12. august 2025

MAT 1080

Matematisk grunnlag for maskinlæring

Geir Dahl og Arne B. Sletsjøe

Høsten 2025

Innhold

Forord/Introduksjon **iii**

1 Funksjoner **1**
1.1 Funksjonsbegrepet 1
1.2 Tallfølger 7
1.3 Aritmetiske, alternerende og geometriske følger 13
1.4 Funksjoner i flere variable 14
1.5 Nivåmengder 15
1.6 Ekstremalpunkter 17
1.7 Algoritmer 19
1.8 Oppgaver 21

2 Derivasjon **23**
2.1 Derivasjon av funksjoner i én variabel 23
2.2 Ekstremalpunkter 31
2.3 Kritiske punkter 32
2.4 Globale ekstremalpunkter 37
2.5 Newtons metode 40
2.6 Krumning og vendepunkter 42
2.7 L'Hôpitals regel 46
2.8 Partiell Derivasjon 49
2.9 Kritiske punkter 52
2.10 Globale ekstremalpunkter 55
2.11 Minste kvadraters metode 56
2.12 Oppgaver 59

3 Integrasjon **68**
3.1 Bruk av integrasjon 68
3.2 Ubestemt integrasjon 69
3.3 Bestemt integrasjon 73
3.4 Fundamentalteoremet 74

3.5 Riemannsummer 80
3.6 Trapesmetoden 83
3.7 Oppgaver 86

4 Vektorer, matriser og lineære likninger 89
4.1 Vektorer 89
4.2 Matriser 95
4.3 Kvadratiske matriser 100
4.4 Likninger og determinanter 107
4.5 Lineære likningssystemer 114
4.6 Spektralteori 118
4.7 Lineære rom 122
4.8 Lineære avbildninger 131
4.9 Oppgaver 142

5 Matriser og konveksitet 152
5.1 Kvadratiske funksjoner og kvadratiske former 152
5.2 Klassifikasjon av matriser og kvadratiske former 156
5.3 Konveksitet 158
5.4 Konvekse funksjoner 163
5.5∗ Variabelskifte i kvadratiske former 170
5.6 Oppgaver 173

6 Bibliografi 181

Forord/Introduksjon

Dette kompendiet er skrevet som læremateriale til emnet

MAT1080 – Matematisk grunnlag for maskinlæring

for bruk i det nye bachelorprogrammet

Informatikk: maskinlæring og kunstig intelligens.

MAT1080 er obligatorisk og taes i første semester.

Emnet har vært planlagt over lengre tid, og gode kolleger ved IFI har kommet med en rekke ønsker og gode innspill til innholdet i emnet. Takk til IFI-kolleger

- **Stephan Oepen, Ole Christian Lingjærde, Andreas Austeng, Geir Kjetil Sandve, Ragnhild Kobro Runde, Jan Tore Lønning og Anne Schistad Solberg**

for samtaler og innspill om MAT1080 og studieprogrammet i sin helhet.

Hovedtanken er å gi studentene et godt grunnlag i matematikk for å lære seg maskinlæring (ML) og KI. Det innebærer at stoffutvalget er tilpasset dette behovet, og at det derfor er noe annerledes enn i enkelte andre av våre grunnemner i matematikk. Videre er en forutsetning at dette emnet bygger på R1 (ikke R2) fra videregående skole.

I arbeidet med dette kompendiet tok vi et utgangspunkt i kompendiene [3, 4] skrevet av Arne B.S., og gjorde utvalg, endringer, tilføyelser etc. Noe stoff ble hentet fra notater fra MAT1120 skrevet av Erik Bedos og Geir D. Takk til Tom Lindstrøm for gode og viktige kommentarer til innhold og struktur på kompendiet. Takk til Tom også for å stille opp som foreleser i dette emnet høsten 2025; da er dette i trygge hender! En stor takk til masterstudent Oscar Brovold som har lest korrektur og laget en rekke oppgaver i heftet. Oscar spesialiserer seg i maskinlæring, og en del av oppgavene han laget knytter matematikken i heftet godt opp mot bruk i maskinlæring.

Filosofien er å dekke flere grunnleggende temaer rimelig bra, med vekt på idéer, forståelse og bruk, og dermed ikke være så detaljorientert som vi er i

andre matematikk emner. Dette innebærer at vi bare av og til går inn på bevis, og at vi ikke kan dekke alle tradisjonelle temaer så godt. F.eks. ut fra behov i ML og KI vil vi ikke gjøre så mye ut av integrasjonsteori som man ser andre steder. Derimot er det rimelig grundig behandling av funksjoner, deriverte, optimering og lineær algebra; alle disse feltene danner et matematisk grunnlag for ML. Under gjengir vi emnebeskrivelsen for MAT1080.

Vi ser for oss at forelesningene dekker det aller meste av innholdet her, men at man gjør noen tilpasninger, og høster erfaringer for fremtidige forbedringer.

Noen viktige ord til slutt:

Til hver student: Velkommen til MAT1080!!

- **Vi gjør det vi kan for å gi deg et best mulig undervisningstilbud!**
- **Kom på forelesning og øvingstimer, og jobb godt med stoffet!**
- **Da får du et flott grunnlag i dine ML og KI studier, og kunnskaper du kan ha glede av hele livet!**
- **Du trenger ikke bruke KI for å lære matematikk. Men du trenger matematikk for å lære KI!**

Vår 2025
Geir Dahl og Arne B. Sletsjøe

****

Emnebeskrivelsen for MAT1080:

Kort om emnet:

Emnet gir det matematiske grunnlaget for studier innen maskinlæring. Det tar for seg elementær funksjonslære og derivasjon i en, så vel som flere variable. I emnet vil det også bli gitt en innføring i grunnleggende begreper og resultater innen lineær algebra, med vekt på vektor- og matriseregning. Emnet bygger på R1 eller S1 + S2 fra videregående skole.

Hva lærer du?

Etter å ha fullført emnet:

- har du god innsikt i det matematiske funksjonsbegrepet i en eller flere variable, og kjenner til de vanlige elementære funksjonene, som polynomer, rasjonale funksjoner, trigonometriske funksjoner og eksponensialfunksjoner
- har du forståelse av grensebegrepet og du vet hva en kontinuerlig funksjon er
- kjenner du til begrepet algoritme og hvordan det kan brukes til å løse regneproblemer
- har du god innsikt i og gode regneferdigheter innen derivasjon, i en og flere variable. Du kjenner også til tolkninger av den deriverte inne ulike fagområder og hva som ligger i begrepet anti-derivasjon
- har du kjennskap til anvendelser av differensialregningen, spesielt rettet inn mot optimering og anvendelser innen maskinlæring
- kjenner du til ulike metoder for å løse lineære likningssystemer
- har du god innsikt i matrisebegrepet og du kan regne med matriser og vektorer
- vet du hva underrom og lineær uavhengighet er, og du kjenner til naturlige underrom knyttet til matriser
- har du innsikt i begrepene basis og dimensjon i vektorrom
- kjenner du til begrepene egenverdier og egenvektorer og relevante anvendelser av disse

Kapittel 1

Funksjoner

1.1 Funksjonsbegrepet

En **funksjon** er en regel som til et reelt tall tilordner et annet reelt tall etter en bestemt oppskrift. Denne oppskriften kan være veldig presist formulert, eller den kan være gitt mer vagt. Oppskriften kan i store trekk være formulert på en eller flere av følgende fire måter:

- Verbal beskrivelse
- Numerisk beskrivelse, tabellform
- Visuelt ved en graf
- Analytisk ved en formel

Noen funksjoner kan beskrives på alle fire måter, mens andre mangler beskrivelse på noen av de fire.

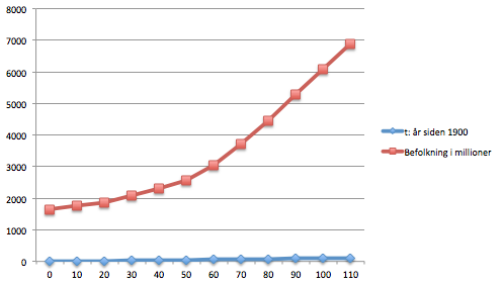
Eksempel 1.1.1. *Arealet av en sirkel som funksjon av radius i sirkelen, er analytisk gitt ved $A(r) = \pi r^2$, der r er radius. Den verbale versjonen er at arealet av en sirkel vokser proporsjonalt med kvadratet av radius, og hvor vi kaller forholdet for π*

$$\frac{A(r)}{r^2} = \pi.$$

Det har som regel liten hensikt å beskrive denne funksjonen med en graf eller tabell.

Eksempel 1.1.2. *Verbal beskrivelse: La $P(t)$ være verdens befolkning ved tidspunktet t.*

t: år siden 1900	Befolkning i millioner
0	1650
10	1750
20	1860
30	2070
40	2300
50	2560
60	3040
70	3710
80	4450
90	5280
100	6080
110	6870



Figur 1.1: Befolkningsdata.

Fig. 1.1 over viser utdrag av funksjonen i tabellform, og en graf hvor vi har trukket rette linjer mellom datapunktene. Det viser seg at funksjonen kan approksimeres relativt godt ved

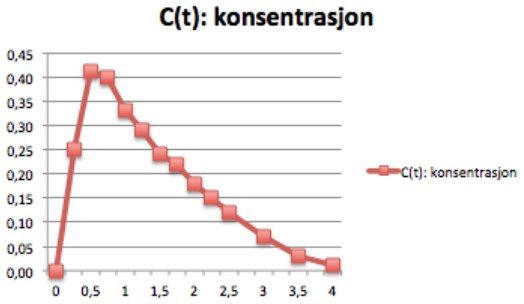
$$P(t) \approx f(t) = (1.43653 \times 10^9) \cdot (1.01395)^t$$

Dette er et eksempel på bruk av alle fire måtene å beskrive funksjoner.

Eksempel 1.1.3. *Verbal beskrivelse: La $C(t)$ være konsentrasjonen av alkohol i blodet ved tiden t ved et inntak av alkohol ved tiden $t = 0$.*

Figurene under viser utdrag av funksjonen i tabellform, og en graf hvor vi har trukket rette linjer mellom datapunktene. Her er det naturlig å hoppe over den analytiske beskrivelsen, siden det er vanskelig å peke på noen funksjon som har akkurat den gitte grafen.

t: timer	C(t): konsentrasjon
0	0,00
0,25	0,25
0,5	0,41
0,75	0,40
1	0,33
1,25	0,29
1,5	0,24
1,75	0,22
2	0,18
2,25	0,15
2,5	0,12
2,75	
3	0,07
3,25	
3,5	0,03
3,75	
4	0,01



Figur 1.2: Alkoholkonsentrasjon.

Noen viktige begreper knyttet til funksjoner:

- **Voksende**: En funksjon f er voksende dersom $f(x_1) \leq f(x_2)$ når $x_1 < x_2$. Funksjonen $P(t)$ i Eksempel 2 er voksende.

- **Avtagende:** Motsatt av voksende. Funksjonen $C(t)$ i Eksempel 3 er avtagende fra $t = 0.5$ og oppover.

- **Periodisk:** En funksjon som gjentar seg i det uendelige, $f(t+P) = f(t)$ for alle t. Tallet P kalles perioden til funksjonen. Typiske eksempler er de trigonometriske funksjonene $\cos t$ og $\sin t$. Disse har begge periode 2π, dvs. $\cos(x + 2\pi) = \cos x$.

- **Symmetri:** Funksjoner der hvor grafen har en form for symmetri. F.eks. vil funksjoner som oppfyller $f(-x) = f(x)$ være symmetriske om y-aksen, mens funksjoner som oppfyller $f(-x) = -f(x)$ vil være symmetriske om origo. Grafen til en funksjon som er symmetrisk om y-aksen vil ha to deler som er speiling av hverandre, på hver sin side av y-aksen.

- **Argument/definisjonsområde:** Det vi putter inn i funksjonen.

- **Ekstremalpunkter:** Punkter der funksjonen enten har mindre verdi eller har større verdi enn alle nærliggende punkter. Begrepet omfatter både maksimums- og minimums-punkter. Funksjonen $C(t)$ i Eksempel 3 har et maksimumspunkt for $t = 0.5$, med maksimumsverdi $C(0.5) = 0.41$.

- **Kontinuerlige funksjoner:** Funksjoner hvor grafen er sammenhengende. De fleste vanlige funksjoner som vi kommer bort i er kontinuerlige. Dette skyldes at veldig mange av naturens prosesser, og andre sammenhenger, er kontinuerlige.

Det finnes mange forskjellige funksjoner. Her er en oversikt over noen vanlige, med beskrivelser og bruksområder.

Polynomer:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$$

hvor n er et helt tall og $a_0, a_1, \ldots, a_n$ er reelle tall som kalles polynomets koeffisienter. Polynomer utgjør de viktigste byggestenene for mer generelle funksjoner. Deres viktigste egenskap er at de lar seg beregne; polynomer reflekterer kun de vanlige regneartene. F.eks. kan vi regne ut verdien $f(2)$ for funksjonen $f(x) = 2x^3 - x - 5$ ved å sette inn

$$f(2) = 2 \cdot 2^3 - 2 - 5 = 16 - 2 - 5 = 9.$$

Rasjonale funksjoner:

Rasjonale funksjoner er brøker av polynomer laget på samme måte som man lager rasjonale tall fra de hele tallene. Et eksempel på en rasjonal funksjon er

$$f(x) = \frac{x^2 + 1}{2x - 3}.$$

Rasjonale funksjoner er i prinsippet like beregnbare som polynomer, siden det eneste ekstra som introduseres er divisjon. Vi kan regne ut verdien i et vilkårlig punkt;

$$f(\tfrac{1}{2}) = \frac{(\frac{1}{2})^2 + 1}{2 \cdot \frac{1}{2} - 3} = \frac{\frac{5}{4}}{-2} = -\frac{5}{8}.$$

Eksponensialfunksjoner:

Eksponensialfunksjoner er funksjoner av typen $f(x) = a^x$ for et reelt **grunntall** a. Eksempel med $a = 2$: $f(x) = 2^x$ og da er f.eks. $f(5) = 2^5 = 32$. Den "mest perfekte" eksponensialfunksjonen har grunntall

$$e = 2.718281828459045\ldots.$$

Eksponensialfunksjoner er grunnpilaren i alle vekstmodeller. For en populasjon med konstant vekst vil antallet individer være gitt ved en eksponensialfunksjon. Vi har

$$P(t) = P_0 e^{kt}$$

der P_0 er størrelsen på populasjonen ved tiden $t = 0$, og k er en konstant som bestemmes av vekstraten til populasjonen.

Trigonometriske funksjoner:

Trigonometriske funksjoner er funksjoner med opphav i cosinus og sinus, dvs. forholdstallene mellom kateter og hypotenus i en rettvinklet trekant, som funksjon av en vinkel. Trigonometriske funksjoner er periodiske og de er velegnet til å beskrive bølger, svingninger eller andre periodiske størrelser.

Rotfunksjoner:

Rotfunksjoner er funksjoner av typen $f(x) = \sqrt{x}$, eller kombinsjoner av slike med f.eks. polynomer. Funksjonen $f(x) = \sqrt{x}$ tilordner til et positivt reelt tall x et annet positivt reelt tall $\sqrt{x}$ som er slik at $\sqrt{x}^2 = x$. Tilsvarende uttrykker $\sqrt[n]{x}$, eller også skrevet $x^{\frac{1}{n}}$ et tall som ganget med seg selv n ganger gir oss x tilbake.

Eksempel 1.1.4. *Vi betrakter et rektangel med sidekanter x og y. Omkretsen av rektangelet er $O = 2x + 2y$, og arealet er xy. Anta at vi har bestemt oss for*

at rektangelet skal ha areal A, dvs. $xy = A$. Det betyr at dersom lengden x av den ene sidekanten er gitt, så vil den andre være bestemt ved $y = \frac{A}{x}$. Det gir oss omkretsen av rektangelet gitt som en funksjon av den variable sidekanten x:

$$O = O(x) = 2x + 2\frac{A}{x}.$$

I dette tilfellet er det klart at vi har en begrensning på hvilke x vi kan bruke. Hvis det skal ha mening å snakke om x som sidekanten i et rektangel, må vi nødvendigvis ha $x > 0$. Vi sier at funksjonen O har definisjonsmengde gitt ved $x > 0$, eller skrevet på en av formene

$$\mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\} = \langle 0, \infty \rangle.$$

Definisjonsområdet til en funksjon kan være gitt ved intervaller på tallinjen. Vi bruker notasjonen

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$$

for **lukkede** intervaller, dvs. intervaller som inneholder sine endepunkter,

$$\langle a, b \rangle = \{x \in \mathbb{R} : a < x < b\}$$

for **åpne** intervaller (inneholder ingen av sine endepunkter) og

$$\langle a, b] = \{x \in \mathbb{R} : a < x \leq b\}$$
$$[a, b\rangle = \{x \in \mathbb{R} : a \leq x < b\}$$

for **halvåpne** intervaller. Merk at $\mathbb{R}$ står for de reelle tallene.

Den formelle definisjonen av en funksjon er som følger:

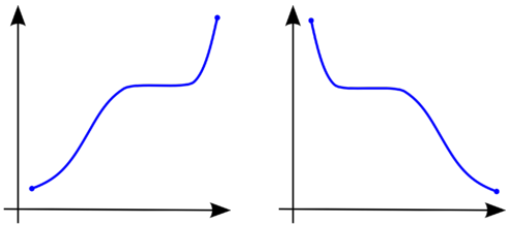
Definisjon 1.1.1. *La $A \subset \mathbb{R}$ være en delmengde av de reelle tallene. En funksjon $f : A \to \mathbb{R}$ er en regel som til et hvert element $x \in A$ tilordner nøyaktig ett element $f(x) \in \mathbb{R}$. Mengden A kalles **definisjonsområdet** til funksjonen.*

Vi kan dele egenskapene til en funksjon i to grupper, lokale og globale egenskaper.

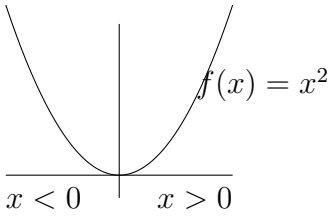
Lokale egenskaper til en funksjon er egenskaper som observeres i små intervaller om punkter i definisjonsområdet til funksjonen. Monotoni-egenskapene til en funksjon, dvs. hvor den vokser og hvor den avtar, er en lokal egenskap, det samme gjelder krumning. Kontinuitet, deriverbarhet og maksimums- og minimumspunkter er også lokale egenskaper. Vi skal komme tilbake til disse begrepene etter hvert.

Definisjon 1.1.2. *En funksjon $y = f(x)$ er* ***(strengt) voksende*** *på et intervall I dersom $f(x_1) \leq f(x_2)$ $(f(x_1) < f(x_2))$ for alle $x_1 < x_2$ i intervallet. Funksjonen er* ***(strengt) avtagende*** *dersom $f(x_1) \geq f(x_2)$ $(f(x_1) > f(x_2))$ for alle $x_1 < x_2$ i intervallet.*

Forskjellen på voksende og strengt voksende funksjoner er at voksende funksjoner kan ha områder hvor de er konstante. Det tillater vi ikke for strengt voksende funksjoner.



Figur 1.3: Grafene til to monotone funksjoner, voksende til venstre og avtagende til høyre.

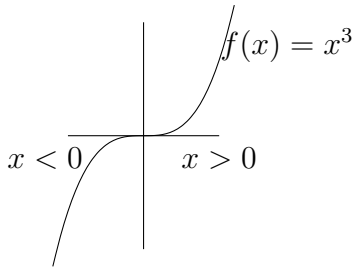


Figur 1.4: Kvadratfunksjon.

Eksempel 1.1.5. *Funksjonen $f(x) = x^2$ er strengt voksende på intervallet $[0, \infty\rangle$, siden $x_1^2 < x_2^2$ når $x_1 < x_2$. Tilsvarende er funksjonen strengt avtagende på intervallet $\langle -\infty, 0]$ siden vi har at $x_1^2 > x_2^2$ for to negative tall x_1 og x_2 hvor $x_1 < x_2$.*

Eksempel 1.1.6. *Funksjonen $f(x) = x^3$ er strengt voksende overalt, siden $x_1^3 < x_2^3$ når $x_1 < x_2$.*

I motsetning til de lokale egenskapene som beskriver hvordan funksjonen ser ut i et lite intervall, karakteriserer de globale egenskapene funksjonen mer som en helhet. Et god illustrasjon er en funksjon som til et hvert punkt på



Figur 1.5: Kubisk funksjon.

Norgeskartet beskriver høyden over havet. Alle små og store fjelltopper vil være lokale maksimumsverdier, siden det ikke er noen steder i nærheten som er høyere. Den globale maksimumsverdien finner vi kun ett sted, nemlig på toppen av Galdhøpiggen. I et senere kapittel skal vi se nærmere på de globale egenskapene.

1.2 Tallfølger

Definisjon 1.2.1. *En* ***tallfølge*** $\{a_n\}$ *er en funksjon som til ethvert naturlig tall (indeksen n) tilordner et reelt tall* a_n*. Alternativt kan vi skrive tallfølgen som en sekvens av tall, ordnet etter indeksen:*

$$a_1, a_2, a_3, \ldots$$

Eksempel 1.2.1. *Et eksempel på en tallfølge er*

$$1, 2, 3, 4, 5, 6, \ldots$$

Et annet eksempel er

$$7, 4, 2, 5, 3, 3, 9, 2, 5, 23, 289, 14, \ldots$$

I det første tilfellet er det lett å tenke seg hva det neste (og neste deretter) leddet i følgen er, mens i det andre tilfellet er det ingen åpenbar systematikk.

Hvis vi har gitt en tallfølge eller en funksjon som beskriver et naturlig forløp, er det interessant å se hva som skjer med systemet i det lange løp. I det lange løp betyr for en tallfølge $\{x_n\}$ at vi forsøker å finne hvordan følgen oppfører seg for store n, det vil si når $n \to \infty$.

Som et første eksempel kan vi se på følgen

$$1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \tfrac{1}{5}, \ldots$$

Siden leddene blir mindre og mindre og vi kan få leddene så små vi bare vil ved å velge stor nok nevner i brøken, er det rimelig å si at følgen går mot 0. Vi skriver dette som

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

En følge som nærmer seg en grense på denne måten sies å være **konvergent**. Det motsatte av konvergens er **divergens**. Det er mange måter for en følge å være divergent. Her er tre forskjellige eksempler på divergente følger:

$$2, 3, 5, 7, 11, 13, \ldots$$
$$1, -1, -1, 1, -1, -1, 1, -1, -1, \ldots$$
$$3, -4, 7, 2, 1, 0, -3, 7, -3, 5, 4, 0, \ldots$$

Den første følgen divergerer fordi leddene bare vokser og vokser uten noen øvre grense (primtallene), den andre følgen har en viss systematikk, men fordi den hopper mellom to verdier er den ikke konvergent. Den har imidlertid konvergente **delfølger**. Den siste følgen er jo bare rotete.

Definisjon 1.2.2. *En følge $\{x_n\}$ sies å ha **grenseverdien** (eller bare **grense**) L og vi skriver*

$$\lim_{n\to\infty} x_n = L \quad \textit{eller} \quad x_n \to L \ \textit{ når } \ n \to \infty$$

*dersom differansen mellom x_n og L blir så liten vi bare vil bare n er stor nok. Dersom en grenseverdi eksisterer sier vi at følgen **konvergerer** (eller er **konvergent**). I motsatt fall sier vi at følgen **divergerer** (eller er **divergent**).*

I det ene tilfellet av divergens, der følgen bare vokser og vokser, bruker vi notasjonen

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

For eksempel vil vi ha

$$\lim_{n\to\infty} n^p = \infty$$

for alle $p > 0$. For negative eksponenter har vi konvergens;

$$\lim_{n\to\infty} \frac{1}{n^p} = 0.$$

Grenseverdier oppfyller de fleste vanlige regnereglene, slik som at grenseverdien av en sum er summen av grenseverdiene, grenseverdien av et produkt er produktet av grenseverdiene, osv.

Eksempel 1.2.2. *Vi skal regne ut*

$$\lim_{n\to\infty} \frac{n^2+3n-2}{2n^2+8n+12}.$$

Vi deler teller og nevner med høyeste potens av n, i dette tilfellet 2;

$$\lim_{n\to\infty} \frac{n^2+3n-2}{2n^2+8n+12} = \lim_{n\to\infty} \frac{1+\frac{3}{n}-\frac{2}{n^2}}{2+\frac{8}{n}+\frac{12}{n^2}}.$$

Deretter bruker vi regnereglene for grenseverdier;

$$\begin{aligned}\lim_{n\to\infty} \frac{n^2+3n-2}{2n^2+8n+12} &= \lim_{n\to\infty} \frac{1+\frac{3}{n}-\frac{2}{n^2}}{2+\frac{8}{n}+\frac{12}{n^2}} \\ &= \frac{\lim\limits_{n\to\infty} 1+\lim\limits_{n\to\infty}\frac{3}{n}-\lim\limits_{n\to\infty}\frac{2}{n^2}}{\lim\limits_{n\to\infty} 2+\lim\limits_{n\to\infty}\frac{8}{n}+\lim\limits_{n\to\infty}\frac{12}{n^2}} \\ &= \frac{1+0-0}{2+0+0} = \frac{1}{2}.\end{aligned}$$

Denne definisjonen sier at dersom en følge konvergerer mot en grenseverdi L, så vil avstanden mellom L og leddene i følgen bli så liten vi bare vil bare vi går langt nok ut i følgen.

Eksempel 1.2.3. *Følgen*

$$0.3, 0.33, 0.333, \ldots$$

konvergerer mot $\frac{1}{3}$*, mens følgene*

$$1, 2, 3, 4, 5, \ldots$$

og

$$1, -1, 1, -1, 1, -1, \ldots$$

begge divergerer, den første fordi den vokser over alle grenser, mens den andre alternerer mellom to ulike verdier.

Følgen

$$0.3, 0.33, 0.333, \ldots$$

som vi kan skrive

$$a_n = \frac{3}{10} + \frac{3}{100} + \cdots + \frac{3}{10^n}$$

avviker fra grenseverdien $\frac{1}{3}$ med mindre enn $\frac{1}{10^m}$ fra ledd nummer $m+1$ og utover.

Til nå har vi sett på grenseverdier for tallfølger. Vi kan med små modifikasjoner bytte ut tallfølger med funksjoner.

Definisjon 1.2.3. *La f være en funksjon definert på et intervall $\langle a, \infty \rangle$. Da betyr*

$$\lim_{t\to\infty} f(t) = L$$

at verdien av $f(t)$ kan komme så nær L vi bare vil, bare t er stor nok.

Tilsvarende som for tallfølgene har vi også for funksjoner følgende to resultater:

$$\lim_{t\to\infty} t^p = \infty$$

for alle $p > 0$. For negative eksponenter har vi konvergens;

$$\lim_{t\to\infty} \frac{1}{t^p} = 0.$$

Dette kan vi bruke til å regne ut mer sammensatte grenseverdier.

Eksempel 1.2.4. *Vi skal regne ut*

$$\lim_{t\to\infty} \frac{8t}{1+4t^2}.$$

Vi deler med høyeste potens av t og får

$$\lim_{t\to\infty} \frac{8t}{1+4t^2} = \lim_{t\to\infty} \frac{\frac{8}{t}}{\frac{1}{t^2}+4} = \frac{0}{4} = 0.$$

Eksempel 1.2.5. *En modell for bakterievekst er gitt som en funksjon av konsentrasjonen av f.eks. glykose i omgivelsene. Modellen er gitt ved*

$$R(N) = \frac{SN}{c+N}$$

der N er konsentrasjonen av glykose, R er reproduksjonsraten til bakterien og S og c er positive konstanter. I dette eksemplet vil vi ha

$$\lim_{N\to\infty} R(N) = \lim_{N\to\infty} \frac{SN}{c+N} = \lim_{N\to\infty} \frac{S}{\frac{c}{N}+1} = S.$$

Et par andre grenseverdier det kan være nyttig å huske på, er eksponensialfunksjonene, med henholdsvis positiv og negativ eksponent.

Vi har

$$\lim_{t\to\infty} e^t = \infty,$$
$$\lim_{t\to-\infty} e^t = 0.$$

Den siste formuleringen er ekvivalent med

$$\lim_{t\to\infty} e^{-t} = 0.$$

Eksempel 1.2.6. *I en biologisk modell er størrelsen P gitt som en funksjon av tiden ved*

$$P(t) = \frac{64}{1+31e^{-0.7944t}}.$$

Den initiale verdien av modellen er

$$P(0) = \frac{64}{1+31e^{-0.7944\cdot 0}} = \frac{64}{1+31e^0} = 2$$

og den stabile verdien i det lange løp er

$$\lim_{t\to\infty} P(t) = \lim_{t\to\infty} \frac{64}{1+31e^{-0.7944t}} = \frac{64}{1+0} = 64.$$

Et eksempel på anvendelse av grenseverdier er begrepet fart. Fart er noen som måles på et bestemt tidspunkt, men som ikke kan beregnes uten at vi vet hva som foregikk rett før og rett etter tidspunktet. Vi beregner fart på et tidspunkt ved å måle gjennomsnittsfarten over et lite tidsrom som inneholder det gitte tidspunktet. Denne størrelsen betrakter vi som en funksjon, f.eks. på bredden av det lille intervallet. Så lar vi denne størrelsen gå mot 0, og grenseverdien vil være farten.

Vi har hittil sett på grenseverdier når argumentet går mot ∞. I eksemplet over avvek vi fra dette ved at vi lot argumentet gå mot 0. Vi kan formalisere dette.

Definisjon 1.2.4. *La $f(x)$ være en funksjon som er veldefinert i nærheten av et punkt $x = a$. Da skriver vi*

$$\lim_{x\to a} f(x) = L, \quad \textit{eller} \quad f(x) \to L \ \textit{ når } x \to a,$$

når vi mener at funksjonsverdien $f(x)$ kan komme så nær L vi bare vil, bare x er nær nok a.

Eksempel 1.2.7. *Vi skal regne ut verdien av*

$$\lim_{x\to 1} \frac{x-1}{x^2-1}.$$

Det er nærliggende å bare sette inn $x = 1$, men det gir $\frac{0}{0}$ som ikke sier oss noen ting. Imidlertid kan vi faktorisere nevneren: $x^2 - 1 = (x-1)(x+1)$. Dermed kan vi forkorte bort leddet $x - 1$ og vi sitter igjen med

$$\lim_{x\to 1} \frac{x-1}{x^2-1} = \lim_{x\to 1} \frac{x-1}{(x-1)(x+1)} = \lim_{x\to 1} \frac{1}{x+1} = \frac{1}{2}.$$

Grenseverdier som involverer ∞ er opphavet til et begrep; **asymptoter**. Det finnes minst to typer asymptoter. Dersom $f(x)$ er en funksjon som oppfyller $\lim_{x\to\infty} f(x) = L$ eller $\lim_{x\to-\infty} f(x) = L$, så sier vi at $y = L$ er en **horisontal** asymptote. Dersom $\lim_{x\to a} f(x) = \pm\infty$, sier vi at $x = a$ er en **vertikal** asymptote.

Eksempel 1.2.8. *Funksjonen*

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

har en vertikal asymptote for $x = 1$ siden $f(x) \to \pm\infty$ når $x \to 1$. Her ser vi at hvis x nærmer seg 1 fra høyre (så $x > 1$), vil $f(x)$ gå mot ∞, mens hvis x nærmer seg 1 fra venstre (så $x < 1$), vil $f(x)$ gå mot $-\infty$.

Eksempel 1.2.9. *Funksjonen*

$$f(x) = \frac{2x+1}{x-1}$$

har en horisontal asymptote for $y = 2$ siden $f(x) \to 2$ når $x \to \infty$.

I en matematisk modell for en prosess i naturen svarer horisontale asymptoter til stabile verdier i det lange løp. Vertikale asymptoter vil veldig sjelden dukke opp i naturen, så for oss vil de kun ha teoretisk interesse.

Til sist i dette avsnittet tar vi med den formelle definisjonen av en kontinuerlig funksjon.

Definisjon 1.2.5. *En funksjon $f(x)$ er **kontinuerlig** i et punkt $x = a$ dersom*

$$\lim_{x\to a} f(x) = f(a).$$

De fleste prosesser i naturen er kontinuerlige, i den forstand at ting ikke skjer i sprang, men heller endrer seg gradvis. Definisjonen av kontinuitet for funksjoner reflekterer akkurat denne egenskapen.

Likevel er det mange situasjoner der vi ikke har kontinuitet, f.eks. i valgsitusjoner der vi har få (eller i alle fall endelig mange) alternativ. Dette skjer i fenomener som er diskrete, altså endelig på en eller annen måte. I slike problemer vil små endringer som vi gjorde over ikke ha mening, og man må bruke andre metoder i analyser.

1.3 Aritmetiske, alternerende og geometriske følger

En følge slik som $1, 2, 3, 4, \ldots$ der $a_{n+1} = a_n + b$ for et fast tall b (i dette tilfellet er $b = 1$) kalles en **aritmetisk følge**, mens en følge der leddene veksler mellom positive og negative tall kalles en **alternerende følge**.

Et viktig resultat sier at en voksende, oppad begrenset følge, eventuelt en avtagende, nedad begrenset følge, alltid vil konvergere. Følger som konsekvent vokser eller konsekvent avtar kaller vi med en fellesbetegnelse for **monotone** følger.

La $\{a_n\}$ være en monoton følge av reelle tall, dvs. enten voksende eller avtagende. Da vil følgen konvergere mot en grense hvis og bare hvis følgen er begrenset.

Bevis. Vi antar at $\{a_n\}$ er voksende og oppad begrenset. Den såkalte **minste-øvre-grense**-egenskapen til de reelle tallene sier at blant alle øvre grenser[1] for følgen, så finnes det en minste, som vi kaller c. For ethvert valg av et lite tall $\epsilon > 0$ må det finnes en N slik at $c - \epsilon < a_n \leq c$ for alle $n > N$. Dette skyldes at følgen vokser og at c er den minste øvre grensen. Dette innebærer at c er grensen for følgen. Et tilsvarende argument kan brukes når følgen avtar. □

Eksempel 1.3.1. *Følgen*

$$1, \frac{3}{2}, \frac{7}{4}, \frac{15}{8}, \ldots, \frac{2^n - 1}{2^{n-1}}, \ldots$$

er oppad begrenset (alle leddene er mindre enn 2) og strengt voksende. Det betyr at følgen konvergerer.

[1] Øvre grense kalles også øvre skranke (alle tallene i følgen er mindre enn eller lik denne skranken).

En spesiell type følge er de såkalte **geometriske følgene**. Dette er følger av typen

$$a, ar, ar^2, ar^3, ar^4, \ldots$$

for to reelle tall a og r. Følgende resultat gir den viktigste egenskapen til de geometriske følgene for positive verdier av r;

$$\lim_{n\to\infty} r^n = \begin{cases} 0 & \text{for } 0 < r < 1 \\ 1 & \text{for } r = 1 \\ \infty & \text{for } r > 1 \end{cases}$$

Definisjon 1.3.1. *En tallfølge $\{s_n\}$ gitt ved en følge av summer*

$$s_n = a + ar + ar^2 + \cdots + ar^n$$

der a og r er reelle tall, kalles en ***geometrisk rekke****.*

Når $r \neq 1$, kan man kan vise at tallene s_n i definisjonen over kan skrives som

$$s_n = a + ar + ar^2 + \cdots + ar^n = a \cdot \frac{1 - r^{n+1}}{1 - r}.$$

For $-1 < r < 1$ vil summen av en (uendelig) geometrisk rekke, dvs. tallfølgen $\{s_n\}$ der

$$s_n = a + ar + ar^2 + \cdots + ar^n$$

være gitt ved

$$\lim_{n\to\infty} s_n = \frac{a}{1 - r}.$$

Eksempel 1.3.2. *Vi har*

$$1 + \frac{1}{2} + \frac{1}{4} + \cdots = \frac{1}{1 - \frac{1}{2}} = 2.$$

1.4 Funksjoner i flere variable

Vi er ferdig med én-variabel-teorien, og vi skal begynne å jobbe med funksjoner i flere variable. Det første vi skal gjøre er å gå gjennom den samme analysen vi gjorde for funksjoner i en variabel, og se hvordan dette ser ut når vi har flere variable. Hovedfokus er introduksjon av partiell derivasjon. Det skal vi bruke til å finne maksimums- og minimumspunkter og til å definere kritiske punkter. I tillegg skal vi se hvordan vi kan generalisere andre-derivert-testen

for funksjoner i en variabel til funksjoner i to variable, gjennom det som kalles Hesse-determinanten.

To variable x_1, x_2 beskrives ved et ordnet par (x_1, x_2), tre variable ved et (ordnet) trippel (x_1, x_2, x_3), og generelt vil n variable $x_1, x_2, \ldots, x_n$ beskrives ved et *n-tuppel* $(x_1, x_2, \ldots, x_n)$. Merk at rekkefølgen spiller en rolle, så f.eks. er $(2, 5)$ ulik $(5, 2)$. Vi lar $\mathbb{R}^n$ være (den uendelige) mengden av alle reelle n-tupler. Dette er en viktig mengde for oss framover, det er "arbeidsrommet"! Vi kaller gjerne $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ en *vektor*.

Definisjon 1.4.1. *En avbildning*

$$f : \mathbb{R}^n \to \mathbb{R}$$

kalles en funksjon i n variable. Så f tilordner til hver vektor $\mathbf{x} \in \mathbb{R}^n$ nøyaktig ett tall $f(\mathbf{x})$.

Eksempel 1.4.1. *Funksjonen $f(x_1, x_2) = x_1^3 + x_1 x_2 - 2x_2^2$ er en funksjon i to variable.*

Eksempel 1.4.2. *Funksjonen $T(x_1, x_2, x_3)$ som til hvert punkt i rommet tilordner temperaturen i akkurat det punktet er en funksjon i tre variable.*

Merk at i tilfellene hvor vi har 2 eller 3 variable, så bruker vi ofte x, y eller x, y, z som navn på variablene. Dersom vi har flere enn 3 variable vil vi normalt kalle variablene $x_1, x_2, \ldots$

Grafen til en funksjon f i n variable er gitt ved punktene

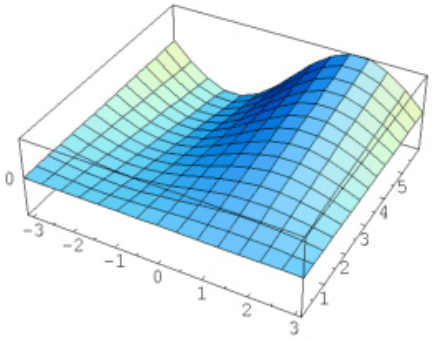
$$(x_1, \ldots, x_n, f(x_1, \ldots, x_n)) \in \mathbb{R}^{n+1}$$

der variablene gjennomløper definisjonsmengden. Grafen til en funksjon i to variable illustrerer vi med en flate i rommet.

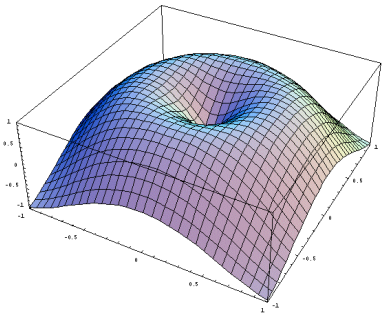
1.5 Nivåmengder

En måte å illustrere funksjoner i to (eller flere) variable er ved å se på funksjonens **nivåmengder**. Nivåmengdene til en flate er kurver langs hvilke funksjonen er konstant. Et eksempel på nivåmengder er høydekurvene på et kart. I det tilfellet måler funksjonen i et punkt høyden over havet i punktet. Nivåmengdene er kurver hvor vi har konstant høyde. Det betyr at dersom vi beveger oss langs en høydekurve, så går vi verken opp eller ned, men holder oss på samme nivå.

Her er en formell definisjon av nivåmengder.



Figur 1.6: Grafen til funksjonen $z = f(x, y) = \frac{y}{6} \cdot \sin x$.

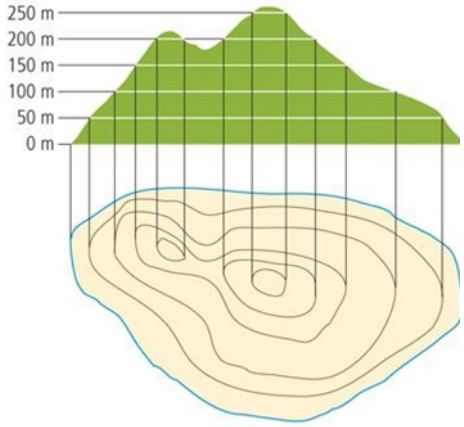


Figur 1.7: Grafen til funksjonen $z = f(x, y) = \sin(\pi \cdot \sqrt{x^2 + y^2})$.

Definisjon 1.5.1. *En* ***nivåmengde*** *for en funksjon* $f(x_1, x_2, \ldots, x_n) = f(\mathbf{x})$ *er delmengden av* $\mathbb{R}^n$ *gitt ved* $\{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = C\}$ *for et reelt tall* C*. Nivåmengdene kalles* ***nivåkurver*** *når* $n = 2$ *og* ***nivåflater*** *når* $n = 3$*. Her lar vi* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$*.*

Eksempel 1.5.1. *La* (x, y) *være et punkt på kartet og la* $f(x, y)$ *måle høyden over havet i dette punktet. Høydekurvene på kartet er nivåmengdene for denne funksjonen, slik det er illustrert i eksempelet over.*

Eksempel 1.5.2. *Vi skal beskrive nivåkurvene til funksjonen* $f(x, y) = \cos(2x) - y^2$*. Vi setter* $f(x, y) = \cos(2x) - y^2 = C$*, dvs.* $\cos(2x) = y^2 + C$ *. For varierende* x *vil* $\cos(2x)$ *alltid ligge mellom* -1 *og* 1*, og* $y^2 = \cos(2x) - C$ *må derfor ligge mellom* $-1 - C$ *og* $1 - C$ *. Siden* $y^2 \geq 0$ *følger det at nivåkurvene forsvinner for* $1 - C < 0$*, dvs. for* $C > 1$*. Dersom den minste verdien,* $-1 - C > 0$*, altså* $C < -1$*, så kan vi for alle verdier av* x *finne to verdier av*



Figur 1.8: Høydekurver.

y slik at $\cos(2x) - y^2 = C$ *. Dermed gjenstår intervallet* $-1 \leq C \leq 1$*. I dette området kan vi løse* $\cos(2x) - y^2 = C$ *med hensyn på y for bare noen verdier av x. Dette forklarer de lukkede (sirkulære) nivåkurvene kontra de utstrukne.*

1.6 Ekstremalpunkter

Vi starter med den lokale teorien, dvs. at vi skal analysere funksjonenes lokale egenskaper. Lokale egenskaper til en funksjon er egenskaper som måles i punkter og små omegner om dem. Vi skal gjøre denne teorien for funksjoner i to variable, men det er ikke noe særlig forskjell å jobbe med flere variable.

Definisjon 1.6.1. *Det er to typer lokale ekstremalpunkter:*

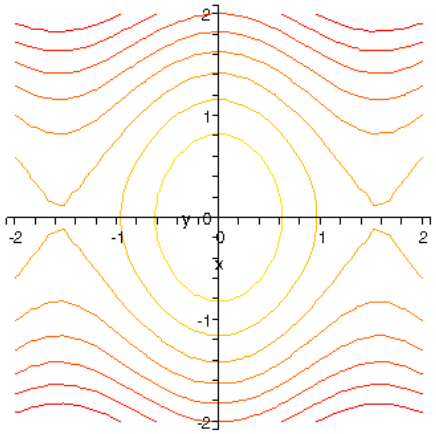
i) Funksjonen $z = f(x, y)$ *har et* ***lokalt minimum*** *i* (a, b) *dersom* $f(a, b)$ *er mindre enn eller lik* $f(x, y)$ *for alle* (x, y) *i nærheten av* (a, b)*.*

ii) Funksjonen $z = f(x, y)$ *har et* ***lokalt maksimum*** *i* (a, b) *dersom* $f(a, b)$ *er større enn eller lik* $f(x, y)$ *for alle* (x, y) *i nærheten av* (a, b)*.*

Merk at vi bruker samme notasjon for et punkt (a, b) og et åpent intervall (a, b). Det vil (forhåpentligvis) alltid framgå av sammenhengen hva vi snakker om.

Vi kan presisere hva vi mener med i nærheten i dette tilfellet. En sirkelskive D (eller disk) med sentrum i punktet (a, b) og radius r er gitt ved

$$D = \{(x, y) \in \mathbb{R}^2 : \sqrt{(x-a)^2 + (y-b)^2)} \leq r\}$$



Figur 1.9: Nivåkurver for $f(x, y) = \cos(2x) - y^2$

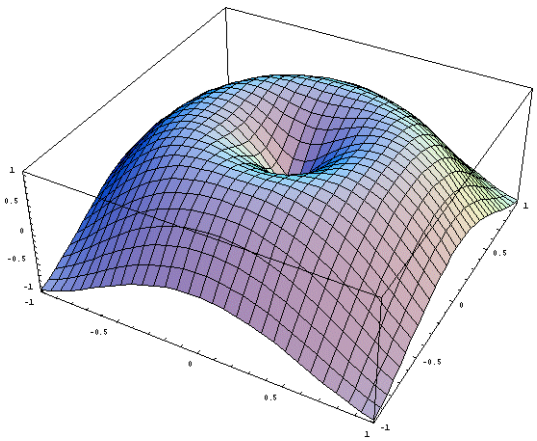
Når vi sier at et punkt $P = (x, y) \neq (a, b)$ ligger i nærheten av (a, b) mener vi at vi kan finne en disk D med en (muligens) liten radius $r > 0$ slik at $P \in D$, og hvor den egenskapen til funksjonen vi ønsker å studere gjelder.

Eksempel 1.6.1. *Funksjonen $f(x, y) = x^2 + y^2$ har et lokalt minimum i $(0, 0)$ siden $f(0, 0) = 0$ og $x^2 + y^2 > 0$ når x og y ikke begge er 0. Det betyr at uansett hvilken origo-sentrert sirkelskive vi ser på, så vil origo være minimum for funksjonen på denne sirkelskiven.*

Funksjonen $f(x, y) = x^2 + y^2$ har ikke noen lokale maksimumspunkter siden funksjonsverdien vokser over alle grenser når vi beveger oss vekk fra origo.

Eksempel 1.6.2. *Funksjonen $z = g(x, y) = \sin(\pi \cdot \sqrt{x^2 + y^2})$ har et lokalt minimum i $(0, 0)$ og lokale maksima i $x^2+y^2 = \frac{1}{2}$ (pluss mange flere som vi ikke ser på figuren). Dette følger siden $g(0, 0) = 0$ og $g(x, y) = \sin(\pi \cdot \sqrt{x^2 + y^2}) > 0$ i nærheten av origo. På sirkelen $x^2 + y^2 = \frac{1}{4}$ har vi at $z = \sin(\pi \cdot \sqrt{\frac{1}{4}}) = \sin\frac{\pi}{2} = 1$, som opplagt er maksimumsverdi for en sinus-funksjon.*

På samme måte som vi for en funksjon i én variabel leter etter ekstremalpunkter der den deriverte av funksjonen er 0, kan vi lete etter ekstremalpunkter for en funksjon i flere variable der alle de partielle deriverte er 0.



