0$, slik at

$$\mathbf{z}(t) = \mathbf{Y}(t)\mathbf{Y}^{-1}(t_0)\mathbf{y}_0 = \mathbf{y}(t).$$

Dette er såpass viktig at vi skriver det opp som et teorem.

**Det finnes bare en løsning og bra er det**

Et initalverdiproblem

$$\dot{\mathbf{y}} = A\mathbf{y} \qquad \mathbf{y}(t_0) = \mathbf{y}_0,$$

har entydig løsning.

**Eksempel 12.10.** Den spesielle løsningen

$$\mathbf{y}(t) = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} e^{4t} + \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} e^{9t} + \begin{pmatrix} -4 \\ 1 \\ 1 \end{pmatrix}$$

til systemet i forrige eksempel, starter i punktet

$$\begin{pmatrix} -3 \\ 2 \\ 4 \end{pmatrix}$$

ved $t = 0$. △

# Noen løsninger i planet

Løsninger av diagonaliserbare $2 \times 2$-systemer kan klassifiseres ganske greit. Vi skal også ta med et tilfelle der $A$ ikke er diagonaliserbar, for å gi en smakebit på den generelle teorien. Det er gunstig å dele inn i forskjellige tilfeller basert på egenverdiene til $A$, se på hva som skjer når $t \to \infty$, og plotte noen løsninger i et *fasediagram*.

## Reelle og distinkte egenverdier

Løsningen er

$$\mathbf{y}(t) = c_1 \mathbf{x}_1 e^{\lambda_1 t} + c_2 \mathbf{x}_2 e^{\lambda_2 t},$$

der både $c_1$, $c_2$, $\mathbf{x}_1$, $\mathbf{x}_2$ og $\lambda_1 \neq \lambda_2$ er relle. Vi illustrerer hva som kan skje med fire eksempler.

**Eksempel 12.11.** La

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

slik at

$$\mathbf{y} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{t}.$$

Merk at uansett hvilke kombinasjoner av $c_1$ og $c_2$ vi har (så lenge ikke begge er 0), vil alle løsninger reise mot uendelig når $t \to \infty$, altså vekk fra den eneste likevektsløsningen $\mathbf{y} = 0$. Vi sier derfor at $\mathbf{y}$ er en *ustabil likevektsløsning*. Nedenfor er plot av løsningskurver for et par tusen tilfeldige verdier av $c_1$ og $c_2$. △

**Eksempel 12.12.** La

$$A = -\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

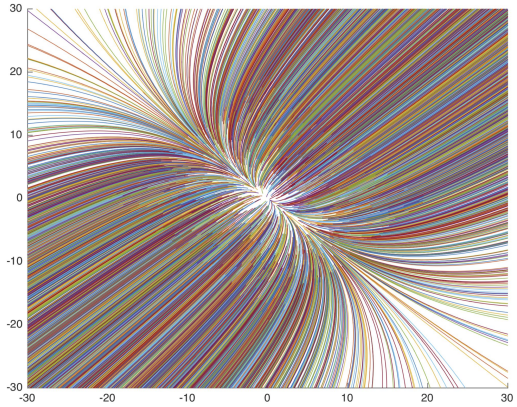
slik at

$$\mathbf{y} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-t}.$$

Uansett hvilke kombinasjoner av $c_1$ og $c_2$ vi har, vil alle løsninger søke mot origo når $t \to \infty$, slik at $\mathbf{y}$ er en såkalt *stabil likevektsløsning*. △

**Eksempel 12.13.** La

$$A = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$$

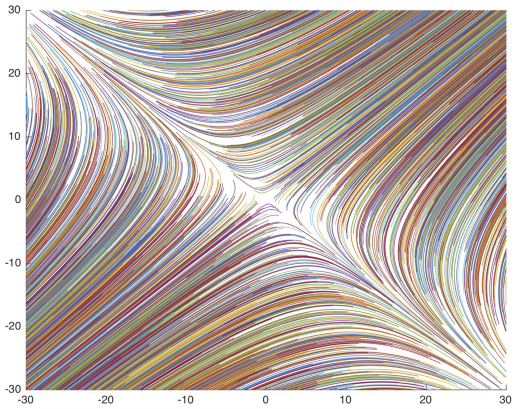



Eksempel 12.11

slik at

$$\mathbf{y} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{-t}.$$

Dersom $c_1 \neq 0$, vil alle løsninger gå mot uendelig når $t \to \infty$, altså inn mot likevektsløsningen $\mathbf{y} = 0$. Men dersom $c_1 = 0$ og $c_2 \neq 0$, vil løsningen søke mot origo. Likevektsløsningen $\mathbf{y} = 0$ kalles en *ustabil sadel*. △




Eksempel 12.13

**Eksempel 12.14.** La

$$A = -\frac{1}{2}\begin{pmatrix} 3 & 3 \\ 3 & 3 \end{pmatrix}$$

slik at

$$\mathbf{y} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{-3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Så lenge $c_1 \neq 0$, vil alle løsninger søke mot likevektsløsningen

$$\mathbf{y} = c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

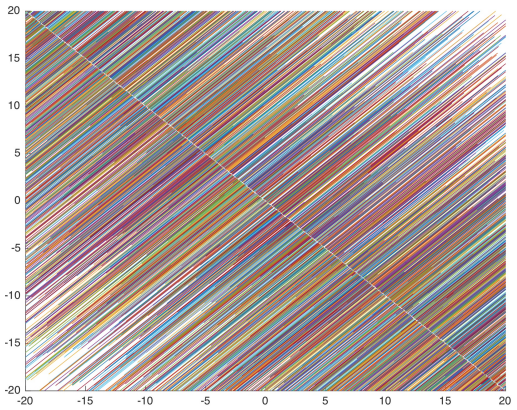
og ingen mot $\mathbf{y} = 0$ når $t \to \infty$. Denne har et litt kjedelig faseplott. △

**Eksempel 12.15.** La

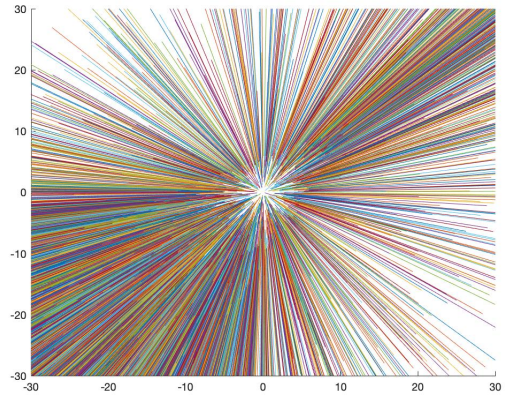
$$A = -\frac{1}{2}\begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$$

slik at

$$\mathbf{y} = e^{3t}\left(c_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right). \quad \triangle$$



Eksempel 12.14



Eksempel 12.15

## Komplekse egenverdier

Vi har i øvingsopplegget vist at dersom en reell matrise har komplekse egenverdier, opptrer disse i kompleksekonjugerte par. Du har kanskje lagt merke til at dette også gjelder for de egenvektorene, siden

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Longleftrightarrow \quad A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}.$$

Dette skal vi benytte oss av for å plukke ut relle løsninger. La $\lambda = \alpha + \beta i$ ha egenvektor

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + i \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

og husk at $e^{\alpha+\beta i} = e^{\alpha}(\cos\beta + i\sin\beta)$, slik at

$$\begin{aligned}\mathbf{y}(t) &= c_1 e^{\lambda t}\mathbf{x} + c_2 e^{\overline{\lambda} t}\overline{\mathbf{x}} \\ &= c_1 e^{\alpha t}\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + i\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right)(\cos\beta t + i\sin\beta t) \\ &\quad + c_2 e^{\alpha t}\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - i\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right)(\cos\beta t - i\sin\beta t).\end{aligned}$$

Denne løsningen er pen på papiret, men vi ønsker å kunne visualisere litt, og da hadde det vært praktisk å finne en løsning som var reell istedet.

Siden $e^{\lambda t}\mathbf{x}$ og $e^{\overline{\lambda} t}\overline{\mathbf{x}}$ er lineært uavhengige for alle $t$, utgjør de en basis for $\mathbb{C}^2$. La oss søke en reell basis istedet. Vi kaller den nye basisen $\mathbf{v}_1$ og $\mathbf{v}_2$. Velg først $c_1 = c_2 = \frac{1}{2}$, og sett

$$\mathbf{v}_1 = e^{\alpha t}\left(\cos\beta t\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - \sin\beta t\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right).$$

Så velger vi $c_1 = -c_2 = \frac{1}{2i}$, og setter

$$\mathbf{v}_2 = e^{\alpha t}\left(\cos\beta t\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \sin\beta t\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right).$$

Nå kan vi skrive

$$\begin{aligned}\mathbf{y}(t) &= d_1\mathbf{v}_1 + d_2\mathbf{v}_2 \\ &= d_1 e^{\alpha t}\left(\cos\beta t\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} - \sin\beta t\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right) \\ &\quad + d_2 e^{\alpha t}\left(\cos\beta t\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} + \sin\beta t\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right).\end{aligned}$$

Merk at siden $\mathbf{y}_1$ og $\mathbf{y}_2$ er lineært uavhengige, og forholdet mellom disse og $\mathbf{v}_1$ og $\mathbf{v}_2$ er gitt ved

$$\begin{pmatrix}\mathbf{y}_2 & \mathbf{y}_2\end{pmatrix}\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} = 2\begin{pmatrix}\mathbf{v}_2 & \mathbf{v}_2\end{pmatrix},$$

er også $\mathbf{v}_1$ og $\mathbf{v}_2$ lineært uavhengige for alle $t$. Fordelen med den nye basisen er at vi nå enkelt kan skille ut alle reelle løsninger ved å holde oss til relle $d_1$ og $d_2$.

**Eksempel 12.16.** La

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

som har egenverdier $\pm i$ og egenvektorer

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

Den generelle løsningen til systemet blir

$$\begin{aligned}\mathbf{y}(t) &= d_1\left(\cos t\begin{pmatrix} 1 \\ 0 \end{pmatrix} - \sin t\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) \\ &\quad + d_2\left(\cos t\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \sin t\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) \\ &= d_1\begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} + d_2\begin{pmatrix} \sin t \\ \cos t \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.\end{aligned}$$

Vi ser at denne løsningen starter i punktet $\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ ved $t = 0$, og kjører deretter i en pen sirkulær bane om origo. Merk at kurven er traversert med klokken. △

**Eksempel 12.17.** La

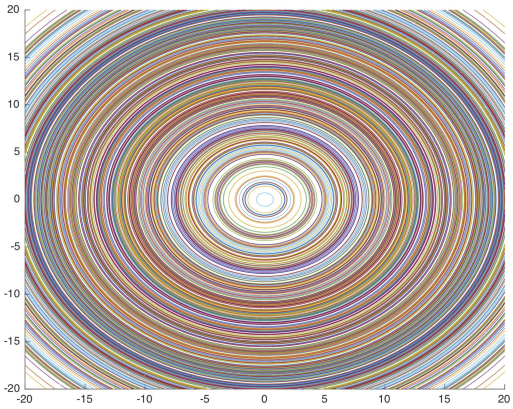
$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

som har egenverdier $1 \pm i$ og de samme egenvektorene

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

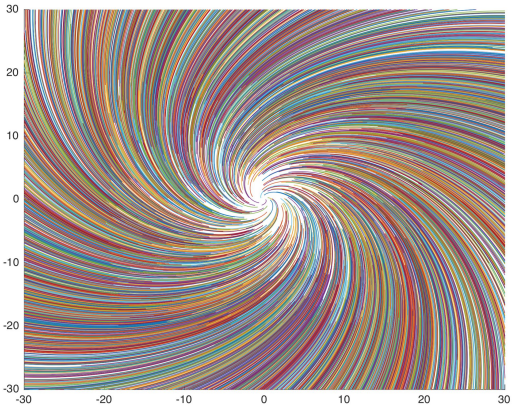
På samme vis som i forrige eksempel blir den generelle løsningen

$$\begin{aligned}\mathbf{y}(t) &= d_1 e^t\begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} + d_2 e^t\begin{pmatrix} \sin t \\ \cos t \end{pmatrix} \\ &= e^t\begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.\end{aligned}$$