Figur 1.10: Graf til g.

1.7 Algoritmer

En algoritme er en *oppskrift* for å gjøre noe, gjerne gitt i form av en rekke steg som skal gjennomføres. Et hverdagseksempel er en matoppskrift (!), f.eks. for å lage vafler:

Algoritme (vafler):

1. `Ha mel, sukker, etc i en bolle.`
2. `Tilsett melk, og rør om.`
3. `Tilsett egg, smeltet smør, la røren svelle.`
4. `Stek vaflene og legg på rist.`

Et annet eksempel: Jeg tenker på et heltall x mellom 0 og 50; still meg ja/nei spørsmål for å bestemme tallet. En mulig algoritme for å løse problemet er

Algoritme (finn heltallet x):

1. Er $x \geq 25$?
2. a) Hvis JA, spør om $x \geq 38$.
   b) Hvis NEI, spør om $x \geq 13$.
3. Fortsett slik, ved å dele det aktuelle intervallet på midten, og spørre.

4. Til slutt har vi bare én mulighet, nemlig x.

I maskinlæring er det en lang rekke algoritmer. I matematikk er algoritmer sentralt for å regne ut de størrelsene man ser på, eller for å besvare spørsmål om egenskaper til et objekt. Allerede på grunnskolen var vi i gang med algoritmer, f.eks. for å legge sammen to tall.

Dette var en nokså uformell beskrivelse. Det som kjennetegner en algoritme er gjerne

- Den har (ofte) input, og ender med å levere output.
- Den er en skrittvis prosedyre: første det, så det osv.
- Den er presis.
- Typisk blir ett eller flere skritt gjentatt.
- Den stopper til slutt opp (terminerer).

La oss se på en algoritme for et problem man ofte kan møte. Vi har gitt n distinkte (ulike) tall

$$a_1, a_2, \ldots, a_n$$

og ønsker å ordne disse i voksende rekkefølge. Dette er altså sortering. Da trenger vi en algoritme, og en mulighet er følgende:

Algoritme (sortering):

Input: $a_1, a_2, \ldots, a_n$.

1. Finn, om mulig, et tall som er større enn det neste tallet.
2. Hvis et slikt tall finnes, bytt om de to tallene, og gjenta forrige steg. Ellers må tallene være i voksende rekkefølge, og dette blir output.

For eksempel, for tallfølgen $4, 1, 3, 2$ kan algoritmen gi

$$4, 1, 3, 2 \to 1, 4, 3, 2 \to 1, 3, 4, 2 \to 1, 3, 2, 4 \to 1, 2, 3, 4.$$

Dette er en enkel sorteringsalgoritme, og langt fra den beste.

Hva er viktige egenskaper ved en algoritme? Først og fremst er det om den virker, altså at den løser den oppgaven den skal. I forrige eksempel er dette tilfelle, men det krever et argument. For i prinsippet kunne det hende at algoritmen fortsetter i det uendelige, og aldri kommer fram til den voksende rekkefølgen. Men dette kan ikke skje av følgende grunn:

- hvis tallfølgen ikke er voksende, så må det finnes to påfølgende tall i gal rekkefølge (det første er større enn det andre)

- hver gang vi bytter to påfølgende tall reduserer vi antall *inversjoner* med én. En inversjon er to tall i følgen som er i gal rekkefølge. Så, til slutt er det ingen inversjoner, og da er jobben gjort!

Så dette viser altså at denne algoritmen er *korrekt*.

En annen viktig egenskap ved algoritmer er *effektivitet*, typisk hvor mange skritt (f.eks, sammenlikninger av to tall) den maksimalt vil kreve. Det er andre former for effektivitet slik som gjennomsnittlig antall skritt osv. Det spiller også en rolle hvor mye lagringsplass en algoritme krever.

Uansett, det er klart at algoritmer er viktig, vi lever i det vi muligens kan kalle *algoritmenes tidsalder*. I flere andre emner på IFI kan man lære mer om algoritmer, både generelt og innen mer spesielle anvendelsesfelt.

I neste kapittel skal vi se på en mye brukt algoritme for å løse likninger, nemlig Newton's metode. Denne metoden er basert på den deriverte av en funksjon, så det temaet må behandles først.

1.8 Oppgaver

Oppgave 1. *Skisser noen nivåkurver for de oppgitte funksjonene over rektangelet* $[-2,2] \times [-2,2]$. *Forsøk også å tegne grafen.*

a) $f(x,y) = x + 2$

b) $f(x,y) = 9 - x^2 - y^2$

c) $f(x,y) = y^2 - x^2$

Oppgave 2. *Skisser noen nivåkurver for de oppgitte funksjonene over rektangelet* $[-1,1] \times [-1,1]$.

a) $f(x,y) = e^{xy}$

b) $f(x,y) = \sin^2(\pi x) + \cos(\pi y)$

Oppgave 3. *Lag en skisse av nivåkurvene til Cobb-Douglas produksjonsfunksjon*

$$f(x,y) = x^{\frac{1}{3}} y^{\frac{2}{3}}$$

Oppgave 4. *Skisser noen nivåkurver for funksjonen* $f(x,y) = \frac{x}{y}$ *over det halvåpne rektangelet* $[-1,1] \times (0,1]$.

Oppgaver relatert til maskinlæring:

Oppgave 5. *La* $s(x) = 2x + 1$ *og* $p(x) = \frac{1}{1+e^{-s(x)}}$. $p(x)$ *kalles en sigmoid-funksjon og brukes ofte i maskinlæring.*

a) Regn ut $s(x)$ *og* $p(x)$ *for* $x \in \{-1, 0, 1\}$ *og skisser grafene.*

b) Forklar at $p(x)$ *er strengt voksende på sitt definisjonsområde.*

c) Finn x *slik at* $p(x) = 0.5$.

Oppgave 6. *Vi skal vise at* $p(x) = \frac{1}{1+e^{-s(x)}}$ *alltid ligger mellom* 0 *og* 1.

a) Forklar at $e^{-s(x)} > 0$ *for alle* x *og videre at* $1 + e^{-s(x)} > 1$.

b) Bruk dette til å vise at $0 < p(x) < 1$.

Oppgave 7. *La* $s(x) = \frac{1}{2}x - 1$ *og* $f(x) = \ln\big(1 + e^{s(x)}\big)$. $f(x)$ *kalles softplus, og brukes i maskinlæring.*

a) Regn ut $s(x)$ *og* $f(x)$ *for* $x \in \{-1, 0, 1\}$ *og skisser grafene.*

b) Forklar at $f(x)$ *er strengt voksende på sitt definisjonsområde.*

c) Finn x *slik at* $f(x) = \ln 2$.

Oppgave 8. *Skisser en algoritme som finner det minste tallet i en liste* `L`. *Test algoritmen på* `L = [7, 2, 9, 2, 5]`.

Oppgave 9. *For å finne ut hvor god en ML-modell er, har vi ofte en liste med hva modellen trodde var riktig og en liste med hva som faktisk var riktig. Ved binær klassifikasjon (JA/NEI = 1/0) kan det se slik ut:*

$$\texttt{y_true} = [1, 0, 1, 1, 0], \qquad \texttt{y_pred} = [1, 1, 1, 0, 0].$$

Skisser en algoritme som returnerer andelen riktige *(accuracy).*

Kapittel 2

Derivasjon

2.1 Derivasjon av funksjoner i én variabel

Monotoni kan måles ved den **deriverte** til funksjonen. For å definere den deriverte av en funksjon trenger vi først begrepet **grenseverdi**. Grenseverdien til en funksjon er den verdien funksjonen $f(x)$ nærmer seg når x nærmer seg en bestemt verdi.

Her gir vi definisjonen, som er presis nok for våre formål her.

Definisjon 2.1.1. *La $f : \mathbb{R} \to \mathbb{R}$ være en funksjon, og $a \in \mathbb{R}$ et punkt på tallinja. Hvis vi kan få $f(x)$ vilkårlig nær L ved å velge x nær nok a, men ikke lik a (hva som skjer i $x = a$ er uten betydning), så sier vi at f har **grenseverdien** L når x går mot a. Vi skriver da*

$$\lim_{x \to a} f(x) = L$$

eller alternativt

$$f(x) \to L \quad \text{når} \quad x \to a.$$

For de fleste funksjoner vi har å gjøre med vil det være slik at

$$\lim_{x \to a} f(x) = f(a)$$

Dersom dette er oppfylt sier vi at funksjonen f er **kontinuerlig** i punktet $x = a$. Polynomialfunksjoner, trigonometriske funksjoner og eksponensialfunksjoner er alle eksempler på funksjoner som er kontinuerlige overalt. Et eksempel på en funksjon med en **diskontinuitet** (punkt hvor funksjonen ikke er kontinuerlig) er

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

som ikke er kontinuerlig i punktet $x = 0$. Hvis $x < 0$, så vil funksjonsverdien uansett være $f(x) = -1$. Det betyr at grenseverdien når x går mot 0 (fra venstre) også vil være -1. Tilsvarende har vi for $x > 0$. Funksjonsverdien er $f(x) = 1$ og grenseverdien av en funksjon som i det aktuelle området er konstant vil ha denne konstanten som sin grenseverdi. Men funksjonsverdien i $x = 0$ er 0 og dermed ulik begge grenseverdiene, både den fra venstre (negative x) og den fra høyre (positive x).

Funksjoner som beskriver ting som opptrer i naturen er stort sett kontinuerlige. Ting kan ikke plutselig endre seg, og hvis de gjør det er det ofte bare tilsynelatende. Det er gjerne en kontinuerlig prosess som går, men den går så fort at det virker som om vi har å gjøre med en diskontinuitet.

Vi skal bruke grenseverdier til å definere den deriverte til en funksjon $f(x)$ i punktet $x = a$. Ideen med den deriverte er å finne et mål for endringen av en funksjon i et bestemt punkt. For å måle denne endringen ser vi på funksjonsverdien i to nærliggende punkter og regner ut gjennomsnittlig endring mellom de to punktene. Ved å la de to punktene komme stadig nærmere hverandre er tanken at grensen for den gjennomsnittlige endringen måler endringen til funksjonen i ett punkt.

Litt mer presist beskrevet, skal vi finne endringen til funksjonen $f(x)$ i punktet $x = a$. De to punktene vi studerer er $x = a$ og $x = a + h$. Den gjennomsnittlige endringen i intervallet $[a, a + h]$ er gitt ved

$$\frac{f(a+h) - f(a)}{(a+h) - a} = \frac{f(a+h) - f(a)}{h}$$

Denne størrelsen regner vi ut, og lar så h gå mot 0.

Definisjon 2.1.2. *La $f : \mathbb{R} \to \mathbb{R}$ være en kontinuerlig funksjon. Den **deriverte** $f'(a)$ av f i punktet $x = a$ er gitt ved*

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

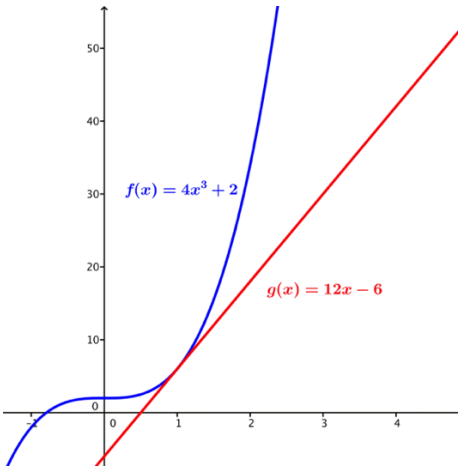
Merk at i definisjonen av den deriverte krever vi at h skal kunne være negativ så vel som positiv. Dersom grenseverdien bare eksisterer fra den ene siden sier vi at vi har en **ensidig grenseverdi**. Ensidige grenseverdier er ikke godt nok for definisjonen av den deriverte.

En funksjon sies å være **deriverbar** i et punkt $x = a$ dersom den deriverte av funksjonen eksisterer i punktet, dvs. at grenseverdien over eksisterer, altså at de to ensidige grenseverdiene er like. En funksjon er deriverbar på et intervall I dersom den er deriverbar i alle punkter $a \in I$ i intervallet. Det er lett å se at

de vanlige funksjonene, polynomer, cosinus og sinus og eksponensialfunksjoner er deriverbare overalt.

Den deriverte av en funksjon er selv en funksjon, og den gir oss et uttrykk for endringen til funksjonen $f(x)$ lokalt. Et eksempel er konstantfuksjonen $f(x) = C$. Den deriverte til denne funksjonen er $f'(x) = 0$, noe som sier at funksjonen ikke endrer seg. Og det er jo tilfellet for en konstant funksjon.

Hvis vi tegner grafen til en funksjon, så vil den deriverte til funksjonen i et punkt gi oss **stigningstallet** til funksjonen i punktet. Stigningstallet til funksjonen i et punkt måler vi ved å se på **tangenten** til funksjonen i punktet. Tangenten er en rett linje som **tangerer** grafen i punktet. Stigningstallet til en rett linje uttrykker hvor mange enheter vi går opp (eller ned) når vi går en enhet mot høyre.



Figur 2.1: Tangent.

Eksempel 2.1.1. *Vi skal regne ut den deriverte av funksjonen* $f(x) = x^2$. *Vi har*

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h\to 0} \frac{x^2 + 2xh + h^2 - x^2}{h} \\ &= \lim_{h\to 0} (2x + h) = 2x. \end{aligned}$$

Tilsvarende regning for $f(x) = x^3$ gir

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{(x+h)^3 - x^3}{h} \\ &= \lim_{h\to 0} \frac{x^3 + 3x^2h + 3xh^2 + h^3 - x^3}{h} \\ &= \lim_{h\to 0} (3x^2 + 3xh + h^2) = 3x^2. \end{aligned}$$

På samme måte får vi for $f(x) = x^n$ for et heltall $n > 0$ at $f'(x) = nx^{n-1}$.

Eksempel 2.1.2. *I eksemplet over så vi på potensfunksjoner med positiv eksponent. Man kan vise at nøyaktig samme formel gjelder for negative eksponenter $n < 0$. Som et eksempel kan vi regne ut den deriverte av funksjonen $f(x) = x^{-1} = \frac{1}{x}$. Vi har*

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\frac{1}{x+h} - \frac{1}{x}}{h} = \lim_{h\to 0} \frac{x - (x+h)}{hx(x+h)} \\ &= \lim_{h\to 0} \frac{-h}{hx(x+h)} = -\frac{1}{x^2} = -x^{-2} \end{aligned}$$

som stemmer godt med formelen $f'(x) = nx^{n-1}$ (sett $n = -1$).

Eksempel 2.1.3. *Vi skal regne ut den deriverte til eksponensialfunksjonen $f(x) = e^x$. For å gjøre det trenger vi å vite at for små verdier av h, så kan vi approksimere eksponensialfunksjonen ved $e^h \approx 1 + h$.*

$$f'(x) = \lim_{h\to 0} \frac{e^{x+h} - e^x}{h} = \lim_{h\to 0} \frac{e^x(e^h - 1)}{h} = \lim_{h\to 0} \frac{e^x h}{h} = e^x.$$

Eksponensialfunksjonen er altså lik sin egen derivert. Denne egenskapen er (med unntak av $f(x) = 0$) unik for eksponensialfunksjonen med e som grunntall.

Eksempel 2.1.4. *Hva med de deriverte av sinus og cosinus? Her bruker vi andre approksimasjoner; for små verdier av h har vi at $\sin h \approx h$, og at $1 - \cos h$ er liten sammenliknet med h. Det gir for $f(x) = \sin x$,*

$$\begin{aligned} f'(x) &= \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} \\ &= \lim_{h\to 0} \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} \\ &= \lim_{h\to 0} \frac{\cos x \sin h + \sin x(\cos h - 1)}{h} = \cos x \end{aligned}$$

hvor vi har brukt formelen for sinus til en sum av to vinkler.

Tilsvarende for $g(x) = \cos x$,

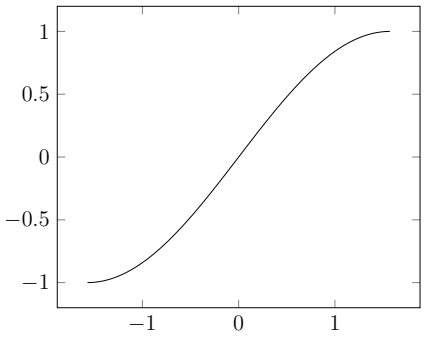
$$\begin{aligned} g'(x) &= \lim_{h\to 0} \frac{\cos(x+h) - \cos x}{h} \\ &= \lim_{h\to 0} \frac{\cos x \cos h - \sin x \sin h - \cos x}{h} \\ &= \lim_{h\to 0} \frac{-\sin x \sin h + \cos x(\cos h - 1)}{h} = -\sin x. \end{aligned}$$

Vi kan bruke den deriverte til å studere monotoniegenskapene til en funksjon. Generelt vil det være klar sammenheng mellom voksende funksjoner og positiv derivert, og avtagende funksjoner og negativ derivert.

Teorem 2.1.3. *Dersom* $f'(x) > 0$ *(henholdsvis* $f'(x) < 0$*) på et intervall, så er* $f(x)$ *voksende (henholdsvis avtagene) på intervallet.*

Eksempel 2.1.5. *Vi ser på funksjonen* $f(x) = x^2$*. Den deriverte er gitt ved* $f'(x) = 2x$*, som betyr at funksjonen vokser på intervallet* $[0, \infty\rangle$ *og avtar på intervallet* $\langle -\infty, 0]$*.*

Eksempel 2.1.6. *Betrakt funksjonen* $f(x) = e^x$*. Den deriverte er gitt ved* $f'(x) = e^x > 0$*, som betyr at funksjonen vokser overalt.*



Figur 2.2: Sinusfunksjon.

Eksempel 2.1.7. *Funksjonen* $f(x) = \sin x$*, hvor* $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ *har derivert* $f'(x) = \cos x$*. På intervallet* $[-\frac{\pi}{2}, \frac{\pi}{2}]$ *har vi* $\cos x \geq 0$ *og funksjonen vil derfor være voksende på dette intervallet.*

Hvis vi skal regne ut den deriverte til en funksjon kan vi alltids bruke definisjonen og regne ut grenseverdien. Men det er vanligvis ganske tungvint. Derfor har man etablert et sett av regneregler for derivasjon. Den første sier at derivasjon er en lineær operator, dvs. at den deriverte av en sum er summen av de deriverte, og at en konstant kan flyttes ut.

Teorem 2.1.4. *La $f(x)$ og $g(x)$ være to deriverbare funksjoner på et område $A \subset \mathbb{R}$, og la $h(x) = f(x) + g(x)$. Da har vi*

$$h'(x) = f'(x) + g'(x).$$

La videre $a \in \mathbb{R}$ være et reelt tall og $k(x) = af(x)$. Da har vi

$$k'(x) = af'(x).$$

Eksempel 2.1.8. *La $f(x) = 2x^3 - 3x^4$. Da har vi*

$$f'(x) = 6x^2 - 12x^3$$

ved å kombinere regelen over (linearitet) med derivasjonsreglene for potensfunksjoner.

Det er ikke like rett fram å derivere et produkt av to funksjoner. Regnereglen er gitt ved følgende resultat:

Teorem 2.1.5. *La $f(x)$ og $g(x)$ være to deriverbare funksjoner på et område $A \subset \mathbb{R}$, og la $h(x) = f(x)g(x)$. Da har vi*

$$h'(x) = f'(x)g(x) + f(x)g'(x).$$

Denne regelen omtaler vi ofte som **produktregelen** eller **Leibniz regel**.

Eksempel 2.1.9. *La $f(x) = x^2 \sin x$. Da har vi*

$$f'(x) = 2x \sin x + x^2 \cos x.$$

Eksempel 2.1.10. *La $f(x) = x^2$ og $g(x) = \frac{1}{x^2}$. Da har vi $h(x) = f(x)g(x) = x^2 \cdot \frac{1}{x^2} = 1$, og $h'(x) = 0$. Ved produktregelen får vi*

$$h'(x) = 2x \cdot (\frac{1}{x^2}) + x^2(-2\frac{1}{x^3}) = \frac{2}{x} - \frac{2}{x} = 0.$$

Den siste i rekken av generelle regler for derivasjon gjelder for sammensatte funksjoner.

Teorem 2.1.6. *La $g(x)$ være en deriverbar funksjon på et område $A \subset \mathbb{R}$, og $f(x)$ en deriverbar funksjon på $f(A) = \{f(x) : x \in A\} \subset \mathbb{R}$. La videre $h(x) = f(g(x))$ være den sammensatte funksjonen. Da har vi*

$$h'(x) = f'(g(x)) \cdot g'(x).$$

Denne regneregelen for derivasjon kalles **kjerneregelen**. Begrepet kommer av at funksjonen $g(x)$ omtales som **kjernen** i den sammensatte funksjonen.

Kjerneregelen viser seg å være svært viktig i maskinlæring, i forbindelse med nevrale nett. Da ønsker man å minimere en funksjon i mange variable som involverer sammensatte funksjoner. Algoritmer for å minimere beregner (tilnærmet) deriverte og bruker da kjerneregelen. Du vil lære mer om dette i emner i maskinlæring.

Eksempel 2.1.11. *Betrakt den sammensatte funksjonen $h(x) = \cos(x^2)$. Vi lar $f(x) = \cos x$ og $g(x) = x^2$, så vi har $h(x) = f(g(x))$. Det gir*

$$h'(x) = f'(g(x)) \cdot g'(x) = -\sin(x^2) \cdot 2x = -2x \sin(x^2).$$

Eksempel 2.1.12. *Betrakt funksjonen $f(x) = \frac{1}{x}$. Vi kan sette denne funksjonen sammen med seg selv. Det gir*

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

Vi kan regne ut den deriverte av denne funksjonen på to ulike måter. For det første har vi åpenbart at $g'(x) = 1$. Hvis vi bruker kjerneregelen får vi, siden $f'(x) = -\frac{1}{x^2}$:

$$g'(x) = f'(f(x)) \cdot f'(x) = -\frac{1}{(\frac{1}{x})^2} \cdot (-\frac{1}{x^2}) = 1.$$

Vi har dermed verifisert kjerneregelen i dette tilfellet.

Eksempel 2.1.13. *Vi kan bruke kjerneregelen til å regne ut den deriverte av logaritmefunksjonen $f(x) = \ln x$. Siden logaritmefunksjonen er den omvendte funksjonen til eksponensialfunksjonen $g(x) = e^x$ har vi $h(x) = g(f(x)) = e^{\ln x} = x$ og $h'(x) = 1$. Vi deriverer ved å bruke kjerneregelen*

$$h'(x) = g'(f(x)) \cdot f'(x) = e^{\ln x} \cdot f'(x) = x \cdot f'(x) = 1$$

og det følger at

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

siden $g'(x) = e^x$.

Logaritmefunksjonen $\ln x$ er bare definert for positive x. Vi kan utvide definisjonsområdet ved heller å se på funksjonen $f(x) = \ln|x|$. Denne funksjonen er definert for alle reelle tall forskjellig fra 0, og er gitt ved

$$f(x) = \begin{cases} \ln(-x) & x < 0, \\ \ln x & x > 0. \end{cases}$$

Vi kan derivere denne funksjonen ved å derivere hvert av de to områdene hver for seg og får

$$f'(x) = \begin{cases} \frac{1}{-x} \cdot -1 = \frac{1}{x} & x < 0, \\ \frac{1}{x} & x > 0. \end{cases}$$

Resultatet er at den deriverte av $\ln|x|$ er $\frac{1}{x}$.

Nå kan vi kombinere regneregler for å utvide verktøykassen. Kombinasjon av potensregelen og kjerneregelen gir at den deriverte av $h(x) = \frac{1}{g(x)}$ er gitt ved

$$h'(x) = -\frac{1}{g(x)^2} \cdot g'(x).$$

Hvis vi kombinerer dette videre med produktregelen, får vi formelen for den deriverte til en rasjonal funksjon:

Teorem 2.1.7. *La $f(x)$ og $g(x)$ være to deriverbare funksjoner på et område $A \subset \mathbb{R}$, og anta at $g(x) \neq 0$ på A. La $h(x) = \frac{f(x)}{g(x)}$. Da har vi*

$$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

Eksempel 2.1.14. *Vi skal derivere kvotienten*

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

Etter formelen over får vi

$$f'(x) = \frac{(1-x)\cdot 1 - x(-1)}{(1-x)^2} = \frac{1}{(1-x)^2}.$$

Eksempel 2.1.15. *Vi kan også bruke denne formelen til å vise potensformelen for negative potenser. La*

$$f(x) = \frac{1}{x^n} = x^{-n}$$

Da har vi

$$\begin{aligned} f'(x) &= \frac{x^n \cdot 0 - 1 \cdot nx^{n-1}}{(x^n)^2} = -n\frac{x^{n-1}}{x^{2n}} \\ &= -n\frac{1}{x^{n+1}} = -nx^{-n-1}. \end{aligned}$$

2.2 Ekstremalpunkter

Vi er interessert i å finne punkter der funksjonen antar sin største/minste verdi. I første omgang leter vi etter lokale maksimums- og minimumspunkter, dvs. punkter der funksjonen er større (mindre) enn eller lik alle andre punkter i nærheten. Dette trenger ikke å gi oss de største (minste) verdiene for funksjonen (globalt).

Vi skal bruke begrepet *i nærheten av* x om punkter som er forskjellig fra x, men hvor avstanden til x er så liten som det er hensiktsmessig i det aktuelle tilfellet. Denne avstanden kan være 1 eller den kan være 0.00001, eller en enda mindre verdi. Det viktigste er at det finnes en fast avstand slik at påstanden vi kommer med er sann for alle y som ikke er lenger fra punktet x enn denne avstanden.

Et eksempel på bruk av begrepet *i nærheten av* er som følger: La $f(x)$ være en kontinuerlig funksjon, dvs. at $\lim_{x\to a} f(x) = f(a)$ for alle a i definisjonsområdet til f. F.eks. kan vi se på $f(x) = x - x^2$. For $x = 0.9$ har vi $f(0.9) = 0.9 - (0.9)^2 = 0.09 > 0$. En kontinuerlig funksjon som er ekte større enn 0 i et punkt, vil alltid være ekte større enn 0 i *nærheten* av punktet. La nå $x = 0.9 + h$ for en liten verdi for h, dvs. vi ser på et punkt som er i *nærheten* av 0.9. Da har vi at

$$\begin{aligned} f(0.9+h) &= 0.9 + h - (0.9+h)^2 \\ &= 0.9 + h - 0.81 - 1.8h - h^2 \\ &= 0.09 - 0.8h - h^2 \end{aligned}$$

Vi ser at dersom $|h| < 0.1$ så har vi $f(0.9 + h) > 0$. Det betyr at i dette eksemplet vil *i nærheten av* av 0.9 naturlig bety at avstanden til 0.9 er ekte mindre enn 0.1.

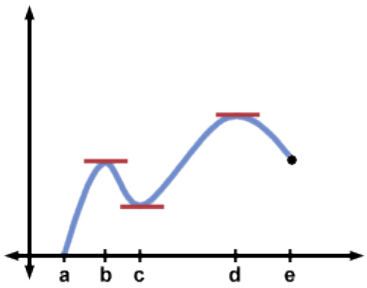
Definisjon 2.2.1. *Det er to typer lokale ekstremalpunkter:*

i) Funksjonen $y = f(x)$ har et ***lokalt minimum*** *i $x = c$ dersom $f(c) \leq f(x)$ for alle x i et åpent intervall om $x = c$.*

ii) Funksjonen $y = f(x)$ har et ***lokalt maksimum*** *i $x = c$ dersom $f(c) \geq f(x)$ for alle x i et åpent intervall om $x = c$.*

En fellesbetegnelse på punkter hvor funksjonen har lokale maks- eller minimumspunkter er **lokale ekstremalpunkter**.

Eksempel 2.2.1. *Funksjonen $f(x) = x^2$ definert på hele $\mathbb{R}$ har et lokalt minimumspunkt i $x = 0$ siden $f(x) > f(0)$ for alle $x \neq 0$.*



Figur 2.3: Lokale maksimums- og minimumspunkter.

Eksempel 2.2.2. *Funksjonen $f(x) = \cos x$ definert på hele $\mathbb{R}$ har et lokalt maksimumspunkt i $x = 0$ siden $\cos x < \cos 0 = 1$ for alle $x \neq 0$ i nærheten av 0. Denne funksjonen er periodisk med periode 2π (betyr at $f(x + 2\pi) = f(x)$) og vi har derfor også maksimumspunkter for $x = 2k\pi$, for alle hele tall k. Tilsvarende har vi lokale minimumspunkter for $x = \pi + 2k\pi$ for alle hele tall k, siden $\cos(\pi + 2k\pi) = \cos \pi = -1 \leq \cos x$ for alle verdier av x.*

Eksempel 2.2.3. *Funksjonen $f(x) = e^x$ har ingen lokale ekstremalpunkter siden dens deriverte $f'(x) = e^x > 0$ for alle x. Funksjonen vokser derfor i hele sitt definisjonsområde, noe som ikke er forenlig med å ha et ekstremalpunkt.*

2.3 Kritiske punkter

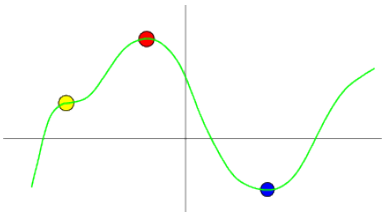
Den deriverte til en funksjon $y = f(x)$ i et punkt x_0 gir oss stigningstallet til tangenten til funksjonen i punktet. I et maksimums- eller minimumspunkt for en deriverbar funksjon vil tangenten være horisontal, dvs. tangentlinja har stigningstall 0. Det betyr at $f'(x_0) = 0$. Slike punkter er så viktige at de har fått et eget navn.

Definisjon 2.3.1. *Et punkt $x = c$ i definisjonsområdet til en funksjon $f(x)$ kalles et **kritisk punkt** for funksjonen dersom $f'(c) = 0$ eller $f'(c)$ ikke eksisterer.*

Merk at vi bruker betegnelsen kritisk punkt både om punkter der den deriverte er 0, og om punkter der funksjonen ikke har noen derivert, f.eks. i *knekkpunkter* på grafen. Knekkpunkter er punkter der de ensidige deriverte er forskjellige. Det betyr at stigningstallet til grafen er forskjellig på de to sidene av punktet. Grafen ser dermed ut som en spiss i dette punktet. Dersom funksjonen $f(x)$ er definert på et lukket intervall $[a, b]$, inkluderer

vi endepunktene $x = a$ og $x = b$ blant funksjonens kritiske punkter. Dette passer med definisjonen siden den deriverte av en funksjon formelt sett ikke er definert i endepunktene av et lukket definisjonsområde.

Vi bruker betegnelsen **indre punkter** om punktene i et intervall som ikke er endepunkter.



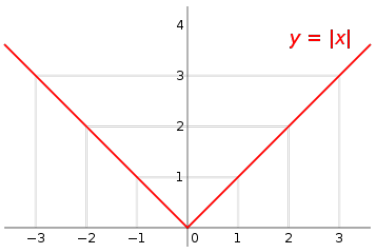
Figur 2.4: De markerte punktene viser de indre kritiske punktene.

Eksempel 2.3.1. *Funksjonen $f(x) = |x|$ har et kritisk punkt i $x = 0$ siden $f'(0)$ ikke er definert i det punktet. Absoluttverdifunksjonen er ikke deriverbar i $x = 0$ siden grenseverdiene for de ensidige deriverte fra høyre og venstre side ikke er like. For positive verdier av h har vi*

$$\lim_{h\to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0} \frac{|h|}{h} = 1$$

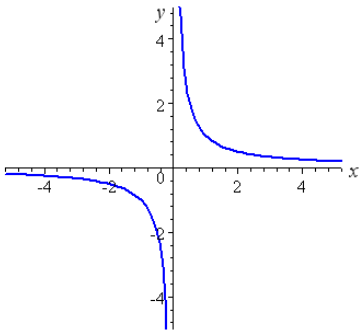
og for negative verdier av h,

$$\lim_{h\to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0} \frac{|h|}{h} = -1.$$



Figur 2.5: Grafen til absoluttverdifunksjonen, $f(x) = |x|$ med et kritisk punkt i $x = 0$.

Eksempel 2.3.2. *Alle potensfunksjoner $g(x) = x^n$ for $n \geq 2$ har kritisk punkt i $x = 0$ siden $f'(0) = n \cdot 0^{n-1} = 0$.*



Figur 2.6: Grafen til $f(x) = \frac{1}{x}$.

Eksempel 2.3.3. *Funksjonen $h(x) = \frac{1}{x}$ har **ikke** et kritisk punkt i $x = 0$ siden funksjonen ikke er definert i dette punktet.*

I mange tilfeller er det sammenfall mellom ekstremalpunkter og kritiske punkter, men det er ikke alltid slik. Imidlertid er alle ekstremalpunkter også kritiske punkter, mens det motsatte ikke nødvendigvis er riktig.

Teorem 2.3.2. *Dersom $f(x)$ har et lokalt maksimums- eller minimumspunkt i $x = c$, så er c et kritisk punkt for f.*

Bevis. Anta at funksjonen er deriverbar i punktet $x = c$. Hvis ikke er punktet pr. definisjon et kritisk punkt. Vi har

$$f'(c) = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h} \approx \frac{f(x) - f(c)}{x - c}$$

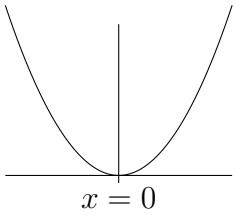
hvor vi i det siste uttrykket lar $x = c + h$ være i nærheten av c, dvs. at h er liten. Anta at $f(x)$ har et lokalt maksimum i $x = c$. Dersom $f'(c) > 0$, så vil $f(x) > f(c)$ for $x > c$. Dette er ikke mulig siden c er et lokalt maksimumspunkt. Dersom $f'(c) < 0$, så vil $f(x) > f(c)$ for $x < c$, som på tilsvarende måte er umulig og vi har en motsetning. Samme type argument gjelder for lokale minimumspunkter. □

Teorem 2.3.3. *Hvis $f(x)$ har et kritisk punkt i $x = c$, så er dette*

i) et lokalt minimum dersom $f'(x)$ skifter tegn fra negativ til positiv i $x = c$.

ii) et lokalt maksimum dersom $f'(x)$ skifter tegn fra positiv til negativ i $x = c$.

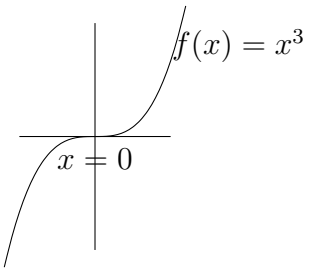
Bevis. Følger nokså direkte fra definisjonen av den deriverte ved å analysere fortegnet til uttrykkene som inngår. □



Figur 2.7: Grafen til funksjonen $f(x) = x^2$.

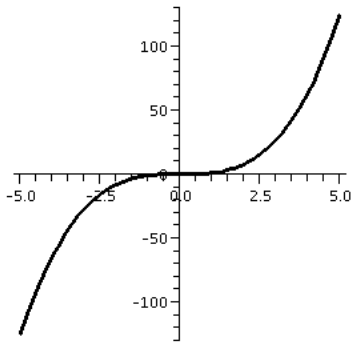
Eksempel 2.3.4. *Vi ser på funksjonen $f(x) = x^2$. Den deriverte er gitt ved $f'(x) = 2x$, som betyr at funksjonen har et kritisk punkt i $x = 0$. Dette er også et lokalt minimumspunkt siden den deriverte til $f(x)$ skifter tegn fra negativt til positivt i $x = 0$.*

Grunnen til at det ikke er en enentydig sammenheng mellom kritiske punkter og ekstremalpunkter, er at det finnes kritiske punkter som ikke er ekstremalpunkter. Det at den deriverte til en funksjon er 0 i et punkt, trenger ikke å bety at punktet er et maksimums- eller minimumspunkt for funksjonen. Det er fullt mulig at den deriverte til funksjonen er 0 i et punkt, samtidig med at funksjonen vokser overalt i nærheten av punktet (se eksemplet under).



Figur 2.8: Grafen til funksjonen $f(x) = x^3$.

Eksempel 2.3.5. *Betrakt funksjonen $g(x) = x^3$. Den deriverte er gitt ved $g'(x) = 3x^2$, som betyr at funksjonen har et kritisk punkt for $x = 0$. Dette er imidlertid ikke noe ekstremalpunkt. Funksjonens deriverte skifter ikke fortegn i dette punktet, vi har $g'(x) > 0$ for alle $x \neq 0$.*



Figur 2.9: Terrassepunkt.

Definisjon 2.3.4. *Et punkt $x = a$ hvor den deriverte $f'(x)$ er 0, men ikke skifter fortegn kalles et* ***terrassepunkt****.*

En typisk problemstilling vi er opptatt av, handler om å analysere monotoni-egenskapene og å finne lokale ekstremalpunkter for en funksjon. Det neste eksemplet illustrerer framgangsmåten, hvor fortegnsskjema er et sentralt hjelpemiddel. Et **fortegnsskjema** er en tabell hvor vi deler opp definisjonsområdet til funksjonen i intervaller hvor den deriverte har samme fortegn (negativ eller positiv).

Eksempel 2.3.6. *Vi skal bestemme monotomi-egenskaper og lokale ekstremalpunkter for funksjonen*

$$f(x) = 3x^4 - 4x^3 + 3.$$

Vi regner ut den deriverte, faktoriserer og setter den lik 0;

$$f'(x) = 12x^3 - 12x^2 = 12x^2(x-1) = 0.$$

Vi tegner fortegnsskjema for $f'(x)$:

	$x < 0$	*0*	$0 < x < 1$	1	$x > 1$
$12x^2$	$+$	0	$+$	$+$	$+$
$x - 1$	$-$	$-$	$-$	0	$+$
$f'(x)$	$-$	0	$-$	0	$+$
	$\searrow$		$\searrow$	$\smile$	$\nearrow$

Hver faktor i uttrykket for den deriverte har sin egen rad og ±-tegnene angir fortegnet til faktoren i det angitte intervallet. Fortegnene multipliseres sammen i raden til $f'(x)$ og i nederste rad er konklusjonen gitt, om hvorvidt funksjonen vokser eller avtar.

Det betyr at vi har et lokalt minimumspunkt for $x = 1$, at funksjonen er avtagende for $x \leq 1$ og at den er voksende for $x \geq 1$. Vi har ikke noe ekstremalpunkt for $x = 0$, selv om $f'(x) = 0$. Den deriverte skifter ikke fortegn i dette punktet.

2.4 Globale ekstremalpunkter

I motsetning til de lokale egenskapene, som sier noe om funksjonen i en liten omegn om et punkt, vil de **globale egenskapene** beskrive funksjonen på hele definisjonsområdet sett under ett.

Definisjon 2.4.1. *(Globale maksimums- og minimumspunkter)*

*i) Funksjonen $y = f(x)$ har et **globalt minimum** i $x = c$ dersom $f(c) \leq f(x)$ for alle x i definisjonsområdet.*

*ii) Funksjonen $y = f(x)$ har et **globalt maksimum** i $x = c$ dersom $f(c) \geq f(x)$ for alle x i definisjonsområdet.*

*Vi vil omtale funksjonsverdien i de globale minimums- og maksimumspunktene som funksjonens **minimums-** og **maksimumsverdier** og samlet som funksjonenes **ekstremalverdier**.*

Funksjoner trenger verken å ha lokale eller globale ekstremalpunkter. Et eksempel er funksjonen $f(x) = x$ definert på hele tallinja. Den har ingen ekstremalpunkter fordi for alle punkter $x = a$ på tallinja så vil det finnes punkter x i nærheten av a slik at $f(x) > f(a)$ og andre som oppfyller $f(x) < f(a)$. Funksjonen $f(x) = x$ har heller ingen ekstremalverdier dersom definisjonsområdet er et åpent intervall (a, b). På et lukket intervall derimot, har funksjonen $f(x) = x$ både et globalt maksimumspunkt og et globalt minimumspunkt, nemlig endepunktene i intervallet.

Et åpent intervall som $(0, 1)$ inneholder ikke sine endepunkter. Vi kan finne tall i intervallet som er så nær opp til 1 som vi måtte ønske. Men uansett hvilket tall vi velger, vil det alltid finnes tall mellom det valgte tallet og 1. Alle tall i det åpne intervallet $(0, 1)$ har andre tall på begge sider. Dette er karakteristisk for åpne intervaller. I det lukkede intervallet $[0, 1]$ finnes det to tall, 0 og 1, som kjennetegnes ved at de har naboer i intervallet kun på sin ene side.

En kontinuerlig funksjon vil alltid oppnå sine ekstremalverdier dersom definisjonsområdet er et lukket intervall. Dette er innholdet i **ekstremalverditeoremet** som vi gjengir uten bevis.

Teorem 2.4.2. *En funksjon $f(x)$ som er kontinuerlig på et lukket, begrenset intervall $[a, b]$, vil oppnå sine maksimums- og minimumsverdier i intervallet.*

Eksempel 2.4.1. *La $f(x) = -x^2 + 4x - 3$. Vi skal finne maksimumsverdien til $f(x)$ på intervallet $[-1, 1]$.*

Vi observerer at $f'(x) = -2x + 4 = 0$ når $x = 2$. Men $x = 2$ er ikke i intervallet, så dette punktet er vi ikke interessert i. Det betyr at det er ingen kritiske punkter i det indre av definisjonsområdet, og de eneste punktene vi trenger å sjekke er endepunktene: $f(-1) = -8$ og $f(1) = 0$. Så den største verdien til $f(x)$ på $[-1, 1]$ er $f(1) = 0$.

Eksempel 2.4.2. *Vi skal finne ekstremalverdiene til funksjonen $f(x) = 7 + |x - 2|$ på intervallet $1 \leq x \leq 4$. Den deriverte $f'(x) \neq 0$ for alle x. Men den er ikke definert i punktet $x = 2$, så vi regner ut verdien her, $f(2) = 7$. For endepunktene har vi $f(1) = 8$ og $f(4) = 9$. Det minste av disse tre verdiene er $f(2) = 7$, som derfor er minimumsverdien, mens den største er $f(4) = 9$ som gir oss maksimumsverdien på intervallet $[-1, 4]$.*

Eksempel 2.4.3. *Av alle rektangler med areal 25, og sidekanter kortere enn 10, hvilket har den minste og hvilket har den største omkretsen?*

Vi starter med å formulere problemet mer matematisk. Vi lar x være lengden på den ene sidekanten. Da vil den andre sidekanten ha lengde $\frac{25}{x}$ siden arealet skal være 25. Omkretsen blir da gitt ved funksjonen

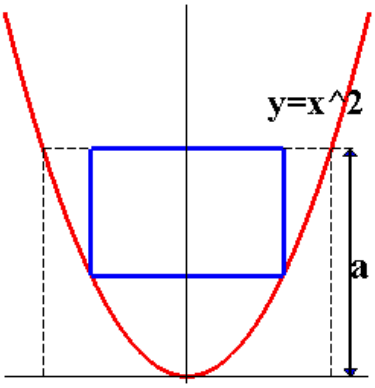
$$f(x) = 2x + 2 \cdot \frac{25}{x}$$

over intervallet $(0, 10]$. Denne funksjonen skal vi finne maksimumsverdien til. Vi deriverer og får

$$f'(x) = 2 - \frac{50}{x^2}.$$

Løser vi likningen $f'(x) = 0$ får vi $x = \pm 5$. Punktet $x = -5$ er ikke med i definisjonsområdet, så det forkaster vi. Dermed står vi igjen med ett kritisk punkt, $x = 5$. Funksjonen er ikke definert i punktet $x = 0$ (rektanglet kollapser) og vi må sjekke de to kritiske punktene $x = 5$ og $x = 10$. Innsetting gir $f(5) = 20$ og $f(10) = 25$. Dette gir kandidater for maksimum og minimum. Men vi er ikke ferdige enda. Siden definisjonsområdet ikke er lukket (men halvåpent) kan vi ikke bruke ekstremalverdisetningen, og vi må sjekke hva som skjer når $x \to 0$. Det er opplagt at omkretsen vil vokse over alle grenser

når den ene sidekanten blir veldig liten, og den andre blir veldig stor. Derfor vil ikke $f(10) = 25$ *være noe globalt maksimum, men* $f(5) = 20$ *er fortsatt minimumsverdien.*



Figur 2.10: Rektangel i en parabel.

	-1	$-1 < x < 0$	0	$0 < x < 2$	2	$2 < x < 3$	3
x		-	0	+	+	+	
$2 - x$		+	+	+	0	-	
e^{-x}		+	+	+	+	+	
$f'(x)$		-	0	+	0	-	
	⌒	↘	⌣	↗	⌒	↘	⌣
$f(x)$	**e**		**0**		$\frac{4}{e^2}$		$\frac{9}{e^3}$

Figur 2.11: Fortegnsskjema: $f'(x) = x(2-x)e^{-x}$.

Eksempel 2.4.4. *Vi skal finne det største rektangelet vi kan legge inne grafen til parabelen* $y = x^2$ *når* y *skal være mindre enn en fast verdi* a*. Se Fig.* 2.4.

La $A(x)$ *være arealet av rektangelet med grunnlinje* $2x$ *(begge sider av y-aksen). Høyden i rektangelet vil være* $a - x^2$*. Det gir areal* $A(x) = 2x(a - x^2) = 2xa - 2x^3$ *hvor* x *ligger i intervallet* $[0, \sqrt{a}]$*. De kritiske punktene vil være endepunktene og punktene der* $A'(x) = 0$*. Den deriverte av* $A(x)$ *er gitt ved*

$$A'(x) = 2a - 6x^2.$$

Setter vi denne lik 0 *får vi* $x = \pm\sqrt{\frac{a}{3}}$*. Vi forkaster den negative løsningen siden den ikke er i definisjonsområdet. Innsetting gir* $A(0) = A(\sqrt{a}) = 0$ *og*

$A(\sqrt{\frac{a}{3}}) = (\frac{4}{9})\sqrt{3}a^{\frac{3}{2}}$. *Det gir minimumsareal (selvfølgelig)* 0 *og maksimumsareal* $(\frac{4}{9})\sqrt{3}a^{\frac{3}{2}}$.

Eksempel 2.4.5. *Vi ser på funksjonen*

$$f(x) = x^2e^{-x} \qquad -1 \leq x \leq 3.$$

For å finne ut hvor funksjonen vokser, hvor den avtar og hvor den har sine ekstremalpunkter, må vi se på dens deriverte. I dette tilfellet får vi

$$f'(x) = 2xe^{-x} - x^2e^{-x} = x(2-x)e^{-x}.$$

Denne informasjonen bruker vi til å tegne et fortegnsskjema for $f'(x)$. *Skjemaet viser at funksjonen avtar på intervallene* $(-1, 0)$ *og* $(2, 3)$ *og vokser på intervallet* $(0, 2)$. *Den har lokale maksimumspunkter i* $x = -1$ *og i* $x = 2$, *og lokale minimumspunkter for* $x = 0$ *og for* $x = 3$. *For å finne ut hvilke punkter som gir oss de globale ekstremalverdiene regner vi ut funksjonsverdien i de lokale ekstremalpunktene, og i tillegg i endepunktene til det lukkede intervallet. Verdiene står i nederste linje. Av disse ser vi at* $x = -1$ *gir den største verdien, altså et globalt maksimumspunkt og* $x = 0$ *gir minste verdi og dermed svarer til et globalt minimumspunkt.*

2.5 Newtons metode

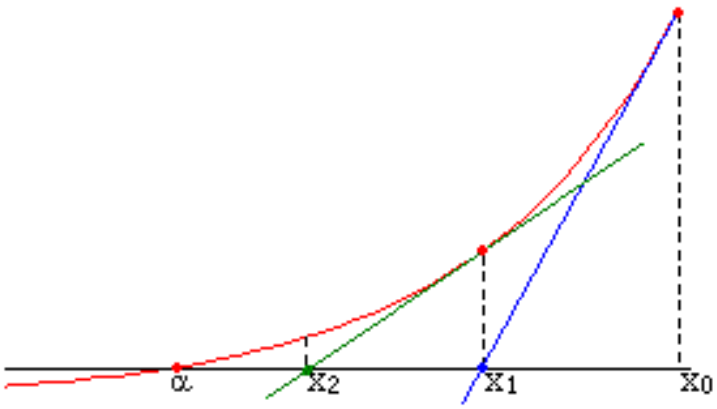
En viktig algoritme i numeriske beregninger er **Newton's metode**. Den brukes til å finne nullpunkt for en funksjon f, altså finne en løsning av likningen $f(x) = 0$. Metoden brukes også i optimering for å finne kritiske punkter, som er nullpunkt for den deriverte av funksjonen f (hvis f er deriverbar). I dette tilfellet ønsker vi altså å løse likningen $f'(x) = 0$.

Betrakt en deriverbar funksjon f (eller $f(x)$) fra $\mathbb{R}$ til $\mathbb{R}$. Vi ser altså på én-variabel situasjonen, men kan nevne at metoden kan generaliseres til å løse et system av n (ulineære) likninger i n variable. Dette innebærer å løse $F(\mathbf{x}) = \mathbf{0}$ der F en deriverbar funksjon fra $\mathbb{R}^n$ til $\mathbb{R}^n$. Metoden ble introdusert av Isaac Newton for visse funksjoner, men metoden ble brukt bl.a for å finne kvadratrøtter langt tidligere. Vi nøyer oss her med å forklare metoden for tilfellet $n = 1$, altså for funksjonen f nevnt over.

Betrakt likningen $f(x) = 0$. Idéen bak metoden er å forenkle problemet. Hvis f er en lineær (eller mer korrekt, affin) funksjon, er det greit: la $f(x) = ax + b$ der vi antar $a \neq 0$. (Hvis $a = 0$, har vi enten ingen eller uendelig mange løsninger, avhengig av om b er 0 eller ikke). Så vi får

$$f(x) = 0 \Leftrightarrow ax + b = 0 \Leftrightarrow x = -b/a.$$

Så, hvis f ikke er lineær, foretar vi en forenkling (approksimasjon) ved å se på tangenten i et punkt og finne hvor denne skjærer x-aksen, dvs. finne nullpunktet for den tilhørende lineære funksjonen.



Figur 2.12: Newton's metode (Kilde: Wikipedia)

Algoritmen beregner en følge $x_0, x_1, x_2, \ldots$. Velg et startpunkt x_0. Vi beskriver hvordan vi går fra x_k til neste punkt x_{k+1}. Tangenten til grafen til f i punktet $(x_k, f(x_k))$ er gitt ved

$$y(x) = f'(x_k)(x - x_k) + f(x_k).$$

Dette er en lineær funksjon av x. Anta at $f'(x_k) \neq 0$ (så tangenten er ikke vannrett). Ved å sette $y(x) = 0$ og løse for x får vi

$$x = x_k - \frac{f(x_k)}{f'(x_k)}.$$

Derfor lar vi neste punkt være

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \tag{2.1}$$

som gir Newton's metode. Man kan vise at under visse antakelser vil denne følgen $\{x_k\}$ konvergere mot et nullpunkt for f.

Eksempel 2.5.1. *Vi skal bruke metoden til å beregne kvadratrot. La $a > 0$, og vi skal beregne $\sqrt{a}$. Vi vil finne $x = \sqrt{a}$ og ser da på $f(x) = x^2 - a$. Så et positivt nullpunkt for f må være $\sqrt{a}$. Her blir $f'(x) = 2x$, så (2.1) gir*

$$x_{k+1} = x_k - \frac{x_k^2 - a}{2x_k}.$$

Dette kan vi også skrive som

$$x_{k+1} = \frac{1}{2}\big(x_k + \frac{a}{x_k}\big).$$

Denne metoden ble faktisk funnet opp av Heron av Alexandria for ca 2000 år siden! For eksempel, la $a = 5$, så vil vil finne $\sqrt{5}$ (tilnærmet). La oss starte med $x_1 = 5$. Da blir

$$x_2 = \frac{1}{2}(5 + \frac{5}{5}) = \frac{1}{2} \cdot 6 = 3$$

og de første tallene vi får med Heron's metode er:

n	x_n
1	5.000000
2	3.000000
3	2.333333
4	2.238095
5	2.236069
6	2.236068
7	2.236068

Når vi skriver med 6 desimaler er alle de neste tallene også lik $x_6 = 2.236068$, og dette er nær $\sqrt{5}$; alle disse sifrene er korrekte! Faktisk er

$$\sqrt{5} = 2.236067977499790\ldots$$

Hvis man er "tilstrekkelig nær" nullpunktet vil Newton's metode konvergere svært raskt. Vi har såkalt kvadratisk konvergens som betyr at antall riktige sifre typisk dobles for hvert skritt. Vi avbryter metoden når f.eks. endringen $|x_{k+1} - x_k|$ er under en viss liten toleranse.

2.6 Krumning og vendepunkter

I dette avsnittet skal vi studere funksjonenes krumningsegenskaper.

Definisjon 2.6.1. *En deriverbar funksjon $y = f(x)$* ***krummer opp*** *(resp.* ***ned****) i et intervall (a, b) dersom $f'(x)$ er voksende (resp. avtagende) i intervallet.*

Krumning innebærer at stigningstallet til funksjonen endrer seg. Krumning oppover betyr at funksjonen blir brattere, som betyr at stigningstallet øker. Stigningstallet til funksjonen er gitt ved den deriverte, så økende stigningstall betyr at den deriverte av den deriverte er positiv. Den deriverte til den deriverte til en funksjon kalles den dobbelt-deriverte og vi skriver $f''(x)$.

Teorem 2.6.2. *La $f(x)$ være to ganger deriverbar på et intervall I, dvs. at den deriverte $f'(x)$ også er deriverbar.*

i) Dersom $f''(x) \geq 0$ på I, så krummer grafen oppover.

ii) Dersom $f''(x) \leq 0$ på I, så krummer grafen nedover.

Bevis. Bruk Teorem 2.1.3 på den deriverte. □

Det at grafen krummer oppover kalles **konveks** (funksjonen er konveks), mens å krumme nedover kalles **konkav**. Dette er viktige egenskaper som vi vender tilbake til.

Vi kan bruke resultatet over til å bestemme type ekstremalpunkter for en funksjon $f(x)$. Dersom vi har et kritisk punkt x_0 med $f'(x_0) = 0$, så vil positiv krumning bety at vi har et minimumspunkt og negativ krumning bety at vi har et maksimumspunkt (se figur).

Dette leder oss fram til **andre-derivert-testen**.

Teorem 2.6.3. *La $f : \mathbb{R} \to \mathbb{R}$ være en to ganger deriverbar funksjon på et intervall I. Anta at $x_0 \in I$ er et kritisk punkt for f, dvs. $f'(x_0) = 0$. Da har vi*

i) Dersom $f''(x_0) > 0$ så er x_0 et lokalt minimumspunkt for f.

ii) Dersom $f''(x_0) < 0$ så er x_0 et lokalt maksimumspunkt for f.

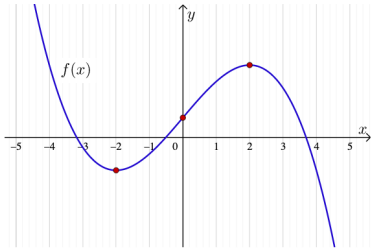
Definisjon 2.6.4. *Et punkt $x = c$ der funksjonen $f(x)$ skifter krumning fra opp til ned eller motsatt, kalles et **vendepunkt** for f.*

På samme måte som at ekstremalpunkter finnes der den deriverte skifter fortegn, finner vi vendepunkter der den dobbelt-deriverte skifter fortegn.

Teorem 2.6.5. *Dersom $x = c$ er et vendepunkt for $f(x)$, og $f''(c)$ er veldefinert, så er $f''(c) = 0$.*

Bevis. Et vendepunkt er et ekstremalpunkt for den deriverte, og Teorem 2.3.2 gir at $x = c$ er et kritisk punkt for $f'(x)$. Siden den deriverte av $f'(x)$ eksisterer i $x = c$, så følger det at $f''(c) = 0$. □

	$x < 0$	0	$0 < x < \frac{2}{3}$	$\frac{2}{3}$	$x > \frac{2}{3}$
$12x$	-	0	+	+	+
$3x - 2$	-	-	-	0	+
$f''(x)$	+	0	-	0	+
	opp	vendepunkt	ned	vendepunkt	opp



Figur 2.13: Lokalt minimum, lokalt maksimum og vendepunkt.

Eksempel 2.6.1. *Funksjonen $f(x) = x^3$ har et vendepunkt i $x = 0$ siden $f''(0) = 0$ og den dobbeltderiverte $f''(x) = 6x$ skifter tegn i $x = 0$.*

Merk at vi kan ha $f''(c) = 0$ uten at $x = c$ er et vendepunkt, slik som tilfellet er for funksjonen $f(x) = x^4$ i punktet $x = 0$. Den dobbelt-deriverte er gitt ved $f''(x) = 12x^2$. Den har et nullpunkt for $x = 0$, men $f''(x) > 0$ for både positive og negative verdier for x. Den dobbelt-deriverte $f''(x)$ skifter derfor ikke tegn i dette punktet, og $x = 0$ er ikke noe vendepunkt.

Eksempel 2.6.2. *Vi skal studere krumningsegenskapene til funksjonen*

$$f(x) = 3x^4 - 4x^3 + 3.$$

Vi regner ut den deriverte og den dobbelt-deriverte og finner deres nullpunkter. Dette bruker vi til å finne faktoriseringer:

$$f'(x) = 12x^3 - 12x^2 = 12x^2(x - 1),$$
$$f''(x) = 36x^2 - 24x = 12x(3x - 2).$$

For å finne monotoniegenskaper tegner vi fortegnsskjema for den deriverte:

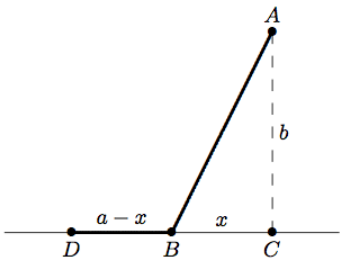
	$x < 0$	0	$0 < x < 1$	*1*	$x > 1$
$12x^2$	$+$	0	$+$		$+$
$x - 1$	$-$		$-$	0	$+$
$f'(x)$	$-$	0		0	$+$
	$\searrow$		$\searrow$	$\smile$	$\nearrow$

Vi ser at $x = 0$ er et kritisk punkt, men ikke noe ekstremalpunkt.

For å finne krumningsegenskapene tegner vi fortegnsskjema for den dobbelderiverte. Konklusjonen står i nederste linje.

I mange anvendelser betrakter vi størrelser som varierer med tiden. Et eksempel kan være en bil som kjører på en vei. Vi betegner avstanden bilen har kjørt fra et startpunkt ved funksjonen $s(t)$. Ved tiden $t = 0$ setter vi $s(0) = 0$. Den deriverte til funksjonen $s'(t)$ betegner farten til bilen ved tidspunktet t, mens den dobbelt-deriverte $s''(t)$ gir et uttrykk for akselerasjonen til bilen. Dersom $s'(t) = 0$ står bilen stille, mens $s''(t) = 0$ uttrykker at akselerasjonen er 0, eller at bilen beveger seg med konstant fart.

Eksempel 2.6.3. *Anta at vi skal gå til et punkt A ute i sanden og at vi befinner oss ved punktet D på veien illustrert ved linja DC (se figuren). Veien er rett og b gir avstanden fra A til det nærmeste punktet C på veien. Avstanden fra punktet D til punktet C kaller vi a. Anta at vi kan holde farten v langs veien og $w < v$ ute i sanden. Hvor langt skal vi gå langs veien før vi svinger ut i sanden for å bruke kortest mulig tid mellom D og A?*



Figur 2.14: Sand.

Vi kaller punktet hvor vi svinger ut i sanden for B og lar avstanden fra B til C være x. Det betyr at vi beveger oss en avstand $a - x$ langs veien, og ved Pythagoras vil avstanden fra B til A være $\sqrt{x^2 + b^2}$. Tid er gitt ved avstand delt med fart, og vi får samlet tid gitt ved funksjonen

$$f(x) = \frac{a - x}{v} + \frac{\sqrt{x^2 + b^2}}{w}.$$

Vi skal finne minimum av $f(x)$ når x ligger mellom 0 og a. Vi setter $f'(x) = 0$ og får

$$0 = f'(x) = -\frac{1}{v} + \frac{x}{w\sqrt{x^2 + b^2}}.$$

Det gir

$$w\sqrt{x^2 + b^2} = vx.$$

Vi kvadrerer begge sider og får

$$w^2(x^2 + b^2) = v^2x^2$$

som ved litt enkel algebra gir

$$x = \frac{wb}{\sqrt{v^2 - x^2}}.$$

Dette gir oss det kritiske punktet vi leter etter og (nokså opplagt) minimumspunktet.

Vi merker oss at a ikke inngår i dette uttrykket, noe som betyr at vi alltid skal gå til det samme punktet, uavhengig av utgangspunkt, men under en forutsetning, at det kritiske punktet ligger inne i intervallet $[0, a]$. Dersom dette ikke er tilfelle har vi ikke noe kritisk punkt inne i intervallet, og da vil minimumsverdien bli å finne i et av endepunktene. De to verdiene der er gitt ved

$$f(0) = \frac{a}{v} + \frac{b}{w}$$
$$f(a) = \frac{\sqrt{a^2 + b^2}}{w}.$$

Nå har vi

$$\begin{aligned} f(a)^2 - f(0)^2 &= \frac{a^2}{v^2} + \frac{2ab}{vw} + \frac{b^2}{w^2} - \frac{a^2 + b^2}{w^2} \\ &= \frac{a^2}{v^2} + \frac{2ab}{vw} - \frac{a^2}{w^2} \\ &= a^2(\frac{1}{v^2} - \frac{1}{w^2}) + \frac{2ab}{vw} \\ &= \frac{a^2(w^2 - v^2) - 2abvw}{v^2 w^2} \end{aligned}$$

Siden $w < v$ vil dette uttrykket alltid være negativt, noe som betyr at $f(0) > f(a)$ og minimum vil være i $x = a$. Oppsummert, vi skal alltid gå langs veien til et punkt som ligger i avstand $\frac{wb}{\sqrt{v^2-x^2}}$ fra C, og deretter gjennom sanden. Dersom vi starter nærmere C enn denne avstanden skal vi gå rett ut gjennom sanden.

2.7 L'Hôpitals regel

Anta at vi skal regne ut en grenseverdi $\lim_{x\to a} \frac{f(x)}{g(x)}$ der enten $f(x)$ og $g(x)$ begge går mot 0 eller begge går mot ∞. Da finnes det ikke noen generell regel for hva grenseverdien blir, men det finnes en metode som ofte hjelper oss å finne svaret. Den heter **L'Hôpitals regel** etter den franske matematikeren og adelsmannen Guillaume Francois Antoine, marquis de L'Hôpital (1661-1704).

Teorem 2.7.1 (L'Hôpitals regel)**.** *Anta at enten*

(i) $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = 0$

eller

(ii) $\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = \infty$.

Da er

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$$

forutsatt at den siste grensen eksisterer (den kan godt være ∞ *eller* $-\infty$*).*

L'Hôpitals regel finner altså ikke grenseverdien for oss, men fører oss videre til en annen grenseverdi som ofte er lettere å regne ut. Noen eksempler vil gjøre dette klarere.

Eksempel 2.7.1. *Regn ut*

$$\lim_{x\to 0} \frac{1-e^x}{x}.$$

Her går både teller og nevner mot 0 *(vi sier gjerne at vi har et "*$\frac{0}{0}$*"-uttrykk), og vi kan bruke L'Hôpitals regel:*

$$\lim_{x\to 0} \frac{1-e^x}{x} \overset{L'H}{=} \lim_{x\to 0} \frac{-e^x}{1} = \frac{-1}{1} = -1.$$

Legg merke til at vi i den andre grenseverdien $\lim_{x\to 0} \frac{-e^x}{1}$ *ikke har et "*$\frac{0}{0}$*"-uttrykk og derfor bare kan sette inn* $x = 0$*.*

Eksempel 2.7.2. *Regn ut*

$$\lim_{x\to\infty} \frac{e^x}{x}.$$

Her går både teller og nevner mot ∞ *(vi sier gjerne at vi har et "*$\frac{\infty}{\infty}$*"-uttrykk), og vi kan igjen bruke L'Hôpitals regel:*

$$\lim_{x\to\infty} \frac{e^x}{x} \overset{L'H}{=} \lim_{x\to\infty} \frac{e^x}{1} = \infty$$

(husk at L'Hôpitals regel også virker når grenseverdien er uendelig).

Det er ikke uvanlig at man må bruke L'Hôpitals regel flere ganger for å komme frem til svaret.

Eksempel 2.7.3. *Regn ut*

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}.$$

Vi bruker L'Hôpitals regel to ganger:

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}\overset{L'H}{=}\lim_{x\to 0}\frac{\sin x}{2x}\overset{L'H}{=}\frac{\cos x}{2}=\frac{1}{2}.$$

Når vi bruker L'Hôpitals regel flere ganger etter hverandre på denne måten, må vi huske på å sjekke at vi virkelig har et "$\frac{0}{0}$"- eller "$\frac{\infty}{\infty}$"-uttrykk hver gang vi bruker regelen.

Ofte kan vi omdanne andre typer uttrykk til "$\frac{0}{0}$"- eller "$\frac{\infty}{\infty}$"-uttrykk og så bruke L'Hôpitals regel til å finne grenseverdien. Her er et typisk eksempel:

Eksempel 2.7.4. *Regn ut*

$$\lim_{x\to 0} x\ln|x|.$$

I dette uttrykket går den ene faktoren, x, mot 0 mens den andre, $\ln|x|$, går mot $-\infty$, og det er ikke lett å se hvilken som "vinner". Omformer vi uttrykket til $\frac{\ln|x|}{\frac{1}{x}}$, får vi imidlertid et "$\frac{\infty}{\infty}$"-uttrykk vi kan bruke L'Hôpitals regel på (strengt tatt er det et "$\frac{-\infty}{\infty}$"-uttrykk, men regelen virker like godt på det). Dermed får vi

$$\lim_{x\to 0} x\ln|x|=\lim_{x\to 0}\frac{\ln|x|}{\frac{1}{x}}\overset{L'H}{=}\lim_{x\to 0}\frac{\frac{1}{x}}{-\frac{1}{x^2}}.$$

Hvis vi nå ganger med $-x^2$ over og under brøkstreken, ser vi at

$$\lim_{x\to 0} x\ln|x|=\lim_{x\to 0}\frac{\frac{1}{x}}{-\frac{1}{x^2}}=\lim_{x\to 0}(-x)=0.$$

Det kan være nyttig å vite at L'Hôpitals regel også gjelder for ensidige grenser, dvs. for uttrykk av typen $\lim_{x\to a^+}\frac{f(x)}{g(x)}$ og $\lim_{x\to a^-}\frac{f(x)}{g(x)}$ (hadde vi ikke hatt tallverdien i Eksempel 2.7.4, måtte vi ha nøyd oss med å regne ut den ensidige grensen $\lim_{x\to 0^+} x\ln x$ siden $\ln x$ bare er definert for $x>0$).

Vi har ikke redskaper til å bevise L'Hôpitals regel i dette kurset, men vi kan vise en forenklet versjon som sier at hvis $f(a)=g(a)=0$ og brøken $\frac{f'(a)}{g'(a)}$ eksisterer, så er $\lim_{x\to a}\frac{f(x)}{g(x)}=\frac{f'(a)}{g'(a)}$. Dette er bare et smart regnestykke der vi utnytter definisjonen av den deriverte:

$$\lim_{x\to a}\frac{f(x)}{g(x)}=\lim_{x\to a}\frac{f(x)-f(a)}{g(x)-g(a)}=\lim_{x\to a}\frac{\frac{f(x)-f(a)}{x-a}}{\frac{g(x)-g(a)}{x-a}}=\frac{f'(a)}{g'(a)}$$

Beviset for den fulle regelen er en litt mer sofistikert versjon av den samme ideen.

2.8 Partiell Derivasjon

Vi kan generalisere begrepet derivasjon til funksjoner i flere variable. Ideen er at vi tenker oss at alle bortsett fra én av de variable er konstanter. Vi betrakter så denne funksjonen som en funksjon i én variabel. Vanlig derivasjon med hensyn på denne variablen kalles **partiell derivasjon**. Partiell derivasjon med hensyn på en variabel x_i, uttrykker funksjonens endring i retningen langs med x_i-aksen.

Definisjon 2.8.1. *La $f(x_1, \ldots, x_n)$ være en funksjon i n variable. Den i'te* ***partielt deriverte*** *av f er gitt ved*

$$\begin{aligned}\frac{\partial f}{\partial x_i} &= \frac{\partial}{\partial x_i} f \\ &= \lim_{\Delta x \to 0} \frac{f(x_1, \ldots, x_i + \Delta x, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{\Delta x}\end{aligned}$$

dvs. at vi lar alle x_j, $j \neq i$ opptre som konstanter og deriverer med hensyn på x_i.

Eksempel 2.8.1. *La $f(x, y) = \cos(2x) - y^2 + xy$. Da har vi*

$$\frac{\partial f}{\partial x} = -2\sin(2x) + y, \quad \frac{\partial f}{\partial y} = -2y + x.$$

Nå kan vi la $x = 0$, og se på funksjonen $g(y) = f(0, y) = 1 - y^2$. Den deriverte av denne funksjonen er $g'(y) = -2y$. Dette er det samme som vi får ved å sette inn $x = 0$ i $\frac{\partial f}{\partial y}$. Tilsvarende kan vi se på $h(x) = f(x, 1) = \cos(2x) - 1 + x$. Den deriverte av denne er $h'(x) = -2\sin(2x) + 1$, som er det samme som den partiellderiverte av f med hensyn på x, innsatt $y = 1$.

Eksempel 2.8.2. *La $f(x, y) = \cos(xy)$. Da har vi*

$$\frac{\partial f}{\partial x} = -\sin(xy) \cdot y, \qquad \frac{\partial f}{\partial y} = -\sin(xy) \cdot x$$

Nå kan vi fortsette å derivere funksjonen i eksempelet, det gir oss fire muligheter

$$\begin{aligned}\frac{\partial}{\partial x}\frac{\partial f}{\partial x} &= -\cos(xy) \cdot y^2, \\ \frac{\partial}{\partial y}\frac{\partial f}{\partial x} &= -\cos(xy) \cdot xy - \sin(xy), \\ \frac{\partial}{\partial x}\frac{\partial f}{\partial y} &= -\cos(xy) \cdot xy - \sin(xy), \\ \frac{\partial}{\partial y}\frac{\partial f}{\partial y} &= -\cos(xy) \cdot x^2.\end{aligned}$$

Vi ser at de to i midten er like.

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x} = -\cos(xy)\cdot xy - \sin(xy),$$
$$\frac{\partial}{\partial x}\frac{\partial f}{\partial y} = -\cos(xy)\cdot xy - \sin(xy).$$

Tilfeldighet? Nei. Under visse (svake) antakelser er kryssderivasjon er uavhengig av rekkefølge:

$$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}.$$

Vi skal bruke notasjonen

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}$$

og hvis $x_i = x_j$;

$$\frac{\partial^2 f}{\partial x_i \partial x_i} = \frac{\partial^2 f}{\partial x_i^2}$$

Eksempel 2.8.3. *La $f(x,y) = x^3 + xy - y^2$. Da har vi*

$$\frac{\partial f}{\partial x} = 3x^2 + y, \qquad \frac{\partial f}{\partial y} = x - 2y.$$

Høyere ordens deriverte

$$\frac{\partial^2 f}{\partial x^2} = 6x, \quad \frac{\partial^2 f}{\partial x \partial y} = 1, \quad \frac{\partial^2 f}{\partial y^2} = -2.$$

Vi har til nå sett på $\mathbb{R}$-valuerte funksjoner i flere variable. Vi kan generalisere dette til vektor-valuerte funksjoner i flere variable.

Definisjon 2.8.2. *En avbildning*

$$F : \mathbb{R}^n \to \mathbb{R}^m$$

*for to heltall $n, m \geq 1$ kalles en **vektor-valuert** funksjon i n variable og vi skriver*

$$F(x_1, \ldots, x_n) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n))$$

hvor $f_i : \mathbb{R}^n \to \mathbb{R}$ for alle i.

Funksjonen er en sammenstilling av m funksjoner i n variable og derivasjon gjøres komponentvis. Vi skal få bruk for dette i det neste resultatet.

Eksempel 2.8.4. *Funksjonen $F : \mathbb{R}^2 \to \mathbb{R}^3$ gitt ved*

$$F(x_1, x_2) = (x_2 + x_1, 1 - x_2^2, 3x_1)$$

er en funksjon i to variable, hvor funksjonsverdien er en vektor i rommet. F.eks. er $F(2, 1) = (3, 0, 6)$.

Partiell derivasjon følger i store trekk de samme regnereglene som vanlig derivasjon. Den deriverte av en sum er summen av de deriverte, den deriverte av en konstant er 0 og produktregelen gjelder. Den regelen som trenger noe mer presisering er kjerneregelen. Vi skal formulere resultatet i den konteksten vi kommer til å trenge det, men det lar seg også generalisere ytterligere.

Teorem 2.8.3. *La $f : \mathbb{R}^n \to \mathbb{R}$ og $g = (g_1, \ldots, g_n) : \mathbb{R} \to \mathbb{R}^n$. La $h = f \circ g : \mathbb{R} \to \mathbb{R}$ være den sammensatte funksjonen, dvs. $h(t) = f(g(t))$ er en funksjon i en variabel t. Da har vi*

$$h'(t) = \frac{\partial f}{\partial x_1}\frac{dg_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dg_n}{dt}.$$

Bevis. Vi skriver

$$\begin{aligned} g(t + \Delta t) &= (g_1(t + \Delta t), \ldots, g_n(t + \Delta t)) \\ &\approx (g_1(t) + \Delta g_1, \ldots, g_n(t) + \Delta g_n) \end{aligned}$$

hvor $\Delta g_i \to 0$ når $t \to 0$. Det gir

$$\begin{aligned} \frac{h(t + \Delta t) - h(t)}{\Delta t} &= \frac{f(g(t + \Delta t)) - f(g(t))}{\Delta t} \\ &= \frac{f(g_1(t) + \Delta g_1, \ldots, g_n(t) + \Delta g_n) - f(g(t))}{\Delta t} \\ &= \frac{f(g(t) + (\Delta g_1, \ldots, \Delta g_n)) - f(g(t))}{\Delta t}. \end{aligned}$$

Vi skal bruke notasjonen

$$\Delta g^j = (\Delta g_1, \ldots, \Delta g_j, 0, \ldots, 0).$$

Da har vi

$$\begin{aligned} &f(g(t)+(\Delta g_1, \ldots, \Delta g_n)) - f(g(t)) \\ &\qquad = \sum_{i=1}^{n} f(g(t) + \Delta g^i) - f(g(t) + \Delta g^{i-1}) \end{aligned}$$

og derfor

$$\frac{h(t+\Delta t)-h(t)}{\Delta t} = \frac{f(g(t+\Delta t))-f(g(t))}{\Delta t}$$
$$= \sum_{i=1}^{n} \frac{f(g(t)+\Delta g^i)-f(g(t)+\Delta g^{i-1})}{\Delta g_i} \cdot \frac{g_i(t+\Delta t)-g_i(t)}{\Delta t}.$$

Når vi lar $\Delta t \to 0$ vil dette uttrykket gå mot summen som er gitt i teoremet. □

Eksempel 2.8.5. *Funksjonen $g : \mathbb{R} \to \mathbb{R}^2$ er gitt ved $g(t) = (t^2, t^3)$ og $f : \mathbb{R}^2 \to \mathbb{R}$ er gitt ved $f(x, y) = \sqrt{x+y}$. Vi har $h(t) = f(g(t)) = \sqrt{t^2+t^3}$ og for den deriverte*

$$h'(t) = \frac{1}{2}(t^2+t^3)^{-\frac{1}{2}}(2t+3t^2).$$

Dersom vi bruker formelen for den utvidede kjerneregelen, får vi

$$h'(t) = \frac{1}{2\sqrt{x+y}}2t + \frac{1}{2\sqrt{x+y}}3t^2$$

som ved innsetting for x og y gir oss samme uttrykk som over.

Eksempel 2.8.6. *La $g(t) = (1, t, e^t)$ og $f(x, y, z) = xy + \ln z$. Sett $h(t) = f(g(t))$. Da har vi*

$$h'(t) = y(t) \cdot 0 + x(t) \cdot 1 + \frac{1}{z(t)} \cdot e^t$$
$$= t \cdot 0 + 1 \cdot 1 + \frac{1}{e^t} \cdot e^t = 2$$

som stemmer overens med det vi får dersom vi regner ut uttrykket for h før vi deriverer. I det tilfellet har vi $h(t) = 1 \cdot t + \ln e^t = 2t$.

2.9 Kritiske punkter

Vi generaliserer begrepet kritisk punkt til teorien for funksjoner i flere variable.

Definisjon 2.9.1. *Et punkt $(x, y) = (a, b)$ kalles et **kritisk punkt** dersom de partielt deriverte i punktet enten ikke eksisterer eller er lik 0.*

Eksempel 2.9.1. *Vi deriverer funksjonen* $f(x,y) = \sin(\pi \cdot \sqrt{x^2+y^2})$ *og får*

$$\frac{\partial f}{\partial x} = \cos(\pi \cdot \sqrt{x^2+y^2}) \cdot \frac{\pi x}{\sqrt{x^2+y^2}}$$
$$\frac{\partial f}{\partial y} = \cos(\pi \cdot \sqrt{x^2+y^2}) \cdot \frac{\pi y}{\sqrt{x^2+y^2}}.$$

Det er lett å se at begge de partiellderiverte er 0 *når* $\sqrt{x^2+y^2} = \frac{1}{2}$*, mens det ikke er fullt så enkelt å se at de ikke eksisterer når* $x = y = 0$*. Men det gjør de ikke og vi har kritiske punkter både for* $\sqrt{x^2+y^2} = \frac{1}{2}$ *og for* $(x,y) = (0,0)$*.*

Ofte vil det være sammenfall mellom ekstremalpunkter og kritiske punkter, men ikke alltid.

Teorem 2.9.2. *Dersom* $f(x,y)$ *har et lokalt maksimum eller lokalt minimum i* (a,b)*, så er* (a,b) *et kritisk punkt.*

Bevis. Anta at $f(x,y)$ har et lokalt maksimum i (a,b). Det betyr at $f(x,y) - f(a,b) \leq 0$ for alle (x,y) i nærheten av (a,b). Anta videre

$$\frac{\partial f}{\partial x}(a,b) \approx \frac{f(a+h,b) - f(a,b)}{h} > 0$$

for alle små $h > 0$. Dette gir en motsigelse siden vi har antatt at $f(a+h,b) - f(a,b) \leq 0$. Anta så at

$$\frac{\partial f}{\partial x}(a,b) \approx \frac{f(a+h,b) - f(a,b)}{h} < 0$$

denne gangen for alle små $h < 0$. Igjen får vi en motsigelse. □

I motsetning til en-variabel-teorien kan vi i flervariabel-teorien ha et slags blandet tilfelle:

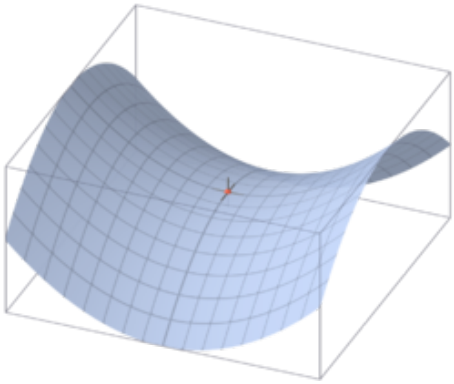
Eksempel 2.9.2. *Vi ser på grafen til funksjonen* $f(x,y) = x^2 - y^2$*.*

De partielt deriverte er gitt ved

$$\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = -2y.$$

Punktet $(0,0)$ *er et kritisk punkt. Det er et minimumspunkt dersom vi beveger oss gjennom det langs med* x*-aksen, og et maksimumspunkt langs med* y*-aksen. Dette er det lett å se ved å studere funksjonene*

$$f(x,0) = x^2, \quad f(0,y) = -y^2.$$



Figur 2.15: Et sadelpunkt.

Slike punkter kalles **sadelpunkter**, oppkalt etter sadelen på en hesterygg. Et sadelpunkt er et kritisk punkt på grafen til f, som verken er et lokalt maksimumspunkt eller et lokalt minimumspunkt.

Det gir oss tre forskjellige typer kritiske punkter for en funksjon $f(x, y)$, lokale minima, lokale maksima og sadelpunkter. For å finne ut hva slags punkt vi har, skal vi se på de andrederiverte til funksjonen i det aktuelle punktet. For et punkt (a, b) lar vi $H(a, b)$ betegne uttrykket

$$H(a,b) = \frac{\partial^2 f}{\partial x^2}(a,b)\frac{\partial^2 f}{\partial y^2}(a,b) - \big(\frac{\partial^2 f}{\partial x \partial y}(a,b)\big)^2$$

den såkalte **Hesse-determinanten**. Betegnelsen "determinant" her vil vi komme tilbake til senere. Vi har følgende resultat (som vi gir uten å føre et bevis):

Teorem 2.9.3. *La $f(x, y)$ være en to ganger deriverbar funksjon i to variable og anta at (a, b) er et kritisk punkt for f. Da har vi*

Dersom $H(a, b) > 0$ og $\frac{\partial^2 f}{\partial x^2}(a, b) > 0$, så er (a, b) et lokalt minimumspunkt.

Dersom $H(a, b) > 0$ og $\frac{\partial^2 f}{\partial x^2}(a, b) < 0$, så er (a, b) et lokalt maksimumspunkt.

Dersom $H(a, b) < 0$, så er (a, b) et sadelpunkt.

Dersom $H(a, b) = 0$ sier ikke denne testen oss noe.

Eksempel 2.9.3. *Vi studerer funksjonen* $f(x,y) = x^2 - y^2$. *De partielt deriverte er gitt ved*

$$\frac{\partial f}{\partial x} = 2x, \quad \frac{\partial f}{\partial y} = -2y.$$

Det betyr at vi har et kritisk punkt for $(x,y) = (0,0)$.

De dobbeltderiverte er gitt ved

$$\frac{\partial^2 f}{\partial x^2} = 2, \quad \frac{\partial^2 f}{\partial y^2} = -2.$$

Det gir for Hesse-determinanten i punktet $(0,0)$*;*

$$H = 2 \cdot (-2) - 0^2 = -4 < 0$$

som sier at det kritiske punktet er et sadelpunkt.

Eksempel 2.9.4. *Se på funksjonen* $f(x,y) = x^3 - y^2 + xy$. *Vi setter de partielt deriverte* $\frac{\partial f}{\partial x} = 3x^2 + y$ *og* $\frac{\partial f}{\partial y} = -2y + x$ *til å være lik* 0*, noe som gir oss kritiske punkter* $(0,0)$ *og* $(-\frac{1}{6}, -\frac{1}{12})$. *Vi regner ut*

$$H(x,y) = 6x \cdot (-2) - 1^2 = -12x - 1$$

Dette gir $H(0,0) = -1$ *og* $H(-\frac{1}{6}, -\frac{1}{12}) = 1$. *Samtidig har vi* $\frac{\partial^2 f}{\partial x^2}(-\frac{1}{6}, -\frac{1}{12}) = -1 < 0$. *Det betyr at* $(0,0)$ *er et sadelpunkt og* $(-\frac{1}{6}, -\frac{1}{12})$ *er et lokalt maksimum.*

2.10 Globale ekstremalpunkter

I motsetning til de lokale egenskapene, som sier noe om funksjonen i en liten omegn om et punkt, vil de **globale egenskapene** beskrive funksjonen på hele definisjonsområdet sett under ett. Vi fortsetter med å studere teorien i to variable, fordi det fortsatt er nokså klart hvordan det kan generaliseres til flere variable.

Definisjon 2.10.1. *(Globale maksimums- og minimumspunkter)*

i) Funksjonen $z = f(x,y)$ *har et* ***globalt minimum*** *i* (a,b) *dersom* $f(a,b) \leq f(x,y)$ *for alle punkter* (x,y) *i definisjonsområdet.*

ii) Funksjonen $z = f(x,y)$ *har et* ***globalt maksimum*** *i* (a,b) *dersom* $f(a,b) \geq f(x,y)$ *for alle punkter* (x,y) *i definisjonsområdet.*

Vi vil omtale funksjonsverdien i de globale minimums- og maksimumspunktene som funksjonens ***minimums-*** *og* ***maksimumsverdier*** *og samlet som funksjonenes* ***ekstremalverdier****.*

Funksjoner trenger ikke å ha globale ekstremalpunkter. Et eksempel er funksjonen $f(x, y) = x$ definert på hele (x, y)-planet. Den har ingen globale ekstremalpunkter fordi for alle punkter (a, b) i planet så vil det finnes punkter (x, y) i nærheten av (a, b) slik at $f(x, y) > f(a, b)$ og andre som oppfyller $f(x, y) < f(a, b)$. Funksjonen $f(x) = x$ har heller ingen ekstremalverdier dersom definisjonsområdet er et åpent område, dvs. et område som ikke inneholder sine randpunkter. På et lukket område derimot, dvs. et område som inneholder alle sine randpunkter, og som i tillegg er begrenset, så har funksjonen $f(x, y) = x$ både et globalt maksimumspunkt og et globalt minimumspunkt. Begrensede områder i planet er områder der alle punkter har begrenset avstand til origo (dvs. det finnes et tall $M > 0$ slik at $\sqrt{x^2 + y^2} < M$ for alle punkter (x, y) i området).

En kontinuerlig funksjon vil alltid oppnå sine ekstremalverdier dersom definisjonsområdet er et lukket og begrenset område. Dette er innholdet i **ekstremalverdi-teoremet** som vi gjengir uten bevis.

Teorem 2.10.2. *En funksjon $f(x)$ som er kontinuerlig på et lukket, begrenset område $D \subseteq \mathbb{R}^2$, vil oppnå sine maksimums- og minimums-verdier i området.*

Eksempel 2.10.1. *La $f(x, y) = x + y + 2$. Vi skal finne maksimumsverdien til $f(x, y)$ på enhetskvadratet $[0, 1] \times [0, 1]$.*

De partielt deriverte til funksjonen er konstanter og det finnes derfor ikke noen lokale ekstremalpunkter inne i det indre av området. De fire kantene til enhetskvadratet er gitt ved $(x, 0)$, $(1, y)$, $(x, 1)$ og $(0, y)$, hvor x og y ligger mellom 0 og 1. På de fire kantene vil funksjonen være $f(x, 0) = x + 2$, $f(1, y) = y + 3$, $f(x, 1) = x + 3$ og $f(0, y) = y + 2$. Det er enkelt å se at når x og y ligger mellom 0 og 1, så vil funksjonen ligge mellom 2 og 4. Det betyr at dette også er globale maksimums- og minimums-verdier.

2.11 Minste kvadraters metode

En interessant, men kanskje noe uventet anvendelse av partiell derivasjon er innen teorien for det som kalles **lineær regresjon** og da spesielt ved bruk av **minste kvadraters metode**. Vi skal se på dette via et konkret eksempel.

Forventet levealder for nyfødte jentebarn i Norge har utviklet seg slik de

siste årene:

2001	81.53
2002	81.52
2003	81.93
2004	82.33
2005	82.52
2006	82.66
2007	82.66
2008	82.95

Vi skal bruke minste kvadraters metode til å finne en formel som beskriver disse tallene. Vi lar x være år og y være forventet levealder. Hypotesen vår er at det er en lineær sammenheng mellom disse, dvs. $y = ax + b$. Vi indekserer de 8 dataparene fra 1 til 8, slik at f.eks. $(x_3, y_3) = (2003, 81.93)$. For hvert par beregner vi avviket fra den antatte rette linja, gitt ved $ax_i + b - y_i$. Disse tallene kvadrerer vi og summerer over $i = 1, 2, \ldots, 8$. Dernest skal vi finne de koeffisientene a og b slik at denne kvadratsummen blir minst mulig. Idéen er at vi betrakter kvadratsummen av avvikene fra den rette linja som en funksjon i de to konstantene a og b og bruker teorien til å finne minimumspunktet til kvadratsum-funksjonen uttrykt ved alle dataene vi har tilgjengelig.

For å gjøre regningene litt enklere bruker vi år 1,2,3 osv. Kvadratsummen blir

$$\begin{aligned}\sum_{i=1}^{8}(ax_i + b - y_i)^2 =& a^2\sum_{i=1}^{8}x_i^2 + 2ab\sum_{i=1}^{8}x_i + \sum_{i=1}^{8}b^2 \\ & - 2a\sum_{i=1}^{8}x_iy_i - 2b\sum_{i=1}^{8}y_i + \sum_{i=1}^{8}y_i^2.\end{aligned}$$

Vi setter inn verdiene og får dette kvadratiske avviket som en funksjon i a og b gitt ved

$$Q(a, b) = 204a^2 + 72ab + 8b^2 - 5941a - 1316b + 54139$$

Vi skal finne minimum for denne funksjonen. Vi partiellderiverer med hensyn på a og b og setter uttrykkene lik 0. Det gir

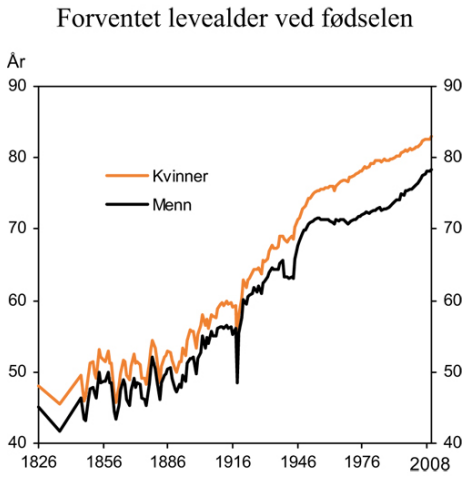
$$\begin{aligned}\frac{\partial Q}{\partial a} &= 408a + 72b - 5941, \\ \frac{\partial Q}{\partial b} &= 72a + 16b - 1316.\end{aligned}$$

som gir $a = 0.226$ og $b = 81.23$ og forventet levealder som en funksjon av årstall blir

$$y = 0.226(x - 2000) + 81.23 = 0.226x - 370.77.$$

Sammenlikner vi med de oppgitte tallene får vi

2001	81.53	81.46
2002	81.52	81.68
2003	81.93	81.91
2004	82.33	82.13
2005	82.52	82.36
2006	82.66	82.59
2007	82.66	82.81
2008	82.95	83.04



Figur 2.16: Utvikling av forventet levealder.

Vi ser at modellen gir oss en meget bra tilnærming tildataene.

Vi kan sette inn $x = 1866$ som er SSBs første registrering. Det gir forventet levealder $y = 50,95$, mens SSB oppgir 50,65 som gjennomsnitt for perioden 1866-1870. Vår lineære tilnærming ser altså ut til å stemme godt. I den motsatte tidsretning ser vi at $y = 100$ gir $x = 2083$, så med samme utvikling vil jentebarn født i 2083 ha en forventet levealder på 100 år.

Gjør vi den samme analysen i en generell setting med datapunkter

$$(x_i, y_i) \quad i = 1, 2, \ldots, n$$

får vi ut likninger for a og b;

$$\begin{aligned} a &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \\ b &= \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n} \end{aligned}$$

som kan brukes i alle eksempler.

Eksempel 2.11.1. *Gitt følgende data for korresponderende høyde og vekt for jentebarn i alderen 5-10 år. Vi skal finne en best mulig lineær sammenheng.*

x	*105*	*110*	*118*	*123*	*128*	*132*	*138*	*144*
y	*17.1*	*19.0*	*20.8*	*23.0*	*25.0*	*27.5*	*30.5*	*35.0*

Vi regner ut $\sum x_i = 998$, $\sum y_i = 197.9$, $\sum x_i^2 = 125766$ *og* $\sum x_i y_i = 25247.9$. *Setter vi* $\sum x_i = 998$, $\sum y_i = 197.9$, $\sum x_i^2 = 125766$ *og* $\sum x_i y_i = 25247.9$ *inn i formelen får vi*

$$\begin{aligned} a &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \\ &= \frac{8 \cdot 25247.9 - 998 \cdot 197.9}{8 \cdot 125766 - (998)^2} \approx 0.4424 \end{aligned}$$

$$a = 0.4424$$

og

$$\begin{aligned} b &= \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n} \\ &= \frac{197.9 - 0.4424 \cdot 998}{8} = -30.5. \end{aligned}$$

Det gir sammenhengen $y = 0.4424x - 30.5$ *for dataene. For hver centimeter jentene i denne aldersgruppen vokser i høyden, så blir de i gjennomsnitt* 442.4 *gram tyngre.*

2.12 Oppgaver

Oppgave 1. *Regn ut den deriverte av funksjonene.*

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

b) $f(x) = 3x^4 - 4x^3$

c) $f(x) = x + \sin x$

d) $f(x) = e^{2x}$

e) $f(x) = e^x \cos x$

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

g) $f(x) = x \ln x$

h) $f(x) = \frac{\sin x}{\cos x}$

Oppgave 2. *Regn ut alle partiellderiverte for funksjonene*

a) $f(x, y) = x^2 + y^3$

b) $f(x, y) = x + \sin x + 1$

c) $f(x, y) = x^2y^2 + xy$

Oppgave 3. *Regn ut alle partiellderiverte for funksjonene*

a) $f(x, y) = x \sin y + y \sin x$

b) $f(x, y) = e^{xy}$

c) $f(x, y) = \frac{x}{y} \quad y \neq 0$

Oppgave 4. *Regn ut alle partiellderiverte for funksjonene*

a) $f(x, y, z) = xy + yz + zx$

b) $f(x_1, x_2, x_3) = x_1x_2x_3 - \frac{1}{2}x_1^2x_3$

c) $f(x_1, x_2, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2$

Oppgave 5. *Finn de dobbeltderiverte av funksjonene og regn ut* $H(x, y)$.

a) $f(x, y) = x^2 + y^3$

b) $f(x, y) = xy$

c) $f(x, y) = x \sin y$

Oppgave 6. *Finn de dobbeltderiverte av funksjonene og regn ut* $H(x, y)$.

a) $fx, y) = e^{x+y}$

b) $f(x, y) = \sin(xy)$

Oppgave 7. *Finn de kritiske punktene til funksjonene*

a) $f(x, y) = x^2 - 2xy$

b) $f(x, y) = (y + 1)e^x - e^y$

c) $f(x, y) = 1 - x^2$

Oppgave 8. *Finn de kritiske punktene til funksjonene*

a) $f(x, y) = x^3 - 3xy^2$

b) $f(x, y) = x^2y^2 - 2xy$

c) $f(x, y) = \ln xy + x - y$

Oppgave 9. *Finn de kritiske punktene til funksjonene*

a) $f(x, y) = x^3 + y^3 - 3xy$

b) $f(x, y) = e^{2x+3y}(8x^2 - 6xy + 3y^2)$

c) $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$

Oppgave 10. *Vis at blant alle rektangler med et gitt areal, så har kvadratet den minste omkretsen.*

Oppgave 11. *Gitt $S > 0$. Vis at blant alle tall x og y slik at $x + y = S$, så er summen $x^2 + y^2$ minst når $x = y$.*

Oppgave 12. *Dersom et firma produserer x artikler av en bestemt vare vil de kunne selge alle til en pris $p = 5 - 0,0005x$, $0 \leq x \leq 8000$. Produksjonskostnaden ved å produsere x artikler er $C(x) = 500 + x$. Fortjenesten blir da $f(x) = (5 - 0,0005x)x - (500 + x)$.*

a) Hvor stor fortjeneste kan firmaet ha med disse varene?