Eksempel 12.16

Denne løsningen starter i punktet $\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ ved $t = 0$, og kjører deretter i en særdeles vakker sirkulær og utadgående spiral. △



Eksempel 12.17

**Eksempel 12.18.** La

$$A = \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix}$$

som har egenverdier $-1 \pm i$ og de samme egenvektorene

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

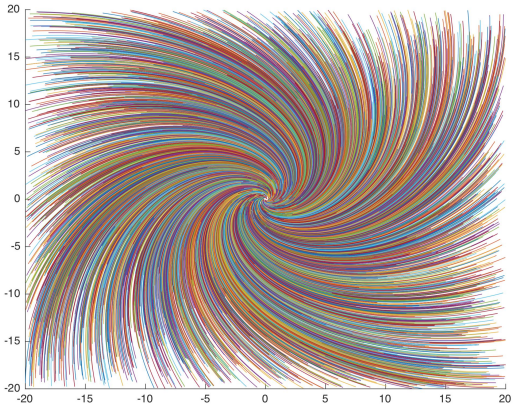
På samme vis som i de to forrige eksemplene blir den generelle løsningen

$$\begin{aligned} \mathbf{y}(t) &= d_1 e^{-t} \begin{pmatrix} \cos t \\ -\sin t \end{pmatrix} + d_2 e^{-t} \begin{pmatrix} \sin t \\ \cos t \end{pmatrix} \\ &= e^{-t} \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}. \end{aligned}$$

Denne løsningen starter i punktet $\begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ ved $t = 0$, og kjører deretter i en innadgående sirkulær spiral. △

## Defekt egenverdi - for spesielt interesserte!

Tilfellet at $A$ ikke er diagonaliserbar, kan vi egentlig ikke analysere med teorien vi har lært til nå, så du skal slippe å kunne det til eksamen. Men vi tar en smakebit på hva som skjuler seg utenfor pensum.



Eksempel 12.18

**Eksempel 12.19.** La

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix}$$

som har dobbel egenverdi 1, men bare en egenvektor $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Hva gjør vi nå? △

For å løse knipen fra forrige eksempel, må vi gjøre noe artig, nemlig introdusere *generalisert egenvektor*. Egenvektoren til $\lambda$ finner man ved å finne nullrommet til $A - \lambda I$. For en $2 \times 2$-matrise med defekt egenverdi, er en generalisert egenvektor en vektor i nullrommet til $(A - \lambda I)^2$.

**Eksempel 12.20.** La $A$ være som i forrige eksempel. Nullrommet til

$$(A - I)^2 = \left( \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \right)^2 = \left( \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right)$$

er alle vektorer i $\mathbb{C}^2$. Altså er alle vektorer i $\mathbb{C}^2$ generaliserte egenvektorer til matrisen $A$. Vi velger oss en tilfeldig vektor som ikke er parallell med $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, for eksempel $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Hvis vi ganger denne inn i $A - I$, får vi

$$\begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

som er en egenvektor. Hm. △

Hvordan bruker vi dette til å løse systemet?

**Eksempel 12.21.** Vektorene $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ og $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ er et eksempel på en *kjede av generaliserte egenvektorer*. Løsningen som korresponderer til den genaliserte egenvektorkjeden $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ og $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ er

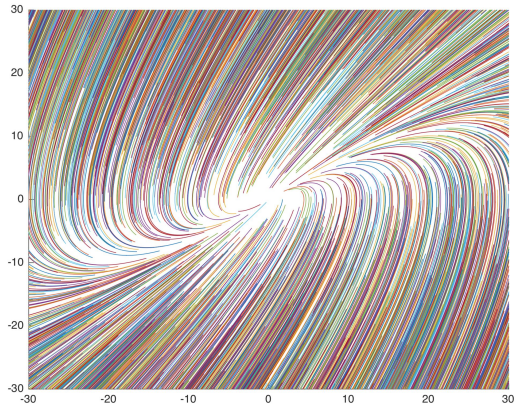
$$\mathbf{y}_2(t) = c_2 e^t \left( t \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right).$$

Løsningen som korresponderer til egenvektoren vi fant tidligere, er

$$\mathbf{y}_1(t) = c_1 e^t \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Dette er to lineært uavhengige løsninger, og den generelle løsningen til systemet er

$$\mathbf{y}(t) = c_1 e^t \begin{pmatrix} 1 \\ 1 \end{pmatrix} + c_2 e^t \left( t \begin{pmatrix} 2 \\ 2 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right). \quad \triangle$$




Eksempel 12.21

# Inhomogene systemer

Differensiallikningssystemet

$$\dot{\mathbf{y}} = A\mathbf{y}$$

har den fordel at det er lett å skrive opp alle mulige løsninger, alt vi trenger er egenverdiene og egenvektorene til $A$. Nå skal vi se litt på *førsteordens inhomogent lineært system av differensiallikninger med konstante koeffisienter*. Dette er et likningssystem på formen

$$\dot{\mathbf{y}} = A\mathbf{y} + \mathbf{f},$$

der $\mathbf{f}$ er en spesifisert vektorfunksjon. Tittelen forkorter vi til *inhomogent system*.

Først kan vi merke oss at dersom vi har en løsning $\mathbf{z}$ til det inhomogene systemet, og en løsning $\mathbf{y}$ til det homogene systemet

$$\dot{\mathbf{y}} = A\mathbf{y},$$

vil $\mathbf{z}+\mathbf{y}$ løse det inhomogene systemet. Det er derfor vanlig å splitte løsninger i

$$\mathbf{y} = \mathbf{y}_h + \mathbf{y}_p$$

der $\mathbf{y}_h$ er den generelle løsningen til det homogene systemet, og $\mathbf{y}_p$ er en løsning til det inhomogene systemet. Den første kalles *den homogene løsningen*, mens den siste kalles enten *den inhomogene løsningen* eller *partikulærløsningen*. Ofte går det an å gjette formen på partikulærløsningen.

**Eksempel 12.22.** Vi løser likningssystemet

$$\dot{\mathbf{y}} = A\mathbf{y} + \mathbf{f}$$

der

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

og

$$\mathbf{f} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Den homogene løsningen er som kjent

$$\mathbf{y}_h = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^t.$$

Men hva med den inhomogene? Siden $\mathbf{f}$ er en konstant vektor, er det ikke utenkelig at $\mathbf{y}_p$ også er det. Vi prøver. La

$$\mathbf{y}_p = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Vi setter denne inn i likningen, og får

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Dette går fint dersom

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

som gir

$$\mathbf{y}_p = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} -1/3 \\ -1/3 \end{pmatrix}$$

Løsningen er med andre ord

$$\begin{aligned} \mathbf{y} &= \mathbf{y}_h + \mathbf{y}_p \\ &= c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^t + \begin{pmatrix} -1/3 \\ -1/3 \end{pmatrix}. \end{aligned} \quad \triangle$$

Med litt erfaring kan man ofte gjette formen på $\mathbf{y}_p$, og på internettet er det mulig å finne lange tabeller med hvordan partikulærløsningen $\mathbf{y}_p$ ser ut for forskjellige $\mathbf{f}$. Dette er litt kjedelig å lære seg, men greit å vite om.

Vi skal heller utlede en formel som i prinsippet alltid kan finne $\mathbf{y}_p$. Strategien baserer seg på å ta utgangspunkt i $\mathbf{y}_h$. Vi lar $\mathbf{Y}$ være fundamentalmatrisen. Siden kolonnene i $\mathbf{Y}(t)$ alltid utgjør en basis for $\mathbb{R}^n$, er det rimelig å forvente at en kontinuerlig deriverbar partikulærløsning, dersom den finnes, kan skrives

$$\mathbf{y}_p(t) = \mathbf{Y}(t)\mathbf{c}(t).$$

der $\mathbf{c}$ er en vektorfunksjon. Merk at $\mathbf{c}$ ikke kan være en konstant funksjon, for da blir $\mathbf{y}_p$ en homogen løsning. Akkurat som i beviset for teorem **??**, ser vi at $\mathbf{c}$ må være kontinuerlig deriverbar. Vi kan derfor trygt skrive $\mathbf{y}_p t = \dot{\mathbf{y}}\mathbf{c} + \mathbf{Y}\mathbf{c}$, og sette dette inn i den inhomogene likningen:

$$\dot{\mathbf{y}}\mathbf{c} + \mathbf{Y}\mathbf{c} = A\mathbf{Y}\mathbf{c} + \mathbf{f}.$$

Siden

$$\dot{\mathbf{y}}{=}A\mathbf{Y},$$

kan likningen forkortes til

$$\mathbf{Y}\mathbf{c} = \mathbf{f},$$

og siden $\mathbf{Y}$ er inverterbar for alle $t$, kan vi invertere

$$\mathbf{c} = \mathbf{Y}^{-1}\mathbf{f}$$

og integrere komponentvis

$$\mathbf{c}(t) = \int^t \mathbf{Y}^{-1}(s)\mathbf{f}(s)\, ds.$$

Denne løsningsformelen gir korrekt løsning, dersom det er mulig å utføre integralet på høyre side. Integralet skal tolkes komponentvis, og det trengs ingen nedre integrasjonsgrense, siden denne bare legger til en lineærkombinasjon av homogene løsninger.

**Eksempel 12.23.** La nok en gang

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

slik at

$$\mathbf{Y} = \begin{pmatrix} e^{3t} & e^t \\ e^{3t} & -e^t \end{pmatrix}$$

og la

$$\mathbf{f} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Vi beregner

$$\mathbf{Y}^{-1} = \frac{-1}{2e^{4t}} \begin{pmatrix} -e^t & -e^t \\ -e^{3t} & e^{3t} \end{pmatrix}$$

og

$$\mathbf{Y}^{-1}\mathbf{f} = \frac{-1}{2e^{4t}} \begin{pmatrix} -e^t & -e^t \\ -e^{3t} & e^{3t} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} e^{-3t} \\ 0 \end{pmatrix}$$

slik at

$$\mathbf{c}(t) = \int_0^t \mathbf{Y}^{-1}(s)\mathbf{f}(s)\, ds = -\frac{1}{3}\begin{pmatrix} e^{-3t} \\ 0 \end{pmatrix}$$

og

$$\mathbf{y}_p(t) = -\frac{1}{3}\begin{pmatrix} e^{3t} & e^t \\ e^{3t} & -e^t \end{pmatrix}\begin{pmatrix} e^{-3t} \\ 0 \end{pmatrix} = -\frac{1}{3}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$ △

Det er bare noen ytterst få differensiallikninger vi kan løse med penn og papir. I dette kapitlet skal vi se på et par numeriske metoder for differensiallikningssystemer. Da er det nyttig å vite at alle differensiallikninger kan skrives om til førsteordens systemer, for dette reduserer antall metoder man må lære seg betraktelig.

**Eksempel 12.24.** Differensiallikningen for en pendel er

$$\ddot{y} + \sin y = 0.$$

Vi skriver denne om til et system ved å sette $z = \dot{y}$, slik at

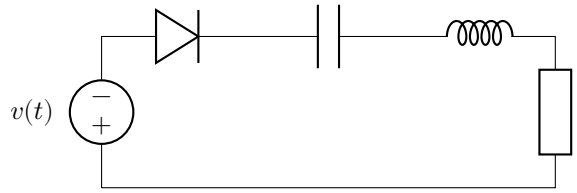
$$\begin{aligned} \dot{y} &= z \\ \dot{z} &= -\sin y \end{aligned}$$ △

**Eksempel 12.25.** Summer du spenningsfallet over kretsen under og deriverer alt, får du

$$\dot{v}(t) = \frac{V_0\dot{i}(t)}{i_0 + i(t)} + \frac{i(t)}{C} + L\ddot{i}(t) + R\dot{i}(t)$$

Vi skriver også denne om til et system ved å sette $y = i$ $z = \dot{i}$, slik at

$$\begin{aligned} \dot{y} &= z \\ L\dot{z} &= \dot{v}(t) - \frac{V_0 z}{i_0 + y} - \frac{y}{C} - Rz \end{aligned}$$ △




**Eksempel 12.26.** Noen modeller må skrives opp direkte som differensiallikningssystemer. Lotka-Volterra-systemet

$$\begin{aligned} \dot{y} &= y(2 - z) \\ \dot{z} &= z(y - 1) \end{aligned}$$

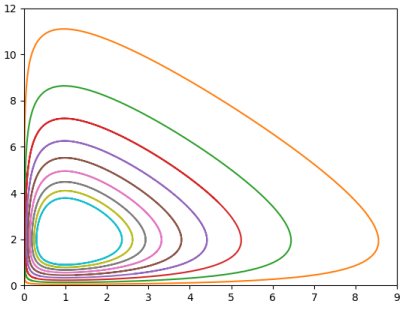
beskriver to dyrepopulasjoner, der den ene driver med predasjon på den andre. Dersom det er mange mus ($y$) på Revneset, får jorduglen ($z$) rikelig med mat til ungene sine, og kan legge opp til 14 egg. Men jorduglen er en trekkfugl, og er det ikke smågnagerår, gidder den ikke hekke i Norge engang. Hvis vi deler likningene på hverandre, samler $y$ og $z$ på hver sin side:

$$\dot{y}\left(1 - \frac{1}{y}\right) = \dot{z}\left(\frac{2}{z} - 1\right)$$

og integrerer, får vi

$$y - \log y = 2\log z - z + C.$$

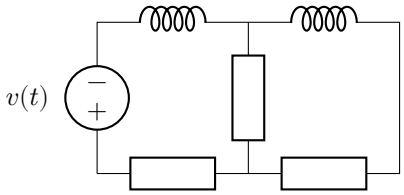
Her har vi en implisitt likning som alle løsninger må tilfredsstille. Forskjellige verdier av $C$ gir forskjellige lukkede løsningskurver, men vi har ikke funnet en funksjon som løser problemet. Se figuren under og her:
`https://mc-stan.org/users/documentation/case-studies/lotka-volterra-predator-prey.html#data-lynx-and-hare-pelts-in-canada` △



**Eksempel 12.27.** Andre situasjoner må ikke nødvendigvis modelleres som systemer av differensiallikninger, men det er veldig praktisk å gjøre det. Summer du spenningsfallet over hver sløyfe i kretsen under (med $L = 1$ og $R = 1$), får du:

$$\begin{aligned}\dot{y} &= -2y + z + v\\ \dot{z} &= \quad y - 2z\end{aligned}$$

Her er $y$ og $z$ de ukjente maskestrømmene, mens $v$ er gitt. △




**Eksempel 12.28.** En likning som opprinnelig dukket opp i radiorørteknologi, kalles van der Pols likning:

$$\ddot{x} - \mu\left(1 - x^2\right)\dot{x} + x = 0$$

Siden har det blitt oppdaget at den kan brukes til andre ting, for eksempel modellering nevroner eller sprekker mellom kontinentalplater. Skriver vi $y = x$ og $z = \dot{x}$, får vi systemet

$$\begin{aligned}\dot{y} &= z\\ \dot{z} &= \mu\left(1 - y^2\right)z - y.\end{aligned}$$ △

I høst så differensiallikningen slik ut:

$$\dot{y} = f(y)$$

Nå ser det slik ut

$$\begin{aligned}\dot{y} &= f(y, z)\\ \dot{z} &= g(y, z)\end{aligned}$$

De numeriske metodene fungerer ganske likt. Vi skriver opp fire av dem. Eksplisitt Euler:

$$\begin{aligned}y_{i+1} &= y_i + hf(y_i, z_i)\\ z_{i+1} &= z_i + hg(y_i, z_i)\end{aligned}$$

Implisitt Euler:

$$\begin{aligned}y_{i+1} &= y_i + hf(y_{i+1}, z_{i+1})\\ z_{i+1} &= z_i + hg(y_{i+1}, z_{i+1})\end{aligned}$$

Symplektisk Euler (denne finnes kun for systemer):

$$\begin{aligned}y_{i+1} &= y_i + hf(y_i, z_i)\\ z_{i+1} &= z_i + hg(y_{i+1}, z_i)\end{aligned}$$

Trapesmetoden:

$$\begin{aligned}y_{i+1} &= y_i + h\left(\frac{f(y_i, z_i) + f(y_{i+1}, z_{i+1})}{2}\right)\\ z_{i+1} &= z_i + h\left(\frac{g(y_i, z_i) + g(y_{i+1}, z_{i+1})}{2}\right)\end{aligned}$$

**Eksempel 12.29.** Differensiallikningen for en pendel er

$$\begin{aligned}\dot{y} &= z\\ \dot{z} &= -\sin y\end{aligned}$$

Eksplisitt Euler:

$$\begin{aligned}y_{i+1} &= y_i + hz_i\\ z_{i+1} &= z_i - h\sin y_i\end{aligned}$$

Imsplisitt Euler:

$$\begin{aligned}y_{i+1} &= y_i + hz_{i+1}\\ z_{i+1} &= z_i - h\sin y_{i+1}\end{aligned}$$

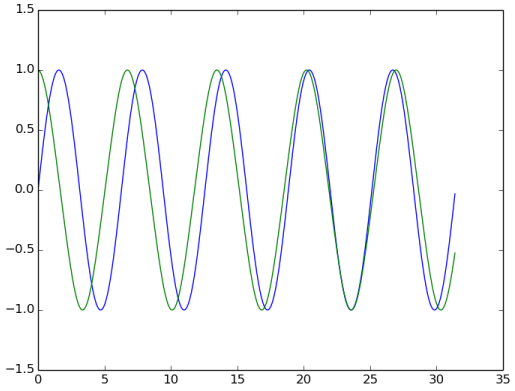
Symplektisk Euler

$$\begin{aligned}y_{i+1} &= y_i + hz_i\\ z_{i+1} &= z_i - h\sin y_{i+1}\end{aligned}$$

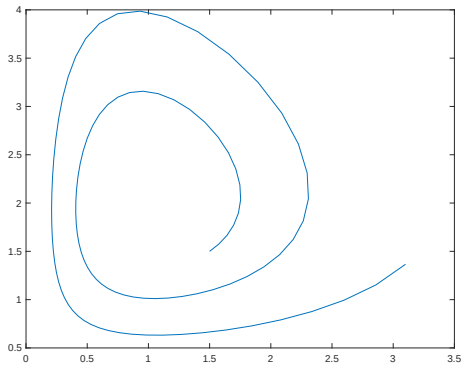
Trapesmetoden:

$$\begin{aligned}y_{i+1} &= y_i + \frac{h}{2}(z_i + z_{i+1})\\ z_{i+1} &= z_i - \frac{h}{2}(\sin y_i + \sin y_{i+1})\end{aligned}$$

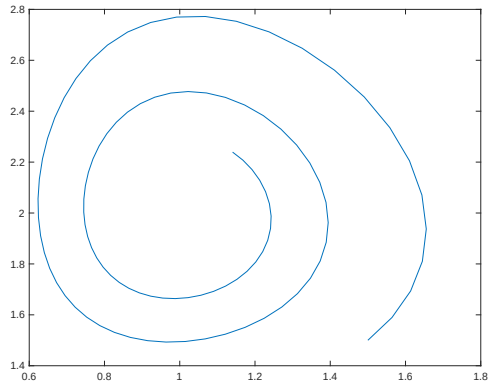
På de implisitte metodene må man kjøre en eller annen flerdimensjonal likningsløser for å finne $(y_{i+1}, z_{i+1})$ i hvert steg. Symplektisk Euler har noen artige egenskaper som vi skal komme tilbake til. Under er et plot av noen løsninger. △



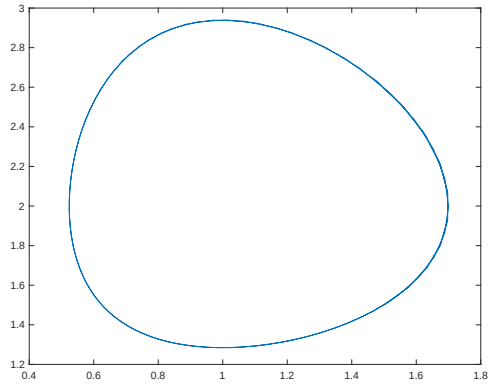

**Eksempel 12.30.** Vi løser Lotka-Volterra-systemet med eksplisitt Euler og $h = 0.1$, og får figuren under. Løsningskurven starter i $(1.5, 1.5)$. Eksplisitt Euler lager en utadgående spiral, men dette må være tull, for vi vet jo at løsningene skal ligge på en lukket kurve. △



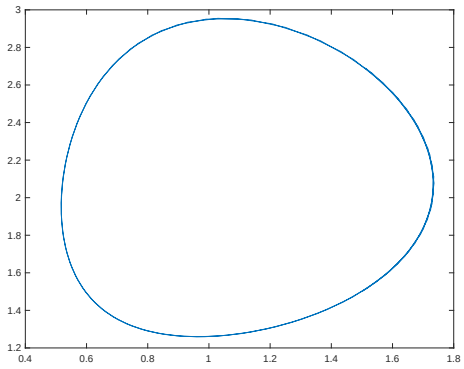
**Eksempel 12.31.** Nå kjører vi implisitt euler med $h = 0.1$, og får får følgende figur. Nå lager metoden en innadgående spiral! △



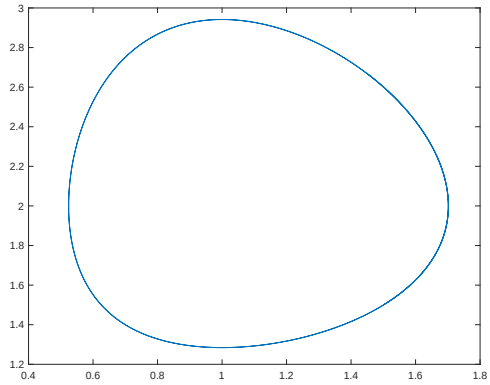
**Eksempel 12.32.** Her er trapesmetoden. Den klarer seg visst mye bedre, selv med $h = 0.1$. △



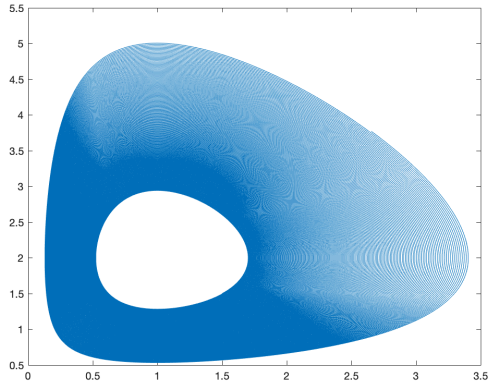
**Eksempel 12.33.** *Symplektisk Euler* er litt spesiell fordi den ikke har noen ekvivalendt i envariabel kalkulus. Trikset er veldig enkelt, vi nyttiggjør oss $y_{i+1}$ i den andre likningen så snart den er beregnet i den første. Denne metoden kalles semi-implisitt, og produserer en like pen figur som trapesmetoden. △



**Eksempel 12.34.** Vi løser Lotka-Volterra-systemet med eksplisitt Euler og $h = 0.0001$, og får figuren under. Man kan her bli forledet til å tro at liten nok $h$ løser alle problemer, men... △



**Eksempel 12.35.** nå tar vi eksplisitt Euler med $h = 0.001$, men for $t \in [0, 1000]$. FIguren under viser hva som skjer. Den numeriske løsningen holder seg ikke på de lukkede banene vi vet den skal, men spiraliserer utover. Det er ingen bug i koden. Det er faktisk bare eksplisitt Euler som ikke greier å løse dette problemet numerisk på en fornuftig måte. △



Løsningene til L-V-systemet skal definitivt være lukkede baner, og det ser ut til at trapesmetoden og symplektisk Euler klarer noe som ikke eksplisitt og implisitt Euler klarer. Numeriske metoder for differensialliknigner er mer enn bare presisjon og desimaler. Dette er for øvrig også et forskningsfelt som kalles *geometrisk integrasjon* - studiet om artige fenomener i numerisk løsning av differensiallikninger.