Anta nå at to firmaer produserer samme vare til samme pris og konkurerer om de samme kundene. Vi antar at de to firmaene produserer henholdsvis x og y artikler. Fortjenesten til firma 1 er gitt ved

$$f_1(x, y) = (5 - 0,0005(x + y))x - (500 + x)$$

og for firma 2

$$f_2(x, y) = (5 - 0,0005(x + y))y - (500 + y)$$

b) Anta at begge firmaene maksimerer sin fortjeneste. Hvor stor er den maksimale fortjeneste for hvert firma nå?

c) Dersom firmaene opptrer som en enhet utad og bare deler fortjenesten, hvor mye ville de da kunne tjene? Kommenter forskjellen på svarene i b) og c).

Oppgave 13. *Et rektangel har hjørner* $(0,0)$, $(0,y)$, $(x,0)$ *og* (x,y). *Vis at forholdet mellom kvadratet av omkretsen og arealet tar sin største verdi når* $x = y$. *Hvor stor er denne verdien? Sammenlikn også svarene med forholdet mellom kvadratet av omkretsen og arealet av en sirkel.*

Oppgave 14. *Den generelle gassloven sier at*

$$p = \frac{mRT}{V}$$

der p *er trykket i gassen,* m *og* V *er gassens masse og volum,* T *er temperaturen og* R *er en konstant. Regn ut de partielt deriverte av* P *med hensy på* V *og* T.

Oppgave 15. *I denne oppgaven skal vi se på funksjonen*

$$f(x,y) = x^3 - xy + y^2 - x.$$

a) Finn de kritiske punktene til f.

b) Avgjør om de kritiske punktene er sadelpunkter, lokale minimumspunkter, eller lokale maksimumspunkter.

Oppgave 16. *Betrakt funksjonen*

$$f(x,y) = x^2 + 2xy - 6x - \tfrac{4}{3}y^3$$

definert over hele $\mathbb{R}^2$.

a) Finn de kritiske punktene til f.

b) Avgjør om de kritiske punktene er sadelpunkter, lokale minimumspunkter, eller lokale maksimumspunkter.

Oppgave 17. *Vi antar at folkemengden* $N = N(t)$ *i Norge vokser eksponensielt, etter formelen*

$$N(t) = Be^{at}$$

der a *og* B *er konstanter og* t *er tiden. I vårt eksempel regner vi tiden i tiår og at 1. januar 1900 er 0. Vi skal se på logaritmen til denne funksjonen, gitt ved*

$$\ln(N) = \ln(B) + at$$

Antakelsen om at veksten er eksponensiell gir at $\ln(N)$ *vokser lineært med tiden. Følgende tall for befolkningsmengden er observert:*

t	*1940*	*1950*	*1960*	*1970*	*1980*	*1990*	*2000*
N	*2.96*	*3.25*	*3.57*	*3.68*	*4.08*	*4.23*	*4.48*
$y = \ln(N)$	*1.09*	*1.18*	*1.27*	*1.30*	*1.41*	*1.44*	*1.50*

der befolkningstallene er i hele millioner. ($\sum t = 49$, $\sum t^2 = 371$, $\sum y = 9.19$, $\sum ty = 66.21$)

Bruk minste kvadraters metode til å finne en lineær sammenheng mellom tallene i første og tredje rad og bruk dette til å finne en formel for befolkningsstørrelsen. Når vil vi med denne modellen passere 5 millioner?

Oppgave 18. *Et firma har de første 9 årene etter oppstart hatt omsetning slik det er beskrevet i tabellen:*

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0.8	*1.0*	*1.4*	*1.9*	*2.0*	*2.6*	*3.5*	*4.4*	*4.7*

hvor omsetningen er i hele millioner.

Bruk minste kvadraters metode til å finne en lineær sammenheng mellom omsetningen og antall år etter oppstart. Hva sier modellen om omsetningen i år 10?

Oppgave 19. *Avgjør hvor funksjonen* $f(x) = x^3 + 1$ *vokser og hvor den (eventuelt) avtar. Finn de kritiske punktene til funksjonen. Er de ekstremalpunkter for funksjonen?*

Oppgave 20. *Betrakt funksjonen*

$$g(x) = \tan x = \frac{\sin x}{\cos x}$$

definert på det åpne intervallet $(0, \frac{\pi}{2})$*. Vis at funksjonen er strengt voksende i hele sitt definisjonsområde.*

Oppgave 21. *Vis at funksjonen* $f(x) = \frac{1}{x}$ *definert på intervallet* $(0, \infty)$ *ikke har noen ekstremalpunkter.*

Oppgave 22. *Finn de kritiske punktene til funksjonen* $f(x) = \cos 2x$*. Hvilke av dem er lokale maksimumspunkter og hvilke er lokale minimums-punkter?*

Oppgave 23. *Finn de kritiske punktene, bestem monotoni-egenskapene og finn de lokale ekstremalpunktene for funksjonene* $f(x) = x^2 - 3x + 2$, $g(x) = (x-1)^2(x+2)$ *og* $h(x) = x + \frac{1}{x^2}$, *alle definert over hele* $\mathbb{R}$.

Oppgave 24. *I de følgende oppgavene, i) finn alle kritiske punkter til* $f(x)$, *ii) beskriv monotoni-egenskapene til funksjonene ved å se på fortegnet til* $f'(x)$, *iii) finn lokale ekstremalpunkter, iv) bestem de intervallene der funksjonene krummer, henholdsvis opp og ned, v) finn funksjonenes vendepunkter, vi) finn globale maks og min*

a) $f(x) = x^3 - 4x$, $x \in [-1, 2]$

b) $f(x) = \frac{(x^2-4)}{(x^2-9)}$, $x \in [0, 2]$

c) $f(x) = x - \sin x$, $x \in [0, 2\pi]$

d) $f(x) = \sin^2 x$, $x \in [-\pi, \pi]$

Oppgave 25. *La funksjonen* f *være definert for alle relle tall* x *ved*

$$f(x) = e^{-\frac{1}{2}(x-a)^2}$$

hvor a *er en konstant.*

a) Avgjør hvor f *vokser og hvor den avtar. Finn eventuelle ekstremalpunkter for* f.

b) Finn eventuelle vendepunkter for f *og undersøk hvor grafen til* f *krummer opp og hvor den krummer ned.*

Oppgave 26. *Funksjonen* $f(t)$ *er gitt ved*

$$f(t) = t \ln t - t, \qquad \tfrac{1}{4} \leq t \leq 3$$

Avgjør hvor funksjonen f *vokser og hvor den avtar og finn ut hvor i definisjonsområdet den antar sin største/minste verdi.*

Oppgave 27. *Finn dimensjonene til rektangelet med størst areal når omkretsen er gitt lik 100.*

Oppgave 28. *Du har 60 meter gjerde til rådighet og skal bruke det til å gjerde inn et rektangel langs med en rett fjellside (hvor du ikke trenger å sette opp gjerde). Hva er det største arealet du kan gjerde inn?*

Oppgave 29. *Et budfirma har følgende begrensning på hvilke rektangulære pakker de kan levere: Summen av lengden og omkretsen må ikke overstige 108cm. Anta at du skal sende en pakke som er kvadratisk i den ene enden. Hva er det maksimale volumet til en akseptabel (for budfirmaet) pakke?*

Oppgave 30. *Et firma produserer små notisblokker med festlige motiv på forsiden. Dersom prisen for hver blokk er 10 kroner vil firmaet ikke få solgt noen, men for hver krone de reduserer prisen vil de kunne selge 500 blokker. De faste kostnadene ved produksjonen (uavhengig av antall) er kr. 3000, og produksjonsprisen pr. blokk er 2 kroner. Hva bør prisen være for at firmaet skal tjene mest mulig på blokkene?*

Oppgave 31. *La f være funksjonen gitt ved*

$$f(x) = x\sqrt{8 - x^2}\,, \quad -2\sqrt{2} \leq x \leq 2\sqrt{2}$$

a) Finn alle kritiske punkter for funksjonen f.

b) Bestem hvor funksjonen vokser og hvor den avtar, og finn maksimums- og minimumspunktene til f.

Oppgave 32. *Bruk L'Hôpitals regel til å vise at disse grensene er riktige:*

a) $\lim\limits_{x\to 0} \frac{\sin 3x}{x} = 3.$

b) $\lim\limits_{x\to 0} \frac{x}{e^{2x}-1} = \frac{1}{2}.$

c) $\lim\limits_{x\to 0} \frac{e^x-1-x}{x^2} = \frac{1}{2}.$

d) $\lim\limits_{x\to \infty} \frac{e^x}{x} = \infty.$

e) $\lim\limits_{x\to \frac{\pi}{2}} \frac{\cos x}{2x-\pi} = -\frac{1}{2}.$

f) $\lim\limits_{x\to 0} \frac{\tan x}{x} = 1.$

g) $\lim\limits_{x\to 0} \frac{\sin x - x}{x^3} = -\frac{1}{6}.$

h) $\lim\limits_{x\to 0} x^2 \ln|x| = 0.$

i) $\lim\limits_{x\to \infty} x^2 e^{-x} = 0.$

Oppgaver relatert til maskinlæring:

Oppgave 33. *Regn ut den deriverte av disse kjente funksjonene hentet fra maskinlæring.*

a) $\sigma(x) = \frac{1}{1+e^{-x}}$

b) $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$

c) $\mathit{Softplus}(x) = \ln(1 + e^x)$

d) $L_{BCE}(x) = -(y \ln x + (1-y)\ln(1-x))$

Oppgave 34. *Betrakt den veldig kjente funksjonen i maskinlæring, ReLU-funksjonen:*

$$\mathrm{ReLU}(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

a) Forklar hva denne funksjonen gjør. Finn deretter den deriverte, og forklar hva som skjer i $x = 0$.

b) En glatt tilnærming til $\mathrm{ReLU}(x)$ *er den såkalte softplus-funksjonen gitt ved*

$$s(x) = \ln\bigl(1 + e^x\bigr)$$

Finn den deriverte, og vis at den deriverte til $s(x)$ *er en annen veldig kjent funksjon i maskinlæring, nemlig sigmoid, gitt ved:*

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

c) Vis videre at den andre deriverte til $s(x)$*, er gitt ved* $s''(x) = \sigma'(x) = \sigma(x)(1 - \sigma(x))$.

d) Avgjør til slutt hvor s *vokser og hvor den avtar. Undersøk også hvor* s *krummer opp og hvor den krummer ned.*

Oppgave 35. *For å trene nevrale nettverk bruker vi gradientbaserte metoder. I denne oppgaven ser vi på et minimalt eksempel med en lineær funksjon fulgt av en ikke-linearitet:*

$$z = w\,x + b, \qquad a = s(z), \qquad L = (a - y)^2.$$

Her er $x, y, w, b \in \mathbb{R}$*,* $s : \mathbb{R} \to \mathbb{R}$ *er deriverbar (med derivert* $s'(z)$*), og* L *er et tall som måler "feil" (kvadratisk tap). Bruk kjerneregelen til å vise at*

$$\frac{\partial L}{\partial w} = 2\,(a - y)\,s'(z)\,x,$$

og

$$\frac{\partial L}{\partial b} = 2\,(a-y)\,s'(z).$$

Kapittel 3

Integrasjon

3.1 Bruk av integrasjon

Matematiske modeller som beskriver naturvitenskapelige fenomener er i mange tilfeller ensbetydende med å gi en eller flere **differensiallikninger.** For å løse differensiallikninger er utfordringen veldig ofte å finne funksjoner med en gitt derivert, dvs. det vi kaller anti-derivasjon. Et eksempel på dette er hastighet og akselerasjon. Akselerasjon er definert som endring av hastighet, og måles med den deriverte av hastighetsfunksjonen. Akselerasjonen til et legeme er styrt av kraftlover som vi tenker oss at vi kjenner fullt ut, og oppgaven blir å finne en hastighetsfunksjon med en derivert funksjon som passer med den aktuelle kraftloven. Dette kalles å anti-derivere funksjonen eller ubestemt integrasjon.

Samme type problemstilling dukker opp i forbindelse med **sannsynlighet**, og spiller en stor rolle i bruk innen **statistikk og data science**. Da ønsker man ofte å beregne forventningsverdien $\mathbb{E}X$ av en tilfeldig variabel X; dette er en slags teoretisk gjennomsnittsverdi for X. $\mathbb{E}X$ er definert som et integral av en bestemt funksjon som beskriver sannsynlighetsfordelingen til X. Du lærer mer om dette i kurs i statistikk/data science, og vil senere se at det dukker opp i maskinlæring.

I dette kapitlet skal vi studere **ubestemte og bestemte integral**. Ubestemte integral er det samme som anti-derivasjon, mens bestemte integral beregnes ved å evaluere ubestemte integral over et intervall slik at svaret blir et tall og ikke en funksjon. Dette tallet kan vi for en positiv funksjon tolke som arealet mellom grafen og x-aksen, eller i mer generelle tilfeller som den akkumulerte verdien for en funksjon over et tidsrom, strekning eller annet valg av argumentet x (eller t).

Et svært viktig resultatet er fundamentalteoremet (introdusert av Newton

og Leibniz) som knytter sammen derivasjon og integrasjon.

Når vi har gitt en funksjon, kan vi alltid regne ut dens deriverte, men det er ikke nødvendigvis mulig å finne et enkelt uttrykk for den anti-deriverte til funksjonen. Vi skal se på hvordan vi kan bruke numeriske beregninger til å bestemme verdien av et bestemt integral når vi ikke kan finne en anti-derivert til den aktuelle funksjonen. Videre skal vi ta for oss noen ulike fysiske tolkninger av det bestemte integral.

Det er ikke alltid at definisjonsområdet til en funksjon er et lukket intervall og da må vi gjøre noen modifikasjoner mht. definisjonen av det bestemte integralet. Det kan også hende at funksjonen vi skal integrere er ubegrenset, men vi kan likevel regne ut et areal under grafen. Dette vil involvere bruk av grenseverdier.

3.2 Ubestemt integrasjon

Vi starter med å gi en kort innføring i regnereglene for anti-derivasjon. Disse reglene framkommer i stor grad ved å se på tilsvarende regneregler for derivasjon, og bruke dem baklengs.

Definisjon 3.2.1. *Dersom $F(x)$[1] er en funksjon med derivert lik $F'(x) = f(x)$, så sier vi at $F(x)$ er en anti-derivert av $f(x)$, og vi skriver*

$$\int f(x)\,dx = F(x) + C.$$

Familien $F(x) + C$ for varierende konstanter C kalles det ***ubestemte integralet*** *av funksjonen f.*

I formelen over inngår det et uttrykk dx. Dette kalles et differensial og har en viktig betydning. For å gjøre det enkelt kan vi tenke på differensialet som en angivelse av hvilken variabel vi anti-deriverer med hensyn på. Bakgrunnen for leddet C på høyre side i formelen er som følger: Det er enkelt å vise at den deriverte av en konstant er 0, men det motsatte er også tilfelle.

En nyttig egenskap er at hvis en funksjon $f(x)$ oppfyller $f'(x) = 0$ for alle x i et intervall, så er funksjonen konstant, $f(x) = C$ i intervallet.

Vi skal ikke vise dette, men det virker kanskje intuitivt: straks en funksjon ikke er konstant, må det være et sted den vokser eller avtar. Fra dette resultater får vi følgende:

[1]Man skriver ofte $F(x)$ for funksjonen, men strengt tatt er F funksjonen og $F(x)$ en funksjonsverdi.

Teorem 3.2.2. *Dersom $F_1(x)$ og $F_2(x)$ begge er anti-deriverte til samme funksjon f på et intervall I, så er de like på en konstant nær, dvs. $F_1(x) - F_2(x) = C$.*

Dette er grunnen til at vi alltid må legge til en konstant når vi anti-deriverer. Vi kaller denne konstanten for en **integrasjonskonstant**.

Det finnes mange gode regneregler for ubestemt integrasjon, og vi skal presentere de viktigste.

I likhet med derivasjon er anti-derivasjon en lineær operasjon. Det innebærer at vi har de to formelene

$$\int (f(x) + g(x))\,dx = \int f(x)\,dx + \int g(x)\,dx$$

og

$$\int af(x)\,dx = a\int f(x)\,dx$$

for funksjoner $f(x)$ og $g(x)$, og reelle tall $a \in \mathbb{R}$. Dermed kan vi alltid anti-derivere ledd for ledd.

Eksempel 3.2.1. *Vi har*

$$\int (x^2 + 2x)\,dx = \int x^2\,dx + 2\int x\,dx.$$

Videre har vi en del *spesielle* integrasjonsregler:

Eksempel 3.2.2. *Det ubestemte integralet av potensfunksjonen er gitt ved*

$$\int x^n\,dx = \frac{1}{n+1}x^{n+1} + C \quad (n \neq -1).$$

Eksempel 3.2.3. *For det spesielle tilfellet der $n = -1$ har vi*

$$\int \frac{1}{x}\,dx = \ln x + C$$

der $\ln x$ er den naturlige logaritmefunksjonen.

Den naturlige logaritmefunksjonen er den inverse funksjonen til eksponensialfunksjonen e^x, dvs. vi har

$$e^{\ln x} = x, \quad \ln e^x = x.$$

Eksempel 3.2.4. *Det ubestemte integralet av eksponensialfunksjonen er gitt ved*

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax} + C.$$

Eksempel 3.2.5. *Det ubestemte integralet av de trigonometriske funksjonene er gitt ved*

$$\int \sin x \, dx = -\cos x + C$$

og

$$\int \cos x \, dx = \sin x + C.$$

I tillegg til de spesielle integrasjonsreglene har vi mer generelle integrasjonsregler. Den første bygger på formelen for derivasjon av et produkt av to funksjoner og kalles **delvis integrasjon**. Produktregelen sier at

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

Dersom vi flytter litt rundt på leddene og integrerer begge sider av likhetstegnet, får vi

$$\begin{aligned}\int f'(x)g(x)\,dx &= \int (f(x)g(x))'\,dx - \int f(x)g'(x)\,dx \\ &= f(x)g(x) - \int f(x)g'(x)\,dx\end{aligned}$$

siden derivasjon og integrasjon er omvendte operasjoner, dvs.

$$\int f'(x)\,dx = f(x) + C.$$

Dette viser seg å være en nyttig regel, og vi skal se på et eksempel på hvordan vi bruker den.

Eksempel 3.2.6. *Vi skal regne ut det ubestemte integralet av funksjonen* $x\sin x$*. Vi lar* $f'(x) = \sin x$ *og* $g(x) = x$ *i formelen over. Da får vi at* $f(x) = -\cos x$ *(Vi trenger ikke å ta med integrasjonskonstanten* C *ennå). Setter vi dette inn i formelen får vi*

$$\int x\sin x\,dx = -x\cos x - \int (-\cos x \cdot 1)\,dx$$

$$\boxed{\begin{array}{cc} g = x & f' = \sin x \\ g' = 1 & f = -\cos x \end{array}}$$

$$\begin{aligned}&= -x\cos x + \int \cos x\,dx \\ &= -x\cos x + \sin x + C\end{aligned}$$

Ved siste integrasjon tar vi med en integrasjonskonstant.

Eksempel 3.2.7. *Det ubestemte integralet av funksjonen* $f(x) = \ln x$ *er gitt ved*

$$\int \ln x \, dx = x \ln x - x + C$$

siden den deriverte av $x \ln x - x + C$ *med hensyn på* x *ved produktregelen er* $\ln x + x\frac{1}{x} - 1 = \ln x$.

Men vi kan også se dette ved å bruke delvis integrasjon. I så fall setter vi $f'(x) = 1$ *og* $g(x) = \ln x$. *(Dette er et lite triks, siden det på ingen måte er noe opplagt at man skal betrakte* $\ln x$ *som* $1 \cdot \ln x$.*) Dette gir*

$$\begin{aligned}\int \ln x \, dx &= \int 1 \cdot \ln x \, dx \\ &\boxed{\begin{array}{cc} g = \ln x & f' = 1 \\ g' = \frac{1}{x} & f = x \end{array}} \\ &= x \ln x - \int x \frac{1}{x} \, dx \\ &= x \ln x - x + C.\end{aligned}$$

En annen generell regneregel er **substitusjon**. Substitusjon baserer seg på baklengs bruk av kjerneregelen. Kjerneregelen sier noe om den deriverte til en sammensatt funksjon:

$$(f(g(x)))' = f'(g(x)) \cdot g'(x)$$

og regelen gir da at

$$\begin{aligned}\int f'(g(x)) \cdot g'(x) \, dx &= \int (f(g(x)))' \, dx \\ &= f(g(x)) + C.\end{aligned}$$

For å bruke denne regneregelen baklengs må vi identifisere de to leddene på høyre side av formelen med faktorer i den oppgitte integranden.

Eksempel 3.2.8. *Vi skal integrere funksjonen* $f(x) = 2x(x^2 + 1)^3$. *Hvis vi setter* $y = g(x) = x^2 + 1$ *og* $h(x) = \frac{1}{4}x^4$, *så ser vi at*

$$h(g(x)) = h(y) = \frac{1}{4}y^4 = \frac{1}{4}(x^2 + 1)^4$$

Det gir at $h'(y) = y^3$ *og dermed* $h'(g(x)) = (x^2+1)^3$. *Videre har vi* $g'(x) = 2x$. *Dermed følger det at*

$$\begin{aligned}\int 2x(x^2+1)^3\,dx &= \int h'(g(x))g'(x)\,dx \\ &= \int (h(g(x))'\,dx \\ &= h(g(x)) + C \\ &= \frac{1}{4}(x^2+1)^4 + C.\end{aligned}$$

Vi skal komme tilbake til flere eksempler på bruk av disse regnereglene.

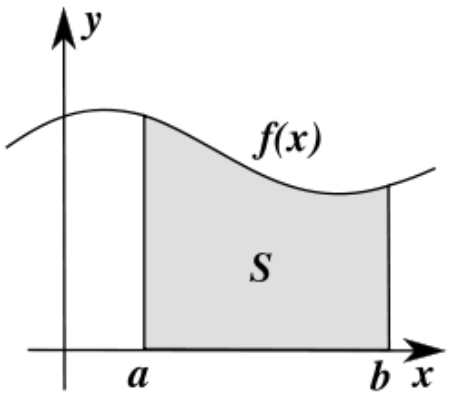
3.3 Bestemt integrasjon

Vi starter med å definere det bestemte integralet av en positiv funksjon over et intervall.

Definisjon 3.3.1. *La* $y = f(x)$ *være en positiv funksjon definert på et intervall* $[a, b]$. *Da er*

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

definert som arealet over intervallet $[a, b]$ *mellom* x*-aksen og grafen til* f.



Figur 3.1: Bestemt integral tolket som areal.

Man kan vise at følgende generelle regler gjelder for bestemte integral:

Linearitet ($\alpha, \beta \in \mathbb{R}$):

$$\int_a^b (\alpha f(x) + \beta g(x))dx \\ = \alpha \int_a^b f(x)dx + \beta \int_a^b g(x)dx.$$

Additivitet av definisjonsområde:
($[a, c] = [a, b] \cup [b, c]$)

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx.$$

Degenerert areal:

$$\int_a^a f(x)dx = 0.$$

De to siste er nokså opplagte ut i fra definisjonen, mens den første følger fra et tilsvarende resultat for ubestemte integral. I tillegg til disse tre definerer vi et bestemt integral beregnet i motsatt retning,

$$\int_a^b f(x)dx = -\int_b^a f(x)dx.$$

For å beregne det bestemte integralet til en vilkårlig funksjon (ikke kun positiv), deler vi opp intervallet i delintervaller slik at funksjonen enten er positiv eller negativ på delintervallene. For en negativ funksjon $g(x) < 0$ setter vi

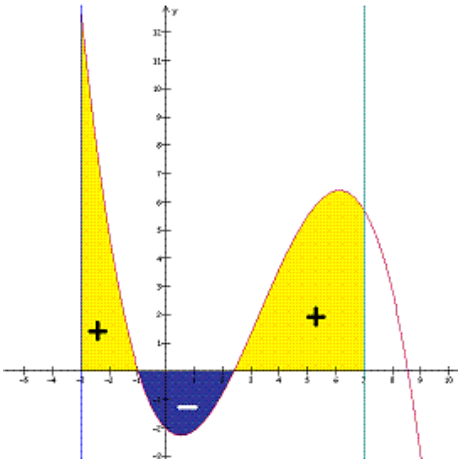
$$\int_a^b g(x)dx = -\int_a^b (-g(x))dx$$

hvor $-g(x)$ nå er en positiv funksjon.

3.4 Fundamentalteoremet

Det er flere måter å definere hva vi skal mene med en **integrerbar funksjon**. Vi skal ikke gå nærmere inn på det her, men nøye oss med å fastslå at alle kontinuerlige funksjoner er integrerbare.

Neste skritt er å finne metoder for å beregne bestemte integral. Det viktigste hjelpemiddelet vil være det såkalte fundamentalteoremet for differensial- og integralregningen. Dette resultatet setter i system det vi har sagt så langt om derivasjon og anti-derivasjon og gir oss linken mellom ubestemt og bestemt integrasjon.



Figur 3.2: Integralet er summen av arealene med positivt tegn over x-aksen og negativt tegn under.

Teorem 3.4.1. *La $f(x)$ være en kontinuerlig funksjon på et intervall $[a, b]$, og la $F(x)$ være en anti-derivert til $f(x)$. Da har vi*

$$F'(x) = \frac{d}{dx}\int_a^x f(t)\,dt = f(x)$$

og

$$\int_a^b f(x)dx = [F(x)]_a^b = F(b) - F(a)$$

Merk: Dette er den mest fundamentale egenskapen ved integralet og faktisk den egenskapen som er grunnlaget for hele differensial- og integralregningen.

Merk at vi også bruker notasjonen

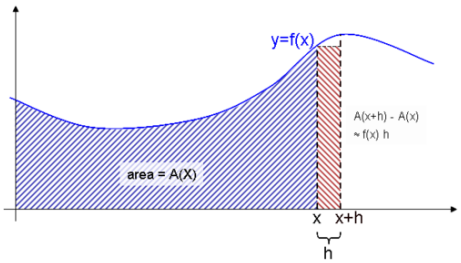
$$[F(x)]_a^b = F(b) - F(a).$$

I tillegg har vi her introdusert den ofte hensiktsmessige notasjonen

$$\frac{d}{dx}g(x) = g'(x).$$

Begge deler vil bli brukt i det som følger.

Figuren illustrerer fundamentalteoremet. Vi lar $A(x)$ være arealet under grafen fra $x = 0$ og ut til en vilkårlig x. Den relative tilveksten fra x til $x + h$ er gitt ved (se figur)



Figur 3.3: Fundamentalteoremet for differensial- og integralregningen.

$$\frac{A(x+h)-A(x)}{h} \approx \frac{f(x)\cdot h}{h} = f(x)$$

hvor uttrykkene har samme grenseverdi når $h \to 0$. Den deriverte av det bestemte integralet som arealfunksjon er altså funksjonen selv.

En praktisk nytte av fundamentalteoremet er at vi kan beregne areal ved å anti-derivere funksjoner. Ved å kombinere fundamentalteoremet med regnereglene for ubestemte integral får vi følgende liste over spesielle integrasjonsregler for bestemt integrasjon:

$$\begin{aligned}
\int_a^b x^k dx &= \frac{1}{k+1}(b^{k+1}-a^{k+1}),\\
\int_a^b \frac{1}{x} dx &= \ln b - \ln a = \ln\frac{b}{a},\\
\int_a^b e^{kx} dx &= \frac{1}{k}(e^{kb}-e^{ka}),\\
\int_a^b \sin x\, dx &= -\cos b + \cos a,\\
\int_a^b \cos x\, dx &= \sin b - \sin a.
\end{aligned}$$

Eksempel 3.4.1. *Vi skal beregne det bestemte integralet*

$$\int_0^{2\pi} \sin x\, dx.$$

Siden $\sin x$ *er positiv på intervallet* $[0,\pi]$ *og negativ på intervallet* $[\pi, 2\pi]$*, så deler vi opp integralet i to deler*

$$\int_0^{2\pi} \sin x\, dx = \int_0^{\pi} \sin x\, dx + \int_{\pi}^{2\pi} \sin x\, dx.$$

Vi vet at $\sin x = -\sin(x-\pi)$. *Det betyr at for hver* x *mellom* π *og* 2π *så finnes en* $u = x - \pi$ *mellom* 0 *og* π *hvor funksjonen* $\sin x$ *har akkurat samme verdi, men med motsatt fortegn. Dermed vil arealet mellom grafen og* x*-aksen for de to delene være nøyaktig likt, men med motsatt fortegn. Til sammen vil derfor integralet bli nullet ut og vi har*

$$\int_0^{2\pi} \sin x \, dx = 0.$$

Når vi beregner et bestemt integral trenger vi ikke å bry oss om integrasjonskonstanten. Grunnen er at dersom $F(x)$ er en anti-derivert til funksjonen f, og vi velger en annen anti-derivert $F(x) + C$, så vil

$$\begin{aligned}\int_a^b f(x)\,dx &= [F(x) + C]_a^b \\ &= F(b) + C - (F(a) + C) \\ &= F(b) - F(a) = [F(x)]_a^b\end{aligned}$$

så svaret blir det samme.

Eksempel 3.4.2. *Vi kan bruke dette til å beregne integralet av* $\sin x$ *over intervallet* $0 \le x \le 2\pi$ *som vi har sett på tidligere.*

$$\begin{aligned}\int_0^{2\pi} \sin x \, dx &= [-\cos x]_0^{2\pi} \\ &= -\cos(2\pi) + \cos 0 \\ &= -1 + 1 = 0.\end{aligned}$$

I tillegg til de spesielle integrasjonsreglene har vi de generelle integrasjonsteknikkene, *substitusjon* og *delvis integrasjon*. Disse er helt analoge med tilsvarende teknikker for å regne ut ubestemte integral.

Substitusjon er aktuelt hvis integralet har følgende form

$$\int_a^b f(g(x))g'(x)dx$$

for funksjoner f og g. Her har vi altså en sammensatt funksjon $f(g(x))$ multiplisert med $g'(x)$. Vi kaller da funksjonen $g(x)$ for *kjernen* og innfører variabelen $u = g(x)$; denne avhenger av x. Den deriverte av u med hensyn på x er

$$u' = g'(x) = \frac{du}{dx}.$$

Det siste uttrykket er en skrivemåte for den deriverte via såkalte "differensialer" (uendelig små størrelser) du og dx; se også slutten av seksjonen. Legg merke til at skrivemåten fanger opp definisjonen av derivert, som grenseverdi for brøken $\Delta u/\Delta x$, der Δu is endring i funksjonsverdi og Δx er endring i argumentverdi. Hvis vi faktisk tenker på $\frac{du}{dx}$ som en brøk, så får vi fra uttrykket over at

$$du = g'(x)dx.$$

Da har vi $u = g(x)$, $du = g'(x)\,dx$, som gir

$$\begin{aligned}\int_a^b f(g(x))g'(x)dx &= \int_{g(a)}^{g(b)} f(u)\,du \\ &= F(g(b)) - F(g(a)),\end{aligned}$$

hvor $F(x)$ er en anti-derivert til $f(x)$. Integrasjonsgrensene må endres slik fordi $x = a$ svarer til $u = g(a)$, og $x = b$ svarer til $u = g(b)$. Denne bruken av differensialer kan virke noe merkelig, men man kan vise matematisk at det blir korrekt.

Når det gjelder delvis integrasjon har vi

$$\int_a^b f(x)g'(x)dx = [f(x)g(x)]_a^b - \int_a^b f'(x)g(x)dx.$$

Eksempel 3.4.3. *Vi skal beregne det bestemte integralet*

$$\int_0^\pi x\cos x\,dx.$$

Vi bruker delvis integrasjon

$$\begin{aligned}\int_0^\pi x\cos x\,dx &= [u\cdot v]_0^\pi - \int_0^\pi u'\cdot v\,dx \\ &\qquad \boxed{\begin{array}{cc} u = x & v' = \cos x \\ u' = 1 & v = \sin x \end{array}} \\ &= [x\sin x]_0^\pi - \int_0^\pi 1\cdot\sin x\,dx \\ &= [x\sin x + \cos x]_0^\pi \\ &= \pi\sin\pi + \cos\pi - 0 - \cos 0 = -2.\end{aligned}$$

Eksempel 3.4.4. *Vi skal beregne det bestemte integralet*

$$\int_0^2 xe^{x^2}\,dx.$$

Vi bruker substitusjon

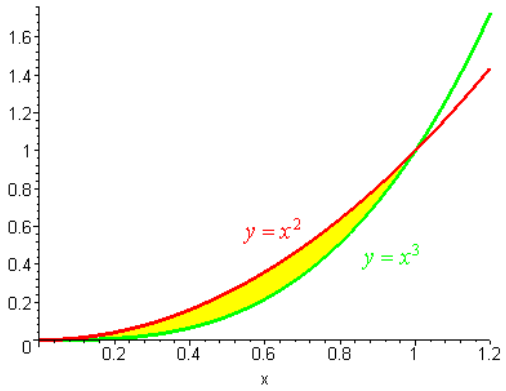
$$\int_0^2 xe^{x^2}\,dx = \int_0^2 \tfrac{1}{2}e^{x^2}\,2x\,dx$$

$u = x^2$	$du = 2x\,dx$
$x = 0$ *gir* $u = 0$	$x = 2$ *gir* $u = 4$

$$= \int_0^4 \tfrac{1}{2}e^u\,du$$
$$= [\tfrac{1}{2}e^u]_0^4 = \tfrac{1}{2}(e^4 - 1).$$

Vi kan bruke det bestemte integralet til å beregne arealet mellom to kurver $y = f(x)$ og $y = g(x)$, der $f(x) \geq g(x)$ på hele intervallet $[a, b]$:

$$\int_a^b (f(x) - g(x))dx = [\textit{Arealet mellom } f \textit{ og } g \textit{ på } [a, b]]$$



Figur 3.4: Arealet mellom grafene til $y = x^2$ og $y = x^3$.

Eksempel 3.4.5.

$$\int_0^1 (x^2 - x^3)dx = [\tfrac{1}{3}x^3 - \tfrac{1}{4}x^4] = \tfrac{1}{3} - \tfrac{1}{4} = \tfrac{1}{12}.$$

Eksempel 3.4.6.

$$\int_0^1 (x^{\frac{1}{3}} - x^{\frac{1}{2}})dx = [\tfrac{3}{4}x^{\frac{4}{3}} - \tfrac{2}{3}x^{\frac{3}{2}}] = \tfrac{3}{4} - \tfrac{2}{3} = \tfrac{1}{12}.$$

Det er ikke tilfeldig at de to siste eksempene gir samme svar. Hvorfor?

I alle integral-uttrykk inngår et *differensial* dx eller dt eller tilsvarende. Denne notasjonen ble innført av Gottfried Leibniz og har flere ulike tolkninger. Den fysiske tolkningen, slik Newton så det, er som en ørliten del av størrelsen x eller t. Differensialet dx har ingen utstrekning, men er heller ikke 0. Man kan tenke på den som en liten størrelse som vi lar gå mot 0. Vi kommer tilbake til denne tolkningen i avsnittet om Riemannsummer.

En annen tolkning kommer fra Leibniz selv. Han betraktet dx som et formelt uttrykk, som ikke skal tillegges noen fysisk størrelse, differensialet lever i en annen verden enn den fysiske størrelsen x. Operasjonen $x \mapsto dx$ tilordner til en størrelse x et differensial dx. Den motsatte operasjonen kalte Leibniz $\int$. Den tilordner en størrelse til et differensial, Altså $\int dx = x$, eller Leibniz sin versjon av fundamentalteoremet.

3.5 Riemannsummer

Vi skal se på en måte å beregne et areal under grafen til en funksjon og som også kan være med på å illustrere fundamentalteoremet. Vi tenker oss at vi deler opp intervallet $[a, b]$ i delintervaller med delingspunkter (kalles ofte en **partisjon** av intervallet)

$$a = a_0 < a_1 < \cdots < a_n = b$$

og at vi velger ut et punkt x_i^* i hvert delintervall, altså $x_i^* \in [a_i, a_{i+1}]$. Da vil summen (som vi kaller en Riemannsum, etter den tyske matematikeren Bernhard Riemann),

$$\sum_{i=0}^{n-1} f(x_i^*)(a_{i+1} - a_i)$$



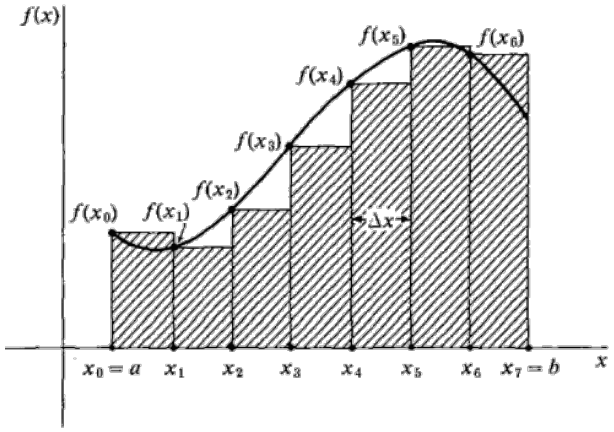
Bernhard Riemann, (1826-1866)

være en god approksimasjon til integralet $\int_a^b f(x)dx$, og bedre jo flere delingspunkter vi velger. Vi sier at vi approksimerer funksjonen ved hjelp av **trappefunksjoner**. Trappefunksjoner er funksjoner som er stykkvis konstante og derfor ser litt ut som trappetrinn. Fordelen med å bruke trappefunksjoner er at det er enkelt å regne ut arealet under grafen. Samtidig kan vi lage dem på en slik måte at når vi gjør en finere og finere oppdeling, dvs. at trinnene blir flere og flere og smalere og smalere, så vil grenseverdien gi oss det arealet vi egentlig vil beregne. Vi kan faktisk i mange tilfeller

definere

$$\int_a^b f(x)dx = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(x_i^*)(a_{i+1} - a_i).$$

Vi skal i alle eksempler holde oss til uniforme oppdelinger, dvs. der $a_{i+1} - a_i = \Delta x$ er den samme for alle i.



Figur 3.5: Trappefunksjon som approksimerer en funksjon, med x_i som venstre endepunkt i hvert delintervall. Merk at partisjonen her er skrevet $a = x_0 < x_1 < \cdots < x_7 = b$.

I noen tilfeller kan vi bruke trappefunksjoner direkte til å beregne et bestemt integral, men ofte vil dette involvere veldig mye regning. For en datamaskin er ikke mye regning noe problem og gode numeriske tilnærminger til et bestemt integral kan derfor lett gjennomføres med datamaskiner.

I det neste eksempelet skal vi beregne arealet av en trekant, med grunnlinje g og høyde h. For enkelthet skyld lar vi trekanten være rettvinklet. En vilkårlig trekant kan alltid deles opp i rettvinklede trekanter og ved å summere areal kan vi lett generalisere formelen til å gjelde for alle trekanter.

Eksempel 3.5.1. *Vi lar trekanten ligge med den rette vinkelen inn mot x- og y-aksen. Grunnlinjen vil da ligge langs x-aksen fra 0 til g, og høyden opp til hypotenusen fra et punkt x på grunnlinjen vil være gitt ved $y = \frac{h}{g}(g - x)$. Dette er akkurat likningen til en rett linje som går gjennom punktene $(g, 0)$ og $(0, h)$. Det betyr at vi må beregne det bestemte integralet*

$$\int_0^g \frac{h}{g}(g - x)dx$$

Vi deler opp grunnlinjen i n like store deler, med $\Delta x = \frac{g}{n}$*, som gir oss en partisjon:*

$$0 < \frac{g}{n} < \frac{2g}{n} < \frac{3g}{n} < \cdots < \frac{(n-1)g}{n} < \frac{ng}{n} = g.$$

Vi skal beregne

$$\sum_{i=1}^{n} f(x_i^*)\Delta x$$

med x_i^* *som høyre endepunkt i delintervallene, dvs.* $x_i^* = \frac{ig}{n}$ *og med* $f(x^*) = \frac{h}{g}(g - x^*)$*. Dette gir oss arealet under trappefunksjonen gitt ved*

$$\begin{aligned}
&\frac{h}{g}(g - \frac{g}{n}) \cdot \frac{g}{n} + \cdots + \frac{h}{g}(g - \frac{(n-1)g}{n}) \cdot \frac{g}{n} \\
&= \frac{h}{g}\frac{g}{n}((n-1)g - \frac{g}{n}(1 + 2 + \cdots + (n-1))) \\
&= \frac{gh}{n}(n - 1 - \frac{1}{n} \cdot \frac{1}{2}n(n-1)) \\
&= \frac{gh}{n}(\frac{1}{2}n - \frac{1}{2}) \\
&= \frac{gh}{2} - \frac{gh}{2n} \to \frac{gh}{2} \qquad \textit{når } n \to \infty
\end{aligned}$$

dvs. arealet er gitt ved $\frac{gh}{2}$*, som forventet.*

I det neste eksemplet skal vi bruke den velkjente summeformelen

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}.$$

Eksempel 3.5.2. *Vi skal beregne*

$$\int_0^1 x^2 dx$$

ved hjelp av Riemannsummer. Partisjon av enhetsintervallet, med $\Delta x = \frac{1}{n}$*:*

$$0 < \frac{1}{n} < \frac{2}{n} < \frac{3}{n} < \cdots < \frac{n-1}{n} < \frac{n}{n} = 1$$

Vi skal beregne

$$\sum_{i=1}^{n} f(x_i^*)\Delta x$$

med x_i som høyre endepunkt i delintervallene, dvs. $x_i^* = \frac{i}{n}$ og $f(x) = x^2$. Dette gir oss arealet under trappefunksjonen gitt ved

$$\begin{aligned}(\frac{1}{n})^2 \cdot \frac{1}{n} + (\frac{2}{n})^2 \cdot \frac{1}{n} + (\frac{3}{n})^2 \cdot \frac{1}{n} + \cdots + (\frac{n}{n})^2 \cdot \frac{1}{n} \\ &= \frac{1}{n^3}(1^2 + 2^2 + 3^2 + \cdots + (n)^2) \\ &= \frac{1}{n^3}\frac{(n+1)n(2n+1)}{6} \\ &= \frac{2n^3 + 3n^2 + n}{6n^3} \\ &= \frac{1}{3} + \frac{1}{2n} + \frac{1}{6n^2} \to \frac{1}{3} \qquad \text{når } n \to \infty\end{aligned}$$

dvs.

$$\int_0^1 x^2 dx = \frac{1}{3}.$$

3.6 Trapesmetoden

Vi fortsetter med å se på det bestemte integralet, spesielt på hvordan vi kan bruke numeriske beregninger til å bestemme verdien når vi ikke nødvendigvis kan finne en anti-derivert.

Vi skal bare se på én metode for å gjøre slike numeriske tilnærminger av et bestemt integral; det finnes ulike metoder som dekkes i andre emner på MI. Metoden kalles **trapesmetoden** (tilnærminger med trapeser eller lineær approksimasjon). Vi deler intervallet $I = [a, b]$ i n like store deler

$$a = a_0 < a_1 < \cdots < a_n = b.$$

Da har vi følgende formel,

$$\begin{aligned}\int_a^b f(x)dx = (\frac{f(a_0)}{2} + f(a_1) + f(a_2) + \ldots \\ + f(a_{n-1}) + \frac{f(a_n)}{2}) \cdot \Delta x + E_T\end{aligned}$$

der E_T er et restledd som gir avviket i tilnærmingen, dvs. det som mangler for at summen på høyresiden skal være lik integralet på venstresiden.

Teorem 3.6.1. *Anta at $f''(x)$ eksisterer og er kontinuerlig på intervallet $[a, b]$ og at $|f''(x)|$ har maksimumsverdi K på intervallet. Da er feilen E_T i trapesmetoden med n like store delingsintervaller begrenset ved*

$$|E_T| \leq \frac{K(b-a)^3}{12n^2}.$$

Det er et viktig poeng ved dette restleddet. Siden nevneren i uttrykket er proporsjonalt med n^2, så vil en dobling av antall punkter i tilnærmingen gi en feil som er rundt en fjerdedel. Tredobling av antall punkter vil gi en feil som er rundt en ni-del osv. Det betyr at jo finere oppdeling vi har, dvs. jo flere delingspunkter, jo bedre blir approksimasjonen.

Trapesmetoder baserer seg på at vi tilnærmer integranden med stykkvis lineære funksjoner, og så regner ut integralet av disse funksjonene i stedet for den originale funksjonen. Dersom integranden er en lineær funksjon i utgangspunktet burde den selv være sin egen beste approksimasjon, og trapesmetoden burde gi oss et eksakt svar.

Eksempel 3.6.1. *Vi skal bruke trapesmetoden på integralet* $\int_0^1 x\,dx$ *med* n *delingspunkter. Det gir*

$$\begin{aligned}\int_0^1 x\,dx &\approx \big(\frac{1}{2}\cdot 0 + \frac{1}{n} + \frac{2}{n} + \cdots + \frac{n-1}{n} + \frac{1}{2}\cdot 1\big)\frac{1}{n} \\ &= \frac{1}{n}\big(\frac{1}{2}(n-1)n\frac{1}{n} + \frac{1}{2}\big) = \frac{1}{2}\end{aligned}$$

som er det samme som vi ville fått ved å finne en anti-derivert og så sette inn.