# Kapittel 13

# Partielle differensiallikninger I

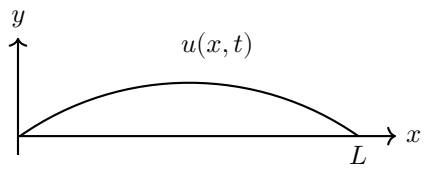
Fysiske situasjoner der det er behov for mer enn en uavhengig variabel, beskrives gjerne av partielle differensiallikninger. Vi skal ta for oss tre av de mest grunnleggende likningene, nemlig bølgelikningen, varmelikningen, og Laplaces likning.

## Bølgelikningen

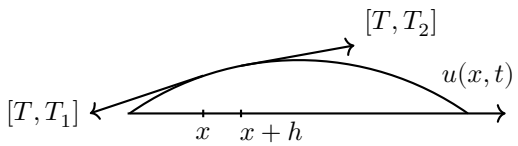
Bølgelikningen er en matematisk beskrivelse av en vibrerende streng, eller en stående luftbølge i en orgelpipe. Vi skal ta for oss selve likningen, hvor den kommer fra, og to forskjellige løsningsteknikker - en for et intervall på $x$-aksen, og en for hele $x$-aksen.

### Utledning

Vi tenker at vi har en vibrerende streng som er spent opp i $x = 0$ og $x = L$. La $u(x,t)$ være en funksjon som for hvert tidspunkt $t$ og hvert punkt $x$ beskriver utslaget fra likevektslinjen, som ligger langs $x$-aksen. Strengen har konstant massetetthet $\rho\,[kg/m]$.




Vi tar en nærmere titt på strekkreftene på et lite stykke av strengen. Vi antar at tyngdekraften er neglisjerbar, og at strengen er helt elastisk, slik at strengestrekket, som virker parallelt med strengen, er eneste kraft. Vi antar at hvert punkt på strengen kun beveger seg loddrett, og at den horisontale komponenten av strengestrekket er konstant lik $T$.




Vi setter opp Newtons andre lov for den lille biten fra $x$ til $x+h$. Massen til en bit med lengde $h$ er $h\rho$, og akselerasjonen til strengen i punktet $x$ er $u_{tt}(x,t)$. Netto kraft på biten er gitt ved $T_2 + T_1$, slik at

$$h\rho u_{tt}(x,t) = T_2 + T_1,$$

eller

$$\begin{aligned}\frac{\rho}{T}u_{tt}(x,t) &= \frac{T_2/T + T_1/T}{h} \\ &= \frac{u_x(x+h,t) - u_x(x,t)}{h},\end{aligned}$$

siden stigningstallet til tangenten til strengen er gitt ved $u_x$. Lar vi nå $h \to 0$, får vi bølgeligningen

$$u_{tt}(x,t) = c^2 u_{xx}(x,t),$$

der $c^2 = \frac{T}{\rho}$.

### Oppstilling av problem

Det er ikke nok med en differensialligning som beskriver strengens bevegelse. Vi må også ha informasjon om hvordan bevegelsen blir satt igang, og hvor strengen er spent opp. Et fullstendig oppstilt problem er:

$$u_{tt}(x,t) = c^2 u_{xx}(x,t), \tag{13.1}$$

med randkrav

$$u(0,t) = u(L,t) = 0, \tag{13.2}$$

og initialkrav

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x). \tag{13.3}$$

Selve differensialligningen (13.1) forteller oss hva slags fysiske lover som skal tilfredsstilles (i dette tilfelle Newtons andre lov), eller hva slags oppførsel vi kan forvente av løsningen, for eksempel at det er en vibrerende streng det er snakk om. Randkravene (13.2) forteller oss at strengen er spent opp i $x = 0$ og $x = L$, slik at løsningen står helt i ro der. Initialkravene (13.3) forteller oss noen om hvordan bevegelsen settes i gang; $f$ angir strengens posisjon ved $t = 0$, mens $g$ angir strengens fart ved $t = 0$. Når man spiller en tone på en gitar ved å dra i strengen og slippe den, slik man vanligvis gjør, er $g = 0$. Randkravene

$$u(0,t) = u(L,t) = 0, \tag{13.4}$$

kalles dirichletrandkrav.

### Løsning på intervall - separasjon av variable

Et stort geni har engang tenkt at løsningen på bølgeligningen kan skrives

$$u(x,t) = F(x)G(t).$$

Han hadde rett. Innsetting i (13.1) gir

$$F(x)G''(t) = c^2F''(x)G(t).$$

Vi deler på $c^2F(x)G(t)$ og får

$$\frac{F''(x)}{F(x)} = \frac{G''(t)}{c^2G(t)}.$$

Siden $x$ og $t$ skal kunne varieres uavhengig av hverandre, må vi ha at

$$\frac{F''(x)}{F(x)} = \frac{G''(t)}{c^2G(t)} = k$$

der $k$ er en foreløbig ubestemt konstant. Vi ganger opp med $c^2F(x)G(t)$ og bytter fortegn på $k$, slik at

$$F''(x) + kF(x) = 0$$

og

$$G''(t) + kc^2G(t) = 0.$$

Vi skal først prøve å finne ut hva $k$ kan være. Vi kan bruke $F$ og randkravene

$$u(0,t) = u(L,t) = 0,$$

til dette. La oss kikke på

$$F''(x) + kF(x) = 0.$$

Dersom $k = 0$, får vi

$$F''(x) = 0$$

som gir

$$F(x) = Ax + B.$$

Er dette en interessant løsning? Vel, nei. Dersom

$$F(0)G(t) = u(0,t) = 0,$$

må enten $F(0) = 0$ eller $G(t) = 0$. At $G(t) = 0$, slik at $u(x,t) = 0$, er en gyldig løsning av bølgeligningen, som også tilfredsstiller randkravene. Men dette er åpenbart ikke en spesielt interessant løsning, så jeg tror vi går for $F(0) = 0$, som impliserer $B = 0$. Det andre randkravet

$$u(L,t) = 0$$

gir likeledes at $F(L) = 0$, altså at $AL = 0$, som impliserer at $A = 0$. Dette impliserer igjen at $u(x,t) = 0$, vi konkluderer at $k = 0$ og $F(x) = 0$ ikke er en interessant løsning av problemet.

La oss prøve $k < 0$. I så fall løses

$$F''(x) + kF(x) = 0$$

av

$$F(x) = Ae^{\sqrt{-k}x} + Be^{-\sqrt{-k}x}$$

Dersom vi nå bruker randkravene, får vi de to likningene

$$\begin{aligned} A + B &= 0 \\ Ae^{\sqrt{-k}L} + Be^{-\sqrt{-k}L} &= 0 \end{aligned}$$

Dersom $k \neq 0$ (som vi jo allerede vet), er determinanten til dette systemet gitt ved

$$e^{\sqrt{-k}L} + e^{-\sqrt{-k}L} \neq 0$$

og vi konkluderer med at $A = B = 0$, slik at $F(x) = 0$. Altså er heller ikke $k < 0$ en interessant løsning.

Dersom $k > 0$, går alt så meget bedre, og vi får

$$F(x) = A\cos\sqrt{k}x + B\sin\sqrt{k}x.$$

Bruker vi randkravet $u(x,t) = 0$, får vi

$$A = 0,$$

og krever vi

$$F(L) = B\sin(\sqrt{k}L)G(t) = 0,$$

kan dette oppnås ved å sette $B = 0$, som er uinteressant siden da blir $u(x,t) = 0$, eller ved å kreve

$$\sqrt{k}L = n\pi.$$

slik at

$$k = \left(\frac{n\pi}{L}\right)^2.$$

Vi ser også at $n > 0$, for dersom $n < 0$ byttes bare fortegnet på $B$, som ennå er ubestemt. Ved å ta en titt på det endelige løsningen av problemet nedenfor, ser man at $B$ kommer til å bli overflødig, så vi velger $B = 1$.

Ligningene

$$G''(t) + c^2\left(\frac{n\pi}{L}\right)^2 G(t) = 0.$$

løses av

$$G_n(t) = A_n\cos c\frac{n\pi}{L}t + B_n\sin c\frac{n\pi}{L}t,$$

så de generelle løsningene til bølgeligningen med randkrav (13.2) blir

$$\begin{aligned} u_n(x,t) &= F(x)G_n(t) \\ &= \left(A_n\cos c\frac{n\pi}{L}t + B_n\sin c\frac{n\pi}{L}t\right)\sin\frac{n\pi}{L}x. \end{aligned}$$

Vi har ennå ikke tatt stilling til initialkravene (13.3). Det kan vi klare ved å skrive

$$\begin{aligned} u(x,t) &= \sum_{n=1}^{\infty} u_n(x,t) \\ &= \sum_{n=1}^{\infty}\left(A_n\cos c\frac{n\pi}{L}t + B_n\sin c\frac{n\pi}{L}t\right)\sin\frac{n\pi}{L}x. \end{aligned}$$

Dersom vi nå krever

$$f(x) = u(x,0) = \sum_{n=1}^{\infty} u_n(x,0) = \sum_{n=1}^{\infty} A_n\sin\frac{n\pi}{L}x,$$

ser vi at summen til høyre bør være fourierrekken til den odde utvidelsen til $f$, og følgelig bør

$$A_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi}{L}x\,dx = \frac{2}{L}\int_0^L f(x)\sin\frac{n\pi}{L}x\,dx.$$

På samme vis, dersom

$$g(x) = u_t(x,0) = \sum_{n=1}^{\infty} B_n c\frac{n\pi}{L}\sin\frac{n\pi}{L}x,$$

bør summen til høyre være fourierekken til den odde utvidelsen til $g$, og følgelig må

$$B_n c\frac{n\pi}{L} = \frac{2}{L}\int_0^L g(x)\sin\frac{n\pi}{L}x\,dx.$$

slik at

$$B_n = \frac{2}{cn\pi}\int_0^L g(x)\sin\frac{n\pi}{L}x\,dx.$$

Når vi skal oppsummere, må vi være litt forsiktige, for vi har ikke sagt noe om deriverbarhet. En løsning av bølgelikningen bør helst være to ganger kontinuerlig deriverbar, men dersom du drar i en gitarstreng med fingeren og slipper strengen, vil $f$ har form som en sagtann, og ikke være deriverbar i punktet der fingeren din holder strengen. (Ihvertfall hvis strengen er elastisk, og fingen din er uendelig tynn.) Av disse grunnene bør man kanskje ikke kalle oppsummeringen for et teorem, men det får få en rød ramme siden det ligger et visst resonnement bak.

**Løsningsformel for dirichletrandkrav**

Bølgeligningen

$$u_{tt}(x,t) = c^2 u_{xx}(x,t),$$

med randkrav

$$u(0,t) = u(L,t) = 0,$$

og initialkrav

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x),$$

løses av

$$u(x,t) = \sum_{n=1}^{\infty}\left(A_n\cos c\frac{n\pi}{L}t + B_n\sin c\frac{n\pi}{L}t\right)\sin\frac{n\pi}{L}x,$$

der

$$A_n = \frac{2}{L}\int_0^L f(x)\sin\frac{n\pi}{L}x\,dx$$

og

$$B_n = \frac{2}{cn\pi}\int_0^L g(x)\sin\frac{n\pi}{L}x\,dx.$$

### Løsning for fløyte

En stående trykkbølge inne i fløyte, beskrives av problemet

$$u_{tt}(x,t) = c^2 u_{xx}(x,t),$$

med randkrav

$$u_x(0,t) = u_x(L,t) = 0,$$

og initialkrav

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x).$$

Konstanten $c$ avhenger nå av trykket og lydhastigheten i mediet der lydbølgene produseres. (Dersom du spiller på fløyten inni en gassballong full av helium, blir $c$ høyere enn i luft.) Randkravene kalles *von-Neumann-randkrav*, og $L$ er lengden på fløyten.

Vi løser problemet på samme måte som for den vibrerende strengen. Den eneste forskjellen blir i forbindelse med løsning av

$$F''(x) + kF(x) = 0.$$

På samme vis må $k > 0$, slik at

$$F(x) = A\cos\sqrt{k}x + B\sin\sqrt{k}x.$$

Siden

$$F'(x) = -A\sqrt{k}\sin\sqrt{k}x + B\sqrt{k}\cos\sqrt{k}x$$

impliserer

$$u_x(0,t) = F'(0) = 0,$$

at $B = 0$, mens

$$u_x(L,t) = F'(x) = -A\sqrt{k}\sin\sqrt{k}L = 0,$$

gir som før at

$$\sqrt{k}L = n\pi.$$

Merk at $n = 0$ er en gyldig løsning her siden en konstant funksjon løser både bølgelikningen og von-Neumann-randkravene. Resten blir som før, men vi må bruke cosinusrekkene til $f$ og $g$ istedet for sinusrekkene. Vi tar ikke resten av regningen.

**Løsningsformel for fløyte**

Bølgeligningen

$$u_{tt}(x,t) = c^2 u_{xx}(x,t),$$

med randkrav

$$u_x(0,t) = u_x(L,t) = 0,$$

og initialkrav

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x),$$

løses av

$$u(x,t) = A + \sum_{n=1}^{\infty}\left(A_n\cos c\frac{n\pi}{L}t + B_n\sin c\frac{n\pi}{L}t\right)\cos\frac{n\pi}{L}x,$$

der $A$ er en vilkårlig konstant,

$$A_n = \frac{2}{L}\int_0^L f(x)\cos\frac{n\pi}{L}x\,dx$$

og

$$B_n = \frac{2}{cn\pi}\int_0^L g(x)\cos\frac{n\pi}{L}x\,dx.$$

### Løsning på hele $x$-aksen - D'alembert

Hvis man ikke ønsker å bruke bøglelikningen til å beskrive en oppspent streng, men heller bølgene fra

et steinkast på et endimensjonalt og uendelig langt hav, må man studere bølgelikningen

$$u_{tt}(x,t) = c^2 u_{xx}(x,t),$$

med initialkrav

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x),$$

og ingen randkrav. Her skal løsningen gjelde for alle $x$, ikke bare på intervallet $[0, L]$

I noen tilfeller kan dette gjøre alt vanskeligere, men akkurat i tilfellet bølgelikningen, blir alt fryktelig enkelt. Det er enkelt å sjekke at funksjonen

$$u(x,t) = \phi(x+ct) + \psi(x-ct)$$

passer i likningen

$$u_{tt}(x,t) = c^2 u_{xx}(x,t).$$

uansett hva $\phi$ og $\psi$ er, så lenge de er to ganger kontinuerlig deriverbare. Spørsmålet blir bare da hvordan vi skal klare å innpasse initialkravene.

Dersom man bruker

$$u(x,0) = f(x)$$

får man

$$\phi(x) + \psi(x) = f(x)$$

og dersom man bruker

$$u_t(x,0) = g(x),$$

får man

$$c\phi'(x) - c\psi'(x) = g(x),$$

eller

$$\phi(x) - \psi(x) = \frac{1}{c}\int^x g(t)\,dt + C$$

Vi har nå et lineært $2\times 2$-likningssystem for $\phi$ og $\psi$. Legger vi likningene sammen, får vi

$$2\phi(x) = f(x) + \frac{1}{c}\int^x g(t)\,dt + C$$

og trekker vi dem fra hverandre, får vi får vi

$$2\psi(x) = f(x) - \frac{1}{c}\int^x g(t)\,dt - C$$

Vi setter nå alt sammen igjen

$$\begin{aligned}
u(x,t) =& \phi(x+ct) + \psi(x-ct)\\
=& \frac{1}{2}\left(f(x+ct) + f(x-ct)\right)\\
&+ \frac{1}{2c}\left(\int^{x+ct} g(t)\,dt - \int^{x-ct} g(t)\,dt\right)\\
=& \frac{1}{2}\left(f(x+ct) + f(x-ct)\right) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(t)\,dt,
\end{aligned}$$

og oppsummerer.

**D'Alemberts løsningsformel**

Bølgeligningen

$$u_{tt}(x,t) = c^2 u_{xx}(x,t),$$

på hele $x$-aksen, med initialkrav

$$u(x,0) = f(x) \qquad u_t(x,0) = g(x),$$

løses av

$$u(x,t) = \frac{1}{2}\left(f(x+ct) + f(x-ct)\right) + \frac{1}{2c}\int_{x-ct}^{x+ct} g(t)\,dt.$$

# Varmelikningen

Varmelikningen beskriver kort og godt varmeflyt i materialer. Vi skal ta for oss den enkleste varianten, der varmeledningsevnen er identisk overalt og i alle retninger i materialet.

**Lesestoff**

`https://www.feynmanlectures.caltech.edu/II_02.html`
`https://www.feynmanlectures.caltech.edu/II_03.html`

### Utledning

I en ideell gass er det et lineært forhold mellom temperatur og opplagret bevegelsesenergi, gitt ved

$$\frac{1}{2}mv^2 = \frac{3}{2}kT$$

der $k = 1.38064852(79)\times 10^{-23}$ J/K er Boltzmanns konstant, $v$ er gjennomsnittsfarten til gassmolekylene, og $T$ er temperaturen målt i Kelvin. Vi kan altså tenke på temperatur som et mål på opplagret kinetisk energi. I andre materialer enn ideelle gasser kan den kinetiske energien også lagres opp som rotasjon og vibrasjon.

La oss tenke at temperaturen er gitt ved funksjonen $u(\mathbf{x}, t)$, der $\mathbf{x} = (x_1, x_2, x_3)$ er de romlige koordinatene og $t$ er tidspunktet. La $\Omega$ være et område i $\mathbb{R}^3$. Den totale opplagrede varmeenergien i $\Omega$ er gitt ved

$$\rho c \iiint_\Omega u(\mathbf{x},t)\,d\mathbf{x},$$

der $\rho$ er massetettheten og $c$ er den spesifikke varmekapasiteten. Dersom vi antar at $\rho$ og $c$ er konstante og legger noen restriksjoner på glattheten til $u$ (for kompliserte for oss), er endringen i denne energien over tid gitt ved

$$\rho c\frac{\partial}{\partial t}\iiint_\Omega u(\mathbf{x},t)\,d\mathbf{x} = \rho c\iiint_\Omega u_t(\mathbf{x},t)\,d\mathbf{x}$$

Transport av varmeenergi er for mange materialer proporsjonal med temperaturgradienten. Dette kalles Fouriers lov, og skrives

$$\mathbf{q} = -\kappa \nabla u$$

der $\kappa$ er den termiske konduktiviteten, og $\mathbf{q}$ er varmefluksen. Endringen i varmeenergi i $\Omega$ skyldes enten varme som produseres inne i $\Omega$ (tenk at det står et stearinlys og brenner inne i $\Omega$), eller varme som slipper inn og ut gjennom randen $\partial\Omega$. Dersom vi antar at $\kappa$ er konstant og at det ikke produseres varme i $\Omega$, må

$$\begin{aligned} -\rho c \iiint_\Omega u_t \, d\mathbf{x} &= \iint_{\partial\Omega} \mathbf{q} \cdot dS \\ &= \iiint_\Omega \nabla \cdot \mathbf{q} \, d\mathbf{x} \\ &= -\kappa \iiint_\Omega \nabla \cdot \nabla u \, d\mathbf{x} = -\kappa \iiint_\Omega \Delta u \, d\mathbf{x} \end{aligned}$$

eller

$$\rho c \iiint_\Omega u_t \, d\mathbf{x} = \kappa \iiint_\Omega \Delta u \, d\mathbf{x}.$$

Dersom dette gjelder for et tilfeldig valgt område $\Omega$, bør integrandene være like:

$$\rho c u_t = \kappa \Delta u$$

Det er vanlig å samle den spesifikke varmekapasiteten, massetettheten og den termiske konduktiviteten i en konstant

$$\alpha = \frac{\kappa}{\rho c}$$

som kalles den termiske diffusiviteten. Likningen

$$u_t = \alpha \Delta u$$

kalles varmelikningen eller diffusjonslikningen. Det er også vanlig å sette $\alpha = 1$, for dersom $u$ løser $u_t = \alpha \Delta u$ må $v(\mathbf{x}, t) = u(\mathbf{x}, t/\alpha)$ løse $v_t = \Delta v$.

### Løsning på begrenset intervall

Nå skal vi tenke oss en isolert stang der temperaturen er gitt av $u(x,t)$. Varme kan slippe ut eller tilføres gjennom endepunktene, men ingen andre steder. Vi skal løse varmeligningen

$$u_t = u_{xx}$$

på stang med lengde $L$. Randkravene

$$u(0,t) = u(L,t) = 0, \tag{13.5}$$

sier at temperaturen holdes konstant lik 0 i endepunktene, og initialkravet

$$u(x,0) = f(x) \tag{13.6}$$

sier at temperaturfordelingen er gitt ved $f(x)$ når tiden begynner å gå.

Vi prøver separasjon av variable

$$u(x,t) = F(x)G(t).$$

På samme vis som for bølgeligningen, får vi

$$F''(x) + kF(x) = 0$$

og

$$G'(t) + kG(t) = 0.$$

Å finne $F$ og $k$ blir eksakt repetisjon av argumentet for bølgeligningen, mens

$$G_n(t) = A_n e^{-(\frac{n\pi}{L})^2 t}$$

slik at de generelle løsningene til bølgeligningen med randkrav $u(0,t) = u(L,t) = 0$ blir

$$u_n(x,t) = F(x)G_n(t) = A_n e^{-(\frac{n\pi}{L})^2 t} \sin \frac{n\pi}{L} x.$$

Likeledes kan initialkravet $u(x,0) = f(x)$ inkorporeres ved å legge sammen alle løsninger

$$u(x,t) = \sum_{n=1}^{\infty} u_n(x,t) = \sum_{n=1}^{\infty} A_n e^{-(\frac{n\pi}{L})^2 t} \sin \frac{n\pi}{L} x,$$

sette $t = 0$

$$f(x) = u(x,0) = \sum_{n=1}^{\infty} A_n \sin \frac{n\pi}{L} x,$$

og observere at vi får dette til dersom $A_n$ er fourierkoeffisientene til den odde utvidelsen til $f$:

$$A_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi}{L} x \, dx$$

**Løsningsformel for varmeledning på gittarstreng**

Varmelikningen

$$u_t(x,t) = u_{xx}(x,t),$$

med randkrav

$$u(0,t) = u(L,t) = 0,$$

og initialkrav

$$u(x,0) = f(x)$$

løses av

$$u(x,t) = \sum_{n=1}^{\infty} A_n e^{-(\frac{n\pi}{L})^2 t} \sin \frac{n\pi}{L} x,$$

der

$$A_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi}{L} x \, dx.$$

### Løsning på hele $x$-aksen

Vi skal nå løse varmeligningen

$$u_t = u_{xx}$$

på en uendelig lang stang. Randkravene skal være

$$\lim_{x \to \pm\infty} u(x,t) = 0, \tag{13.7}$$

mens initialkravet er som før

$$u(x,0) = f(x), \tag{13.8}$$

der vi antar

$$\lim_{x\to\pm\infty} f(x) = 0. \tag{13.9}$$

Vi skal løse dette problemet ved å fouriertransformere varmeligningen med hensyn på $x$. Siden $\lim_{x\to\infty} u(x,t) = \lim_{x\to-\infty} u(x,t) = 0$, satser vi på at det går bra, og vi får

$$\mathcal{F}(u_t) = \mathcal{F}\left(u_{xx}\right).$$

Hvis vi antar at $u$ er kontinuerlig deriverbar i $t$, og fouriertransformen er med hensyn på $x$, kan vi skrive

$$\begin{aligned}\mathcal{F}(u_t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u_t(x,t)e^{-i\omega x}\,dx \\ &= \frac{\partial}{\partial t}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u(x,t)e^{-i\omega x}\,dx = \hat{u}_t(w,t),\end{aligned}$$

og

$$\mathcal{F}(u_{xx}) = -\omega^2\mathcal{F}(u) = -\omega^2\hat{u}(w,t),$$

slik at

$$\hat{u}_t = -\omega^2\hat{u}.$$

Dette er en ordinær differensialligning for $\hat{u}$, med løsning

$$\hat{u}(\omega,t) = A(\omega)e^{-\omega^2 t}.$$

Fouriertransformerer vi initialkavet, ser vi at

$$A(\omega) = \hat{u}(\omega,0) = \mathcal{F}\left(u(x,0)\right) = \mathcal{F}(f(x)) = \hat{f}(\omega),$$

og da er det egentlig bare å inversfouriertransformere, og så har vi løsningen

$$u(x,t) = \mathcal{F}^{-1}\left(\hat{u}\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{-\omega^2 t}e^{i\omega x}\,d\omega.$$