Eksempel 3.6.2. *Vi skal estimere integralet* $\int_0^2 \sqrt{1+x^4}\,dx$ *ved trapesmetoden med fire like store delintervaller. Vi har* $\Delta x = 0.5$ *og derfor* $x_0 = 0$, $x_1 = 0.5$, $x_2 = 1$, $x_3 = 1.5$ *og* $x_4 = 2$. *Det gir*

$$\begin{aligned}\int_0^2 \sqrt{1+x^4}\,dx &\approx \big(\frac{1}{2}\sqrt{1+0^4} + \sqrt{1+0.5^4} \\ &\quad + \sqrt{1+1^4} + \sqrt{1+1.5^4} + \frac{1}{2}\sqrt{1+2^4}\big)\cdot 0.5 \\ &= \big(\frac{1}{2} + \sqrt{1.065} + \sqrt{2} + \sqrt{6.0625} + \frac{1}{2}\sqrt{17}\big)\cdot 0.5 \\ &\approx 3.73.\end{aligned}$$

Setter vi $f(x) = \sqrt{1+x^4}$ *får vi* $f'(x) = \frac{2x^3}{\sqrt{1+x^4}}$ *og* $f''(x) = \frac{6x^2+2x^6}{(1+x^4)^{\frac{3}{2}}}$. *Ved å derivere enda en gang er det relativt enkelt å se at den dobbelt-deriverte har sitt maksimum for* $x = 1$ *og at vi har* $|f''(x)| < 3$. *Det gir* $|E_T| < \frac{1}{8} = 0.125$. *Det betyr at*

$$3.73 - 0.125 \le \int_0^2 \sqrt{1+x^4}\,dx \le 3.73 + 0.125.$$

Eksempel 3.6.3. *Vi skal estimere integralet* $\int_0^1 \frac{1}{1+x^2}\,dx$ *ved trapesmetoden med fire like store delintervaller. Vi har* $\Delta x = \frac{1}{4}$ *og derfor* $x_0 = 0$, $x_1 = \frac{1}{4}$, $x_2 = \frac{1}{2}$, $x_3 = \frac{3}{4}$ *og* $x_4 = 1$. *Det gir*

$$\begin{aligned}\int_0^1 \frac{1}{1+x^2}\,dx &= \left(\frac{1}{2}\cdot\frac{1}{1+0^2} + \frac{1}{1+(\frac{1}{4})^2} + \frac{1}{1+(\frac{1}{2})^2}\right.\\ &\quad \left. + \frac{1}{1+(\frac{3}{4})^2} + \frac{1}{2}\cdot\frac{1}{1+1^2}\right)\cdot\frac{1}{4}\\ &= \left(\frac{1}{2} + \frac{16}{17} + \frac{4}{5} + \frac{16}{25} + \frac{1}{4}\right)\cdot\frac{1}{4} \approx 0.7828.\end{aligned}$$

I dette tilfellet kan man vise at $|f''(x)| < 1$, *og vi får* $|E_T| < \frac{1}{192} < 0.0053$. *Vi kan røpe at den faktiske verdien av integralet er* $\frac{\pi}{4} \approx 0.7854$, *så vi ser at vi holder oss innenfor feilestimatet:*

$$0.7828 - 0.0053 \leq \int_0^1 \frac{1}{1+x^2}\,dx = \frac{\pi}{4} \approx 0.7854 \leq 0.7828 + 0.0053.$$

3.7 Oppgaver

Oppgave 1. *Regn ut det bestemte integralet av potensfunksjonen* $f(x) = x^n$ *over det lukkede intervallet* $[0, 1]$.

Oppgave 2. *Regn ut de bestemte integralene.*

a) $\int_0^3 x^2\,dx$

b) $\int_{-3}^3 x^2\,dx$

c) $\int_0^2 4x^3\,dx$

d) $\int_{-2}^2 4x^3\,dx$

Oppgave 3. *Regn ut de bestemte integralene.*

a) $\int_{-1}^2 (t+1)\,dt$

b) $\int_{-1}^0 (x+1)^2\,dx$

c) $\int_0^1 x^{\frac{1}{3}}\,dx$

d) $\int_0^\pi (x + \sin x)dx$

Oppgave 4. *Regn ut de bestemte integralene*

a) $\int_e^{e^2} \frac{1}{x}(\ln x)\,dx$

b) $\int_1^4 \frac{1}{x} + \sqrt{x}\,dx$

c) $\int_0^{2\pi} \sin t\,\sin(\omega t)\,dt$, $\omega \neq \pm 1$

d) $\int_0^1 xe^{3x^2}\,dx$

e) $\int_0^1 xe^{2x}\,dx$

Oppgave 5. *Finn arealet avgrenset av grafen til* f, x*-aksen og den rette linja* $(x = a)$.

a) $f(x) = 5x$ *og* $x = 1$

b) $f(x) = 3x^2$ *og* $x = 1$

c) $f(x) = \sin x$ *og* $x = \frac{\pi}{2}$

Oppgave 6. *Finn arealet avgrenset av grafen til f, x-aksen og de to rette linjene.*

a) $f(x) = \sqrt{x}$ *og* $x = 1$, $x = 4$

b) $f(x) = x + \frac{1}{x}$ *og* $x = 1$, $x = 2$

c) $f(x) = \frac{1}{x^2}$ *og* $x = 1$, $x = 3$

Oppgave 7. *Regn ut arealet mellom de grafene, avgrenset av de to rette linjene.*

a) $f(x) = x^3$, $g(x) = 3x^2 - 6$ *og* $x = 0$, $x = 2$

b) $f(x) = x^4$, $g(x) = 2x^2 - 1$ *og* $x = -1$, $x = 1$

c) $f(x) = e^x$, $g(x) = e^{-x}$ *og* $x = 0$, $x = \ln 2$

d) $f(x) = \cos x$, $g(x) = \sin x$ *og* $x = 0$, $x = \frac{\pi}{4}$

e) $f(x) = x^2$, $g(x) = x + 2$ *og* $x = -1$, $x = 2$

Oppgave 8. *Regn ut arealet mellom grafene til f og g, avgrenset av de to rette linjene.*

a) $f(x) = 9 - x^2$, $g(x) = x^2 + 1$ *og* $x = -2$, $x = 2$

b) $f(x) = \sqrt{4 + x}$, $g(x) = 21 + \frac{1}{4}x$ *og* $x = 0$, $x = 12$

c) $f(x) = \frac{1}{2}x$, $g(x) = e^{-x}$ *og* $x = 0$, $x = \ln 2$

Oppgave 9. *Regn ut arealet av området mellom kurvene $y = x^2 + x + 1$ og $y = -2x^2 + 4x + 7$ mellom skjæringspunktene for de to kurvene.*

Oppgave 10. *Estimer integralet ved å bruke trapesmetoden, med like store delintervaller og det oppgitte antallet delintervaller.*

a) $\int_0^1 e^{\sqrt{x}}\,dx$, *og* $n = 4$

b) $\int_{-1}^2 \frac{1}{1+x^2}\,dx$, *og* $n = 6$

c) $\int_{-1}^1 \sqrt{1 - x^2}\,dx$, *og* $n = 10$

d) $\int_{-1}^1 \frac{1}{\sqrt{1+x^2}}\,dx$, *og* $n = 8$

Oppgave 11. *Finn en tilnærmet verdi for integralet $\int_1^2 \frac{1}{x}\,dx$ ved Simpsons metode, først med $n = 2$, deretter $n = 4$, og så $n = 8$. Sammenlikn svaret med $\ln 2$. Hvordan endrer restleddet seg?*

Oppgave 12. *Følgende data er kjent for en funksjon $f(x)$.*

x	0	1	2	3	4
$f(x)$	0.1	0.4	0.5	1	-0.1

Vi skal estimere $\int_0^4 f(x)\,dx$

a) Ved trapesmetoden

b) Ved Simpsons metode

Oppgave 13. *a) Gi et estimat for integralet $\int_0^4 x^4 - 9x^3 + 22x^2 - 17x\,dx$ ved å bruke trapesmetoden og Simpsons metode med $\Delta x = 1$.*

b) Beregn integralet $\int_0^4 x^4 - 9x^3 + 22x^2 - 17x\,dx$ eksakt.

Oppgave 14. *Finn en tilnærmet verdi for integralet $\int_0^1 x\,dx$ ved Simpsons metode, først med $n = 2$, deretter $n = 8$.*

Oppgaver relatert til maskinlæring:

Oppgave 15. *La $p(x) = c\,e^{-\lambda x}$ for $x \in [0, 1]$ (med $\lambda > 0$) og $p(x) = 0$ ellers. For at p skal være en gyldig sannsynlighetstetthet må*

$$\int_0^1 p(x)\,dx = 1.$$

Finn c for at $p(x)$ skal være en gyldig sannsynlighetstetthet.

Oppgave 16. *En vanlig måte å evaluere en binær klassifikasjonsmodell på er* AUROC *(Area Under the ROC Curve). ROC-kurven viser sammenhengen mellom* TPR *(true positive rate) og* FPR *(false positive rate) når beslutningsgrensen varierer. AUROC er arealet under denne kurven.*

Anta at ROC-kurven er gitt ved

$$\mathrm{TPR}(u) = \sqrt{u} \qquad \textit{for} \quad \mathrm{FPR} = u \in [0, 1].$$

Beregn

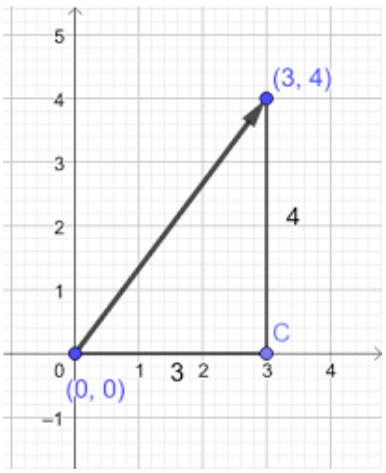
$$\mathrm{AUROC} = \int_0^1 \mathrm{TPR}(u)\,du = \int_0^1 \sqrt{u}\,du.$$

Kapittel 4

Vektorer, matriser og lineære likninger

4.1 Vektorer

Et grunnleggende begrep i lineær algebra er vektorer. En vektor kan illustreres geometrisk som en pil, med gitt retning og lengde. Retningen på vektoren følger retningen på pila. To vektorer er like dersom de har samme lengde og retning. Vektorenes plassering i planet (eller rommet) har ingen betydning. To eksempler på vektorer, hentet fra fysikk, er hastighet og kraft. Både hastighet og fart skjer i vektorens retning og lengden av vektoren illustrerer henholdsvis farten og styrken i krafta.



Figur 4.1: Vektor.

En annen måte å framstille vektorer er som såkalte tall-tupler, f.eks. $(3, 4)$. Den geometriske tolkningen av vektoren $(3, 4)$ er en pil som starter i origo og ender i punktet $(3, 4)$. Dette gir oss en 1-1-sammenheng mellom punkter i planet, vektorer og tallpar. Det fine med å bruke tall-tupler er at denne notasjonen ikke begrenser seg til å knytte vektorer til geometri i planet eller rommet. 5-tuplet $(2, 4, -1, 3, 0)$ har ikke noen klar geometrisk tolkning, men kan likevel være svært nyttig for oss i mange sammenhenger.

Det er vanlig å bruke fet skrift for å betegne en vektor, som i $\mathbf{v}$ og $\mathbf{w}$, eller ved en pil over vektornavnet, som i $\overrightarrow{v}$. For lengden av en vektor bruker vi notasjonen $\|\mathbf{v}\|$. Siden lengden av en vektor $\mathbf{v} = (a_1, a_2)$ i planet,

$$\|\mathbf{v}\| = \|(a_1, a_2)\| = \sqrt{a_1^2 + a_2^2}$$

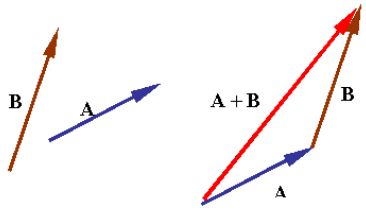
er gitt ved å bruke Pythagoras formel, etteraper vi dette også for mer generelle vektorer, $\mathbf{v} = (a_1, a_2, \ldots, a_n)$ hvor n er et vilkårlig positivt heltall, og sier at lengden av $\mathbf{v}$ er gitt ved

$$\|\mathbf{v}\| = \|(a_1, a_2, \ldots, a_n)\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

Eksempel 4.1.1.

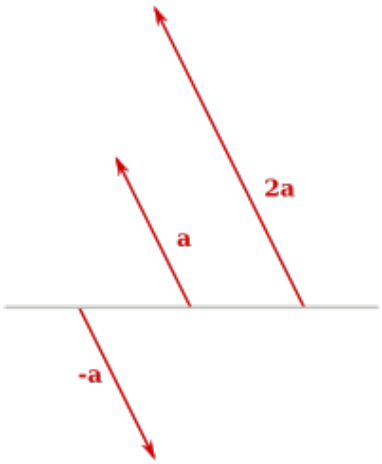
$$\|(2, 4, -1, 3, 0)\| = \sqrt{2^2 + 4^2 + (-1)^2 + 3^2 + 0^2} = \sqrt{30}.$$

Vi kan legge sammen vektorer. Hvis de to vektorene er illustrert ved hver sin pil, finner vi summen av dem ved å legge pilene etter hverandre. **Summen av vektorene** er da gitt ved en pil som starter i startpunktet for den ene vektoren og ender i endepunktet for den andre.



Figur 4.2: Vektoraddisjon.

Hvis vi multipliserer en vektor med et tall, så beholder vi retningen og endrer lengden i henhold til størrelsen på tallet. Vektoren $2\mathbf{v}$ peker dermed i samme retning som $\mathbf{v}$, men er dobbelt så lang. Multiplikasjon med negative



Figur 4.3: Multiplikasjon.

tall gir oss vektorer i motsatt retning, så vektoren $-\mathbf{v}$ er like lang som $\mathbf{v}$, men peker stikk motsatt vei. Spesielt har vi $1 \cdot \mathbf{v} = \mathbf{v}$.

Vi har tilsvarende operasjoner for n-tupler. La

$$\mathbf{v} = (a_1, a_2, \ldots, a_n)$$

Da har vi

$$2\mathbf{v} = (2a_1, 2a_2, \ldots, 2a_n), \quad -\mathbf{v} = (-a_1, -a_2, \ldots, -a_n)$$

Tallene som inngår i en vektor kalles **koordinater** og addisjon og skalarmultiplikasjon skjer **koordinatvis** dvs. vi gjør regneoperasjonene for hver enkelt koordinat for seg;

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$
$$k \cdot (a_1, a_2, \ldots, a_n) = (ka_1, ka_2, \ldots, ka_n)$$

for to vektorer $(a_1, a_2, \ldots, a_n)$ og $(b_1, b_2, \ldots, b_n)$ og et reelt tall $k \in \mathbb{R}$. Den eneste begrensningen her er at **størrelsen** (dvs. antall koordinater) på vektorene må være den samme.

Vi kaller multiplikasjon av en vektor med et tall for **multiplikasjon med skalar**. Multiplikasjon med skalar tilfredsstiller de samme reglene som vanlig multiplikasjon, dvs. $r(s\mathbf{v}) = (rs)\mathbf{v}$ og $(r + s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$ for reelle tall r og s.

Eksempel 4.1.2.

$$\begin{aligned}(2+3)(2,-1,3,0) &= 2(2,-1,3,0)+3(2,-1,3,0)\\ &= (4,-2,6,0)+(6,-3,9,0)\\ &= (10,-5,15,0),\\ 5(2,-1,3,0) &= (10,-5,15,0).\end{aligned}$$

Vektoraddisjon er **kommutativ**

$$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a},$$

og **assosiativ**

$$\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$$

dvs. at rekkefølgen på vektorene og vektoraddisjonene ikke spiller noen rolle. Subtraksjon av to vektorer definerer vi ved å kombinere multiplikasjon av en vektor med et negativt tall med vektoraddisjon. Vi definerer null-vektoren $\mathbf{0}$ som en vektor med bare 0-er

$$\mathbf{0}=(0,0,\ldots,0)$$

Den har lengde 0 og har derfor ikke noen spesiell retning. Videre har vi $0\cdot\mathbf{a}=\mathbf{0}$ for alle vektorer $\mathbf{a}$. Vektoren med kun 1-ere skriver vi som

$$\mathbf{1}=(1,1,\ldots,1).$$

Vi kaller vektoren $\mathbf{e}_i=(0,0,\ldots,0,1,0,\ldots,0)$ med kun nuller bortsett fra en 1-er i i-te koordinat for den i-te **standard enhetsvektor**.

Det viser seg at det er hensiktsmessig å bruke to ulike måter å angi en vektor. Vi kan skrive den som en **radvektor**;

$$[a_1\ a_2\ \cdots\ a_n]$$

eller som en **kolonnevektor**;

$$\begin{bmatrix} a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}.$$

Vi skal også identifisere en slik vektor med n-tuppelet $(a_1,a_2,\ \ldots,a_n)$, men vi vil unngå at ting kan misforståes. Disse skrivemåtene angir essensielt den samme vektoren, men vi skal bruke dem litt forskjellig. Bakgrunnen for denne

forskjellen vil bli klarere når vi kommer til matriser. Overgangen fra den ene til den andre formen har også et eget navn; vi kaller det å **transponere**, og skriver

$$[a_1\ a_2\ \cdots\ a_n]^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

og omvendt. Vektorene er de **transponerte** til hverandre.

Det finnes også et produkt mellom vektorer. Dette produktet gir oss ikke en ny vektor, derimot et tall. Produktet kalles **skalarprodukt** eller **prikkprodukt** av de to vektorene. For å regne ut skalarproduktet av to vektorer må vektorene ha samme størrelse, la oss kalle dem $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ og $\mathbf{w} = (w_1, w_2, \ldots, w_n)$.

Definisjon 4.1.1. *Skalarproduktet av to vektorer* $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ *og* $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ *er gitt ved*

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &= (v_1, v_2, \ldots, v_n) \cdot (w_1, w_2, \ldots, w_n) \\ &= v_1 w_1 + v_2 w_2 + \cdots + v_n w_n. \end{aligned}$$

Skalarproduktet har mange anvendelser som vi etterhvert kommer til, men kanskje den viktigste er at skalarproduktet er nært knyttet til lengden av en vektor:

$$(v_1, v_2, \ldots, v_n) \cdot (v_1, v_2, \ldots, v_n) = v_1^2 + v_2^2 + \cdots + v_n^2 = \|(v_1, v_2, \ldots, v_n)\|^2,$$

eller i kortform

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v}.$$

Eksempel 4.1.3.

$$(2, 1, 3, -1) \cdot (1, 0, 0, -2) = 2 + 0 + 0 + 2 = 4.$$

$$\|(2, 1, 3, -1)\|^2 = 2^2 + 1^2 + 3^2 + (-1)^2 = 15.$$

Vi skal nå definere et svært viktig begrep innen lineær algebra, nemlig lineærkombinasjoner. Hvis vi har gitt vektorer $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ kan vi produsere enda flere vektorer ved å danne **lineærkombinasjoner** av vektorene.

Definisjon 4.1.2. *En **lineærkombinasjon** av vektorene $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ er en sum*

$$\mathbf{w} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m$$

for reelle tall $a_1, \ldots, a_m \in \mathbb{R}$.

Dersom dette er den *eneste måten* vi kan skrive $\mathbf{w}$ som en lineærkombinasjon av $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ på sier vi at $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ er **lineært uavhengig**, eller at mengden med disse vektorene er lineært uavhengig.

Dersom en liste av vektorer er lineært uavhengig, vil denne egenskapen gjelde for alle valg av $\mathbf{w}$, også for nullvektoren $\mathbf{0}$. Den formelle definisjonen av lineær uavhengighet ser derfor slik ut:

Definisjon 4.1.3. *En mengde $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ av vektorer sies å være **lineært uavhengig** dersom*

$$\sum_{i=1}^{m} a_i\mathbf{v}_i = \mathbf{0}$$

*medfører at alle $a_i = 0$, $i = 1, 2, \ldots, m$. En mengde av vektorer som ikke er lineært uavhengig sies å være **lineært avhengig**.*

Eksempel 4.1.4. *Vektorene*

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$$

er lineært avhengige siden

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

mens vektorene

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

er lineært uavhengige siden

$$a_1 \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} + a_3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

kun er mulig dersom $a_1 = a_2 = a_3 = 0$.

Vi skal komme tilbake til dette i et senere kapittel.

4.2 Matriser

Det neste begrepet vi skal se på er **matriser**.

Definisjon 4.2.1. *En reell $m \times n$-**matrise** A er en tabell med m **rader** og n **kolonner**.*

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix},$$

*der **elementene** a_{ij}, $i = 1, \dots, m$, $j = 1, \dots, n$, i matrisen er reelle tall. Vi bruker ofte den kortfattede skrivemåten*

$$A = [a_{ij}].$$

De to tallene m og n kan være hvilke som helst positive heltall. Matrisen

$$\begin{bmatrix} 1 & -1 \\ -3 & 0 \\ 1 & 2 \end{bmatrix}$$

er en 3×2-matrise, mens

$$\begin{bmatrix} 1 & -3 & 1 \\ -1 & 0 & 2 \end{bmatrix}$$

er en 2×3-matrise.

Vi trenger ikke å begrense oss til reelle elementer i matrisen, vi kan gjerne studere komplekse matriser, hvor elementene i matrisen er komplekse tall. Men vi velger å holde oss til reelle tall dersom ikke annet er angitt.

Vi kan også betrakte vektorer som matriser. En **radvektor** $[a_1\ a_2\ \dots\ a_n]$ er en $1 \times n$-matrise, og en **kolonnevektor**

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$$

er en $m \times 1$-matrise. Det vil vise seg at dette er svært hensiktsmessig.

Hvis vi har to matriser av samme størrelse, kan vi legge dem sammen, omtrent som vi gjør med vanlige tall.

Definisjon 4.2.2. *La A og B være to reelle $m \times n$-matriser, $A = [a_{ij}]$ og $B = [b_{ij}]$. Da er summen av dem gitt ved $A + B = [a_{ij} + b_{ij}]$, som igjen er en reell $n \times m$-matrise. Vi kan også gange en matrise med et reelt tall c. Det gjør vi ved å gange alle elementene i matrisen med det samme tallet, $c \cdot [a_{ij}] = [ca_{ij}]$. Vi kaller denne operasjonen for **multiplikasjon med skalar**.*

Eksempel 4.2.1. *La*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 2 \\ 2 & 0 & -1 \end{bmatrix} \text{ og } B = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 1 & -1 \\ 2 & 1 & 0 \end{bmatrix}.$$

Da har vi

$$A + B = \begin{bmatrix} 1-1 & 2+0 & 0+1 \\ 1+1 & -1+1 & 2-1 \\ 2+2 & 0+1 & -1+0 \end{bmatrix} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 4 & 1 & -1 \end{bmatrix}.$$

Eksempel 4.2.2. *La A være som i forrige eksempel. Da har vi*

$$2A = 2\begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 2 \\ 2 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 0 \\ 2 & -2 & 4 \\ 4 & 0 & -2 \end{bmatrix}.$$

Vi kan også multiplisere en matrise med det reelle tallet 0. Det gir oss **0**-matrisen (f.eks. av størrelse 2×3).

$$\mathbf{0} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Eksempel 4.2.3. *Vi har*

$$-\begin{bmatrix} 1 & -1 \\ -3 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} -1 & 1 \\ 3 & 0 \\ -1 & -2 \end{bmatrix}$$

og

$$\begin{bmatrix} 1 & -1 \\ -3 & 0 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ 3 & 0 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Vi kan gi en oppsummering av de formelle egenskapene for matriser vi har studert:

Proposisjon 4.2.3. *La A, B og C være reelle $m \times n$-matriser. Da har vi at:*

i) Addisjon av matriser er assosiativ, dvs.

$$(A + B) + C = A + (B + C).$$

ii) Addisjon av matriser er kommutativ, dvs.

$$A + B = B + A.$$

iii) Null-matrisen er et 0*-element, det vil si* $A + \mathbf{0} = A$.

iv) Enhver matrise har en additiv invers, det vil si en matrise $-A = (-1)A$ *som oppfyller* $A + (-A) = \mathbf{0}$.

v) Skalar multiplikasjon er assosiativ, dvs.

$$c(dA) = (cd)A.$$

vi) Skalar multiplikasjon er distributiv over addisjon av matriser, dvs.

$$(c + d)A = cA + dA$$

og

$$c(A + B) = cA + cB.$$

vii) $1 \cdot A = A$.

Bevisene for alle disse påstandene er mer eller mindre trivielle og vi lar det være opp til leseren å skrive dem ut.

Under forutsetning av at størrelsene passer sammen kan matriser multipliseres med hverandre. Dette er litt komplisert, så vi skal gjøre beskrivelsen i to skritt.

Vi starter med å se på multiplikasjon av en matrise med en kolonnevektor. La A være en $m \times n$ matrise med kolonner $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, så vi kan skrive

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

La x være en $n \times 1$ matrise, altså en kolonnevektor, så

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

Her er første definisjon:

Definisjon 4.2.4. *La* A *være en* $m \times n$ *matrise* A *og* $\mathbf{x}$ *en* $n \times 1$ *matrise (kolonnevektor) som over. Da er produktet* $A\mathbf{x}$ *gitt ved lineærkombinasjonen*

$$A\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n.$$

Dette er dermed en $m \times 1$ *matrise (kolonnevektor).*

Legg merke til at dette er lineærkombinasjonen av $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ med koeffisienter $x_1, x_2, \ldots, x_n$. Altså: *produktet* $A\mathbf{x}$ *av en* $m \times n$ *matrise* A *med en* $n \times 1$ *matrise* $\mathbf{x}$ *(kolonnevektor) er en lineærkombinasjon av kolonnene i* A*, der* $\mathbf{x}$ *inneholder koeffisientene*; dette gir en kolonnevektor med m komponenter. Merk: for at dette skal ha mening må antall kolonner i A være lik antall komponenter i $\mathbf{x}$.

Eksempel 4.2.4. *La*

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

Da er

$$A\mathbf{x} = \begin{bmatrix} -2 \\ 0 \\ 5 \end{bmatrix}$$

som er lineærkombinasjonen gitt ved én ganger første kolonne i A *minus to ganger tredje kolonne i* A*.*

Dette kan vi så bruke til å definere matriseprodukt AB der B kan ha flere kolonner. La A være en $m \times n$ matrise som over og la B være en $n \times p$ matrise med kolonner $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_p$, så

$$B = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_p \end{bmatrix}.$$

Definisjon 4.2.5. *Med notasjon som over definerer vi matriseproduktet* AB *ved*

$$AB = \begin{bmatrix} A\mathbf{w}_1 & A\mathbf{w}_2 & \cdots & A\mathbf{w}_p \end{bmatrix}.$$

Så produktet av en $m \times n$ *matrise og en* $n \times p$ *matrise er en* $m \times p$ *matrise.*

Dette betyr at hver kolonne i matrisen AB er en lineærkombinasjon av kolonnene i A. Koeffisientene som brukes her er elementene i hver kolonne i B.

Ved litt regning kan man se at hvis $AB = [c_{ij}]$ så er

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

som vi gjenkjenner som skalarproduktet av rad i i A og kolonne j i B.

Eksempel 4.2.5. *La*

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ 3 & 0 \\ -1 & -2 \end{bmatrix}$$

være en 3×3*-, henholdsvis* 3×2*-matrise. Da passer størrelsene slik at vi kan definere produktet* AB*.*

$$(3 \times 3) \cdot (3 \times 2) = (3 \times 2)$$

Det blir en ny 3×2*-matrise*

$$AB = \begin{bmatrix} 5 & -2 \\ -3 & 0 \\ 1 & 5 \end{bmatrix}$$

hvor elementene framkommer slik;

$$\begin{aligned} 5 &= 0 \cdot (-1) + 2 \cdot 3 + 1 \cdot (-1) \\ -2 &= 0 \cdot 1 + 2 \cdot 0 + 1 \cdot (-2) \\ -3 &= 2 \cdot (-1) + 0 \cdot 3 + 1 \cdot (-1) \\ &\textit{osv.} \end{aligned}$$

Legg merke til at andre kolonne i AB *er det vi regnet ut i forrige eksempel.*

De formelle egenskapene ved matrise-multiplikasjon er som følger:

Proposisjon 4.2.6. *La* A*,* B *og* C *være reelle matriser av rett størrelse for de aktuelle formålene i denne proposisjonen. Da har vi at:*

i) Matrise-multiplikasjon er assosiativ, dvs.

$$(AB)C = A(BC).$$

ii) Matrise-multiplikasjon er distributiv over addisjon, dvs.

$$(A + B)C = AC + BC.$$

og

$$A(B + C) = AB + AC.$$

Proposisjonen bevises ved å bruke definisjonene direkte og er utelatt.

Eksempel 4.2.6. *La matrisene A og B være som i forrige eksempel og la*

$$C = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}.$$

Da har vi

$$(AB)C = \begin{bmatrix} 5 & -2 \\ -3 & 0 \\ 1 & 5 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -5 \\ -3 & 3 \\ 11 & -1 \end{bmatrix}.$$

På den annen side har vi

$$BC = \begin{bmatrix} -1 & 1 \\ 3 & 0 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 3 & -3 \\ -5 & 1 \end{bmatrix}$$

og dermed

$$A(BC) = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 1 \\ 3 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & -3 \\ -5 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -5 \\ -3 & 3 \\ 11 & -1 \end{bmatrix}$$

som bekrefter punkt i) i proposisjonen.

Så reglene over var stort sett slik vi kjenner det for tall. Men det er et viktig unntak: for matriser gjelder ikke generelt noen kommutativ lov, så AB er ikke nødvendigvis lik BA. Faktisk behøver ikke BA å være definert en gang, selv om AB er det. Prøv å finne 2×2 matriser A og B der $AB \neq BA$!

4.3 Kvadratiske matriser

Kvadratiske matriser er matriser med like mange rader som kolonner, dvs. matriser av størrelse $n \times n$ for et positivt heltall n. Slike matriser har mange fine egenskaper. Blant annet finnes det en **identitetsmatrise**, I (eller I_n dersom vi har behov for åangi størrelsen) gitt ved $n \times n$-matrisen

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

dvs. en diagonal matrise med 1-ere på **hoveddiagonalen**, og 0-ere ellers. Denne matrisen har mange av de samme egenskapene som tallet 1 og kalles

for **identitetsmatrisen**. F.eks. har vi for en kvadratisk $n \times n$-matrise A at $A \cdot I_n = I_n \cdot A = A$.

Hoveddiagonalen i en matrise $A = [a_{ij}]$ er plassene med samme rad- og kolonne-nummer, dvs. elementene som vi angir som a_{ii}.

Det finnes en lang rekke kvadratiske matriser med egne navn. De har egne navn fordi de spiller sentrale roller i ulike sammenhenger.

Definisjon 4.3.1. *En kvadratisk matrise A sies å være* ***øvre triangulær*** *dersom alle elementene under hoveddiagonalen er 0. Matrisen er* ***nedre triangulær*** *dersom alle elementene over hoveddiagonalen er 0. En kvadratisk matrise som er både øvre og nedre triangulær (dvs. alle elementer bortsett fra de på diagonalen er 0) kalles en* ***diagonalmatrise****.*

Eksempel 4.3.1. *Matrisen*

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$$

er et eksempel på en øvre triangulær matrise, mens

$$B = \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ 3 & -2 & 1 \end{bmatrix}$$

er et eksempel på en nedre triangulær matrise.

Et eksempel på en diagonalmatrise er

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Alle tall bortsett fra 0 har multiplikative inverser, dvs. det finnes et annet reelt tall som multiplisert med tallet gir oss 1. For eksempel er $\frac{1}{2}$ den multiplikative inversen til 2 fordi $2 \cdot \frac{1}{2} = 1$. På samme måte *kan* kvadratiske matriser ha inverser. I likhet med tallene har ikke **0**-matrisen noen invers, men ulikt tallene finnes det mange andre kvadratiske matriser som heller ikke har noen invers. Vi skal komme tilbake til dette etter hvert.

Definisjon 4.3.2. *La A være en kvadratisk $n \times n$-matrise. Anta at det finnes en $n \times n$-matrise B slik at $AB = BA = I_n$. Da sier vi at B er den inverse matrisen til A, og vi skriver $B = A^{-1}$. Dersom det finnes en invers til A sier vi at matrisen A er* ***invertibel****.*

Man kan vise at dersom det finnes en B slik at $AB = I$, så vil også $BA = I$, men vi går ikke inn på beviset her.

Eksempel 4.3.2. *Matrisen*

$$A = \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}$$

har invers

$$A^{-1} = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix}$$

fordi

$$AA^{-1} = \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Vi har også

$$A^{-1}A = \begin{bmatrix} -1 & -2 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

De formelle egenskapene for inverse matriser er gitt i den neste proposisjonen.

Proposisjon 4.3.3. *La A være en invertibel $n \times n$-matrise. Da har vi at*

i) Den inverse matrisen er entydig bestemt av A.

ii) $(A^{-1})^{-1} = A$.

iii) $(AB)^{-1} = B^{-1}A^{-1}$.

iv) A^{-1} er også invertibel, og $(A^{-1})^{-1} = A$.

Bevis. For å bevise i) antar vi at det finnes to inverser B_1 og B_2 til A. Bruker vi den definerende egenskapen til den inverse matrisen, sammen med assosiativitet av matrise-multiplikasjon, får vi

$$B_1 = B_1 I_n = B_1(AB_2) = (B_1A)B_2 = I_nB_2 = B_2$$

Med andre ord, de to matrisene vi startet med kan ikke være forskjellige. De to andre punktene følger nå fra entydigheten i i) og ved definisjonen av den inverse. F.eks. er A^{-1} definert ved at

$$(A^{-1})(A^{-1})^{-1} = I$$

Men vi har også at $(A^{-1})A = I$. Entydigheten av inversen til A^{-1} gir da at $(A^{-1})^{-1} = A$. Punkt (iv) følger direkte fra definisjonen. □

Eksempel 4.3.3. *Det er ikke alltid enkelt å regne ut inversen til en matrise. Senere skal vi se på noen forskjellige teknikker som vi kan benytte oss av. Men i noen tilfeller er det relativt enkelt å beregne inverser. Spesielt gjelder det for diagonalmatriser. I det tilfellet kan vi bare invertere elementene på diagonalen (under forutsetning av at disse er* $\neq 0$*). For eksempel:*

Grunnen er at produktet av to diagonalmatriser er en diagonalmatrise bestående av *samplasserte* produkter på diagonalen.

$$\begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix} \begin{bmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & 0 & 0 \\ 0 & a_{22}b_{22} & 0 \\ 0 & 0 & a_{33}b_{33} \end{bmatrix}.$$

Vi har tidligere definert begrepet transponering av en vektor. Vi kan utvide begrepet til å omfatte matriser generelt.

Definisjon 4.3.4. *Den* ***transponerte*** *til en matrise* A*, skrevet* A^T*, finner vi ved å bytte om rader og kolonner, dvs.*

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

gir

$$A^T = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & & \vdots \\ a_{1n} & \cdots & a_{mn} \end{bmatrix}$$

Den transponerte til en $m \times n$*-matrise er en* $n \times m$*-matrise.*

Eksempel 4.3.4. *La* A *være* 3×2*-matrisen gitt ved*

$$A = \begin{bmatrix} 1 & -1 \\ -3 & 0 \\ 1 & 2 \end{bmatrix}.$$

Da har vi at A^T *er en* 2×3*-matrise gitt ved*

$$A^T = \begin{bmatrix} 1 & -3 & 1 \\ -1 & 0 & 2 \end{bmatrix}.$$

Definisjon 4.3.5. *En kvadratisk matrise* A *som er lik sin egen transponerte, kalles en* ***symmetrisk*** *matrise. Vi skriver denne betingelsen som* $A^T = A$*.*

Eksempel 4.3.5. *Symmetriske matriser kan vi speile om hoveddiagonalen uten at de endrer seg. Et eksempel på en symmetrisk matrise er*

$$A = \begin{bmatrix} 1 & 2 & -2 \\ 2 & 0 & -1 \\ -2 & -1 & 1 \end{bmatrix}.$$

Den transponerte av en matrise har noen egenskaper som minner om egenskapene til den inverse.

Lemma 4.3.6. *La A være en $n \times n$-matrise. Da har vi at*

i) $(A^T)^T = A$,

ii) $(AB)^T = B^T A^T$.

Bevis. Følger nokså direkte fra definisjonene av den transponerte og av matriseprodukt. □

Siden invertering og transponering deler noen egenskaper kan vi finne sammenhenger mellom dem.

Definisjon 4.3.7. *En invertibel kvadratisk matrise A som oppfyller $A^{-1} = A^T$ kalles en* ***ortogonal*** *matrise.*

Vi skal komme tilbake til hvorfor vi kaller slike matriser ortogonale.

Eksempel 4.3.6. *Matrisen*

$$A = \begin{bmatrix} 0 & -0.8 & -0.6 \\ 0.8 & -0.36 & 0.48 \\ 0.6 & 0.48 & -0.64 \end{bmatrix}$$

er ortogonal. Spesielt har vi

$$\begin{bmatrix} 0 & -0.8 & -0.6 \\ 0.8 & -0.36 & 0.48 \\ 0.6 & 0.48 & -0.64 \end{bmatrix} \begin{bmatrix} 0 & 0.8 & 0.6 \\ -0.8 & -0.36 & 0.48 \\ -0.6 & 0.48 & -0.64 \end{bmatrix} = I$$

Proposisjon 4.3.8. *Produktet av to ortogonale matriser er en ortogonal matrise.*

Bevis. La A og B være to ortogonale matriser. Da har vi

$$(AB)^{-1} = B^{-1}A^{-1} = B^T A^T = (AB)^T$$

□

Grunnen til at slike matriser kalles ortogonale er at de har følgende viktige egenskap.

Proposisjon 4.3.9. *Prikkproduktet av to forskjellige kolonner i en ortogonal matrise er* 0*, mens prikkproduktet av en søyle med seg selv er* 1*.*

Bevis. Siden den inverse er den transponerte, så har vi $A^T A = I_n$. Hvis vi kaller kolonnene i A for $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$, så sier denne betingelsen at $\mathbf{a}_i \cdot \mathbf{a}_j = 0$ for $i \neq j$, og $\mathbf{a}_i \cdot \mathbf{a}_i = 1$. □

Så en kvadratisk matrise kalles ortogonal fordi kolonnene i matrisen er ortogonale, dvs. står normalt på hverandre. I tillegg har de alle lengde 1.

Det finnes metoder og formler for å regne ut den inverse til en matrise. Vi skal ikke gå inn på det i noen stor generalitet, men vi kan skrive opp formelen i 2×2-tilfellet.

Eksempel 4.3.7. *For en* 2×2*-matrise*

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

har vi den inverse matrisen

$$A^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

under forutsetning av at $a_{11}a_{22} - a_{12}a_{21} \neq 0$*. Hvis* $a_{11}a_{22} - a_{12}a_{21} = 0$ *har ikke matrisen noen invers.*

Eksempel 4.3.8. *Vi skal regne ut inversen til* 2×2*-matrisen*

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

Vi bruker formelen over og får

$$A^{-1} = \frac{1}{1 \cdot 1 - 2(-2)} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}.$$

Vi kan kontrollere svaret ved å regne ut

$$\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \cdot \frac{1}{5} \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Matriseproduktet utgjør en slags generalisering av prikkproduktet. Det innebærer også at vi kan se på prikkproduktet som et matriseprodukt. La $\mathbf{v}$ og $\mathbf{w}$ være to kolonnevektorer, dvs. $n \times 1$-matriser. Da vil $\mathbf{v}^T$ være en

radvektor (dvs. en $1 \times n$-matrise). Disse kan vi multiplisere sammen og få en 1×1-matrise. En 1×1-matrise A består kun av ett tall

$$A = [a_{11}]$$

og vi har derfor det vi kaller en en-til-en-korrespondanse mellom 1×1-matriser og reelle tall. Det betyr at vi kan bruke de to begrepene litt om hverandre, uten at det skal skape noen forvirring. Det vil alltid framgå av sammenhengen om det er et tall eller en 1×1-matrise, dvs. et skalarprodukt (et tall) eller et matriseprodukt (en 1×1-matrise).

$$\mathbf{v}^T\mathbf{w} = \begin{bmatrix} v_1 & v_2 & \cdots & v_p \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix} = v_1w_1 + \cdots + v_pw_p$$

Det betyr at vi kan skrive prikkproduktet som et matriseprodukt;

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T\mathbf{w}$$

Vi har tidligere sett at lengden til en vektor er gitt ved en Pythagoreisk formel. Siden vektorer med flere koordinater enn 3 ikke har noen klar geometrisk tolkning, erstatter vi det geometriske begrepet lengde med et mer ikke-geometrisk begrep, **norm**, men som egentlig er akkurat det samme.

Definisjon 4.3.10. *Vi definerer* ***normen*** $||\mathbf{v}||$ *til en vektor* $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ *ved*

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Vi kan bruke det vi har gjort over til å gi en alternativ definisjon av normen til en vektor;

Definisjon 4.3.11. *Vi definerer* ***normen*** $||\mathbf{v}||$ *til en vektor* $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ *ved*

$$||\mathbf{v}|| = \sqrt{\mathbf{v}^T\mathbf{v}}.$$

Ved å bruke matriseformen for prikkprodukt kan vi definere en stor klasse av ortogonale matriser, omtalt som Householder-transformasjoner. Householder-transformasjoner er oppkalt etter Alston Scott Householder, som introduserte denne konstruksjonen i 1958.

Definisjon 4.3.12. *La* $\mathbf{v}$ *være en ikke-triviell vektor (Ikke-triviell betyr at den er forskjellig fra 0-vektoren). Vi kaller* $n \times n$ *matrisen*

$$Q_{\mathbf{v}} = I - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T$$

for en ***Householder****-transformasjon.*

Uttrykket i denne definisjonen kan se komplisert ut, men det er ikke så vanskelig å regne ut.

Eksempel 4.3.9. *La*

$$\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}.$$

Da har vi

$$\mathbf{v}^T\mathbf{v} = (2^2 + 1^2 + (-2)^2) = 9$$

og

$$Q_{\mathbf{v}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{2}{9}\begin{bmatrix} 2\cdot 2 & 2\cdot 1 & 2\cdot(-2) \\ 2 & 1 & -2 \\ -2\cdot 2 & -2\cdot 1 & -2\cdot(-2) \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 1 & -4 & 8 \\ -4 & 7 & 4 \\ 8 & 4 & 1 \end{bmatrix}.$$

Proposisjon 4.3.13. *Householder-transformasjonen $Q_{\mathbf{v}}$ er en ortogonal matrise.*

Bevis. Vi har

$$Q_{\mathbf{v}}^T = I^T - 2(\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}})^T = I - 2\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} = Q_{\mathbf{v}}$$

Samtidig har vi

$$Q_{\mathbf{v}}^2 = (I - 2\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}})^2 = I - 4\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} + 4(\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}})^2 = I$$

siden vi har

$$(\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}})^2 = \frac{\mathbf{v}\mathbf{v}^T\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}\mathbf{v}^T\mathbf{v}} = \frac{\mathbf{v}^T\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\cdot\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}} = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}.$$

□

Vi ser av beviset at Householder-transformasjonen ikke bare er ortogonal, men den oppfyller i tillegg at $Q_{\mathbf{v}}^2 = I$.

4.4 Likninger og determinanter

En nyttig anvendelse av matriser er ved løsning av lineære likningssystemer.

Vi betrakter følgende problemstilling: *Finn to tall x og y som har sum 20 og differanse 4.* Det er ikke så vanskelig å se at 8 og 12 gir en løsning på problemet. Men det er ønskelig å ha en mer generell metode for å løse slike problemer.

Vi begynner med å stille opp to likninger

$$x + y = 20$$
$$x - y = 4.$$

Nå kan vi enten legge sammen likningene eller trekke dem fra hverandre. I det første tilfellet får vi

$$2x = 24, \quad \text{dvs. } x = 12,$$

og i det andre tilfellet

$$2y = 16, \quad \text{dvs. } y = 8,$$

som til sammen gir oss løsningen.

Vi kan gå et skritt videre og se på et mer generelt likningssystem;

$$a_{11}x + a_{12}y = b_1$$
$$a_{21}x + a_{22}y = b_2.$$

For å løse dette kan vi f.eks. multiplisere den øverste likningen med a_{21} og den nederste med a_{11} og så trekke den øverste fra den nederste. Da forsvinner x-leddene og vi står igjen med likningen

$$(a_{11}a_{22} - a_{12}a_{21})y = a_{11}b_2 - a_{21}b_1$$

Dersom $a_{11}a_{22} - a_{12}a_{21} \neq 0$ kan vi løse denne likningen med hensyn på y. Hvis $a_{11}a_{22} - a_{12}a_{21} = 0$ må vi i tillegg ha $a_{11}b_2 - a_{21}b_1 = 0$ for at det fortsatt skal være mulig å løse likningen. La oss anta at $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Da har vi

$$y = \frac{a_{11}b_2 - a_{21}b_1}{a_{11}a_{22} - a_{12}a_{21}} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}}(-a_{21} \ a_{11}) \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

og vi kan regne ut

$$x = \frac{a_{22}b_1 - a_{12}b_2}{a_{11}a_{22} - a_{12}a_{21}} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}}(a_{22} \ -a_{12}) \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

som gir oss en generell løsning av systemet.

Vi kan alternativt angripe likningssystemet på en annen måte. Med vår kunnskap om matrisemultiplikasjon observerer vi at vi kan skrive likningssystemet som en **matriselikning**;

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

og løsningen kan vi skrive som

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Uttrykket på høyre side kjenner vi igjen som en invers matrise, og løsningen kan skrives som

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}.$$

Hvis vi setter

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \text{ og } \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

får vi en likning

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

med løsning

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

Matrisen A kaller vi **koeffisientmatrisen** til likningssystemet.

Et springende punkt for å finne en løsning av likningssystemet er betingelsen $a_{11}a_{22} - a_{12}a_{21} \neq 0$. Dette uttrykket er en viktig størrelse for likningssystemet og har fått et eget navn;

Definisjon 4.4.1. *La*

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

være en 2×2*-matrise. Da er* ***determinanten*** $\det A$ *til* A *gitt ved*

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

Eksempel 4.4.1.

$$\det \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 2 \cdot 2 - 1 \cdot 1 = 3.$$

Med denne definisjonen kan vi skrive opp den inverse til en 2×2-matrise $A = [a_{ij}]$ på en alternativ form;

Proposisjon 4.4.2. *La* A *være en invertibel* 2×2*-matrise. Da har vi*

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Det er ikke bare 2×2-matriser som har determinanter, alle kvadratiske matriser har determinanter. For n større enn 2 er det ikke like enkelt å regne ut determinanten til en matrise, men det finnes en grei rekursjonsformel. Først tar vi den formelle definisjonen (eller Laplaces formel, som den også kalles).

Definisjon 4.4.3. *La*

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

være en $n \times n$-matrise. Da er determinanten til A gitt ved

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j})$$

der A_{1j} er $(n-1) \times (n-1)$-matrisen som framkommer ved stryke første rad og j-te kolonne i matrisen A.