Vel og bra, men vi kan komme oss et knepp til. Husk at

$$f * g = \mathcal{F}^{-1}\left(\sqrt{2\pi}\,\hat{f}\hat{g}\right),$$

og observer at

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{-\omega^2 t}e^{i\omega x}\,d\omega = \mathcal{F}^{-1}\left(\hat{f}e^{-\omega^2 t}\right).$$

Med andre ord, hvis vi kunne inverstransformere $e^{-\omega^2 t}$ hadde det vært bra. Husk at

$$\mathcal{F}\left(e^{-ax^2}\right) = \frac{1}{\sqrt{2a}}e^{\frac{-\omega^2}{4a}}.$$

Setter vi

$$a = \frac{1}{4t},$$

og ganger og deler litt med konstanter, kan vi si at

$$\mathcal{F}\left(\frac{1}{\sqrt{2t}}e^{\frac{-x^2}{4t}}\right) = e^{-\omega^2 t},$$

som gir

$$\begin{aligned}u(x,t) =& \mathcal{F}^{-1}(\hat{f}e^{-\omega^2 t}) \\ =& \frac{1}{\sqrt{2\pi}}\left(f * \frac{1}{\sqrt{2t}}e^{\frac{-x^2}{4t}}\right) \\ =& \frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} f(v)e^{\frac{-(x-v)^2}{4t}}\,dv\end{aligned}$$

Onde tunger vil kanskje hevde at $u(x,t)$ ikke er definert for $t = 0$. Dette er ikke bra, for vi har jo prøvd å få til at $u(x,0) = f(x)$, som åpenbart ikke kan være helt riktig. Nå kan det heldigvis vises at

$$\lim_{t\to 0} u(x,t) = f(x)$$

men dette er for vanskelig for oss.

**Kjempe lang stang på Vang**

Varmelikningen

$$u_t = u_{xx},$$

på hele $x$-aksen med initialkrav

$$u(x,0) = f(x)$$

løses av

$$u(x,t) = \frac{1}{2\sqrt{\pi t}}\int_{-\infty}^{\infty} f(v)e^{\frac{-(x-v)^2}{4t}}\,dv,$$

og vi har

$$\lim_{t\to 0} u(x,t) = f(x).$$

# Laplaces ligning

For å få en følelse for Laplaces ligning kommer fra, må man betrakte varmeligningen i to romlige dimensjoner

$$u_t = u_{xx} + u_{yy}.$$

Det fysiske bildet du kan ha, er en tynn varmeisolert plate der varmeenergi kan forsvinne ut eller inn gjennom sidekantene. Løsningen til varmeligningen vil da gi temperaturen $u(x,y,t)$ i platen gitt fornuftige rand- og initialkrav. Vi skal ikke løse varmeligningen i to romlige dimensjoner, men derimot betrakte spesialtilfellet der $t \to \infty$, slik at varmestrømmen er konstant, eller steady-state flow som vi sier på norsk. Dersom varmestrømmen er konstant, er $u_t(x,y,t) = 0$, slik at varmeligningen blir

$$u_{xx} + u_{yy} = 0.$$

Dette kalles Laplaces ligning, og den kan vi løse med teknikker vi har lært i kurset. Her er en figur som beskriver en typisk situasjon. Dette er temperaturen i en tynn plate der temperaturen holdes konstant lik null på tre sider, og konstant lik fem eller en eller noe på den siste siden.

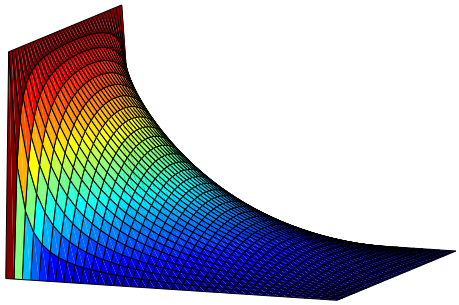
### Løsning på begrenset intervall - separasjon av variable

Vi setter opp problemet

$$u_{xx} + u_{yy} = 0,$$

på rektangelet $[0,a]\times[0,b]$ med randkrav

$$u(x,0) = u(0,y) = u(a,y) = 0$$



og

$$u(x,b) = f(x).$$

Vi separerer i vei

$$u(x,y) = F(x)G(y).$$

På samme vis som for bølgeligningen og varmeligningen, får vi

$$F''(x) + kF(x) = 0$$

og

$$G''(y) - kG(y) = 0.$$

Å finne $F$ og $k$ blir eksakt repetisjon av argumentet for bølgeligningen, mens

$$G_n(y) = A_n e^{-\frac{n\pi}{a}y} + B_n e^{\frac{n\pi}{a}y}$$

slik at de generelle løsningene som tilfredsstiller $u(0,y) = u(b,y) = 0$ blir

$$\begin{aligned} u_n(x,y) &= F(x)G_n(y) \\ &= \left(A_n e^{-\frac{n\pi}{a}y} + B_n e^{\frac{n\pi}{a}y}\right) \sin\frac{n\pi}{a}x. \end{aligned}$$

Krever vi $u(x,0) = 0$, får vi

$$A_n = -B_n,$$

slik at

$$\begin{aligned} u_n(x,y) &= A_n\left(e^{-\frac{n\pi}{a}y} - e^{\frac{n\pi}{a}y}\right)\sin\frac{n\pi}{a}x \\ &= A_n \sinh\frac{n\pi}{a}y \sin\frac{n\pi}{a}x, \end{aligned}$$

og krever vi $u(x,b) = f(x)$, får vi det til ved å legge sammen alle løsninger

$$u(x,y) = \sum_{n=1}^{\infty} u_n(x,y) = \sum_{n=1}^{\infty} A_n \sinh\frac{n\pi}{a}y \sin\frac{n\pi}{a}x,$$

og kreve

$$f(x) = u(x,b) = \sum_{n=1}^{\infty} A_n \sinh\frac{n\pi b}{a}\sin\frac{n\pi}{a}x,$$

slik at

$$A_n = \frac{2}{a\sinh\frac{n\pi b}{a}}\int_0^a f(x)\sin\frac{n\pi x}{a}\,dx.$$

**Løsning i halvplanet**

Vi skal se på Laplaces ligning

$$u_{xx} + u_{yy} = 0,$$

i halvplanet. Randkrav skal være

$$\lim_{x\to\pm\infty} u(x,y) = \lim_{y\to\infty} u(x,y) = 0$$

og

$$u(x,0) = f(x),$$

der

$$\lim_{x\to\pm\infty} f(x) = 0.$$

Siden $\lim_{x\to\pm\infty} u(x,y) = 0$ satser vi på at $u$ er absolutt integrerbar i $x$, slik at vi kan fouriertransformere med hensyn på $x$. Da får vi

$$\mathcal{F}(u_{xx}) + \mathcal{F}(u_{yy}) = \mathcal{F}(0)$$

eller

$$-w^2\hat{u}(w,y) + \hat{u}_{yy}(w,y) = 0.$$

Akkurat som med varmeligningen, later vi som om dette er en ordinær differensialligning i $y$, og får

$$\hat{u}(w,y) = A(w)e^{-|w|y} + B(w)e^{|w[y}.$$

Nå må vi fouriertransformere randkravet $\lim_{y\to\infty} u(x,y) = 0$, og da får vi

$$\lim_{y\to\infty} \hat{u}(w,y) = 0.$$

Bruker vi dette, ser vi at $B(w) = 0$, og følgelig er

$$\hat{u}(w,y) = A(w)e^{-|w|y}.$$

Transformerer vi $u(x,0) = f(x)$, får vi

$$\hat{u}(w,0) = \hat{f}(w),$$

og bruker vi denne, får vi

$$\hat{u}(w,y) = \hat{f}(w)e^{-|w|y}.$$

Nå kan vi inverstransformere og få

$$\begin{aligned} u(x,y) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \hat{f}(w)e^{-|w|y}e^{iwx}\,dw \\ &= \mathcal{F}^{-1}(\hat{f}(w)e^{-|w|y}). \end{aligned}$$

Nå bruker vi samme trikset som for varmeligningen. Siden dette er inversfouriertransformen til produktet mellom $\hat{f}(w)$ og $e^{-|w|y}$, og

$$\mathcal{F}\left(\sqrt{\frac{2}{\pi}}\frac{y}{x^2+y^2}\right) = e^{-|w|y},$$

ser vi at

$$\begin{aligned} u(x,y) &= \frac{1}{\sqrt{2\pi}} f * \left(\sqrt{\frac{2}{\pi}}\frac{y}{x^2+y^2}\right) \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} f(v)\frac{y}{(x-v)^2+y^2}\,dv. \end{aligned}$$

**Noen artige egenskaper**

En funksjon som tilfredsstiller

$$u_{xx} + u_{yy} = 0$$

kalles gjerne en *harmonisk funksjon*. Harmoniske funksjoner har morsomme egenskaper.

**Middelverdisats for harmoniske funksjoner**

**Teorem 13.1.** *La $u$ være harmonisk på $\Omega$, og la $\mathcal{C} \in \Omega$ være en sirkelskive med sentrum i $(x_0, y_0)$ og radius $a$. Da gjelder at*

$$u(x_0, y_0) = \frac{1}{2\pi a} \int_{\partial \mathcal{C}} u \, ds = \frac{1}{\pi a^2} \iint_{\mathcal{C}} u \, dA.$$

Vi lar denne stå ubevist inntil videre. Beviset kommer ikke på eksamen i år. Ta kontakt om du lurer.

**Maksverdisats for harmoniske funksjoner**

**Teorem 13.2.** *La $u$ være harmonisk på $\Omega$, og la $\mathcal{C} \in \Omega$ være en sirkelskive med sentrum i $(x_0, y_0)$ og radius $a$. Da ligger*

$$\max_{(x,y)\in\mathcal{C}} u(x, y)$$

*og*

$$\min_{(x,y)\in\mathcal{C}} u(x, y)$$

*på $\partial\mathcal{C}$.*

Vi beviser ikke denne heller.

## Numerisk derivasjon

Taylorrekker kan brukes til å utlede en stor klasse av numeriske metoder for beregne tilnærminger til den deriverte.

I M1 definerte vi den deriverte til en funksjon $f$ som

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}.$$

Uttrykket

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

er stigningstallet til sekanten til $f$ mellom punktene $x$ og $x + h$. For små $h$ er denne sekanten en grei tilnærming til stigningstallet $f(x)$:

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}.$$

På papiret er det slik at jo mindre $h$, desto bedre tilnærming.

**Eksempel 13.3.** La $f(x) = e^x$, og la $h = 0.1$. Vi beregner

$$f'(1.5) \approx \frac{e^{1.6} - e^{1.5}}{0.1} = 4.7134.$$

Merk at

$$f'(1.5) = e^{1.5} = 4.4817,$$

så denne tilnærmingen bommer med rundt $2 \cdot 10^{-1}$. Vi kan også prøve $h = 0.01$. Da får vi

$$f'(1.5) \approx \frac{e^{1.51} - e^{1.5}}{0.01} = 4.5042.$$

som er noe bedre, nå er feilen på rundt $2 \cdot 10^{-2}$. Vi knekker til med enda en:

$$f'(1.5) \approx \frac{e^{1.501} - e^{1.5}}{0.001} = 4.4839,$$

og får en feil på rundt $2 \cdot 10^{-3}$. △

**Merk.** Feilen i forrige eksempel er tydelig proporsjonal med $h$ - deler du $h$ på 10, deler du feilen på 10. En god illustrasjon av lineær feil.

Det går an å utlede formelen

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}.$$

ved å bruke taylorutviklingen til $f$ i punktet $x$:

$$f(x+h) = f(x) + f'(x)h + \frac{f''(x)}{2}h^2 + \cdots.$$

Stigningen til sekanten kan skrives

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{f''(x)}{2}h + \cdots$$

og vi ser at dette stigninstallet består av den eksakte verdien for $f'(x)$ pluss resten av taylorrekken til $f$

$$\frac{f''(x)}{2}h + \frac{f'''(x)}{6}h^2 + \cdots$$

Denne halen forteller oss noe om feilen. Dersom $h$ er liten nok, vil $h$ være mye større enn $h^2$, og vi skriver

$$\frac{f''(x)}{2}h + \frac{f'''(x)}{6}h^2 + \cdots = O(h)$$

for å signalisere at feilen er proporsjonal med $h$.

Vi kan relativt lett forbedre den lineære tilnærmingen fra forrige avsnitt ved å skrive

$$f(x-h) = f(x) - f'(x)h + \frac{f''(x)}{2}h^2 + \cdots$$

og sette opp tilnærmingen

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + \frac{f'''(x)}{3!}h^2 + \frac{f^5(x)}{5!}h^4 + \cdots$$

Feilen i denne tilnærmingen er

$$\frac{f'''(x)}{3!}h^2 + \frac{f^5(x)}{5!}h^4 + \cdots$$

Hvis $h$ er liten, er det rimelig å anta denne feilen er mye mindre enn for den første tilnærmingen, siden $h^2 << h$.

**Eksempel 13.4.** La $f(x) = e^x$, og la $h = 0.1$. Vi beregner

$$f'(1.5) \approx \frac{e^{1.6} - e^{1.4}}{0.2} = 4.489162287752202,$$

som gir en feil på rundt $7.5 \cdot 10^{-3}$. Mye bedre enn i sted. Vi kan prøver også $h = 0.01$:

$$f'(1.5) \approx \frac{e^{1.51} - e^{1.49}}{0.02} = 4.481763765529401$$

som er noe bedre, nå er feilen på rundt $7.5 \cdot 10^{-5}$. Knekker vi til med $h = 0.001$, får vi

$$f'(1.5) \approx \frac{e^{1.501} - e^{1.499}}{0.002} = 4.481689817286139$$

som gir en feil på $7.5 \cdot 10^{-7}$. △

**Merk.** Legg merke til hvordan feilen i forrige eksempel er nærmest perfekt kvadratisk - vi får to nye desimaler hver gang vi deler $h$ på 10. Feilen deles altså på 100 når $h$ deles på 10.

Hvis du virkelig vil slå på stortrommen, kan du bruke formelen

$$f'(x) = \frac{f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)}{12h}$$

**Eksempel 13.5.** Vi bruker den store formelen med $h = 0.1$, $h = 0.01$ og $h = 0.001$. Då får vi feil på $10^{-5}$, $10^{-9}$, og $10^{-13}$. Fire nye desimaler hver gang $h$ deles på 10. Prøv selv. △

Dersom du trenger en tilnærming for $f''(x)$, kan du bruke den andre ordens sentraldifferansen

$$\frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = f''(x) + O(h^2)$$

Nå bør det etterhvert være klart hvordan en slik derivasjonsformel konstrueres - man søker en lineærkombinasjon av funksjonsverdier $f(x)$, $f(x-h)$ og $f(x+h)$ og lignende ledd, for å oppnå to ting:

- Korrekt tilnærming av den $n$-te deriverte.
- Så høy orden som mulig.

Det går an å kombinere tilnærminger av forskjellig orden til å oppnå høyere ordens tilnærminger. Vi definerer

$$\phi(h) = \frac{f(x+h) - f(x-h)}{2h}.$$

Da har vi at

$$\phi(h) = f'(x) + h^2 \frac{f'''(x)}{6} + h^4 \frac{f^5(x)}{120} + \cdots$$

og

$$\phi\left(\frac{h}{2}\right) = f'(x) + \left(\frac{h}{2}\right)^2 \frac{f'''(x)}{6} + \left(\frac{h}{2}\right)^4 \frac{f^5(x)}{120} + \cdots$$

Her er trikset:

$$\frac{4\phi\left(\frac{h}{2}\right) - \phi(h)}{3} = f'(x) - h^4 \frac{f^5(x)}{480} + \cdots$$

Med andre ord: den rette lineærkombinasjonen av to estimater med forskjellige gitterfinheter kan få et (eller flere) ledd i feilutviklingen til å forsvinne, og da får vi en høyere ordens tilnærming.

**Eksempel 13.6.** Hvis vi setter $h = 0.1$, og tar to tidligere approksimasjoner,

$$\phi(0.1) = \frac{e^{1.6} - e^{1.4}}{0.2} = 4.489162287752202,$$

med en feil på rundt $7.5 \cdot 10^{-3}$, og

$$\phi(0.01) = \frac{e^{1.51} - e^{1.49}}{0.02} = 4.481763765529401$$

med en feil på rundt $7.5 \cdot 10^{-5}$, og og kombinerer dem, får vi

$$\frac{100\phi\left(\frac{h}{10}\right) - \phi(h)}{99} = 4.481689032981695$$

som gir en feil på $-3.73 \cdot 10^{-8} \approx \left(\frac{h}{10}\right)^4$. △

Det går an å formulere presise teoremer som forteller hvordan man skal lineærkombinere tilnærminger til høyere ordens tilnærminger, men vi nøyer oss med denne lille smakebiten.

Vi skal lage numerisk metoder for varmeligningen

$$u_t = u_{xx}$$

med randkrav

$$u(0,t) = u(1,t) = 0$$

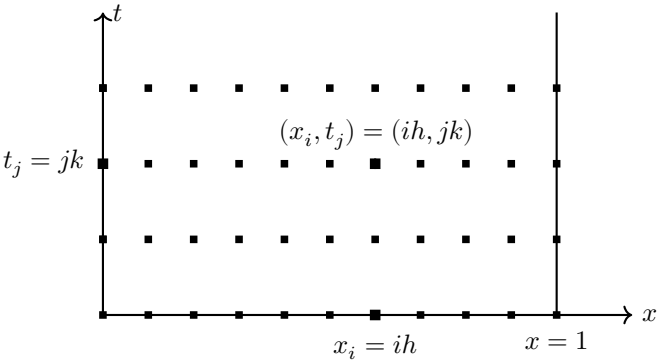
og

$$u(x,0) = f(x).$$

Metodene vi skal lage er basert på ting vi har gjort tidligere, og varmelikningen er kun en illustrasjon. Det er relativt lett å konstruere liknende metoder for andre likninger med de teknikkene vi skal gå gjennom for varmelikningen.

## Gitteret

Når vi skal løse en partiell differensiallikning, må vi holde styr på to gitre - et i $x$-retningen, og et i $t$-retningen. Vi skal finne approksimasjoner til løsningen $u$ i alle gitterpunktene. Vi gitrer opp intervallet $[0,1]$ på $x$-aksen med gitteravstanden $h$, og nummererer punktene slik at $x_0 = 0$ og $x_n = 1$. Den positive $t$-aksen gitrer vi opp med gitteravstanden $k$, og nummerer slik at $t_0 = 0$.



### Diskretisering i $x$

Vi setter først opp en tilnærming for $u_{xx}(x,t)$, basert på den andre ordens endelige differanseformelen for $x$:

$$u_{xx}(x,t) \approx \frac{u(x+h,t) - 2u(x,t) + u(x-h,t)}{h^2}$$

Man kan fint bruke høyere ordens differanseformler, men det skal ikke vi gjøre. Vi kan nå tenke at vi erstatter $u(x,t)$ med $n+1$ envariable funksjoner $u_i(t)$, som beskriver temperaturendringen i hvert sitt punkt $x_i$ på stangen:

$$u_{xx}(x_i,t) \approx \frac{u_{i+1}(t) - 2u_{ij}(t) + u_{i-1}(t)}{h^2},$$

Nå setter vi inn dette uttrykket i varmelikningen:

$$u_t(x_i,t) = u_{xx}(x_i,t) \approx \frac{u_{i+1}(t) - 2u_{ij}(t) + u_{i-1}(t)}{h^2},$$

Siden $u_i$ kun er en funksjon av $t$, passer det å skrive

$$u_i'(t) = \frac{u_{i+1}(t) - 2u_{ij}(t) + u_{i-1}(t)}{h^2},$$

som vi kjenner igjen som et system av ordinære differensiallikninger. Dette systemet har $n-1$ likninger, for i $u_0 = 0$ og $u_n = 0$ kjenner vi; disse er gitt av randbetingelsene.

### Diskretisering i $t$

Nå kan vi løse systemet med en ønsket metode fra forrige kapittel, og vi skal ta i betraktning tre valg. Etter diskretiseringen i $t$, skriver vi approksimasjonen i punktet $(x_i, t_j)$ som $u_{ij}$:

$$u(x_i,t_j) \approx u_{ij}$$

De tre metodene vi skal bruke, er eksplisitt Euler, implisitt Euler, og trapesmetoden. De korresponderende skjemaene for varmelikningen blir:

Eksplisitt:

$$\frac{u_{i,j+1} - u_{ij}}{k} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2}$$

Implisitt:

$$\frac{u_{i,j+1} - u_{ij}}{k} = \frac{u_{i+1,j+1} - 2u_{i,j+1} + u_{i-1,j+1}}{h^2}$$

Crank-Nicholson:

$$\frac{u_{i,j+1} - u_{ij}}{k} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{2h^2} + \frac{u_{i+1,j+1} - 2u_{i,j+1} + u_{i-1,j+1}}{2h^2}$$

**Teorem 13.7.** *Det explisitte og implisitte skjema er av orden 1 i $t$, og av orden 2 i $x$. Crank-Nicolson er av orden 2 i begge variable.*

Beviset er en ganske hårete variant av argumentet for ordenen til eksplisitt Euler i forrige kapittel. Hver av disse skjemaene har sine fordeler, og vi skal behandle dem i tur og orden.

#### Eksplisitt skjema

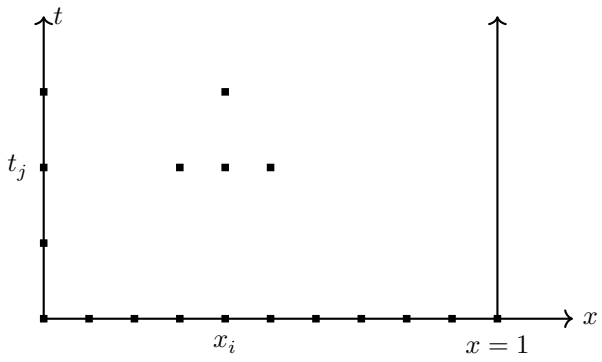
Det eksplisitte skjemaet

$$\frac{u_{i,j+1} - u_{ij}}{k} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2}$$

har egentlig bare en fordel: det er så lett å programmere opp. Vi løser for $u_{i,j+1}$, og får

$$u_{i,j+1} = u_{ij} + \frac{k}{h^2}\left(u_{i+1,j} - 2u_{i,j} + u_{i-1,j}\right)$$

Nå er det vanlig å tegne opp noe som kalles *stensilen*. Dette er en figur som illustrerer hvilke gitterpunkter som er involvert i ligningen man bruker for å beregne nye approksimasjoner.




Anta at du har beregnet $u_{ij}$ for alle $i$ og opp til og med en bestemt $j$. Vi kan da enkelt beregne $u_{i,j+1}$ for alle $i$ ved å bruke formelen.