◇

Siden vi har en enkel formel for determinanten til en 2×2-matrise, kan vi bruke denne formelen rekursivt til å beregne determinanten til større matriser. Merk at vi også har en alternativ notasjon for determinanter. For

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

bruker vi de to ekvivalente skriveformene

$$\det(A) = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Eksempel 4.4.2. *Vi skal regne ut determinanten til 3×3-matrisen*

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{bmatrix}$$

ved hjelp av rekursjonsformelen. Vi har

$$\begin{aligned} \det(A) &= 2 \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix} + 3 \begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix} \\ &= 2(1-0) - 1(1-0) + 3(-1-2) = -8. \end{aligned}$$

I determinanten $\begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix}$ *har vi fjernet første rad og første kolonne i matrisen* A*, i* $\begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix}$ *har vi fjernet første rad og andre kolonne og i* $\begin{vmatrix} 1 & 1 \\ 2 & -1 \end{vmatrix}$ *første rad og tredje kolonne.*

En pussig egenskap ved determinanter er at de kan beregnes ved følgende formel

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \tag{4.1}$$

som holder for hver $i = 1, 2, \ldots, n$. Matrisen A_{ij} er $(n-1) \times (n-1)$ matrisen vi får fra A ved å slette rad i og kolonne j. Definisjonen av $\det(A)$ over svarer til $i = 1$, og for generell i går vi utover i rad i og gjør beregninger på "tilsvarende måte".

I praksis kan man regne ut determinanten på en alternativ måte som er mer effektiv, spesielt på en datamaskin. Det skjer ved **Gauss eliminasjon** eller **elementære radoperasjoner**; mer om dette i neste seksjon. Faktisk er Gauss eliminasjon en god teknikk for å løse lineære likningssystemer, og finne den inverse til en matrise (hvis vi ønsker det).

Det er noen nyttige regneregler knyttet til determinanter. Den kanskje aller viktigste er produktformelen.

Teorem 4.4.4. *La A og B være to kvadratiske matriser av samme størrelse. Da har vi*

$$\det(AB) = \det(A) \cdot \det(B)$$

Vi skal ikke vise det her, det involverer en god del begreper som vi ikke vil ha bruk for til annet enn å bevise formelen. Men vi kan skrive ut formelen i 2×2-tilfellet;

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

Det gir at

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

og

$$\begin{aligned}\det(AB) &= (a_{11}b_{11}+a_{12}b_{21})(a_{21}b_{12}+a_{22}b_{22}) - (a_{11}b_{12}+a_{12}b_{22})(a_{21}b_{11}+a_{22}b_{21}) \\ &= a_{11}b_{11}a_{21}b_{12} + a_{11}b_{11}a_{22}b_{22} + a_{12}b_{21}a_{21}b_{12} + a_{12}b_{21}a_{22}b_{22} \\ &\quad - a_{11}b_{12}a_{21}b_{11} - a_{12}b_{22}a_{22}b_{21} - a_{11}b_{12}a_{22}b_{21} - a_{12}b_{22}a_{21}b_{11} \\ &= a_{11}a_{22}b_{11}b_{22} - a_{11}a_{22}b_{12}b_{21} - a_{12}a_{21}b_{22}b_{11} + a_{12}a_{21}b_{21}b_{12} \\ &= (a_{11}a_{22} - a_{12}a_{21})(b_{11}b_{22} - b_{12}b_{21}) \\ &= \det(A)\cdot\det(B)\end{aligned}$$

Proposisjon 4.4.5. *Determinanten til en matrise med en rad eller kolonne av* 0*-er er 0.*

Bevis. Siden alle leddene som inngår i determinanten inneholder en faktor fra hver rad og hver kolonne, vil i dette tilfelle alle leddene inneholde en faktor 0, og determinanten er derfor 0. □

Proposisjon 4.4.6. *Determinanten til en triangulær matrise er produktet av elementene på diagonalen.*

Bevis. Dette følger fra definisjonen av $\det(A)$ ved induksjon; vi hopper over detaljene. □

Det følger at determinanten til identitetsmatrisen er 1, og determinanten til $k \cdot I$ er k^n.

Eksempel 4.4.3.

$$\det\begin{bmatrix} 2 & 1 & -3 \\ 0 & 3 & -2 \\ 0 & 0 & 1 \end{bmatrix} = 2\cdot 3\cdot 1 = 6.$$

Proposisjon 4.4.7. *La A være en $n \times n$-matrise. Da har vi*

$$\det(k\cdot A) = k^n \cdot \det(A).$$

Bevis. Vi kan skrive

$$k\cdot A = (kI)\cdot A.$$

Dermed følger resultatet fra kommentaren etter proposisjonen over. □

Proposisjon 4.4.8. *Determinanten til en matrise og dens transponerte er like.*

Bevis. Siden determinanten er satt sammen av alle produkter av ett element fra hver rad og hver kolonne, kan vi bytte om rader og kolonner og få de samme leddene. Fortegnene vi introduserer er laget slik at de ikke endrer seg om vi bytter om rader og kolonner. □

Eksempel 4.4.4. *Vi har*

$$\det \begin{bmatrix} 2 & 0 & -1 \\ 1 & 1 & 0 \\ -2 & -1 & 1 \end{bmatrix} = 1,$$

og for den transponerte

$$\det \begin{bmatrix} 2 & 1 & -2 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix} = 1.$$

Proposisjon 4.4.9. *Bytter vi om to rader eller to kolonner i en matrise skifter determinanten tegn.*

Bevis. Bytter vi om to rader (eller to kolonner) vil alle leddene i determinanten bli de samme, men alle tegnene vil endre seg. Igjen bruker vi det at fortegnene er satt slik at determinanten ved denne prosessen vil skifte tegn. □

Eksempel 4.4.5. *Vi har*

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21},$$

mens

$$\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22} = -(a_{11}a_{22} - a_{12}a_{21}).$$

Eksempel 4.4.6. *Vi har*

$$\det \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 3,$$

mens

$$\det \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = -3.$$

Proposisjon 4.4.10. *Determinanten til den inverse matrisen er gitt ved*

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Bevis. Siden vi har $A \cdot A^{-1} = I$ følger det at

$$\det(A) \cdot \det(A^{-1}) = \det(I) = 1.$$

□

Eksempel 4.4.7. *La*

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

Da har vi

$$\det(A) = 5,$$

mens determinanten til den inverse matrisen er

$$\det(A^{-1}) = \det \begin{bmatrix} \frac{1}{5} & \frac{-2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{bmatrix} = \frac{1}{5}^2 - \frac{2}{5}(-\frac{2}{5}) = \frac{1}{5}.$$

Proposisjon 4.4.11. *Determinanten til en ortogonal matrise er* ± 1.

Bevis. En ortogonal matrise er definert ved at $A^{-1} = A^T$. Vi har vist at $\det(A^T) = \det(A)$ og siden vi samtidig har at $\det(A) \cdot \det(A^{-1}) = 1$ følger det at $\det(A)^2 = 1$. □

Eksempel 4.4.8.

$$\begin{vmatrix} 0 & -0.8 & -0.6 \\ 0.8 & -0.36 & 0.48 \\ 0.6 & 0.48 & -0.64 \end{vmatrix}$$

$$= -0.8 \cdot 0.48 \cdot 0.6 - 0.6 \cdot 0.8 \cdot 0.48 - 0.8 \cdot 0.8 \cdot 0.64 - 0.6 \cdot 0.36 \cdot 0.6$$

$$= -0.2304 - 0.2304 - 0.4096 - 0.1296 = -1.$$

4.5 Lineære likningssystemer

Vi har allerede møtt noen lineære likningssystemer og sett på hvordan vi kan løse dem. I dette kapitlet skal vi gjøre en litt mer grundig gjennomgang av det vi kommer til å trenge.

Definisjon 4.5.1. *Et system av lineære likninger*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

kalles et ***lineært likningssystem*** *i n variable. Dersom alle $b_i = 0$ sier vi at likningssystemet er* ***homogent****. I motsatt fall er det* ***inhomogent****.*

Vi kan skrive et lineært likningssystem på matriseform;

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

I hvilken grad vi kan finne løsninger til et lineært likningssystem avhenger av forholdet mellom antall ukjente og antall likninger. Vi tar utgagnspunkt i at løsningene $\mathbf{x} = (x_1, x_2, \dots, x_n)$ har n variable som vi kan velge fritt. Dette omtales ofte som at vi har n **frihetsgrader** eller n **ukjente**. Hver likning gir oss en begrensning på hvilke $\mathbf{x}$ vi kan velge, og reduserer dermed (som oftest!) antall frihetsgrader med 1. Det betyr at med m likninger, så står vi igjen med $n - m$ frihetsgrader, under forutsetning av at dette tallet er større enn eller lik 0. Et system som har flere likninger enn frihetsgrader sies å være **overbestemt**. Det systemet vil normalt ikke ha noen løsninger. Det motsatte tilfellet kalles et **underbestemt** system. Da har vi flere frihetsgrader enn likninger, og vi vil normalt ha mange løsninger. Men et underbestemt system trenger ikke å ha noen løsninger. F.eks. vil ikke systemet

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

ha noen løsninger, selv om vi har 3 ukjente og kun 2 likninger. På den annen side kan et tilsynelatende overbestemt system likevel ha løsninger hvis flere av likningene er like eller eventuelt summer av hverandre. F.eks. er ikke systemet

$$\begin{aligned} x + y &= 2 \\ x - y &= 0 \\ 3x + y &= 4 \end{aligned}$$

overbestemt, på tross av at vi har flere likninger enn ukjente. Den siste likningen $3x + y = 4$ kan skrives som 2 ganger den første likningen pluss den andre. Det betyr at den siste likningen ikke gir oss noe mer informasjon om løsningen enn vi allerede har fra de to første likningene, og derfor heller ikke noen flere begrensniger på hvilke løsninger som er mulig. Likningssystemet er derfor **ekvivalent** med et system med 2 ukjente og 2 likninger.

Likningssystemer med like mange frihetsgrader som likninger kaller vi for **kvadratiske** likningssystemer. Kvadratiske likningssystemer kan være underbestemt, **entydig bestemt** eller overbestemt, avhengig av hvordan

likningene ser ut. Entydig bestemt betyr at likningssystemet har en entydig løsning.

Vi skriver systemet på matriseform;

$$\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

hvor **koeffisientmatrisen**

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$$

er en kvadratisk matrise. Det første spørsmålet er om A er invertibel.

Proposisjon 4.5.2. *La*

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

være et likningsystem med n likninger og n ukjente, dvs. at A er en kvadratisk matrise. Dersom A er invertibel, så har likningssystemet en entydig løsning gitt ved

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

Ved å bruke produktformelen for determinanter følger det at en invertibel matrise må ha determinant ulik 0.

$$\det(A) \cdot \det(A^{-1}) = \det(I) = 1.$$

Når produktet av to tall er 1, eller mer generelt ulik 0, er det ikke mulig at noen av tallene er lik 0. I så fall ville jo produktet av dem være 0. Faktisk er det motsatte også sant; dersom determinanten til en kvadratisk matrise er ulik 0, så er matrisen inverterbar. Vi skal ikke gå inn på beviset for dette resultatet her.

Proposisjon 4.5.3. *En kvadratisk matrise A er invertibel hvis og bare hvis determinanten* $\det(A) \neq 0$.

Det finnes mange ulike måter å løse lineære likningssystemer. Hovedmetoden baserer seg på såkalte **elementære (rad)operasjoner**, slik vi nå ser på. Disse operasjonene er

(R1) *Multiplisere en likning med en konstant ulik 0.*

(R2) *Bytte rekkefølge på to likninger.*

(R3) *Addere et multippel av en likning til en annen likning.*

Disse operasjonene har den viktige egenskapen at de *ikke endrer løsningsmengden* til systemet.

Eksempel 4.5.1. *Vi skal løse det lineære likningssystemet*

$$\begin{array}{rrrcr} x_1 & -x_2 & +x_3 & = & 3 \\ & 3x_2 & +x_3 & = & 2 \\ 2x_1 & -x_2 & -x_3 & = & 0. \end{array}$$

Vi adderer nå (-2) *ganger første likning til tredje likning og får*

$$\begin{array}{rrrcr} x_1 & -x_2 & +x_3 & = & 3 \\ & 3x_2 & +x_3 & = & 2 \\ & x_2 & -3x_3 & = & -6. \end{array}$$

Så bytter vi rekkefølge på andre og tredje likning:

$$\begin{array}{rrrcr} x_1 & -x_2 & +x_3 & = & 3 \\ & x_2 & -3x_3 & = & -6 \\ & 3x_2 & +x_3 & = & 2. \end{array}$$

Vi adderer nå (-3) *ganger andre likning til tredje likning og får*

$$\begin{array}{rrrcr} x_1 & -x_2 & +x_3 & = & 3 \\ & x_2 & -3x_3 & = & -6 \\ & & 10x_3 & = & 20 \end{array}$$

og så ganger vi tredje likning med $\frac{1}{10}$ *som gir*

$$\begin{array}{rrrcr} x_1 & -x_2 & +x_3 & = & 3 \\ & x_2 & -3x_3 & = & -6 \\ & & x_3 & = & 2. \end{array}$$

Nå har vi et såkalt triangulært system, og kan finne løsningen ved å starte med siste likning: vi får $x_3 = 2$*, og ved å sette dett inn i andre likning, får vi* $x_2 = 0$*. Til slutt, fra første likning, får vi* $x_1 = 1$*. Løsningen er altså* $\mathbf{x} = (1, 0, 2)$*.*

Vi kunne valgt andre elementære radoperasjoner, f.eks. kunne vi droppet å bytte de to likningene. Men vi vil uansett få samme løsning.

Metoden over kalles **Gauss-eliminasjon** (eller å løse ved elementære radoperasjoner). Vi jobber altså mot å få et triangulært system, og finner så løsningen ved å starte med siste likning, og sette inn.

Gauss-eliminasjon egner seg fint for å programmere på datamaskin, og dette er gjort i matematiske programsystemer, som f.eks. Matlab. Der finner man løsning(ene) ved en enkel kommando. At vi kaller dette radoperasjoner

er fordi man kan gjøre beregningen på radene i en matrise. Likningssystemet over vil vi da skrive som

$$\begin{bmatrix} 1 & -1 & 1 & 3 \\ 0 & 3 & 1 & 2 \\ 2 & -1 & -1 & 0 \end{bmatrix}$$

ved at vi bare holder styr på tallene uten å skrive x_1 etc. Den første operasjonen vi gjorde svarer da til å legge (-2) ganger første rad til tredje rad. Avslutningen er å lese ut løsning ved å gå tilbake til det tilhørende likningssystemet. Vi dropper detaljene i dette her.

Gauss-eliminasjon er viktig, og gir også en effektiv måte for å beregne determinanten til en kvadratisk matrise. Operasjon R1 endrer determinanten med nettopp den konstanten man multipliserer med, og operasjon R2 endrer fortegnet. Operasjon R3 gir ingen endring. Ved å holde styr på disse endringene finner man determinanten når man har kommet fram til det triangulære systemet.

Vi avslutter dette kapitlet med en liten historisk kommentar. Kanskje litt overraskende, men determinanter ble altså oppfunnet lenge før matrisene. En form av determinanter, som et tall som avgjør om et likningssett har en entydig løsning, dukket opp i Kina allerede i det 3. århundre før vår tid. Gauss introduserte ordet determinant i 1801, og i 1812 formulerte, uavhengig av hverandre, Binet og Cauchy produktformelen for determinanter. Matriser ble først introdusert av Sylvester i 1850.

4.6 Spektralteori

Først gir vi en kommentar om notasjon: I teksten er det hensiktsmessig å skrive en vektor som et n-tuppel, ikke som en kolonnevektor. For eksempel skriver vi $\mathbf{v} = (1, 2)$ og mener da vektoren

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Til en kvadratisk matrise finnes det alltid noen vektorer som har den nyttige egenskapen at når vi ganger dem med matrisen så endrer det ikke på forholdet mellom koordinatene til vektoren. Slike vektorer kalles **egenvektorer** for matrisen og det finnes en svært omfattende teori som tar for seg disse. Vi skal se på den grunnleggende teorien for disse vektorene, en teori som ofte omtales med det klingende navnet **spektralteori**.

La A være en kvadratisk $n \times n$ matrise. En vektor $\mathbf{v} \neq \mathbf{0}$ kalles en **egenvektor** for A dersom

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

for et reelt tall λ. Tallet λ kalles for den tilhørende **egenverdien**[1]. Merk at vi krever at $\mathbf{v}$ ikke er nullvektoren i denne definisjonen. Neste eksempel viser at det er nokså enkelt å sjekke om en gitt vektor er en egenvektor (eller ikke).

Eksempel 4.6.1. *La*

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}.$$

Da er $\lambda = 2$ en egenverdi for A med tilhørende egenvektor $\mathbf{v} = (1, 0)$. Dette kan vi sjekke ved utregning

$$A\mathbf{v} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 2\mathbf{v}.$$

Er $\mathbf{w} = (1, 1)$ en egenvektor for A, for en eller annen egenverdi? Dette sjekkes ved regning

$$A\mathbf{w} = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \end{bmatrix} \neq \lambda \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

for enhver λ. Så $\mathbf{w} = (1, 1)$ er ikke en egenvektor for A.

For en egenvektor $\mathbf{v}$ vil også alle vektorer på formen $a\mathbf{v}$ for et reelt tall a der $a \neq 0$ være en egenvektor med samme egenverdi λ:

$$A(a\mathbf{v}) = aA\mathbf{v} = a\lambda\mathbf{v} = \lambda(a\mathbf{v})$$

En viktig grunn til at man er interessert i egenverdier og egenvektorer er bruken innenfor bl.a. sannsynlighetsregning og statistikk[2]. Løst sagt handler dette om at tilstanden til et system beskrives ved en vektor $\mathbf{v}$ og denne tilstanden endres for hvert tidsskritt ved å multiplisere med matrisen A, så tilstanden etter ett tidsskritt er $A\mathbf{v}$. Tilstanden etter to tidsskritt er $A(A\mathbf{v}) = A^2\mathbf{v}$ osv. Men legg merke til hva som skjer hvis tilstandsvektoren $\mathbf{v}$ er en egenvektor for en egenverdi λ, da er

$$\begin{aligned} A^2\mathbf{v} &= A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda A\mathbf{v} = \lambda \cdot \lambda\mathbf{v} = \lambda^2\mathbf{v}, \\ A^3\mathbf{v} &= \lambda^3\mathbf{v} \text{ osv.} \end{aligned}$$

Dette betyr at for slike startvektorer er utviklingen til systemet over tid svært enkelt å beregne og analysere! Videre kan man fra dette finne utvikling fra enhver startvektor (med noe ekstra regning).

Regningen vi nettopp gjorde gir også følgende resultat.

[1] Omvendt sier vi at λ er en egenverdi med tilhørende egenvektor $\mathbf{v}$.

[2] Statistikk ... alias Data Science ... nesten alias Maskinlæring!

Proposisjon 4.6.1. *La A være en $n \times n$ matrise med egenverdi λ og tilhørende egenvektor $\mathbf{v}$. La $k \geq 1$ være et heltall. Da er $\mathbf{v}$ også egenvektor for A^k med tilhørende egenverdi λ^k.*

Mengden av egenvektorer med samme egenverdi λ, sammen med nullvektoren, kalles **egenrommet** til egenverdien λ. Merk at nullvektoren *ikke* er en egenvektor, men det likevel hensiktsmessig å inkludere $\mathbf{0}$ i egenrommet (fordi man da får et underrom).

Eksempel 4.6.2. *Enhver vektor er en egenvektor for identitetsmatrisen I_n med egenverdi 1 siden*

$$I_n\mathbf{v} = \mathbf{v} = 1\mathbf{v}$$

Det betyr at egenrommet til I_n for egenverdien 1 er hele $\mathbb{R}^n$.

Anta at $\mathbf{v}$ er en egenvektor for en $n \times n$ matrise A, med tilhørende egenverdi λ. Da har vi

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

Dette er ekvivalent med

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

Egenvektorer er alltid antatt å være forskjellige fra nullvektoren. Det betyr at likningssystemet $(A - \lambda I)\mathbf{x} = \mathbf{0}$ skal ha en ikke-triviell løsning, og derfor er matrisen $A - \lambda I$ singular (altså, ikke invertibel). Dette er ekvivalent med at determinanten til $A - \lambda I$ er lik 0. Det omvendte holder også, så hvis $\det(A - \lambda I) = 0$, så er λ en egenverdi for A.

Fra dette følger at det å finne en egenverdi λ for A er ekvivalent med finne en løsning av likningen

$$\det(A - \lambda I) = 0.$$

Dersom vi skriver dette ut får vi en polynomlikning i λ, som vi kaller den **karakteristiske likningen** til A. Polynomet

$$\chi_A(\lambda) = \det(A - \lambda I)$$

kalles det **karakteristiske polynomet** til A. Så egenverdiene er røtter (løsninger) til den karakteristiske likningen.

Eksempel 4.6.3. *Betrakt igjen matrisen*

$$A = \begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix}$$

der vi så at $\lambda = 2$ var en egenverdi. Da er

$$A - \lambda I = \begin{bmatrix} 2-\lambda & 3 \\ 0 & 1-\lambda \end{bmatrix}$$

så det karakteristiske polynomet er

$$\chi_A(\lambda) = \det(A - \lambda I) = (2-\lambda)(1-\lambda).$$

Vi trenger ikke regne ut polynomet fordi vi ser at $\chi_A(\lambda) = 0$ har løsninger $\lambda = 2$ (som vi visste) og $\lambda = 1$. Så A har egenverdier 1 og 2. Det karakteristiske polynomet er $\chi_A(\lambda) = \lambda^2 - 3\lambda + 2$.

Eksempel 4.6.4. *Det karakteristiske polynomet til matrisen*

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

er gitt ved

$$\begin{aligned} \chi_A(\lambda) &= \det(A - \lambda I) \\ &= \det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix} \\ &= (1-\lambda)(1-\lambda) - 2 \cdot 2 \\ &= \lambda^2 - 2\lambda - 3. \end{aligned}$$

For å finne røttene bruker vi abc-formelen;

$$\lambda = \frac{2 \pm \sqrt{2^2 - 4(-3)}}{2} = 1 \pm 2.$$

Egenverdiene blir derfor $\lambda_1 = 3$ og $\lambda_2 = -1$.

Det karakteristiske polynomet til en $n \times n$-matrise vil være et polynom av grad n i λ. Matrisen kan derfor ha inntil n egenverdier, de n røttene til det karakteristiske polynomet.

For større matriser kreves (ofte mye) mer regning å finne egenverdiene, men det finnes numeriske programsystemer (f.eks, Matlab) som beregner egenverdiene og egenvektorene effektivt.

Vi kan bruke det karakteristiske polynomet til å studere egenskaper ved ulike matriser.

Proposisjon 4.6.2. *La A være en invertibel $n \times n$-matrise, og $\mathbf{v}$ en egenvektor for A med tilhørende egenverdi λ. Da er $\lambda \neq 0$ og $\mathbf{v}$ er også egenvektor for A^{-1} med egenverdi $\frac{1}{\lambda}$.*

Bevis. Først observerer vi at $\lambda \neq 0$, for hvis λ var 0, ville $A\mathbf{v} = \lambda\mathbf{v} = \mathbf{0}$. Det betyr at $\mathbf{v}$ er ikke-null løsning av det homogene likningsystemet $A\mathbf{x} = \mathbf{0}$ som strider mot at A er invertibel.

Vi har $A\mathbf{v} = \lambda\mathbf{v}$. Siden A er invertibel kan vi multiplisere med A^{-1} på begge sider av likhetstegnet, og vi får $\mathbf{v} = A^{-1}\lambda\mathbf{v}$. Deler vi begge sider med λ (som er ulik 0) får vi

$$A^{-1}\mathbf{v} = \frac{1}{\lambda}\mathbf{v}.$$

□

Vi har også følgende resultat.

Proposisjon 4.6.3. *En kvadratisk matrise og dens transponerte har samme karakteristiske polynom og derfor de samme egenverdiene.*

Merk at de to matrisene har samme egenverdier, vi sier ikke noe om egenvektorene.

4.7 Lineære rom

Vi begynner med å se på begrepet vektorrom.

Definisjon 4.7.1. *Et reelt* ***vektorrom*** V *(også kalt et* ***lineært rom****) er en mengde med to operasjoner,* ***addisjon****, dvs. at for* $\mathbf{x}, \mathbf{y} \in V$*, så er summen* $\mathbf{x} + \mathbf{y} \in V$*, og* ***skalar multiplikasjon****, dvs. hvis* $\mathbf{x} \in V$ *og* $a \in \mathbb{R}$ *er et reelt tall, så er* $a\mathbf{x} \in V$*. De to operasjonene må oppfylle betingelsene:*

i) Addisjon er ***assosiativ****, dvs.* $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$ *for alle* $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$*.*

ii) Addisjon er ***kommutativ****, dvs.* $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ *for alle* $\mathbf{x}, \mathbf{y} \in V$*.*

iii) Det finnes et ***nøytralt additivt element*** *eller* $\mathbf{0}$*, slik at* $\mathbf{x} + \mathbf{0} = \mathbf{x}$ *for alle* $\mathbf{x} \in V$*.*

iv) Til alle elementer $\mathbf{x} \in V$ *så finnes det et element* $-\mathbf{x}$ *slik at* $\mathbf{x}+(-\mathbf{x}) = 0$*. Vi kaller elementet den* ***additive inversen*** *til* $\mathbf{x}$*.*

v) Skalar multiplikasjon er ***assosiativ****, dvs. for alle reelle tall* $a, b \in \mathbb{R}$ *og* $\mathbf{x} \in V$*, så er* $a(b\mathbf{x}) = (ab)\mathbf{x}$*.*

vi) Vi har $1 \cdot \mathbf{x} = \mathbf{x}$ *for alle* $\mathbf{x} \in V$*.*

vii) De to operasjonene er ***distributive over hverandre****, dvs.* $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$ *og* $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$ *for alle* $a, b \in \mathbb{R}$ *og* $\mathbf{x}, \mathbf{y} \in V$*.*

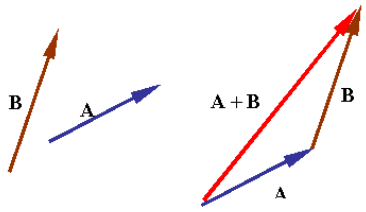
Elementene i et vektorrom kalles ofte for ***vektorer*** *(selv om de ikke ser ut som det vi vanligvis forbinder med vektorer), og de reelle tallene vi ganger med kalles* ***skalarer****.*

De vanligste vektorrommene vi kommer bort i kan vi dele i 3 hovedtyper. Den første er en geometrisk tolkning av vektorrom.

Eksempel 4.7.1. *Mengden av vektorer i planet* $\mathbb{R}^2$ *(eller rommet,* $\mathbb{R}^3$*), skrevet som* $(a, b) \in \mathbb{R}^2$ *danner et vektorrom. Den generelle definisjonen har hentet sitt navn fra dette eksempelet. Addisjon og skalar multiplikasjon av vektorer i planet gjøres komponentvis, dvs.*

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$$
$$c(a, b) = (ca, cb)$$

som svarer til å legge sammen piler, med startpunktet for neste vektor i spissen av den forrige.



Figur 4.4: Vektoraddisjon

Skalar multiplikasjon beholder retningen på en vektor, men forandrer lengden, eventuelt snur pila dersom skalaren er et negativt tall.

I det neste eksemplet fjerner vi oss fra den geometriske tolkingen av vektorer, men holder oss til en tilsvarende skriveform.

Eksempel 4.7.2. *For et vilkårlig helt tall* n *lar vi* $\mathbb{R}^n$ *betegne mengde av alle* n***-tupler*** $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ *med komponentvis addisjon og skalar multiplikasjon,*

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)$$
$$c(a_1, a_2, \ldots, a_n) = (ca_1, ca_2, \ldots, ca_n)$$

$\mathbf{0}$*-elementet er gitt ved* n*-tuplet med kun 0-er, dvs.* $\mathbf{0} = (0, \ldots, 0)$ *og det inverse elementet til* $\mathbf{a} = (a_1, \ldots, a_n)$ *er gitt ved* $-\mathbf{a} = (-a_1, \ldots, -a_n)$*. Mengden av reelle* n*-tupler* $\mathbb{R}^n$ *danner dermed et vektorrom.*

Eksempel 4.7.3. *I det tredje eksempelet lar vi vektorrommet $V = \mathcal{F}(I)$ bestå av funksjoner definert på et intervall $I \subset \mathbb{R}$. Addisjon og skalar multiplikasjon er gitt som ordinær addisjon og multiplikasjon av funksjoner;*

$$(f+g)(x) = f(x) + g(x)$$
$$(cf)(x) = cf(x)$$

for alle $f, g \in \mathcal{F}(I)$ og $c \in \mathbb{R}$ og 0-elementet er konstantfunksjonen $f(x) = 0$.

Det er ikke alltid hensiktsmessig å se på alle mulige funksjoner på et intervall. Vi kan gjerne begrense oss til kontinuerlige funksjoner, eller betrakte en enda mer restriktert mengde.

Eksempel 4.7.4. *Mengden $\{ae^x + be^{-x} \mid a, b \in \mathbb{R}\}$ danner et vektorrom under addisjon og skalar multiplikasjon.*

Eksempel 4.7.5. *La $\mathcal{P}_n = \{a_0 + a_1x + \cdots + a_nx^n\}$ der $a_i \in \mathbb{R}$ for alle i og n er et ikke-negativt heltall. $\mathcal{P}_n$ betegner mengden av alle polynomer i x, med reelle koeffisienter. Mengden $\mathcal{P}_n$ danner et vektorrom under vanlig addisjon og skalar multiplikasjon.*

Siden vi kan legge sammen og skalarmultiplisere elementer i et vektorrom, kan det være interessant å gå den andre veien. Med det mener vi at vi kan fundere på om det finnes noen elementer i vektorrommet som alt springer ut i fra. Kan vi finne en (relativt) liten mengde av elementer som er slik at gjentatt addisjon og skalarmultiplikasjon kan gi oss hele vektorrommet? Et første skritt på veien til å løse dette problemet er å definere begrepene **lineærkombinasjon** og **lineær uavhengighet**.

Definisjon 4.7.2. *En **lineærkombinasjon** av elementer $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ er en sum*

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r$$

for reelle tall $a_1, a_2, \ldots, a_r$.

Enhver lineær kombinasjon av vektorer i et vektorrom er selv et element i vektorrommet.

Definisjon 4.7.3. *En mengde $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ i et vektorrom V sies å være **lineært uavhengig** dersom*

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r = \mathbf{0}$$

*medfører at alle $a_1 = a_2 = \cdots = a_r = 0$. I motsatt fall sier vi at S er **lineært avhengig**. Dersom vi har at $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r = \mathbf{0}$ uten at alle $a_i = 0$, kaller vi ofte r-tuppelet $(a_1, a_2, \ldots, a_r)$, eller lineærkombinasjonen selv, for en **lineær avhengighet** for disse vektorene.*

I situasjonen over der S er lineært uavhengig (evt. avhengig) sier vi også at vektorene $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ er lineært uavhengig (evt. avhengig).

Eksempel 4.7.6. *Betrakt vektorene*

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}.$$

Da har vi

$$2\begin{bmatrix} 2 \\ 1 \end{bmatrix} - 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

og de tre vektorene er lineært avhengig. Lineærkombinasjonen $2\mathbf{v}_1 - 2\mathbf{v}_2 - \mathbf{v}_3$ *gir oss en lineær avhengighet for* $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

Eksempel 4.7.7. *La* $\mathcal{T} = \{\sin t, \cos t\}$. *Hvis vi har*

$$a \sin t + b \cos t = 0$$

som funksjoner, så betyr det at likningen også må stemme for alle mulige verdier av t. *Setter vi inn* $t = 0$ *får vi*

$$a \sin 0 + b \cos 0 = b = 0$$

og setter vi inn $t = \frac{\pi}{2}$ *får vi*

$$a \sin \frac{\pi}{2} + b \cos \frac{\pi}{2} = a = 0$$

Det betyr at for at likheten skal være oppfyllt som en relasjon mellom funksjoner, så må vi ha $a = b = 0$ *og de to funksjonene er derfor lineært uavhengig.*

Vi er spesielt interessert i mengder som **genererer** vektorrommet.

Definisjon 4.7.4. *En mengde* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ *av vektorer i* V *kalles en (ordnet)* ***basis*** *for* V *dersom disse er lineært uavhengig og* ***utspenner*** *hele* V, *dvs. at for alle* $\mathbf{v} \in V$, *så kan vi finne reelle tall* $a_1, a_2, \ldots a_r \in \mathbb{R}$ *slik at*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r.$$

Det er også mulig at en slik basis kan bestå av uendelig mange elementer.

Grunnen til at vi vil at en basis skal være lineært uavhengig er at vi vil at rommet skal være generert av så få elementer som mulig. En basis er en minimal mengde i den forstand at hvis vi fjerner en vektor fra basisen, så vil vektorrommet som basisen utspenner bli ekte redusert. Derimot, hvis de

gitte vektorene er lineært avhengige, er det mulig å fjerne et element uten at mengden som utspennes endres.

I eksemplet over vil vektorene

$$\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}$$

utspenne hele $\mathbb{R}^2$. For et vilkårlig element

$$\begin{bmatrix}a\\b\end{bmatrix} \in \mathbb{R}^2$$

kan vi f.eks. skrive

$$\begin{bmatrix}a\\b\end{bmatrix} = b\begin{bmatrix}2\\1\end{bmatrix} + (a-2b)\begin{bmatrix}1\\0\end{bmatrix}.$$

Fjerner vi den siste vektoren $\begin{bmatrix}2\\2\end{bmatrix}$ fra mengden har vi fortsatt nok vektorer til å spenne ut hele $\mathbb{R}^2$. De to elementene

$$\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}$$

er lineært uavhengige og siden de samtidig utspenner hele $\mathbb{R}^2$ utgjør de en basis for $\mathbb{R}^2$.

Eksempel 4.7.8. *Betrakt vektorene*

$$\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$$

i $\mathbb{R}^3$*. Disse er lineært uavhengig siden*

$$a\begin{bmatrix}1\\0\\0\end{bmatrix} + b\begin{bmatrix}0\\1\\0\end{bmatrix} + c\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$$

kun er mulig dersom $a = b = c = 0$*. Vi ser også at vektorene utspenner hele* $\mathbb{R}^3$ *siden et vilkårlig element*

$$\begin{bmatrix}a\\b\\c\end{bmatrix} = a\begin{bmatrix}1\\0\\0\end{bmatrix} + b\begin{bmatrix}0\\1\\0\end{bmatrix} + c\begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Eksempel 4.7.9. *La*

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

være tre vektorer i $\mathbb{R}^3$*. For å avgjøre om disse er lineært uavhengig setter vi*

$$a\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + c\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} a-b+2c \\ a+b-c \\ a+c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Det gir oss likningssystemet

$$\begin{aligned} a-b+2c &= 0 \\ a+b-c &= 0 \\ a+c &= 0 \end{aligned}$$

som har som eneste løsning $a = b = c = 0$ *og mengden er derfor lineært uavhengig. Vi ser også at vektorene utspenner hele* $\mathbb{R}^3$ *siden et vilkårlig element kan skrives som en lineærkombinasjon av de tre vektorene;*

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = u_1\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + u_2\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + u_3\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}.$$

hvor

$$\begin{aligned} u_1 &= x+y-z \\ u_2 &= -2x-y+3z \\ u_3 &= -x-y+2z. \end{aligned}$$

Betrakt et vektorrom V som har en basis $\mathcal{B}$ med endelig mange elementer. Vi skal se at alle andre basiser for V har like mange elementer som $\mathcal{B}$.

Lemma 4.7.5. *La* $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ *være en basis for et vektorrom* V*. Enhver mengde med minst* $r+1$ *vektorer i* V *vil da være lineært avhengig.*

I en litt annen versjon kalles dette lemmaet vanligvis for **Steinitz' erstatningsresultat**.

Bevis. Anta at mengden $\mathcal{W}$ er lineært uavhengig. Siden $\mathcal{B}$ er en basis kan vi skrive $\mathbf{w}_1 = b_1\mathbf{v}_1 + \cdots + b_r\mathbf{v}_r$ hvor ikke alle $b_j = 0$. Ved å bytte om på

indeksene, kan vi anta at vi har ordnet basiselementene slik at $b_1 \neq 0$. Da har vi

$$\mathbf{v}_1 = \frac{1}{b_1}\mathbf{w}_1 - \frac{b_2}{b_1}\mathbf{v}_2 - \cdots - \frac{b_r}{b_1}\mathbf{v}_r$$

Det betyr at vi kan erstatte $\mathbf{v}_1$ med $\mathbf{w}_1$ i basisen $\mathcal{B}$ og det vil fortsatt være en basis. Vi gjentar denne prosedyren til vi har byttet ut alle $\mathbf{v}_j$ med $\mathbf{w}_j$. (Dette er mulig siden begge mengdene er lineært uavhengige). Vi ender opp med at $\{\mathbf{w}_1, \ldots, \mathbf{w}_r\}$ danner en basis for V. Men det betyr at $\mathbf{w}_{r+1}$ kan skrives som en lineær kombinasjon av $\mathbf{w}_1, \ldots, \mathbf{w}_r$, som strider mot antakelsen om at $\mathcal{W}$ er lineært uavhengig. Altså var antakelsen gal. □

Det følgende resultatet er en direkte konsekvens av Steinitz' erstatningsresultat:

Teorem 4.7.6. *La V være et vektorrom. Anta at $\mathcal{B}_1 = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ og $\mathcal{B}_2 = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_s\}$ er to forskjellige basiser for V. Da har de to basisene like mange elementer, dvs. $r = s$.*

Definisjon 4.7.7. *La V være et vektorrom, og anta at $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ er en basis for V. Vi definerer **dimensjonen** til vektorrommet V til å være antall elementer r i basisen $\mathcal{B}$. Siden alle basiser i henhold til Steinitz' erstatningsresultat må ha like mange elementer, er dimensjonen til V uavhengig av valg av basis $\mathcal{B}$.*

Eksempel 4.7.10. *Dimensjonen til $\mathbb{R}^n$ er n. Dette kan vi se siden alle elementer i $\mathbb{R}^n$ er på formen $(a_1, \ldots, a_n)$ og*

$$(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} a_i \mathbf{e}_i$$

*der $\mathbf{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ er et n-tuppel med en 1-er på plass i og resten 0-er. Vi kaller $\mathcal{S} : \mathbf{e}_1, \ldots, \mathbf{e}_n$ for **standardbasisen** for $\mathbb{R}^n$. At mengden er lineært uavhengig følger av at dersom*

$$(a_1, a_2, \ldots, a_n) = \sum_{i=1}^{n} a_i \mathbf{e}_i = \mathbf{0}$$

så må alle $a_i = 0$.

Det siste eksemplet viser at dimensjonsbegrepet passer med vår intuitive oppfatning av dimensjon, planet $\mathbb{R}^2$ har dimensjon 2 og rommet $\mathbb{R}^3$ har dimensjon 3.

Et vektorrom trenger ikke å ha endelig dimensjon. Vi kan likevel skrive opp en basis, selv om denne mengden ikke er endelig.

Eksempel 4.7.11. *La $\mathcal{P}$ være vektorrommet av polynomer med reelle koeffisienter. Anta at dette rommet har endelig dimensjon, og at vi kan skrive opp en endelig basis for vektorrommet. La n være den høyeste graden av polynomene som inngår i basisen. Vi vet at et polynom av høyere grad enn n ikke kan skrives som en lineærkombinasjon av polynomer av grad mindre enn eller lik n, noe som ville bety at basisen ikke spenner ut hele rommet. Det er imidlertid et krav til en basis, og en endelig mengde av polynomer kan derfor ikke utgjøre en basis for $\mathcal{P}$.*

Proposisjon 4.7.8. *Mengden*

$$\mathcal{B} = \{1, x, x^2, x^3, \dots\}$$

danner en basis for $\mathcal{P}$.

Bevis. Elementene i $\mathcal{P}$ er på formen

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

for et ikke-negativt tall n og $a_i \in \mathbb{R}$. Det betyr at $\mathcal{B}$ spenner ut hele $\mathcal{P}$. Vi må også vise at $\mathcal{B}$ er lineært uavhengig. Anta det motsatte, det vil si at det finnes reelle tall $a_0, a_1, \dots, a_n$ slik at

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0$$

Ingen ikke-trivielle polynomer av grad n kan ha mer enn n røtter. La x være et reelt tall som ikke er blant disse røttene. Det betyr at

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \neq 0$$

og vår antagelse måtte i utgangspunktet ha vært feil. Dermed har vi vist at $\mathcal{B}$ er en basis for $\mathcal{P}$. □

En delmengde av et vektorrom som selv er et vektorrom kalles et **underrom**. For å vise at en delmengde er et underrom holder det å vise at mengden er lukket under addisjon og skalarmultiplikasjon, dvs. at dersom $\mathbf{x}, \mathbf{y} \in V$ og $a, b \in \mathbb{R}$ er reelle tall, så er også $a\mathbf{x} + b\mathbf{y} \in V$. Merk at dette ene kravet inneholder både lukkethet under addisjon (sett $a = b = 1$) og skalar multiplikasjon (sett $b = 0$).

Definisjon 4.7.9. *La V være et vektorrom, og la $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$ være en endelig eller uendelig delmengde av V. La W være det minste vektorrommet som inneholder $\mathcal{V}$. Vi sier at mengden $\mathcal{V}$ genererer W, og vi skriver*

$$W = \langle \mathbf{v}_1, \mathbf{v}_2, \dots \rangle.$$

I dette tilfellet vil W utgjøre et underrom av V.

Eksempel 4.7.12. *La*

$$\mathcal{B} = \{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \}.$$

Da vil $\mathcal{B}$ utspenne et underrom W av $\mathbb{R}^3$ gitt ved

$$W = \{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \ : \ a, b \in \mathbb{R} \}.$$

Et monom er et polynom med kun ett ledd, mono=en, poly=mange.

Eksempel 4.7.13. *Mengden av polynomer av grad mindre enn eller lik n, som vi skriver $\mathcal{P}_n$, utgjør et underrom av vektorrommet av alle polynomer. Rommet er generert av* ***monomene*** *av grad mindre enn eller lik n;*

$$\mathcal{P}_n = \langle 1, x, x^2, x^3, \ldots, x^n \rangle$$

Gitt en matrise A. Mengden av kolonnevektorer i A utspenner et vektorrom som vi kaller **kolonnerommet** til A.

Definisjon 4.7.10. *La A være en matrise. Dimensjonen til kolonnerommet til A kalles* ***rangen*** *til matrisen, vi skriver dette tallet som* $\text{rank}(A)$.

Eksempel 4.7.14. *Betrakt matrisen*

$$\begin{bmatrix} 1 & -1 & 2 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Kolonnerommet er utspent av vektorene

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$$

som vi tidligere har sett er lineært uavhengig. Rangen til matrisen er derfor 3.

Siden det er n kolonner i en $n \times n$-matrise vil dimensjonen til kolonnerommet alltid være mindre enn eller lik n. Dersom kolonnene utgjør en lineært uavhengig mengde vil matrisen ha **maksimal rang**, det vil si n for en $n \times n$-matrise. Maksimal rang betyr at kolonnene utspenner hele $\mathbb{R}^n$.

Proposisjon 4.7.11. *En $n \times n$-matrise A har maksimal rang hvis og bare hvis matrisen er invertibel.*

Bevis. Observer først at dersom en vektor $\mathbf{w}$ ligger i spennet av kolonnevektorene til A, så er det ekvivalent med at det finnes en annen vektor $\mathbf{v}$ slik at $A\mathbf{v} = \mathbf{w}$.

Anta at A er invertibel og $\mathbf{w}$ en vektor i $\mathbb{R}^n$. Sett $\mathbf{v} = A^{-1}\mathbf{w}$. Da har vi

$$A\mathbf{v} = AA^{-1}\mathbf{w} = \mathbf{w}$$

og $\mathbf{w}$ ligger i kolonnerommet. Anta så at matrisen har maksimal rang. Det betyr at for en hver vektor $\mathbf{w}$ i kolonnerommet til A, så finnes det en vektor $\mathbf{v}$ slik at $A\mathbf{v} = \mathbf{w}$. Siden A har maksimal rang, vil standardvektorene $\mathbf{e}_i$ ligge i kolonnerommet til A. La $\mathbf{v}_i$ oppfylle $A\mathbf{v}_i = \mathbf{e}_i$, og la B være $n \times n$-matrisen med $\mathbf{v}_i$ som sine kolonner. Da vil

$$AB = I$$

som viser at $B = A^{-1}$ og A er invertibel. □

4.8 Lineære avbildninger

Vektorrom er karakterisert ved at de er lukket under addisjon og skalar multiplikasjon, og at de derfor inneholder $\mathbf{0}$-elementet og additive inverser. Når vi skal definere hva en avbildning mellom to vektorrom bør være, er det da naturlig å se på avbildninger som respekterer akkurat disse strukturene.

Definisjon 4.8.1. *La V og W være to vektorrom. En avbildning $T : V \to W$ kalles en* ***lineær avbildning*** *dersom*

$$T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y})$$

for alle $\mathbf{x}, \mathbf{y} \in V$ og $a, b \in \mathbb{R}$. Dersom $W = V$, så kaller vi avbildningen $T : V \to V$ en ***lineær operator****.*

Proposisjon 4.8.2. *La V og W være to vektorrom og $T : V \to W$ en lineær avbildning. Da har vi at $T(\mathbf{0}) = \mathbf{0}$ og $T(-\mathbf{v}) = -T(\mathbf{v})$.*

Bevis. Vi har

$$T(\mathbf{0}) = T(0 \cdot \mathbf{v}) = 0 \cdot T(\mathbf{v}) = \mathbf{0}$$

og

$$T(-\mathbf{v}) + T(\mathbf{v}) = T(-\mathbf{v} + \mathbf{v}) = T(\mathbf{0}) = \mathbf{0}$$

Siden $-T(\mathbf{v}) + T(\mathbf{v}) = \mathbf{0}$ og inverser er entydige, så følger resultatet i proposisjonen. □

Eksempel 4.8.1. *En lineær avbildning $T : \mathbb{R} \to \mathbb{R}$ må være på formen $T(x) = ax$ for et vilkårlig reelt tall $a \in \mathbb{R}$.*

Eksempel 4.8.2. *Avbildningen gitt ved*

$$T(x_1, x_2) = (x_1 + x_2, x_2 - 2x_1)$$

er en lineær operator på $\mathbb{R}^2$.

Eksempel 4.8.3. *Vi lar V være vektorrommet av alle deriverbare funksjoner på hele $\mathbb{R}$. Derivasjon er en lineær operator på dette rommet, dvs. hvis vi setter $D(f(x)) = f'(x)$ så har vi*

$$\begin{aligned} D(af(x) + bg(x)) &= \frac{d}{dx}(af(x) + bg(x)) \\ &= af'(x) + bg'(x) \\ &= aD(f(x)) + bD(g(x)). \end{aligned}$$

Eksempel 4.8.4. *La $\mathcal{T} = \langle \cos t, \sin t \rangle$. Da virker derivasjonsoperatoren D som en lineær operator på $\mathcal{T}$ fordi vi har*

$$\begin{aligned} D(\cos t) &= -\sin t \\ D(\sin t) &= \cos t. \end{aligned}$$

Vi betrakter lineæravbildningen $T : \mathbb{R}^2 \to \mathbb{R}^2$ gitt ved $T(x_1, x_2) = (x_1 + x_2, x_2 - 2x_1)$. La $\mathcal{S}$ være standardbasisen for $\mathbb{R}^2$;

$$\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2\}.$$

Da har vi at

$$\begin{aligned} T(\mathbf{e}_1) &= T(1, 0) = (1, -2) \\ T(\mathbf{e}_2) &= T(0, 1) = (1, 1). \end{aligned}$$

Dette kan vi sette opp i en matrise

$$[T]_{\mathcal{S}} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix}.$$

Lineæravbildningen blir nå gitt ved matrisemultiplikasjon;

$$[T]_{\mathcal{S}} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ x_2 - 2x_1 \end{bmatrix}$$

Vi kommer til å bruke notasjonen $[T]$ om en matrise som representerer en lineæravbildning T. Vi sier at $[T]_{\mathcal{S}}$ er **matriseformen til T relativt til basisen $\mathcal{S}$.**

La A være en $n \times n$-matrise og definer avbildningen $T : \mathbb{R}^n \to \mathbb{R}^n$ ved

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

for en vektor $\mathbf{v} = (v_1, \ldots, v_n)$. Da har vi

$$\begin{aligned} T(a\mathbf{v} + b\mathbf{w}) &= A(a\mathbf{v} + b\mathbf{w}) \\ &= aA\mathbf{v} + bA\mathbf{w} \\ &= aT(\mathbf{v}) + bT(\mathbf{w}) \end{aligned}$$

som viser at multiplikasjon med en matrise gir oss en lineæravbildning.

Eksempel 4.8.5. *La D være derivasjonsoperatoren slik vi har definert den tidligere. Vi skal se på D som en lineær operator*

$$D : \mathcal{T} \to \mathcal{T}$$

hvor $\mathcal{T} = \langle \cos t, \sin t \rangle$, med basis $\{\cos t, \sin t\}$. Siden

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

kan vi skrive

$$[D]_{\mathcal{T}} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

hvor

$$[D]_{\mathcal{T}} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} b \\ -a \end{bmatrix}$$

reflekterer at

$$D(a\cos t + b\sin t) = b\cos t - a\sin t.$$

Definisjon 4.8.3. *La A være en $n \times n$-matrise og definer en lineæravbildning $T : \mathbb{R}^n \to \mathbb{R}^n$ ved*

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

Rangen $\operatorname{rank}(T)$ *er definert som rangen til matrisen A.*

Eksempel 4.8.6. *Rangen til derivasjons-operatoren i forrige eksempel er 2, siden matrisen er invertibel og derfor har maksimal rang.*

Eksempel 4.8.7. *Vi betrakter en lineæravbildning* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *gitt ved*

$$T(x_1, x_2, x_3) = (x_1 + x_2 + x_3, x_1 - x_2 + 2x_3, 2x_1 + 3x_3)$$

Vi er interessert i å finne tripler (x_1, x_2, x_3) *slik at* $T(x_1, x_2, x_3) = (0, 0, 0)$. *Det betyr at vi må løse likningssystemet*

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 - x_2 + 2x_3 &= 0 \\ 2x_1 + 3x_3 &= 0 \end{aligned}$$

Her observerer vi at den tredje likningen faktisk er summen av de to første, og derfor er overflødig. På den annen side gir den tredje likningen at

$$x_3 = -\frac{2}{3}x_1$$

og setter vi det inn i den første likningen, får vi

$$x_2 = -x_1 - x_3 = -x_1 + \frac{2}{3}x_1 = -\frac{1}{3}x_1.$$

Samler vi disse opplysningene ser vi at

$$T(x_1, -\frac{1}{3}x_1, -\frac{2}{3}x_1) = (0, 0, 0).$$

Alle vektorer på formen

$$(x_1, -\frac{1}{3}x_1, -\frac{2}{3}x_1), \quad x_1 \in \mathbb{R}$$

utgjør et vektorrom, det såkalte nullrommet til avbildningen T.

Definisjon 4.8.4. *La* $T : V \to W$ *være en lineær-avbildning. Mengden*

$$\mathcal{N}(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$$

kalles ***nullrommet*** *eller* ***kjernen*** *til* T.

Proposisjon 4.8.5. *Nullrommet* $\mathcal{N}(T)$ *til* T *er et underrom av* V.

Bevis. La $\mathbf{v}, \mathbf{w} \in \mathcal{N}(T)$, og $a, b \in \mathbb{R}$. Da har vi

$$T(a\mathbf{v} + b\mathbf{w}) = aT(\mathbf{v}) + bT(\mathbf{w}) = a \cdot \mathbf{0} + b \cdot \mathbf{0} = \mathbf{0}$$

som viser at nullrommet er et vektorrom. □

Eksempel 4.8.8. *La $T : \mathbb{R}^2 \to \mathbb{R}^2$ være gitt ved*

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

der A er matrisen

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Vi har

$$A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

og $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ligger derfor i nullrommet til T. Faktisk vil denne vektoren alene danne en basis for nullrommet til T.

Eksempel 4.8.9. *La $T : \mathbb{R}^3 \to \mathbb{R}^3$ være gitt ved*

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

der A er matrisen

$$A = \begin{bmatrix} 1 & -1 & 2 \\ 1 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Dersom en vektor

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

ligger i nullrommet til T, betyr det at

$$v_1\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + v_2\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + v_3\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} = \mathbf{0}.$$

Men vi har i et tidligere eksempel sett at søylevektorene til A er lineært uavhengig, som betyr at

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \mathbf{0}$$

og nullrommet består kun av nullvektoren.

Proposisjon 4.8.6. *La $T : V \to V$ være en lineær avbildning gitt ved en kvadratisk matrise A;*

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

Da har vi at $\mathcal{N}(T) = \{\mathbf{0}\}$ hvis og bare hvis A er invertibel.

Bevis. Anta A er invertibel og la $T(\mathbf{v}) = A\mathbf{v} = \mathbf{0}$. Det betyr at $\mathbf{v} = A^{-1}\mathbf{0} = \mathbf{0}$.

Anta så at $\mathcal{N}(T) = \mathbf{0}$. La $\mathbf{v} = (v_1, \ldots, v_n)$ være en lineær avhengighet mellom kolonnene i A. Da har vi at $A\mathbf{v} = \mathbf{0}$. Siden nullrommet er trivielt betyr det at $\mathbf{v} = \mathbf{0}$ som igjen betyr at A har maksimal rang, som igjen betyr at A er invertibel. □

Det er en nær sammenheng mellom rangen til en lineær avbildning og dimensjonen til nullrommet. Vi har sett at vi kan uttrykke lineære sammenhenger mellom kolonnene i en matrise ved n-tupler, slik som i et tidligere eksempel: Vektorene

$$\begin{bmatrix}2\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}2\\2\end{bmatrix}$$

er lineært avhengig og vi har

$$2\begin{bmatrix}2\\1\end{bmatrix} - 2\begin{bmatrix}1\\0\end{bmatrix} - \begin{bmatrix}2\\2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}.$$

Setter vi de tre vektorene sammen i en 2×3-matrise

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

så vil den lineære avhengigheten gi oss et en vektor i nullrommet;

$$\begin{bmatrix}2 & 1 & 2\\1 & 0 & 2\end{bmatrix}\begin{bmatrix}2\\-2\\-1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Dette gir oss en korrespondanse mellom nullvektorer og lineære avhengigheter mellom søylene i matrisen. Vi kan formulere dette i et generelt resultat:

Teorem 4.8.7. *En lineæravbildning $T : \mathbb{R}^n \to \mathbb{R}^m$ er gitt ved $T(\mathbf{v}) = A\mathbf{v}$ for en $m \times n$-matrise A. Da har vi*

$$\operatorname{rank}(T) + \dim(\mathcal{N}(T)) = n.$$

Bevis. La $\mathcal{B}_0 = \{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ være en basis for nullrommet til A. Vi kan utvide denne basisen til en basis for hele $\mathbb{R}^n$, $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_{r+1}, \ldots, \mathbf{w}_n\}$. Vi skal vise at $\operatorname{rank}(T) = n - r$. La $\mathbf{u} = T(\mathbf{v})$. Siden $\mathcal{B}$ er en basis for $\mathbb{R}^n$ kan vi skrive

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r + a_{r+1}\mathbf{w}_{r+1} + \cdots + a_n\mathbf{w}_n$$

og

$$\begin{aligned}\mathbf{u} = T(\mathbf{v}) &= a_1T(\mathbf{v}_1) + \cdots + a_rT(\mathbf{v}_r) + a_{r+1}T(\mathbf{w}_{r+1}) + \cdots + a_nT(\mathbf{w}_n) \\ &= a_{r+1}T(\mathbf{w}_{r+1}) + \cdots + a_nT(\mathbf{w}_n)\end{aligned}$$

Det betyr at $T(\mathbf{w}_{r+1}), \ldots, T(\mathbf{w}_n)$ spenner ut $T(\mathbb{R}^n)$. Samtidig har vi at dersom

$$\begin{aligned}&a_{r+1}T(\mathbf{w}_{r+1}) + \cdots + a_nT(\mathbf{w}_n) \\ &\quad = T(a_{r+1}\mathbf{w}_{r+1} + \cdots + a_n\mathbf{w}_n) = \mathbf{0}\end{aligned}$$

så ligger $a_{r+1}\mathbf{w}_{r+1} + \cdots + a_n\mathbf{w}_n$ i nullrommet til A, dvs.

$$a_{r+1}\mathbf{w}_{r+1} + \cdots + a_n\mathbf{w}_n = a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r$$

Dermed kan vi skrive

$$a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r - a_{r+1}\mathbf{w}_{r+1} - \cdots - a_n\mathbf{w}_n = \mathbf{0}$$

og siden $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}_{r+1}, \ldots, \mathbf{w}_n\}$ er en basis for $\mathbb{R}^n$ og derfor danner en lineært uavhengig mengde, så må vi ha $a_{r+1} = \cdots = a_n = 0$ og mengden $\{T(\mathbf{w}_{r+1}), \ldots, T(\mathbf{w}_n)\}$ er lineært uavhengig. Det følger at dimensjonen til $T(\mathbb{R}^n)$ som er det samme som rangen til T, er $n - r$. □

Vi har en litt mer generell formulering av dette resultatet. I stedet for å se på $T : \mathbb{R}^n \to \mathbb{R}^m$ kan vi betrakte lineæravbildninger $T : V \to W$, hvor V er et endelig-dimensjonalt vektorrom. Da ser resultatet ut som følger

$$\operatorname{rank}(T) + \dim(\mathcal{N}(T)) = \dim(V).$$

Eksempel 4.8.10. *La $T : \mathbb{R}^2 \to \mathbb{R}^2$ være gitt ved*

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

der A er matrisen

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Kolonnerommet til A er generert av

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

og nullrommet av

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Dermed har vi

$$\operatorname{rank}(T) + \dim(\mathcal{N}(T)) = 1 + 1 = 2.$$

Eksempel 4.8.11. *La $\mathcal{P}_n$ være vektorrommet av polynomer av grad mindre enn eller lik n. La $D : \mathcal{P}_n \to \mathcal{P}_n$ være derivasjonsoperatoren $D(f(x)) = f'(x)$. Vi vet at de eneste polynomene som deriveres på 0 er konstantpolynomene. Det betyr at nullrommet til D har dimensjon 1, generert av $f(x) = C$. Dimensjonen til $\mathcal{P}_n$ er $n+1$ og det følger at* $\text{rank}T = n$.

Vi definerte tidligere egenvektorer og egenverdier for en matrise. Vi skal se at egenvektorer som svarer til forskjellige egenverdier danner lineært uavhengige mengder.

Proposisjon 4.8.8. *La $T : \mathbb{R}^n \to \mathbb{R}^n$ være en lineæravbildning gitt ved $T(\mathbf{v}) = A\mathbf{v}$ for en $n \times n$-matrise A. La videre $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ være egenvektorer for A svarende til* **ulike** *egenverdier $\lambda_1, \lambda_2, \ldots, \lambda_r$. Da er mengden*

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$$

lineært uavhengig.

Bevis. La først $r = 2$, og anta at $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 = \mathbf{0}$. Anta at det finnes en lineær avhengighet mellom de to vektorene, anta f.eks at $a_1 \neq 0$. Da kan vi skrive

$$\mathbf{v}_1 = -\frac{a_2}{a_1}\mathbf{v}_2$$

og siden $\mathbf{v}_1$ og $\mathbf{v}_2$ begge er egenvektorer får vi

$$T(\mathbf{v}_1) = \lambda_1\mathbf{v}_1 = -\frac{a_2}{a_1}\lambda_1\mathbf{v}_2$$

og samtidig

$$T(\mathbf{v}_1) = T(-\frac{a_2}{a_1}\mathbf{v}_2) = -\frac{a_2}{a_1}T(\mathbf{v}_2) = -\frac{a_2}{a_1}\lambda_2\mathbf{v}_2$$

Det følger at

$$-\frac{a_2}{a_1}\lambda_1 = -\frac{a_2}{a_1}\lambda_2$$

og siden $\lambda_1 \neq \lambda_2$ så må $a_2 = 0$. Siden $\mathbf{v}_1 \neq \mathbf{0}$ følger det at også $a_1 = 0$, som gir oss en motsetning.

Anta så at $r = 3$ og at vi har en lineær avhengighet

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 = \mathbf{0}$$

Vi kan uten videre anta at $a_1 \neq 0$ og ved å dele likningen på a_1 kan vi like godt sette a_1 til å være 1. Det gir

$$\mathbf{v}_1 = -a_2\mathbf{v}_2 - a_3\mathbf{v}_3$$

Bruker vi nå T på dette uttrykket får vi

$$\lambda_1 \mathbf{v}_1 = -a_2\lambda_2\mathbf{v}_2 - a_3\lambda_3\mathbf{v}_3$$

Vi multipliserer uttrykket over med λ_1 og trekker de to likningene fra hverandre. Det gir

$$\mathbf{0} = -a_2(\lambda_2 - \lambda_1)\mathbf{v}_2 - a_3(\lambda_3 - \lambda_1)\mathbf{v}_3$$

Dermed er vi tilbake i situasjonen med to egenvektorer og siden $\lambda_2 - \lambda_1, \lambda_3 - \lambda_1 \neq 0$ følger det at koeffisientene $a_2 = a_3 = 0$, og vektorene er lineært uavhengige. For $r > 3$ følger vi nå nøyaktig samme prosedyre. □

Vi har sett at en $n \times n$-matrise kan ha maksimalt n egenverdier. Anta nå at A har nøyaktig n egenverdier. Det betyr at vi kan finne n egenvektorer som utgjør en lineært uavhengig mengde. Siden vektorrommet vi opererer på i dette tilfellet er av dimensjon n betyr det at vi kan finne en basis av egenvektorer. Vi formulerer dette i en proposisjon.

Proposisjon 4.8.9. *La V være et n-dimensjonalt vektorrom. La $T : V \to V$ være en lineæravbildning gitt ved $T(\mathbf{v}) = A\mathbf{v}$ for en $n \times n$-matrise A. Anta at A har n forskjellige egenverdier. Da finnes en basis for V av egenvektorer for A.*

Eksempel 4.8.12. *La A være 2×2-matrisen*

$$A = \begin{bmatrix} 2 & 0 \\ -2 & 1 \end{bmatrix}.$$

Det karakteristiske polynomet til A er gitt ved

$$\chi_A(\lambda) = \lambda^2 - 3\lambda + 2$$

som gir egenverdier $\lambda_1 = 1$, $\lambda_2 = 2$, med tilhørende egenvektorer

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Vi ser at mengden

$$\{\mathbf{v}_1, \mathbf{v}_2\}$$

er lineært uavhengig og danner en basis for $\mathbb{R}^2$.

Vi har tidligere sett at vektorrom kan bestå av funksjoner. I slike tilfeller er det naturlig å bruke betegnelsen **egenfunksjoner** i stedet for egenvektorer, men definisjonen er den samme.

Definisjon 4.8.10. *La $V = \mathcal{F}(I)$ være et vektorrom av funksjoner, og $T : V \to V$ en lineær operator på dette rommet. En ikke-triviell funksjon f som oppfyller $T(f) = \lambda f$ for et reelt tall λ kalles en* ***egenfunksjon*** *for operatoren.*

Eksempel 4.8.13. *La $V = \{ae^x + be^{-x} \mid a, b \in \mathbb{R}\}$ og la $D : V \to V$ være derivasjonsoperatoren. Da er $f_1(x) = e^x$ en egenfunksjon for D med egenverdi $\lambda_1 = 1$ og $f_2(x) = e^{-x}$ en egenfunksjon med egenverdi $\lambda_2 = -1$.*

Eksempel 4.8.14. *La $\mathcal{T} = \{\sin t, \cos t\}$ og $\Delta = D^2 : \mathcal{T} \to \mathcal{T}$ være den 1-dimensjonale* ***Laplaceoperatoren****. Da er alle lineære kombinasjoner $a \sin x + b \cos x$ egenfunksjoner for Δ med egenverdi $\lambda = -1$;*

$$\begin{aligned} D^2(a\sin x + b\cos x) &= D(a\cos x - b\sin x) \\ &= -a\sin x - b\cos x \end{aligned}$$

En spesiell type matriser dukker opp i forbindelse med sannsynlighet. En *stokastisk matrise* er en $n \times n$ matrise der hvert element er et ikkenegativt tall, og summen av elementene i hver rad er 1. F.eks. er følgende en stokastisk matrise

$$P = \begin{bmatrix} 0.2 & 0.6 & 0.2 \\ 0 & 1 & 0 \\ 0.5 & 0.4 & 0.1 \end{bmatrix}.$$

Her gir hver rad en sannsynlighetsfordeling på tre elementer.

For stokastiske matriser kan vi si noe generelt om egenvektorene. La P være en stokastisk matrise, med egenverdier $\lambda_1 = 1 > \lambda_2 \geq \cdots \geq \lambda_n$ og tilhørende egenvektorer $\mathbf{v}_i$ for $i = 1, 2, \ldots$.

Definer $\sigma(\mathbf{v}) = v_1 + \cdots + v_m$ til å være komponentsummen til vektoren $\mathbf{v}$.

Proposisjon 4.8.11. *Alle egenvektorene til en stokastisk matrise, bortsett fra de med egenverdi 1, har komponentsum lik 0, dvs.*

$$\sigma(\mathbf{v}_i) = 0 \quad i = 2, 3, \ldots, n$$

Bevis. Vi har at $P(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ og siden P er stokastisk, så er $\sum_j p_{ij} = 1$ for alle i. La $\mathbf{v}_i = (v_1, v_2, \ldots, v_m)$. Da har vi

$$\sigma(P(\mathbf{v}_i)) = \sum_{j=1}^{m} \sum_{k=1}^{m} p_{jk} v_k = \sum_{j=1}^{m} v_k = \sigma(\mathbf{v}_i)$$

og samtidig

$$\sigma(P(\mathbf{v}_i)) = \sigma(\lambda \mathbf{v}_i) = \lambda \sigma(\mathbf{v}_i)$$

Det gir at

$$\sigma(\mathbf{v}_i) = \lambda \sigma(\mathbf{v}_i)$$

og derfor $\sigma(\mathbf{v}_i) = 0$ siden $\lambda \neq 1$. $\square$

En størrelse som er nært knyttet til komponentsummen er den såkalte **Manhattan-lengden** (eller ℓ_1-lengden) til en vektor. Manhattan-lengden regnes ut som komponentsummen til vektoren vi får ved å ta absoluttverdien av alle komponentene,

Definisjon 4.8.12. *La* $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ *være en vektor i* $\mathbb{R}^n$*. Vi definerer* ***Manhattan-lengden*** $|\mathbf{v}|_1$ *til* $\mathbf{v}$ *ved*

$$|\mathbf{v}|_1 = \sum_{k=1}^{n} |v_k|.$$

Anta nå at vi har et lukket dynamisk system, gitt ved en stokastisk matrise P. For en tilstand for systemet, dvs. for en vektor $\mathbf{w}$, lar vi $E(\mathbf{w}) = |\mathbf{v}_1|_1$ uttrykke **energien** i tilstanden, mens $S(\mathbf{w}) = |\mathbf{v}_2|_1$, uttrykker **entropien**. Begrepene er hentet fra termodynamikk, ikke uten grunn. I analogien med termodynamikk er det naturlig å knytte den nest største egenverdien λ_2 til tilstandens temperatur gjennom formelen

$$T(P) = \log \frac{1}{|\lambda_2|}.$$

Tilstanden $\mathbf{w}$ kan skrives som en lineærkombinasjon av egenvektorer, som vi skriver om og får

$$\begin{aligned} \mathbf{w} &= \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_m \\ &= E(\mathbf{w})\frac{\mathbf{v}_1}{|\mathbf{v}_1|_1} + S(\mathbf{w})\frac{\mathbf{v}_2}{|\mathbf{v}_2|_1} + \cdots + \mathbf{v}_m. \end{aligned}$$

Multiplikasjon med den stokastiske matrisen P gir

$$\begin{aligned} P\mathbf{w} &= P\mathbf{v}_1 + P\mathbf{v}_2 + \cdots + P\mathbf{v}_m \\ &= E(\mathbf{w})\frac{\mathbf{v}_1}{|\mathbf{v}_1|_1} + \lambda_2 S(\mathbf{w})\frac{\mathbf{v}_2}{|\mathbf{v}_2|_1} + \cdots + \lambda_m \mathbf{v}_m \end{aligned}$$

og dermed

$$S(P\mathbf{w}) = \lambda_2 S(\mathbf{w})$$

eller

$$T(P) = \log \frac{S(\mathbf{w})}{S(P\mathbf{w})}$$

som gir

$$\frac{S(P\mathbf{w})}{S(\mathbf{w})} = e^{-T(P)}$$

eller

$$\frac{S(P\mathbf{w}) - S(\mathbf{w})}{S(\mathbf{w})} = e^{-T(P)} - 1.$$

Dette kan vi tolke som at jo lavere temperatur, dvs. $T(p) \to 0$, så vil entropiendringen også nærme seg 0. Det betyr at det ikke skjer noen endring i systemet. For høye temperaturer vil vi observere en stor endring i entropien, med andre ord systemet vil rask gå mot likevektstilstanden, hvor vi har den laveste entropien (dvs. størst mulig uorden).

Eksempel 4.8.15. *Vi har gitt en stokastisk* 2×2*-matrise*

$$P = \begin{bmatrix} p & q \\ 1-p & 1-q. \end{bmatrix}.$$

Vi kan regne ut egenverdiene ved å finne nullpunktene til det karakteristiske polynomet

$$\det(P - \lambda I) = \lambda^2 - (p + 1 - q)\lambda + (p - q) = 0$$

som gir $\lambda_1 = 1$ *og* $\lambda_2 = p - q$*. De tilhørende egenvektorene er (normert til taxi-lengde 1)*

$$\mathbf{v}_1 = \frac{1}{p+1-q}(q, 1-p), \quad \mathbf{v}_2 = \frac{1}{2}(1, -1).$$

For en tilstandsvektor har vi

$$\mathbf{w} = \begin{bmatrix} x \\ y \end{bmatrix} = (x+y)\mathbf{v}_1 + 2\frac{(1-p)x - qy}{1+q-p}\mathbf{v}_2$$

som gir at

$$E(\mathbf{w}) = x + y$$
$$S(\mathbf{w}) = 2\frac{(1-p)x - qy}{1+q-p}$$

og temperatur

$$T(P) = \log \frac{1}{|p-q|}.$$

4.9 Oppgaver

Oppgave 1. *Betrakt vektorene* $\mathbf{a} = (1, 0)$*,* $\mathbf{b} = (0, 2)$*, og* $\mathbf{c} = (1, 3)$

a) Tegn vektorene i et vanlig koordinatsystem; tegn alle med startpunkt i origo.

b) Tegn deretter de samme vektorene, men med startpunkt i $(2, 2)$.

c) Regn ut normen (lengden) av hver av vektorene, altså $\|\mathbf{a}\|$, $\|\mathbf{b}\|$ *og* $\|\mathbf{c}\|$.

Oppgave 2. *La* $\mathbf{a}$, $\mathbf{b}$ *og* $\mathbf{c}$ *være som i forrige oppgave.*

a) Regn ut $\mathbf{x} = \mathbf{a} + \mathbf{b}$ *og* $\mathbf{y} = \mathbf{b} + \mathbf{c}$.

b) Tegn $\mathbf{x}$ *og* $\mathbf{y}$ *i samme koordinatsystem som foran.*

c) Bruk forrige punkt til å sjekke den geometriske tolkningen av vektoraddisjon.

Oppgave 3. *La* $\mathbf{a}$, $\mathbf{b}$ *og* $\mathbf{c}$ *være som i forrige oppgave.*

a) Regn ut $2\mathbf{a}$, $-\mathbf{b}$ *og* $3\mathbf{c}$.

b) Tegn igjen vektorene fra a) og sjekk geometrisk tolkning.

Oppgave 4. *Finn tre vektorer* $\mathbf{x}$, $\mathbf{y}$ *og* $\mathbf{z}$ *i planet, alle ulik nullvektoren, slik at*

$$\mathbf{x} + \mathbf{y} + \mathbf{z} = \mathbf{0}.$$

Oppgave 5. *La* $\mathbf{v} = (1, 1, 1)$ *og* $\mathbf{w} = (2, 0, 3)$.

a) Beregn $\mathbf{v} + \mathbf{w}$, *og* $3\mathbf{v}$.

b) Beregn normen (lengden) til de to vektorene, dvs. $\|\mathbf{v}\|$ *og* $\|\mathbf{w}\|$.

c) Beregn skalarproduktet $\mathbf{v} \cdot \mathbf{w}$.

d) Bestem den transponerte til $\mathbf{v}$, *altså* $\mathbf{v}^T$.

Oppgave 6. *La igjen* $\mathbf{v} = (1, 1, 1)$ *and* $\mathbf{w} = (2, 0, 3)$.

a) Finn en vektor $\mathbf{x}$ *som løser likningen*

$$\mathbf{v} + 2\mathbf{w} + \mathbf{x} = \mathbf{0}.$$

b) Kan du finne alle *vektorer* $\mathbf{x}$ *som løser likningen over?*

Oppgave 7. *La* $\mathbf{v} = (2, 1, 1)$ *og* $\mathbf{w} = (-1, 1, 0)$.

a) Finn normen (lengden) til $\mathbf{v}$*, altså* $\|\mathbf{v}\|$*.*

b) Betrakt funksjonen $f : \mathbb{R} \to \mathbb{R}$ *gitt ved* $f(s) = \|\mathbf{v} + s\mathbf{w}\|^2$*.*

c) Regn ut $f(s)$ *for et par verdier av* s*. Prøv å finne minimum av* f*.*

Oppgave 8. *La* $\mathbf{v}_1 = (2, 1, 1)$ *og* $\mathbf{v}_2 = (-1, 1, 0)$*.*

a) Beregn lineærkombinasjonene $\mathbf{v}_1 + 2\mathbf{v}_2$*,* $3\mathbf{v}_1 + 2\mathbf{v}_2$*,* $1.5\mathbf{v}_1 - 2\mathbf{v}_2$*.*

b) Betrakt lineærkombinasjonen $a\mathbf{v}_1 + b\mathbf{v}_2$*. Skriv opp likningen*

$$a\mathbf{v}_1 + b\mathbf{v}_2 = \mathbf{0}$$

på komponentform.

c) Bruk forrige punkt til å vise at $\mathbf{v}_1$ *og* $\mathbf{v}_2$ *er lineært uavhengige. (Hint: vis at eneste løsning er* $a = b = 0$*.)*

c) Siden vi her hadde to vektorer kunne vi også ha argumentert slik: $\mathbf{v}_1$ *og* $\mathbf{v}_2$ *er ikke parallelle, så derfor er de lineært uavhengige.*

Oppgave 9. *La* $\mathbf{v}_1 = (1, 0)$*,* $\mathbf{v}_2 = (0, 1))$ *og* $\mathbf{v}_3 = (3, 2)$*.*

a) Kan du finne en lineærkombinasjon $\mathbf{z}$ *av* $\mathbf{v}_1$ *og* $\mathbf{v}_1$ *slik at* $\mathbf{z} = \mathbf{v}_3$*?*

b) Er $\mathbf{v}_1$*,* $\mathbf{v}_2$*,* $\mathbf{v}_3$ *lineært uavhengige?*

Oppgave 10. *La* $\mathbf{v}_1 = (1, 1, 1, 1)$*,* $\mathbf{v}_2 = (0, 1, 1, 1)$*,* $\mathbf{v}_3 = (0, 0, 1, 1)$*,* $\mathbf{v}_4 = (0, 0, 0, 1)$*.*

a) Vis at $\mathbf{v}_1$*,* $\mathbf{v}_2$*,* $\mathbf{v}_3$*,* $\mathbf{v}_4$ *er lineært uavhengige*

Oppgave 11. *La* $\mathbf{v}_1 = (a, 1, 1, 1)$*,* $\mathbf{v}_2 = (0, b, 1, 1)$*,* $\mathbf{v}_3 = (0, 0, c, 1)$*,* $\mathbf{v}_4 = (0, 0, 0, d)$ *der* a, b, c, d *er reelle tall.*

a) Vis at $\mathbf{v}_1$*,* $\mathbf{v}_2$*,* $\mathbf{v}_3$*,* $\mathbf{v}_4$ *er lineært uavhengige hvis alle tallene* a, b, c, d *er ulik 0.*

b) Hva skjer hvis minst ett av tallene a, b, c, d *er lik 0?*

Oppgave 12. *La* $\mathbf{v}_1$ *og* $\mathbf{v}_2$ *være to vektorer i* $\mathbb{R}^n$*, begge ulik nullvektoren. Anta at* $\mathbf{v}_1, \mathbf{v}_2$ *er lineært uavhengige.*

a) Finn en vektor $\mathbf{w} \neq \mathbf{0}$ *slik at* $\mathbf{v}_1, \mathbf{v}_2, \mathbf{w}$ *er lineært avhengige.*

b) Kan du finne alle *slike vektorer* $\mathbf{w}$ *med egenskapene i a)?*

Oppgave 13. *La*

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 4 \\ 0 & 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 1 \end{bmatrix}.$$

a) Beregn $2A$, $0B$, $-B$, $A + B$, $A - B$, $2A + B$.

b) Beregn A^T *og* B^T. *Hva er* $(A + B)^T$ *?*

c) Beregn $A + B + A^T$.

Oppgave 14. *La*

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}.$$

a) Beregn $\mathbf{x}^T$ *og* $(\mathbf{x}^T)^T$.

b) Forklar hvorfor $\mathbf{x} + \mathbf{x}^T$ *ikke har mening.*

Oppgave 15. *La*

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 3 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

a) Hvorfor har produktet $A\mathbf{x}$ *mening? (sjekk dimensjon)*

b) Beregn $A\mathbf{x}$.

c) $A\mathbf{x}$ *er en lineærkombinasjon av hvilke vektorer (ut fra* A*)?*

Oppgave 16. *La*

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 3 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}.$$

La $\mathbf{x}$ *og* $\mathbf{y}$ *være hhv. første og andre kolonne i* B.

a) Beregn $A\mathbf{x}$ *og* $A\mathbf{y}$.

b) Beregn AB.

c) Hva er sammenhengen mellom det du beregnet i a) og b)?

Oppgave 17. *La*

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

a) La I være 3×3 identitetsmatrisen. Hva er I^2, I^3, etc

b) Beregn P^2 og P^3. Hva med P^4 og videre?

c) Hva er $P\mathbf{x}$ der $\mathbf{x} \in \mathbb{R}^3$?

Oppgave 18. *La*

$$A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix}.$$

a) Hvorfor er A invertibel?

b) Finn den inverse $B = A^{-1}$ til A. Sjekk at $AB = BA = I_2$.

Oppgave 19. *La*

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

a) Se oppgave over der du beregnet potenser av P. Hva er P^3?

b) Vis at P er invertibel og finn P^{-1}.

Oppgave 20. *La*

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 4 \\ 3 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}.$$

La $\mathbf{x}$ og $\mathbf{y}$ være hhv. første og andre kolonne i B.

a) Beregn skalarproduktet (prikkproduktet) $\mathbf{x} \cdot \mathbf{y}$.

b) Beregn matriseproduktet $\mathbf{x}^T\mathbf{y}$.

c) Hva blir $\mathbf{x}\mathbf{y}^T$?

Oppgave 21. *La*

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

a) Vis at A er invertibel og finn A^{-1}.

b) Samme spørsmål, nå for diagonalmatrisen med diagonal a, b, c.

Oppgave 22. *La*

$$A = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}.$$

a) Vis at kolonnene til A er ortogonale (regn ut skalarprodukt), og har norm 1.

b) Vis at A er invertibel og at $A^{-1} = A^T$, dvs. at A er en ortogonal matrise.

Oppgave 23. *La θ være et tall; tenker på det som vinkel i radianer.*

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

a) Vis at kolonnene til A er ortogonale (beregn skalarprodukt), og har norm 1.

b) Vis at A er invertibel og at $A^{-1} = A^T$, dvs. at A er en ortogonal matrise.

Oppgave 24. *La*

$$A = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

a) Løs likningssystemet $A\mathbf{x} = \mathbf{b}$ ved å skreve det på komponentform (og bruke eliminasjon).

b) Beregn A^{-1} og sjekk at $A^{-1}\mathbf{b}$ gir løsningen i a).

Oppgave 25. *La*

$$A = \begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}, \ \ B = \begin{bmatrix} 1 & 0 \\ 4 & 2 \end{bmatrix},$$

a) Beregn determinanten til A og til B.

b) Hva er $\det(AB)$?

Oppgave 26. *La A være en invertibel $n \times n$ matrise.*

a) Hvis $\det(A) = 1/4$*, hva er da* $\det(A^{-1})$*?*

Oppgave 27. *Betrakt det lineære likningssystemet*

$$\begin{array}{rrcl} x_1 & -x_2 & = & 3 \\ 2x_1 & -x_2 & = & 0. \end{array}$$

a) Er $\mathbf{x} = (1, 1)$ *en løsning av systemet?*

b) Løs likningssystemet.

c) Skrive likningssystemet på matriseform $A\mathbf{x} = \mathbf{b}$*.*

d) Finn A^{-1}*, og beregn* $A^{-1}\mathbf{b}$*.*

Oppgave 28. *Betrakt likningssystemet*

$$2x_1 + x_2 = 1.$$

Altså: én likning med to ukjente.

a) Finn to ulike løsninger av systemet.

b) Finn alle løsninger av systemet.

Oppgave 29. *Betrakt det lineære likningssystemet*

$$\begin{array}{rrrcl} x_1 & -2x_2 & +2x_3 & = & 10 \\ & 3x_2 & -x_3 & = & 0 \\ & & 2x_3 & = & 6. \end{array}$$

a) Er $\mathbf{x} = (1, 1)$ *en løsning av systemet?*

b) Løs likningssystemet.

c) Skrive likningssystemet på matriseform $A\mathbf{x} = \mathbf{b}$*. Er A invertibel?*

Oppgave 30. *Se på systemet i forrige oppgave.*

a) Skriv systemet på matriseform $A\mathbf{x} = \mathbf{b}$*.*

b) Løs likningssystemet ved radoperasjoner på den utvidede matrisen

$$\left[\begin{array}{cc} A & \mathbf{b} \end{array}\right].$$

Oppgave 31. *Betrakt det lineære likningssystemet*

$$\begin{array}{rrrcl} x_1 & -2x_2 & +2x_3 & = & 2 \\ & 3x_2 & -x_3 & = & 2 \\ 2x_1 & & +x_3 & = & 5. \end{array}$$

a) Løs likningssystemet.

Oppgave 32. *La A være en $m \times n$ matrise og $\mathbf{b} \in \mathbb{R}^m$.*

a) La $\mathbf{x}_1$ og $\mathbf{x}_2$ begge være løsninger av $A\mathbf{x} = \mathbf{0}$. Vis at da er $\mathbf{x}_1 + \mathbf{x}_2$ også en løsning av samme likning.

b) Samme spørsmål som i a), bortsett fra at du skal vise at $t_1\mathbf{x}_1 + t_2\mathbf{x}_2$ også en løsning av $A\mathbf{x} = \mathbf{0}$, for alle $t_1, t_2 \in \mathbb{R}$.

c) Anta at $\mathbf{x}_1$ og $\mathbf{x}_2$ begge er løsninger av $A\mathbf{x} = \mathbf{b}$. Regn ut $A(t_1\mathbf{x}_1 + t_2\mathbf{x}_2)$. For hvilke t_1, t_2 vil $t_1\mathbf{x}_1 + t_2\mathbf{x}_2$ også være en løsning av $A\mathbf{x} = \mathbf{b}$.

Oppgave 33. *Betrakt mengden $H = \{ae^x + bx^2 : a, b \in \mathbb{R}\}$.*

a) Vis at H er et vektorrom under addisjon og skalar multiplikasjon.

a) Hvilket "naturlig" større vektorrom V inneholder H?

Oppgave 34. *La V være et vektorrom, og la $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ være en mengde med lineært uavhengige vektorer i V.*

a) Vis at delmengden $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ også er lineært uavhengig.

b) Prøv å vise det mer generelle resultatet: Enhver delmengde av en lineært uavhengig mengde er lineært uavhengig.

Oppgave 35. *La $S = \{1, x^3\}$ som er en delmengde i vektorrommet av alle funksjoner fra $\mathbb{R}$ til $\mathbb{R}$.*

a) Vis at er lineært uavhengig.

b) La W være underrommet utspent av S. Er funksjonen $4 - 3x^2$ i W? Hva med funksjonen $2 + x$?

Oppgave 36. *La*

$$S = \{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \}.$$

a) Bestem underrommet W av $\mathbb{R}^3$ som S utspenner.

b) Finn en vektor som ligger i W og en som ikke gjør det.

Oppgave 37. *La T være lineæravbildningen $T : \mathbb{R}^3 \to \mathbb{R}^3$ gitt ved*

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

der A er matrisen

$$A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & -1 & 4 \\ 3 & -3 & 12 \end{bmatrix}.$$

a) Vis at andre og tredje kolonne i A begge er et multippel av første kolonne.

b) Hva er $\operatorname{rank}(A)$?

c) Hva er dimensjonen til nullrommet til A.

d) Finn en vektor, ulik nullvektoren, som ligger i nullrommet til A.

Oppgaver relatert til maskinlæring:

Oppgave 38. *Betrakt den lineære funksjonen:*

$$f(x_1, x_2, \ldots, x_D) = \sum_{d=1}^{D} w_d\, x_d + b.$$

a) Skriv om funksjonen med vektornotasjon.

Hint: *Definer først $\mathbf{w} = [\, w_1, \ldots, w_D \,]^\top$ og $\mathbf{x} = [\, x_1, \ldots, x_D \,]^\top$, og vis at $\sum_{d=1}^{D} w_d\, x_d \;=\; \mathbf{w}^\top \mathbf{x}$.*

b) Forklar at mengden $\left\{ \mathbf{x} \in \mathbb{R}^D \;\middle|\; f(\mathbf{x}) = b \right\}$ står ortogonalt på $\mathbf{w}$.

Oppgave 39. *I maskinlæring/AI er vi ofte interessert i hvor like vektorer er. Dette kan for eksempel være for å sammenligne tekst og bilder.*

Gitt to vektorer $\mathbf{v}_1$ og $\mathbf{v}_2$ kan vi for eksempel se på $\|\mathbf{v}_1 - \mathbf{v}_2\|_2$. En mer brukt metode (i hvert fall innenfor tekstdomenet) er såkalt cosinuslikhet, som gir cosinus til vinkelen mellom vektorene, og er gitt ved:

$$\cos(\mathbf{v}_1, \mathbf{v}_2) = \frac{\mathbf{v}_1^\top \mathbf{v}_2}{\|\mathbf{v}_1\|_2\, \|\mathbf{v}_2\|_2}.$$

a) Vis at

$$\cos(\mathbf{v}_1, \mathbf{v}_2) = \frac{\mathbf{v}_1^\top \mathbf{v}_2}{\|\mathbf{v}_1\|_2 \, \|\mathbf{v}_2\|_2} = \left(\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|_2}\right)^\top \left(\frac{\mathbf{v}_2}{\|\mathbf{v}_2\|_2}\right).$$

Altså at vi både kan regne ut indreproduktet først og deretter normalisere, men også normalisere vektorene først og deretter beregne indreproduktet.

b) Betrakt nå vektorene $\mathbf{w}_1 = [1, 0, 3]^\top$ *og* $\mathbf{w}_2 = [2, -1, 1]^\top$*. Finn cosinuslikhet ved å normalisere vektorene først.*

Vi skal nå se på hvordan vi effektivt kan finne cosinuslikhet mellom mange vektorer. Tenk deg at du har 100 tekster; du har brukt en maskinlæringsmodell for å gjøre om alle disse til vektorer av lengde 64. Altså har vi at $\mathbf{v}_i \in \mathbb{R}^{64}$ *er* kolonnevektorer *for* $i = 1, \ldots, 100$.

c) Forklar at vi kan systematisere denne informasjonen i en (100×64)*-matrise* V*, der rad* i *er* $\mathbf{v}_i^\top$*.*

d) Vi gjør nå matrisemultiplikasjonen $VV^\top$*. Forklar at oppføringen på posisjon* (i, j) *er*

$$(VV^\top)_{ij} \;=\; \mathbf{v}_i^\top \mathbf{v}_j.$$

La $\hat{\mathbf{v}}_i = \mathbf{v}_i / \|\mathbf{v}_i\|_2$ *og la* $\hat{V}$ *være* V *der rad* i *er* $\hat{\mathbf{v}}_i^\top$*. Vis at*

$$(\hat{V}\hat{V}^\top)_{ij} \;=\; \hat{\mathbf{v}}_i^\top \hat{\mathbf{v}}_j \;=\; \cos(\mathbf{v}_i, \mathbf{v}_j),$$

slik at $\hat{V}\hat{V}^\top$ *inneholder alle cosinuslikheter.*

e) Regn ut $\sum_{i=1}^{100} (\hat{V}\hat{V}^\top)_{ii}$*.*

f) Anta at du har n *tekster. Hvor mange indreprodukter må du beregne for å fylle hele cosinuslikhetsmatrisen? Hvordan skalerer dette seg når* n *øker?*

Kapittel 5

Matriser og konveksitet

5.1 Kvadratiske funksjoner og kvadratiske former

De enkleste funksjonene, men også mest brukte, er de *lineære*. Et eksempel på en lineær funksjon $f : \mathbb{R}^2 \to \mathbb{R}$ er

$$f(\mathbf{x}) = 2x_1 + 5x_2 \quad \text{der } \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2.$$

Generelt er en **lineær funksjon** fra $\mathbb{R}^n$ til $\mathbb{R}$ på formen

$$\mathbf{x} \to \mathbf{c}^T\mathbf{x} = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

der $c_1, c_2, \ldots, c_n$ er gitte reelle tall. En **affin** funksjon er en lineær funksjon + en konstant (som kan være 0!). Så disse funksjonene har bare førstegradsledd.

Ofte møter vi mer kompliserte funksjoner med kvadratiske ledd og produkt av par av variable. En **kvadratisk funksjon** fra $\mathbb{R}^n$ til $\mathbb{R}$ er på formen

$$H(\mathbf{x}) = \sum_{i,j} a_{ij}x_ix_j + \sum_i c_ix_i + d$$

der a_{ij}, c_i og d er gitte reelle tall (for hver i og j). Det første leddet her, $\sum_{i,j} a_{ij}x_ix_j$, er den ulineære delen og kalles en **kvadratisk form**. Et eksempel på en kvadratisk funksjon er

$$H(x_1, x_2) = 3x_1^2 - x_2^2 + x_1x_2 + 7x_1 - x_2 + 8,$$

og den tilhørende kvadratiske formen er $Q(x_1, x_2) = 3x_1^2 - x_2^2 + x_1x_2$.

Vi skal nå studere kvadratiske former. Disse dukker opp i geometri når man analyserer kjeglesnitt (i planet, når $n = 2$: ellipse, parabel, hyperbel).

Kvadratiske former er også svært viktige i optimering for å avgjøre om et stasjonært punkt er minimum (minimumspunkt), maksimum, eller ingen av delene. Vi minner om at et **stasjonært punkt** for en funksjon f er et punkt $\mathbf{x}$ der alle partielle deriverte er 0, så

$$\partial f(\mathbf{x})/\partial x_i = 0 \text{ for hver } i.$$

Dette er det samme som kritisk punkt (siden funksjonen er deriverbar) I dette tilfellet vil en bestemt kvadratisk form som involverer annen ordens partiellderiverte, og fortegnet av denne, være avgjørende for om vi har f.eks. minimum.

En **kvadratisk form** (på R^n) er en funksjon $Q : \mathbb{R}^n \to \mathbb{R}$ som er en lineærkombinasjon av ledd av typen x_i^2 eller $x_i x_j$ (der $i \neq j$). Her er noen eksempler når $n = 2$.

Eksempel 5.1.1.

$$Q_1(\mathbf{x}) = x_1^2 + x_2^2, \quad Q_2(x) = 5x_1^2 - 3x_2^2, \quad \textit{og}$$
$$Q_3(\mathbf{x}) = x_1^2 - 8x_1x_2 + 3x_2^2.$$

Vi kommer til å være opptatt av fortegnet til kvadratiske former, altså fortegnet til funksjonsverdien for ulike $\mathbf{x}$. I eksempelet over ser vi at $Q_1(\mathbf{x}) \geq 0$ for *alle* $\mathbf{x} \in \mathbb{R}^2$ fordi Q_1 er sum av to kvadrater. Derimot kan Q_2 bli både positiv og negativ, avhengig av $\mathbf{x}$, f.eks. er $Q_2(1, 0) = 5$ og $Q_2(0, 1) = -3$. For Q_3 er det ikke helt opplagt hva som kan være tilfelle, men det finner vi ut av senere.

En generell kvadratisk form for $n = 2$ er

$$Q(\mathbf{x}) = ax_1^2 + bx_1x_2 + cx_2^2$$

der $\mathbf{x} = (x_1, x_2)$ og a, b og c er gitte konstanter.

Vi skal nå knytte kvadratiske former til matriser, og snart se at enhver kvadratisk form på $\mathbb{R}^n$ kan skrives på formen

$$\mathbf{x}^T A\,\mathbf{x}, \quad \text{der } A \text{ er en symmetrisk } n \times n \text{ matrise.} \tag{5.1}$$

Dette leder til at man kan forenkle kvadratiske former slik vi omtaler senere.

La oss se hvordan vi kan skrive de tre kvadratiske formene i Eksempel 5.1.1 via matriser. Vi har

$$Q_1(\mathbf{x}) = x_1^2 + x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

For å se dette kan det være lurt å huske på følgende generelle resultat (som man får ved å regne ut)

$$\begin{aligned}\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 \\ &= a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2.\end{aligned}$$

Fra dette ser vi også at

$$Q_3(\mathbf{x}) = x_1^2 - 8x_1x_2 + 3x_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -4 \\ -4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Legg merke til at vi valgte å bruke en symmetrisk matrise her. Det er fullt mulig å skrive den kvadratiske formen med usymmetriske matriser, men det er ikke så gunstig. Grunnen er at man kan analysere kvadratiske former ved å studere tilhørende symmetriske matriser, og for slike matriser har vi den viktige spektralteorien å støtte oss til.