For å analysere mer, må vi sette opp et matrise-vektorprodukt som beskriver iterasjonen. Vi definerer

$$\mathbf{u}_j = \begin{bmatrix} u_{1j} \\ u_{2j} \\ \vdots \\ u_{n-2,j} \\ u_{n-1,j} \end{bmatrix}$$

Det eksplisitte skjemaet kan skrives kompakt som

$$\mathbf{u}_{j+1} = (I - \frac{k}{h^2}A)\mathbf{u}_j$$

der

$$A = \begin{bmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & \ddots & \ddots & & \\ & & \ddots & \ddots & -1 & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{bmatrix}$$

Det kan vises at denne matrisen har $n-1$ forskjellige egenverdier gitt ved

$$\lambda_k = 4\sin^2\frac{k\pi}{2n} \quad \text{for} \quad 1 \le k \le n-1.$$

Egenverdiene til

$$I - \frac{k}{h^2}A$$

er dermed gitt ved

$$1 - 4\frac{k}{h^2}\sin^2\frac{k\pi}{2n}.$$

Dersom

$$\frac{k}{h^2} < \frac{1}{2}.$$

går dette veldig bra, for alle egenverdiene er mellom 0 og 1. Da vil

$$\mathbf{u}_{j+1} = (I - \frac{k}{h^2}A)\mathbf{u}_j$$

være en kontraksjon, og konvergere mot et fikspunkt. Dersom

$$\frac{k}{h^2} \geq \frac{1}{2}$$

vil egenverdiene sende $\mathbf{u}_i$ mot uendelig, og dette er ikke oppførsel vi ønsker fra et numerisk skjema for varmelikningen. Vi vet jo at temperaturen skal synke mot 0 i denne situasjonen!

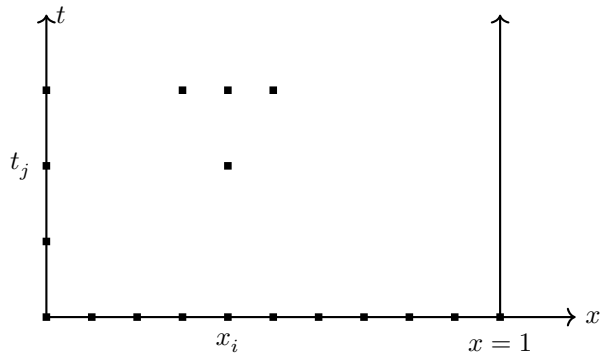
Dette er ganske restriktivt. Kanskje noen andre skjema er enklere å ha med å gjøre.

### Implisitt skjema

Problemet med det eksplisitte skjemaet er at det går til h... med mindre $k/h^2 < 1/2$, og dette gjør at du må ha veldig tett mellom punktene $t$-aksen. Akkurat som for ordinære differensialliknigner, kan vi bøte på dette med å bruke implisitt skjema. Det implisitte skjemaet skriver vi

$$\left(1 + 2\frac{k}{h^2}\right) u_{i,j+1} - \frac{k}{h^2}u_{i+1,j+1} - \frac{k}{h^2}u_{i-1,j+1} = u_{ij}.$$

Stensilen er




I hver tidssteg må vi løse et lineært likningssystem

$$(I + \frac{k}{h^2}A)\mathbf{u}_{j+1} = \mathbf{u}_j$$

der $A$ er den samme matrisen som i sted.

Egenverdiene til $(I + \frac{k}{h^2}A)$ er

$$1 + 4\frac{k}{h^2}\sin^2\frac{k\pi}{2n},$$

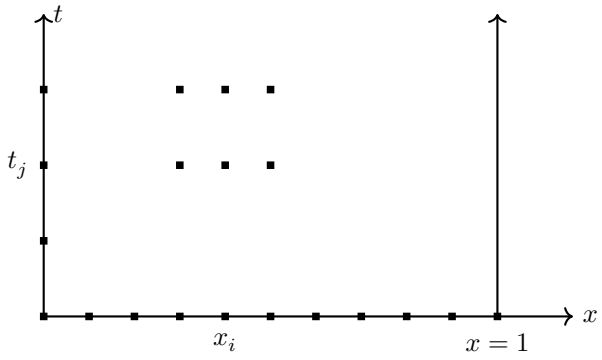
som åpenbart er større enn 1 uansett hva $\frac{k}{h^2}$ er. Dermed vil egenverdiene til $(I + \frac{k}{h^2}A)^{-1}$ alltid være mindre enn 1, og følgelig vil skjemaet konvergere mot 0 uansett.

### Crank-Nicolson

Crank-Nicolson skriver vi

$$2u_{i,j+1} + \frac{k}{h^2}\left(-u_{i+1,j+1} + 2u_{i,j+1} - u_{i-1,j+1}\right) = 2u_{ij} - \frac{k}{h^2}\left(-u_{i+1,j} + 2u_{i,j} - u_{i-1,j}\right).$$

Stensilen er




Også her blir det et lineært ligningssystem

$$(2I + \frac{k}{h^2}A)\mathbf{u}_{j+1} = (2I - \frac{k}{h^2}A)\mathbf{u}_j$$

å løse på hvert tidssteg. Det kan vises at Crank-Nicolson er en stabil metode for varmelikningen, men det dropper vi.

### Laplaces likning

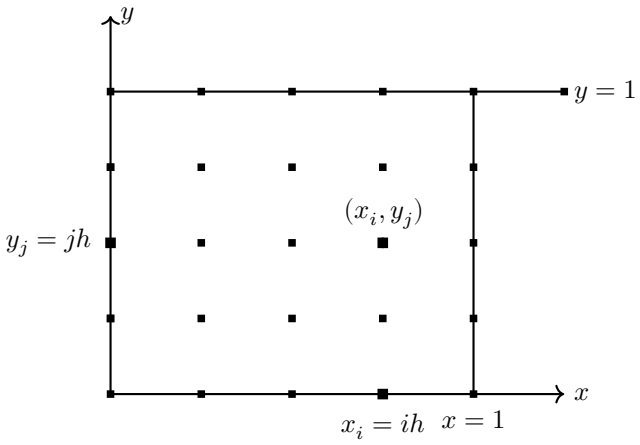
Vi kan også sette opp et numerisk skjema for Laplaces likning

$$u_{xx} + u_{yy} = 0,$$

på kvadratet $[0,1] \times [0,1]$ med randkrav

$$u(x,0) = u(0,y) = u(1,y) = 0 \qquad \text{og} \qquad u(x,1) = f(x).$$

Vi gitrer opp på følgende vis:




og bruker den kjente og kjære differanseformelen

$$u_{xx}(x,t) \approx \frac{u(x+h,t) - 2u(x,t) + u(x-h,t)}{h^2}$$

i både $x$ og $y$, slik at vi får skjemaet

$$4u_{i,j} - u_{i+1,j} - u_{i-1,j} - u_{i,j+1} - u_{i,j-1} = 0.$$

Nå er $u$ kjent på sidekantene til kvadratet, slik at det er kun de indre gitterpunktene som gir opphav til likninger og ukjente. La oss si at det er $n-1$ indre gitterpunkter i hver koordinatretning, slik at antall likninger og ukjente blir $(n-1)^2$. Vi organiserer $u_{ij}$ i en vektor på formen

$$\mathbf{u} = \begin{bmatrix} u_{11} \\ \vdots \\ u_{n-1,1} \\ u_{1,2} \\ \vdots \\ u_{n-1,2} \\ u_{1,3} \\ \vdots \\ u_{n-1,n-1} \end{bmatrix}$$

og får følgelig et likningssystem på formen

$$A\mathbf{u} = \mathbf{f}$$

der

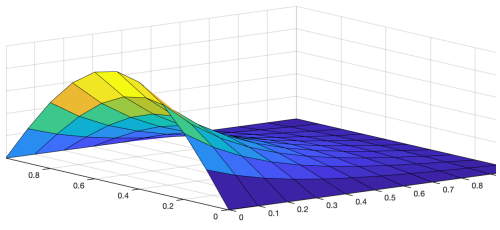
$$B = \begin{bmatrix} 4 & -1 & & & & \\ -1 & 4 & -1 & & & \\ & -1 & \ddots & \ddots & & \\ & & \ddots & \ddots & -1 & \\ & & & -1 & 4 & -1 \\ & & & & -1 & 4 \end{bmatrix}$$

og

$$A = \begin{bmatrix} B & 0 & & & & \\ 0 & B & 0 & & & \\ & 0 & \ddots & \ddots & & \\ & & \ddots & \ddots & 0 & \\ & & & 0 & B & 0 \\ & & & & 0 & B \end{bmatrix} + \begin{bmatrix} 0 & \cdots & 0 & -1 & & 0 \\ \vdots & 0 & & & \ddots & \\ 0 & & \ddots & \ddots & & -1 \\ -1 & & \ddots & \ddots & & 0 \\ & \ddots & & & \ddots & \vdots \\ 0 & & -1 & 0 & \cdots & 0 \end{bmatrix}$$

Den første matrisen i uttrykket for $A$ er en såkalt blokkdiagonal matrise, og den andre matrisen har to sammenhengende super- og subdiagonaler med -1 i alle komponenter, der superdiagonalen begynner i rad $n$, og subdiagonalen begynner i kolonne $n$.

Under er et plot av en løsning der $f(x) = \sin \pi x$.



## Varmelikningen på en plate

Hvis man først har forstått hvordan laplaces likning diskretiseres, er det ikke så vanskelig å forstå hvordan man diskretiserer varmelikningen i to romlige dimensjoner. Temperaturen i platen er gitt ved funksjonen $u(x, y, t)$, men nå tillater vi at temperaturen varierer med tiden, og problemet vi skal løse, er

$$u_t = u_{xx} + u_{yy}$$

med randkrav

$$u(0, y, t) = u(1, y, t) = u(x, 0, t) = 0$$
$$u(x, 1, t) = f(x)$$

og initialkrav

$$u(x, y, 0) = g(x, y).$$

Gitter i $x$ og $y$ blir det samme som for laplaces likning, mens $k$ er tidssteget. Vi lar telleren i tid være $m$. Vi diskretiserer med de samme differanseformlene som før, og får eksplisitt skjema:

$$\frac{u_{i,j,m+1} - u_{i,j,m}}{k} = \frac{u_{i+1,j,m} - 2u_{i,j,m} + u_{i-1,j,m}}{h^2} + \frac{u_{i,j+1,m} - 2u_{i,j,m} + u_{i,j-1,m}}{h^2}$$

implisitt skjema:

$$\frac{u_{i,j,m+1} - u_{i,j,m}}{k} = \frac{u_{i+1,j,m+1} - 2u_{i,j,m+1} + u_{i-1,j,m+1}}{h^2} + \frac{u_{i,j+1,m+1} - 2u_{i,j,m+1} + u_{i,j}}{h^2}$$

og Crank-Nicolson

$$\frac{u_{i,j,m+1} - u_{i,j,m}}{k} = \frac{u_{i+1,j,m} - 2u_{i,j,m} + u_{i-1,j,m}}{h^2} + \frac{u_{i,j+1,m} - 2u_{i,j,m} + u_{i,j-1,m}}{h^2} + \frac{u_{i+1,j,m+1} - 2u_{i,j,m+1} + u_{i-1,j,m+1}}{h^2} + \frac{u_{i,j+1,m+1} - 2u_{i,j,m+1} + u_{i,j}}{h^2}$$

Det går an å skrive disse skjemaene om til matrise-vektor-iterasjoner på samme vis som i det endimensjonale tilfellet. For eksplisitt skjema får vi

$$\mathbf{u}_{m+1} = \left(I - \frac{k}{h^2}A\right)\mathbf{u}_m + \frac{k}{h^2}\mathbf{f}$$

der $A$ er matrisen fra diskretiseringen av Laplaces likning, og

$$\mathbf{f} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ f(x_1) \\ \vdots \\ f(x_{n-1}) \end{bmatrix}$$

Likninger for implisitt skjema og Crank-Nicolson blir henholdsvis

$$\left(I + \frac{k}{h^2}A\right)\mathbf{u}_{m+1} = \mathbf{u}_m + \frac{k}{h^2}\mathbf{f}$$

og

$$\left(I + \frac{k}{2h^2}A\right)\mathbf{u}_{m+1} = \left(\left(I - \frac{k}{2h^2}A\right)\mathbf{u}_m + \frac{k}{2h^2}\mathbf{f}\right) + \frac{k}{2h^2}\mathbf{f}.$$

## Kapittel 14

# Vektorkalkulus

## Kapittel 15

# Funksjoner fra $\mathbb{C}$ til $\mathbb{C}$

## Kapittel 16

# Partielle differensiallikninger II

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}$$
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$
$$\nabla \cdot \mathbf{B} = 0$$
$$\nabla \times \mathbf{B} = \frac{\mathbf{j}}{\epsilon_0} + \frac{\partial \mathbf{E}}{\partial t}$$