Neste eksempel viser hvordan vi håndterer kvadratiske former når $n = 3$.

Eksempel 5.1.2. *La* $\mathbf{x} = (x_1, x_2, x_3) \in \mathbb{R}^3$ *og*

$$Q(\mathbf{x}) = a\,x_1^2 + b\,x_1x_2 + c\,x_2^2 + d\,x_2x_3 + e\,x_3^2 + f\,x_1x_3,$$

Da er

$$Q(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a & b/2 & f/2 \\ b/2 & c & d/2 \\ f/2 & d/2 & e \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Ut fra disse eksemplene skulle det være klart hvordan man gjør samme omskrivning for kvadratiske former på $\mathbb{R}^n$. Vi ser da at vi like gjerne kan definere en kvadratisk form på $\mathbb{R}^n$ som en funksjon $Q : \mathbb{R}^n \to \mathbb{R}$ på formen (5.1), dvs. $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$ for en symmetrisk $n \times n$ matrise A.

Legg merke til at kvadratiske former for diagonalmatriser er "enkle", f.eks.

$$\mathbf{x}^T \begin{bmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \mathbf{x} = 4x_1^2 + x_2^2 + 3x_3^2.$$

Med "enkel" menes at det ikke finnes noen kryssledd av typen x_ix_j der $i \neq j$. Generelt har vi at hvis $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$ og $Q(\mathbf{x}) = \mathbf{x}^T D\mathbf{x}$ der $\mathbf{x} \in \mathbb{R}^n$, så er

$$Q(\mathbf{x}) = d_1x_1^2 + d_2x_2^2 + \cdots + d_nx_n^2.$$

En viktig fordel med slike kvadratiske former Q uten kryssledd er at vi enkelt kan avklare om Q alltid er ≥ 0 (eller alltid ≤ 0, eller kan anta både

positive og negative verdier). Dette spørsmålet er viktig i optimering, og vi vender tilbake til dette.

Det viser seg at *alle* kvadratiske former kan forenkles slik at man ikke har kryssledd! Dette kan gjøres ved å foreta et variabelskifte. La oss se på et eksempel.

Eksempel 5.1.3. *Betrakt den kvadratiske formen*

$$Q(x_1, x_2) = 9x_1^2 + 3x_2^2 - 6x_1x_2$$

der vi altså har kryssledd. Det neste vi gjør kan virke litt magisk! La oss innføre nye variable y_1 og y_2 ved at disse er definert som

$$y_1 = x_1 + x_2, \quad y_2 = 2x_1 - x_2, \tag{5.2}$$

og betrakt en ny kvadratisk form $Q'(y_1, y_2) = y_1^2 + 2y_2^2$. Så, Q' er en kvadratisk form i variablene y_1, y_2, og den er uten kryssledd. Vi setter inn uttrykkene for y_1 og y_2 fra (5.2) og får

$$\begin{aligned} Q'(y_1, y_2) &= y_1^2 + 2y_2^2 \\ &= (x_1 + x_2)^2 + 2(2x_1 - x_2)^2 \\ &= x_1^2 + 2x_1x_2 + x_2^2 + 2(4x_1^2 - 4x_1x_2 + x_2^2) \\ &= 9x_1^2 + 3x_2^2 - 6x_1x_2 \\ &= Q(x_1, x_2). \end{aligned}$$

Dette betyr at

$$Q(x_1, x_2) = Q'(y_1, y_2) \text{ når } (x_1, x_2) \text{ og } (y_1, y_2) \text{ er relatert ved (5.2).}$$

Så med nye variable har vi skrevet om den kvadratiske formen til en ny kvadratisk form uten kryssledd.

Som nevnt i eksempelet kunne det virke nærmest magisk å komme opp med nettopp dette variabelskiftet, og den bestemte nye kvadratiske formen. Neste resultat er viktig, for det sier at en slik forenkling alltid er mulig. *Hvordan* man gjør det, omtaler vi kort under (og detaljer er i stjernemerket seksjon).

Proposisjon 5.1.1. *La $Q(\mathbf{x})$ være en kvadratisk form i variable $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Da finnes det et "lineært skifte" over i nye variable $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ gitt ved*

$$\mathbf{y} = V\mathbf{x}$$

og en kvadratisk form i $\mathbf{y}$, $Q'(\mathbf{y})$, slik at Q' ikke har kryssledd og der

$$Q(\mathbf{x}) = Q'(\mathbf{y}) \text{ når } \mathbf{y} = V\mathbf{x}.$$

Dette resultatet sier altså at kvadratiske former alltid kan forenkles. Beviset for resultatet støtter seg til spektralteoremt for symmetriske matriser, og det gir også en metode for å finne variabelskiftet over. Dette er beskrevet i detalj i (stjernemerket) Seksjon 5.5. Kort fortalt er metoden å finne egenvektorer og egenvektorer for matrisen A (der $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$) og bruke disse til variabelskiftet; man får en ny basis og $\mathbf{y}$ gir koordinatene i denne nye basisen. Egenverdiene $\lambda_1, \lambda_2, \ldots, \lambda_n$ til A dukker opp som koeffisienter foran y_i^2-ene så den nye kvadratiske formen er

$$Q'(\mathbf{y}) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2. \tag{5.3}$$

5.2 Klassifikasjon av matriser og kvadratiske former

La $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$ være en symmetrisk kvadratisk form på $\mathbb{R}^2$, så vi har

$$Q(\mathbf{x}) = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i<j}(a_{ij} + a_{ji})x_i x_j = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i<j} 2a_{ij}x_i x_j$$

Da kan vi bestemme de partielle deriverte. For $i \leq n$ får vi

$$\partial Q(\mathbf{x})/\partial x_i = 2a_{ii}x_i + \sum_{j\neq i} 2a_{ij}x_j = 2(A\mathbf{x})_i.$$

Det betyr at gradienten til Q er

$$\nabla Q(\mathbf{x}) = 2A\mathbf{x}.$$

Dermed ser vi at origo $\mathbf{0}$ er et stasjonært punkt for Q, idet $\nabla Q(\mathbf{0}) = \mathbf{0}$. Et naturlig spørsmål er derfor: hva slags stasjonært punkt er $\mathbf{0}$?

For å svare på dette trenger vi å definere noen viktige begreper. En kvadratisk form Q på $\mathbb{R}^n$ kalles

- **positiv definit** dersom $Q(\mathbf{x}) > 0$ for alle $\mathbf{x} \neq \mathbf{0}$. Da er $\mathbf{0}$ entydig minimumspunkt for Q.
- **positiv semidefinit** dersom $Q(\mathbf{x}) \geq 0$ for alle $\mathbf{x}$. Da er $\mathbf{0}$ minimumspunkt for Q.
- **negativ definit** dersom $Q(\mathbf{x}) < 0$ for alle $\mathbf{x} \neq \mathbf{0}$. Da er $\mathbf{0}$ entydig maksimumspunkt for Q.

- **negativ semidefinit** dersom $Q(\mathbf{x}) \leq 0$ for alle $\mathbf{x}$. Da er $\mathbf{0}$ maksimumspunkt for Q.
- **indefinit** dersom $Q(x)$ antar både positive og negative verdier. Da vil $\mathbf{0}$ være et sadelpunkt for Q.

For en symmetrisk $n \times n$ matrise A, bruker vi akkurat de samme betegnelsene slik at

- A **er positiv definit** dersom $\mathbf{x}^T A\,\mathbf{x} > 0$ for alle $\mathbf{x} \neq \mathbf{0}$.
- A **er positiv semidefinit** dersom $\mathbf{x}^T A\,\mathbf{x} \geq 0$ for alle $\mathbf{x}$.
- A **er negativ definit** dersom $\mathbf{x}^T A\,\mathbf{x} < 0$ for alle $\mathbf{x} \neq \mathbf{0}$.
- A **er negativ semidefinit** dersom $\mathbf{x}^T A\,\mathbf{x} \leq 0$ for alle $\mathbf{x}$.
- A **er indefinit** dersom $\mathbf{x}^T A\,\mathbf{x}$ antar både positive og negative verdier.

Dette betyr at den kvadratiske formen $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$ og A er av samme type, f.eks. er begge positiv semidefinit. Typisk vil man møte disse begrepene for matriser oftere enn for kvadratiske former. Men problemene man ser på kommer gjerne fra kvadratiske funksjoner i flere variable (som er kvadratiske former).

La oss se på disse begrepene for den enkleste typen matriser, nemlig diagonalmatriser.

Eksempel 5.2.1.

$$A = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}$$

der $d_1, d_2 \in \mathbb{R}$. For å avgjøre om f.eks. A er positiv semidefinit, ser vi på den tilhørende kvadratiske formen der $\mathbf{x} = (x_1, x_2)$:

$$Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} = d_1 x_1^2 + d_2 x_2^2.$$

Vi ser at fortegnene på d_1 og d_2 er avgjørende. Hvis $d_1, d_2 \geq 0$, vil $Q(\mathbf{x}) \geq 0$ for alle $\mathbf{x}$ fordi x_1^2 og x_2^2 begge er ≥ 0; så da er A positiv semidefinit.

Tilsvarende, hvis $d_1, d_2 \leq 0$, vil $Q(x) \leq 0$ for alle x; så da er A negative semidefinit. Hvis d_1 og d_2 har motsatt fortegn, en er positiv og den andre er negativ, da vil A være indefinit. Dette ser vi ved å sette inn hhv. $(1, 0)$ og $(0, 1)$.

Endelig, på samme måte ser vi at A er positiv definit hvis $d_1, d_2 > 0$, og A er negativ definit hvis $d_1, d_2 < 0$.

Neste teorem sier oss hvordan egenverdiene til A bestemmer hvilken type kvadratisk form vi har. Vi finner svar på dette ved å tilbakeføre spørsmålet til situasjonen vi nettopp behandlet, altså en diagonalmatrise, og bruke resultatet vi nevnte i forrige seksjon, se (5.3).

Theorem 5.2.1. *La A være en $n \times n$ symmetrisk matrise. Den kvadratiske formen $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$ på $\mathbb{R}^n$, og matrisen A, er*

- *positiv definit $\Leftrightarrow$ alle egenverdiene til A er positive,*
- *negativ definit $\Leftrightarrow$ alle egenverdiene til A er negative,*
- *indefinit $\Leftrightarrow$ A har både positive og negative egenverdier.*

Merk: tilsvarende gjelder det at Q er

- positiv semidefinit $\Leftrightarrow$ alle egenverdiene til A er ikkenegative,
- negativ semidefinit $\Leftrightarrow$ alle egenverdiene til A ikkepositive,

Skisse til bevis: Ved å benytte Proposition 5.1.1 kan vi istedet betrakte

$$Q'(y) = d_1 y_1^2 + d_2 y_2^2 + \cdots + d_n y_n^2$$

der $d_1, d_2, \ldots, d_n$ er egenverdiene til A (med multiplisitet).

Ved å studere fortegnet til dette uttrykket slik som i Eksempel 5.2.1 får vi etter litt drøfting resultat i teoremet. □

Eksempel 5.2.2. *La $Q(x) = 5\,x_1^2 - 4\,x_1x_2 + 5\,x_2^2$.*

Siden $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ har egenverdiene 3 og 7, som begge er positive, så er Q positiv definit. Dermed er $(0,0)$ et entydig minimumspunkt for Q.

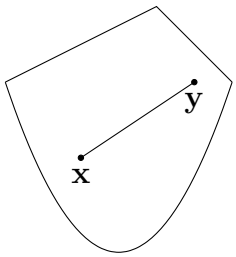
5.3 Konveksitet

Vi skal nå se på noen begreper som er grunnleggende for optimering, nemlig konvekse mengder og konvekse funksjoner. Dyp læring i maskinlæring er en spesiell type optimering, så det er viktig å kjenne noe til dette feltet. Noen referanser om konveksitet er [1, 5].

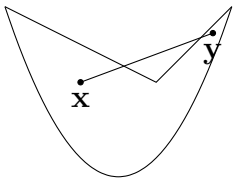
La oss starte med (punkt)mengder i planet, så la $A \subseteq \mathbb{R}^2$. En enkel mengde er et rektangel, f.eks. $A = \{(u,v) \in \mathbb{R}^2 : 0 \le u \le 3,\ 1 \le v \le 3\}$. Velg to punkter i A, $\mathbf{x}$ og $\mathbf{y}$, og se på linjesegmentet $L_{\mathbf{x},\mathbf{y}}$ mellom $\mathbf{x}$ og $\mathbf{y}$. Legg merke

til at *hele* $L_{\mathbf{x},\mathbf{y}}$ ligger i mengden A. Uansett hvilke to punkter $\mathbf{x}$ og $\mathbf{y}$ vi velger i A, så vil det samme være tilfelle. Dette er en viktig egenskap, og vi sier at mengden A er **konveks**. Vi kan tenke på en konveks mengde som en mengde uten noen "hull" og som "buler ut overalt".

Eksempel 5.3.1. *En konveks mengde C i planet er vist i Fig.* 5.3. *Linjesegmentet $L_{\mathbf{x},\mathbf{y}}$ er inneholdt i C. Derimot har vi en ikke-konveks mengde A i Fig.* 5.3 *og linjesegmentet $L_{\mathbf{x},\mathbf{y}}$ inneholder punkter utenfor A.*



Figur 5.1: En konveks mengde i planet.



Figur 5.2: Ikke konveks mengde.

Generelt, for en mengde $C \subseteq \mathbb{R}^n$, sier vi at C er **konveks** dersom for alle $\mathbf{x}, \mathbf{y} \in C$ så er

$$L_{\mathbf{x},\mathbf{y}} \subseteq C$$

der linjesegmentet $L_{\mathbf{x},\mathbf{y}}$ mellom $\mathbf{x}$ og $\mathbf{y}$ er definert ved

$$L_{\mathbf{x},\mathbf{y}} = \{(1-\lambda)\mathbf{x} + \lambda\mathbf{y} : 0 \leq \lambda \leq 1\}.$$

Legg merke til at

$$(1-\lambda)\mathbf{x} + \lambda\mathbf{y} = \mathbf{x} + \lambda(\mathbf{y} - \mathbf{x})$$

som når λ varierer mellom 0 og 1 gir alle punkter på linjesegmentet (slik vi kjenner fra skolen). Punktet (vektoren) $(1-\lambda)\mathbf{x} + \lambda\mathbf{y}$ kalles en **konveks**

kombinasjon av $\mathbf{x}$ og $\mathbf{y}$. Mer generelt kan vi lage konveks kombinasjoner av flere vektorer

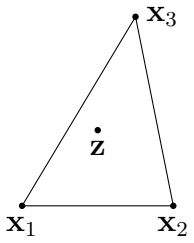
$$\mathbf{z} = \lambda_1 \mathbf{x}^{(1)} + \lambda_2 \mathbf{x}^{(2)} + \cdots + \lambda_k \mathbf{x}^{(k)}$$

der $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)} \in \mathbb{R}^n$ og $\lambda_i \geq 0$ $(i \leq k)$ og $\sum_{i=1}^{k} \lambda_i = 1$. Vi kan tenke på dette som en "blanding" av vektorene, f.eks. for $k = 3$, kan vi ha 30% av $\mathbf{x}^{(1)}$, 50% av $\mathbf{x}^{(2)}$, 20% av $\mathbf{x}^{(3)}$, som svarer til

$$\mathbf{z} = 0.3\mathbf{x}^{(1)} + 0.5\mathbf{x}^{(2)} + 0.2\mathbf{x}^{(3)}.$$

Dette betyr at vi har samme kombinasjon for hver komponent i, så $z_i = 0.3x_i^{(1)} + 0.5x_i^{(2)} + 0.2x_i^{(3)}$ for $i = 1, 2, \ldots, n$.

Man kan vise at en mengde $C \subseteq \mathbb{R}^n$ er konveks hvis og bare hvis den inneholder enhver konveks kombinasjon av (to eller flere) punkter i C. I Fig. 5.3 er $\mathbf{z}$ er en konveks kombinasjon av $\mathbf{x}_1$, $\mathbf{x}_2$ og $\mathbf{x}_3$.



Figur 5.3: En konveks kombinasjon.

La oss se på viktige eksempler på konvekse mengder. Et **lineært system** er et system av lineære likninger og/eller lineære ulikheter i endelig mange variable. Løsningsmengden til et lineært system kalles et **polyeder**. For eksempel er

$$x_1 + x_2 = 3,\ x_1 \geq 0,\ x_2 \geq 0$$

et lineært system i variablene x_1, x_2. Løsningsmengden er punktene $(x_1, 3 - x_1)$ der $0 \leq x_1 \leq 3$.

Dette betyr at ethvert polyeder P i $\mathbb{R}^n$ har formen

$$P = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\} \tag{5.4}$$

der $A = [a_{ij}]$ er en $m \times n$ matrise og $\mathbf{b} \in \mathbb{R}^m$, og der $\leq$ betyr komponentvis ulikhet, altså at $(A\mathbf{x})_i \leq b_i$ for hver $i \leq m$. På komponentform er $A\mathbf{x} \leq b$

følgende lineære system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & \leq & b_1 \\ a_{21}x_1 & + & a_{21}x_2 & + & \cdots & + & a_{2n}x_n & \leq & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & \leq & b_m. \end{array}$$

Vi kunne her ha skrevet et polyheder på formen

$$\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq b,\, C\mathbf{x} = \mathbf{d}\}$$

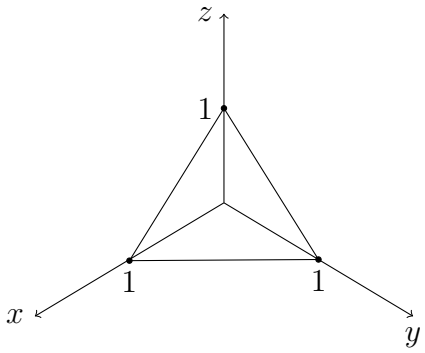
for å fremheve at vi kan ha lineære likninger, ikke bare ulikheter. Men enhver lineær likning kan skrives som to lineære ulikheter (en ulikhet hver vei!), så derfor er (5.4) en generell form.

Et viktig eksempel i $\mathbb{R}^n$ er polyedret

$$S_n = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0 \ (i \leq n),\ \sum_{i=1}^{n} x_i = 1\}.$$

S_n kalles **standardsimpleksen** i $\mathbb{R}^n$, se Fig. 5.4 for tilfellet $n = 3$.

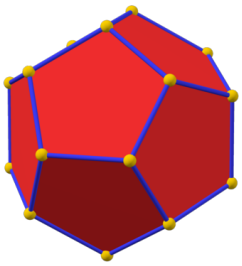
Et annet eksempel er lineære underrom; ethvert lineært underrom L er en konveks mengde. Dette skyldes at L kan skrives som løsningsmengden til et (homogent) lineært likningssystem.



Figur 5.4: Standardsimpleks i $\mathbb{R}^3$.

Polyedre er svært viktige i optimering fordi mengden vi optimerer over, de tillatte punktene, ofte er et polyeder. Dessuten, i mer kompliserte situasjoner, vil man gjerne bruke polyedre til å approksimere mengden av tillatte løsninger.

Eksempel 5.3.2. *Fig.* 5.5 *viser et polyeder i* $\mathbb{R}^3$*, og dette kalles* **dodekaedret** D*. Det er* 20 *hjørner, hvert angitt ved et gult punkt. Et hjørne er et punkt på randen som ikke kan skrives som konveks kombinasjon av andre punkter i mengden ("spissene"). D har* 12 *sideflater som er pentagoner (regulære femkanter) angitt i rødt. Videre har D* 30 *kanter som er linjesegment i blått, som hver forbinder to "nabohjørner". Dodekaedret er et såkalt regulært polyeder, så sideflatene og kantene er like og regelmessige. Det er ett av de fem regulære tre-dimensjonale polyedrene, og disse kalles Platonske legemer.*

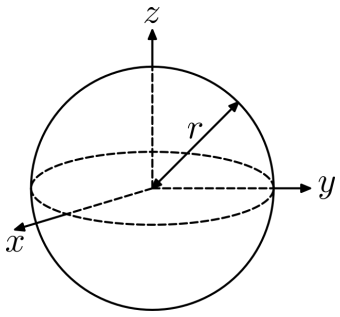


Figur 5.5: Et dodekaeder. (Kilde: Wikimedia)

Vi definerer en **kule** i $\mathbb{R}^n$ med sentrum i $\mathbf{a}$ og radius r ved

$$B(\mathbf{a}; r) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{a}\| \leq r\}.$$

For eksempel, hvis $n = 2$, $\mathbf{a} = (0, 0)$ og $r = 1$, er $B(\mathbf{a}; r)$ enhetssirkelen med sentrum i origo og radius 1 samt alle punkter inne i sirkelen. Dette er det samme som alle punkter med avstand høyst 1 fra origo. Se Fig. 5.6 for tilfellet $n = 3$.



Figur 5.6: Kule. (Kilde: Wikimedia)

Proposisjon 5.3.1. (*i*) *Hvert polyeder er en konveks mengde.*
(*ii*) *Enhver kule* $B(\mathbf{a}; r)$ *er en konveks mengde.*

Bevis. La polyedret P være de punkter $\mathbf{x}$ som oppfyller $A\mathbf{x} \leq \mathbf{b}$, for en $m \times n$ matrise A og $\mathbf{b} \in \mathbb{R}^m$. Anta $\mathbf{x}$ and $\mathbf{y}$ ligger i P, så $A\mathbf{x} \leq \mathbf{b}$, og $A\mathbf{y} \leq \mathbf{b}$. Da er

$$A(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) = \lambda A\mathbf{x} + (1-\lambda)A\mathbf{y} \leq \lambda\mathbf{b} + (1-\lambda)\mathbf{b} = \mathbf{b}.$$

Dette viser at $\lambda A\mathbf{x} + (1-\lambda)A\mathbf{y}$ også ligger i P, så P er konveks.

Betrakt kulen $B(\mathbf{a}; r)$. Vi minner om **trekantulikheten** som sier at

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \text{ når } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n.$$

La $\mathbf{x}, \mathbf{y} \in B(\mathbf{a}; r)$ og$\lambda \in [0, 1]$. Da er

$$\begin{aligned} \|((1-\lambda)\mathbf{x} + \lambda\mathbf{y}) - \mathbf{a}\| &= \|(1-\lambda)(\mathbf{x}-\mathbf{a}) + \lambda(\mathbf{y}-\mathbf{a})\| \\ &\leq \|(1-\lambda)(\mathbf{x}-\mathbf{a})\| + \|\lambda(\mathbf{y}-\mathbf{a})\| \\ &= (1-\lambda)\|\mathbf{x}-\mathbf{a}\| + \lambda\|\mathbf{y}-\mathbf{a}\| \\ &\leq (1-\lambda)r + \lambda r = r. \end{aligned}$$

Så derfor er $B(\mathbf{a}; r)$ konveks. □

Vi nevner også et fundamentalt resultat om polyedre, uten å ta detaljer. Et **ekstrempunkt** i et polyeder P er et punkt $\mathbf{x} \in P$ som *ikke* kan skrives som konveks kombinasjon av andre punkter i P. I polyedret (dodekaedret) i Fig. 5.5 er ekstrempunktene nettopp hjørnene, de gule punktene. Betrakt en av sideflatene (rødt) F som da inneholder 5 ekstempunkter. Da vil ethvert punkt i F kunne skrives som en konveks kombinasjon av disse 5 ekstrempunktene. Tilsvarende for en kant: hvert punkt på en kant er en konveks kombinasjon av de to ekstrempunktene. Og faktisk kan ethvert punkt i P skrives som en konveks kombinasjon av alle ekstrempunktene til P. Dette er eksempler på et generelt resultat om polyedre og ekstrempunkter, men som vi ikke går nærmere inn på her.

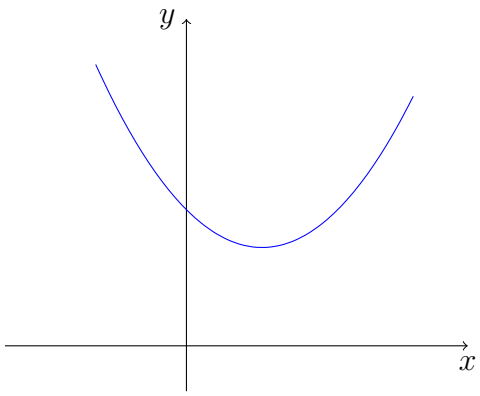
5.4 Konvekse funksjoner

I forrige seksjon så vi på konvekse mengder, og i denne seksjonen går vi videre med konvekse *funksjoner*. Disse funksjonene er viktige for optimering fordi (løst sagt):

- *Vi kan løse minimeringsproblemer for konvekse funksjoner over konvekse mengder på en effektiv måte!*

- Videre kan vi si at optimeringsproblemer som "mangler konveksitet" er typisk vanskelige å løse.

Vi starter med å se på funksjoner fra $\mathbb{R}$ til $\mathbb{R}$. Da kan vi tegne grafen, og hvis grafen "krummer oppover" er funksjonen konveks, se Fig. 5.4. Som definisjon er ikke dette presist nok, så her kommer definisjonen, og den kan brukes også for funksjoner definert på en konveks mengde i $\mathbb{R}^n$.



Figur 5.7: En konveks funksjon.

La $f : C \to \mathbb{R}$ der $C \subseteq \mathbb{R}^n$ er en konveks mengde. Vi sier at f er **konveks** (f er en **konveks funksjon**) dersom

$$f((1-\lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1-\lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}) \quad (\mathbf{x}, \mathbf{y} \in C,\, 0 \leq \lambda \leq 1). \tag{5.5}$$

Dette kan tolkes geometrisk, se Fig. 5.4: betrakt to punkter på grafen til f, $P_1 = (\mathbf{x}, f(\mathbf{x}))$ og $P_2 = (\mathbf{x}, f(\mathbf{y}))$; da ligger grafen til f, restriktert til linjesegmentet $L_{\mathbf{x},\mathbf{y}}$, *under* det rette linjesegmentet mellom disse to punktene P_1 og P_2.

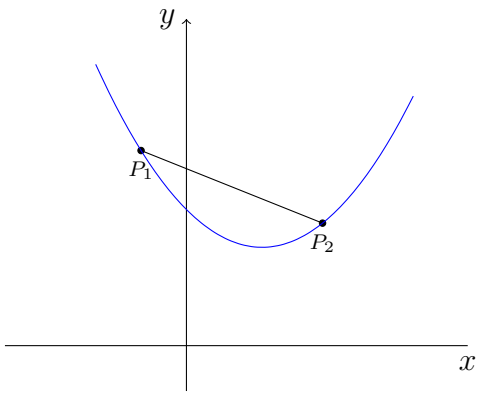
Vi sier at en funksjon f er **konkav** dersom $-f$ er konveks.

Eksempel 5.4.1. *Funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved $f(x) = x^2 + 3x - 1$ er konveks (prøv å tegne grafen!). Mer generelt, funksjonen $f(x) = ax^2 + bx + c$ der $a \geq 0$ er konveks. Hvis $a \leq 0$, er funksjonen konkav.*

Legg merke til at vi krever at definisjonsmengden C for f skal være konveks. Dette er for å sikre at konveks kombinasjonen $(1-\lambda)\mathbf{x} + \lambda\mathbf{y}$ også ligger i C (ellers kunne f være udefinert i dette punktet).

Enhver lineær funksjon, og enhver affine funksjon, er konveks. Her er lineær funksjon en funksjon på formen $\mathbf{x} \to \mathbf{c}^T\mathbf{x}$ der $\mathbf{c} \in \mathbb{R}^n$, mens en affin funksjon også kan ha et konstantledd, så $\mathbf{x} \to \mathbf{c}^T\mathbf{x} + d$ der $d \in \mathbb{R}$.

Noen flere eksempler på konvekse funksjoner fra $\mathbb{R}^n$ til $\mathbb{R}$ er:



Figur 5.8: Tolkning av konveksitet.

(i) $f(\mathbf{x}) = \|\mathbf{x}\|$ (Euklidsk norm).

(ii) $f(\mathbf{x}) = e^{\sum_{j=1}^{n} x_j}$ der $\mathbf{x} = (x, x_2, \ldots, x_n)$.

(iii) $f(\mathbf{x}) = e^{h(\mathbf{x})}$ der $h : \mathbb{R}^n \to \mathbb{R}$ er en konveks funksjon.

(iv) $f(\mathbf{x}) = \max_i g_i(\mathbf{x})$ der $g_i : \mathbb{R}^n \to \mathbb{R}$ er en affin funksjon ($i \leq m$). Dette er punktvis maksimum av affine funksjoner, og slike funksjoner vil typisk ha enkelte punkter der de ikke er deriverbare. Samme resultat gjeler for punktvis maksimum av konvekse funksjoner. Dette er et nyttig resultat for anvendelser.

Neste resultat kalles **Jensen's ulikhet**[1] og handler om konvekse funksjoner og konveks kombinasjoner. Det kan vises ved induksjon.

Theorem 5.4.1 (Jensen's ulikhet)**.** *La $f : C \to \mathbb{R}$ være en konveks funksjon definert på en konveks mengde $C \subseteq \mathbb{R}^n$. Hvis $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)} \in C$ og $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$ oppfyller $\sum_{j=1}^{k} \lambda_j = 1$, så er*

$$f(\lambda_1 \mathbf{x}^{(1)} + \cdots + \lambda_k \mathbf{x}^{(k)}) \leq \lambda_1 f(\mathbf{x}^{(1)}) + \cdots + \lambda_k f(\mathbf{x}^{(k)}). \tag{5.6}$$

Eksempel 5.4.2. *Et eksempel på Jensen's ulikhet når $k = 3$ er*

$$f(0.4\mathbf{x}^{(1)} + 0.5\mathbf{x}^{(2)} + 0.1\mathbf{x}^{(3)}) \leq 0.4 f(\mathbf{x}^{(1)}) + 0.5 f(\mathbf{x}^{(2)}) + 0.1 f(\mathbf{x}^{(3)}).$$

Videre, hvis alle vektene er like, altså $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$, får vi

$$f((1/3)(\mathbf{x}^{(1)} + \mathbf{x}^{(2)} + \mathbf{x}^{(3)})) \leq (1/3)(f(\mathbf{x}^{(1)}) + f(\mathbf{x}^{(2)}) + f(\mathbf{x}^{(3)}))$$

[1] Ulikheten har navn etter den danske matematikeren Johan Jensen (1859–1925)

som betyr at funksjonsverdien i gjennomsnittet av punktene er mindre enn eller lik gjennomsnittet av funksjonsverdiene. *Det samme gjelder for generell k (flere punkter).*

Merk at når $k = 2$ faller Jensen's ulikhet sammen med ulikheten i definisjonen av en konveks funksjon. Denne ulikheten (5.6) kan vi sammenlikne med tilfellet der f er en lineær (eller affin) funksjon: da holder ulikheten med likhet, og vi kan "flytte f inn" på hver $\mathbf{x}^{(j)}$. Men for konvekse funksjoner får vi altså ulikhet når vi gjør dette. Dessuten gjelder dette bare for konveks kombinasjoner (ikke lineær kombinasjoner, som er mer generelt).

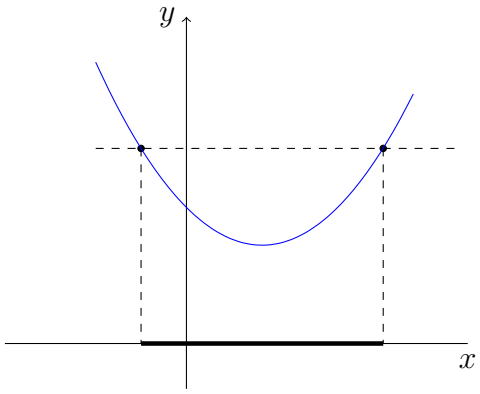
La igjen $f : C \to \mathbb{R}$ være en funksjon der $C \subseteq \mathbb{R}^n$ er en konveks mengde. For optimering, eller minimering, ønsker vi å finne et (globalt) minimumspunkt $\mathbf{x}^*$ som altså betyr at

$$f(\mathbf{x}^*) \leq f(\mathbf{y}) \text{ for alle } \mathbf{y} \in C.$$

En interessant type mengde er **subnivå-mengden** definert ved

$$S_f(a) = \{\mathbf{x} \in C : f(\mathbf{x}) \leq a\}$$

der $a \in \mathbb{R}$. Dette er altså alle mulige $\mathbf{x} \in C$ der funksjonsverdien er høyst a. Et eksempel er vist i Fig. 5.4, der subnivåmengden er linjesegmentet markert med tykk skrift. Spesielt, hvis $a^* = \min\{f(\mathbf{x}) : \mathbf{x} \in C\} = f(\mathbf{x}^*)$ (så vi antar at f har globalt minimumspunkt $\mathbf{x}^*$) er $S_f(a^*)$ lik *mengden av optimale løsninger*.



Figur 5.9: Subnivå mengde.

Optimeringsalgoritmer består gjerne i å finne en sekvens av punkter i C

$$\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots$$

med stadig lavere funksjonsverdier, så

$$f(\mathbf{x}^{(0)}) > f(\mathbf{x}^{(1)}) > f(\mathbf{x}^{(2)}) > \cdots$$

(noen ganger er det $\geq$). Dette betyr at de tilhørende subnivå-mengdene $S_f(\alpha^{(i)})$, der $\alpha^{(i)} = f(\mathbf{x}^{(i)})$, blir stadig mindre mengder, inntil vi forhåpentligvis er nær $S_f(a^*)$, og problemet er løst.

For konvekse funksjoner er disse subnivå mengdene spesielt attraktive, som neste resultat sier.

Proposisjon 5.4.2. *La $C \subseteq \mathbb{R}^n$ være en konveks mengde og betrakt en konveks funksjon $f : C \to \mathbb{R}$. La $a \in \mathbb{R}$. Da er subnivå-mengden $S_f(a)$ en konveks mengde.*

Dette innebærer at hvis vi for eksempel har to optimale løsninger x^* og y^*, så vil også enhver konveks kombinasjon

$$\mathbf{z} = (1 - \lambda)\mathbf{x}^* + \lambda\mathbf{y}^*$$

også være en optimal løsning!

Det er et flott resultat som knytter sammen konvekse mengder og konvekse funksjoner. Vi gir dette uten bevis. For en funksjon $f : \mathbb{R}^n \to \mathbb{R}$ definerer vi **epigrafen** til f som følgende mengde i $\mathbb{R}^{n+1}$ (eller i $\mathbb{R}^n \times \mathbb{R}$, hvis vi vil være pirkete)

$$\text{epi}(f) = \{(\mathbf{x}, y) \in \mathbb{R}^{n+1} : y \geq f(\mathbf{x})\}.$$

Så derfor er epi(f) all punktene som ligger på eller over grafen til f.

Theorem 5.4.3. *Betrakt en funksjon $f : \mathbb{R}^n \to \mathbb{R}$. Da er f konveks hvis og bare hvis* epi(f) *er en konveks mengde.*

Dette gir altså en svært nær sammenheng mellom konvekse funksjoner og konvekse mengder. Det betyr i stor grad at resultater kan "oversettes" fra mengder til funksjoner og omvendt. Vi skal ikke gjøre mer ut av dette samspillet her.

Konvekse funksjoner har en rekke fine egenskaper. For eksempel er de deriverbare nesten overalt (som har en presis betydning). Hva med kontinuitet?

Theorem 5.4.4. *La $f : C \to \mathbb{R}$ være en konveks funksjon definert på en åpen konveks mengde $C \subseteq \mathbb{R}^n$. Da er f kontinuerlig på C.*

Dette er et viktig resultat, men vær oppmerksom på at en konveks funksjon kan være diskontinuerlig på *randen* av definisjonsområdet. For eksempel, funksjonen $f : [0, 1] \to \mathbb{R}$ gitt ved $f(0) = 1$ and $f(x) = 0$ for $x \in (0, 1]$ er konveks, men er diskontinuerlig i $x = 0$.

Et nyttig resultat for å avgjøre om en gitt funksjon er konveks er følgende.

Theorem 5.4.5. *La $f : C \to \mathbb{R}$ være en konveks funksjon definert på en åpen konveks mengde $C \subseteq \mathbb{R}^n$. Anta at f har kontinuerlige annen-ordens partielle deriverte på C.*

Da er f konveks hvis og bare hvis Hesse matrisen $\nabla^2 f(\mathbf{x}) = [\partial f(\mathbf{x})/\partial x_i x_j]$ er positiv semidefinit for hver $x \in C$.

Så her fikk vi en forbindelse tilbake til seksjonen om positiv semidefinite matriser.

Eksempel 5.4.3. *Vi ser igjen på funksjonen $f : \mathbb{R} \to \mathbb{R}$ gitt ved $f(x) = ax^2 + bx + c$ der $a \geq 0$. Hesse matrisen når $n = 1$ er rett og slett den andre-deriverte, $f''(x)$. Vi har*

$$f'(x) = 2ax + b, \quad f''(x) = 2a \geq 0,$$

så f er konveks.

La nå $n = 2$ og betrakt

$$f(x) = 5\,x_1^2 - 4\,x_1x_2 + 5\,x_2^2 = x^T A\,x$$

der

$$A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}.$$

I forrige seksjon så vi at dette var en kvadratisk form der tilhørende matrise A hadde egenverdiene 3 og 7, så A er positiv definit og dermed positiv semidefinit. Da følger det av Teorem 5.4.5 at f er en konveks funksjon.

En viktig klasse av konvekse funksjoner er bestemte kvadratiske funksjoner. La A være en symmetrisk $n \times n$ matrise som er positiv semidefinit og betrakt den kvadratiske funksjonen $f : \mathbb{R}^n \to \mathbb{R}$

$$f(\mathbf{x}) = (1/2)\,\mathbf{x}^T A\,\mathbf{x} - \mathbf{b}^T\mathbf{x} = (1/2)\sum_{i,j} a_{ij}x_i x_j - \sum_{j=1}^{n} b_j x_j.$$

(Hvis A er nullmatrisen, er funksjonen lineær, så la oss se bort fra dette tilfellet.) Litt regning viser at Hesse matrisen til A i punktet $\mathbf{x}$ er A, så denne er uavhengig av punktet $\mathbf{x}$. Det følger da av Teorem 5.4.5 at f er en konveks funksjon.

Tilbake til optimering av en (generell) funksjon f, og anta at en optimeringsalgoritme har kommet til et punkt $\mathbf{x}$ som potensielt kan være et *lokalt* minimumspunkt for f. Hvordan kan vi sjekke om det er tilfelle? Dette kan gjøres ved å beregne Hesse matrisen $H_f(\mathbf{x})$ og sjekke om den er positiv definit. Hvis det er tilfellet, vil x være et lokalt minimumspunkt og algoritmen stopper.

Ellers vil man gjerne finne en søkeretning fra $\mathbf{x}$ der lavere funksjonsverdier kan finnes. Dette kan komme ut fra nevnte sjekk av om $H_f(\mathbf{x})$ er positiv definit.

En viktig klasse av positiv semidefinite matriser dukker ofte opp i anvendelser, nemlig diagonaldominante matriser. Vi sier at en $n \times n$ matrise $A = [a_{ij}]$ er **diagonaldominant** dersom

$$|a_{ii}| \geq \sum_{j \neq i} |a_{ij}| \quad (i \leq n).$$

For eksempel er matrisen A i Eksempel 5.4.3 diagonal dominant og derfor positiv semidefinit. (Den er faktisk *strengt* diagonaldominant, som innebærer ekte ulikhet i definisjonen over. Dette medfører at matrisen er positiv definit.)

Til slutt tar vi med et teorem som gir en svært viktig egenskap ved konvekse funksjoner.

Theorem 5.4.6. *La $f : C \to \mathbb{R}$ være en deriverbar funksjon definert på en åpen konveks mengde $C \subseteq \mathbb{R}^n$. Da er følgende betingelser ekvivalente:*

(i) *f er konveks,*

(ii) $f(\mathbf{x}) \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0) \qquad (\mathbf{x}, \mathbf{x}_0 \in C).$

Egenskap (ii) sier at første-ordens Taylor approksimasjon til f i $\mathbf{x}_0$ (som er høyre side av ulikheten) alltid underestimerer f, se Fig. 5.4.

Her er en viktig konsekvens av dette for optimering.

Corollary 5.4.7. *Under antakelsene i Teorem* 5.4.6 *har vi at $\mathbf{x}^* \in C$ er minimumspunkt for f hvis og bare hvis $\mathbf{x}^*$ er et stasjonært punkt, dvs. $\nabla f(\mathbf{x}^*) = \mathbf{0}$.*

Bevis. La $\mathbf{x}^* \in C$. Vi vet fra tidligere at hvis $\mathbf{x}^*$ er minimumspunkt, så må $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Omvendt, anta $\nabla f(\mathbf{x}^*) = \mathbf{0}$. Ved Teorem 5.4.6 (med $\mathbf{x}_0 = \mathbf{x}^*$) får vi da at

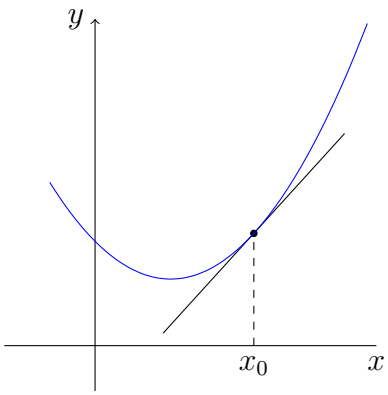
$$f(\mathbf{x}) \geq f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T(\mathbf{x} - \mathbf{x}^*) = f(\mathbf{x}^*) \quad (\mathbf{x} \in C),$$

så $\mathbf{x}^*$ er minimumspunkt. □

Dette betyr altså at

- for en *konveks funksjon definert på en åpen konveks mengde, vil et lokalt minimum også være et globalt minimum.*

Dette resultatet er svært viktig, og blir brukt gang på gang innen optimering. For funksjoner som *ikke* er konvekse, risikerer man at lokalt minimum ikke er globalt minimum. Et godt eksempel her er dyp læring, der funksjonen man minimerer ikke er konveks, så når algoritmen terminerer, har man ikke nødvendigvis funnet globalt minimum, selv om man gjerne har et lokalt minimum. Dette er én av utfordringene i maskinlæring!



Figur 5.10: Taylor approksimasjon, konveks funksjon.

5.5 ∗ Variabelskifte i kvadratiske former

Et variabelskifte svarer til et bytte av koordinatsystem, f.eks. å dreie begge aksene 45° mot urviseren. Dette kan gi enklere formler for å beskrive samme punktmengde i planet.

For en funksjon i to variable, si $f : \mathbb{R}^2 \to \mathbb{R}$ er **nivåmengder** punktmengder der f har samme verdi. Definer

$$N_f(a) = \{\mathbf{x} \in \mathbb{R}^2 : f(\mathbf{x}) = a\},$$

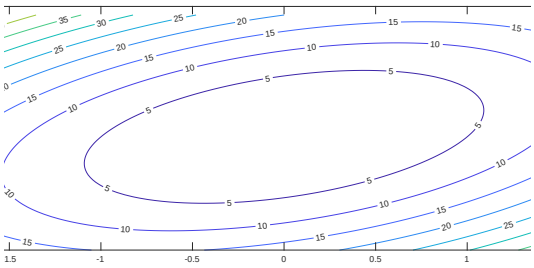
der $\mathbf{x} = (x_1, x_2)$, som nivåmengden til f med nivå a. Nivåmengder[2] er viktige i praktiske sammenhenger, f.eks. for vanlige kart der nivåmengdene kan vise punkter med samme høyde over havet. Nivåmengder for en kvadratisk form (i to variable) er enkle å beskrive når man velger passende koordinatsystem og nivåmengdene man da får er kjente kurver som ellipser eller hyperbler. I Fig. 5.11 vises noen nivåmengder for the kvadratiske formen $Q(\mathbf{x}) = 5x_1^2 - 4x_1x_2 + 5x_2^2$ som vi ser på i eksemplet under. Legg merke til at disse nivåmengdene er ellipser, og disse avtar i størrelse inn mot minimumspunktet $\mathbf{x} = (0, 0)$.

Vi skal også se at kvadratiske former (og symmetriske matriser) kan klassifiseres i noen hovedtyper. Disse typene er viktige f.eks. når man studerer stasjonære punkter til reelle funksjoner på $\mathbb{R}^n$ (ved å se på Hesse-matrisene).

Eksempel 5.5.1. *La* $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$ *og* $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$. *Da er*

$$Q(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 5x_1 - 2x_2 \\ -2x_1 + 5x_2 \end{bmatrix}$$

[2]Ofte brukes betegnelsen nivåkurver når mengdene faktisk er kurver.



Figur 5.11: Nivåmengder.

$$= x_1(5x_1 - 2x_2) + x_2(-2x_1 + 5x_2) = 5x_1^2 - 4x_1x_2 + 5x_2^2.$$

Vi skal nå se at vi kan alltid gjøre om en kvadratisk form til en "enkel" kvadratisk form uten kryssledd ved et passende variabelskifte.

Et variabelskifte svarer til at vi skifter basis (og dermed koordinatsystem): hvis

$$P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

er en $n \times n$ invertibel matrise og vi foretar variabelskiftet

$$\mathbf{y} = P^{-1}\mathbf{x}, \text{ altså } \mathbf{x} = P\mathbf{y}$$

så er y koordinatvektoren til $\mathbf{x}$ mhp. (den ordnede) basisen $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_n\}$ (fordi P^{-1} er koordinatskiftematrisen fra standard basisen til B).

Betrakt en kvadratisk form (på $\mathbb{R}^n$)

$$Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} \quad \text{der } A \text{ er en symmetrisk } n \times n \text{ matrise.}$$

Siden A er symmetrisk vet vi at A er ortogonalt diagonaliserbar: det finnes da en ortogonal $n \times n$ matrise P og en $n \times n$ diagonalmatrise $D = \text{diag}(d_1, d_2, \ldots, d_n)$ slik at

$$A = PDP^{-1} = PDP^T$$

Dette er fordi $P^{-1} = P^T$, og da er $P^TAP = D$.

Kolonnene i P er da en ortonormal basis B for $\mathbb{R}^n$ bestående av egenvektorer for A tilhørende egenverdiene $d_1, d_2, \ldots, d_n$.

La oss foreta variableskiftet $\mathbf{y} = P^{-1}\mathbf{x}$, dvs. $\mathbf{x} = P\mathbf{y}$. Dette gir

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = (P\mathbf{y})^T A(P\mathbf{y}) = \mathbf{y}^T P^T AP\mathbf{y} = \mathbf{y}^T D\mathbf{y}.$$

Nå er

$$Q'(\mathbf{y}) := \mathbf{y}^T D\mathbf{y} = \sum_{i=1}^{n} d_i y_i^2$$

en kvadratisk form uten kryssledd!

Vi har dermed vist følgende resultat.

Theorem 5.5.1. *I koordinatsystemet for $\mathbb{R}^n$ med akser bestemt av en ortonormal egenvektorbasis B for den symmetriske matrisen A, så blir den kvadratiske formen $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ gjort om til en kvadratisk form uten kryssledd.*

Aksene i koordinatsystemet ovenfor kalles ofte hovedaksene (eller prinsipalaksene).

Eksempel 5.5.2. *(Fortsettelse av Eksempel* 5.5.1*) La igjen*

$$A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$$

og $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$. Ved litt regning finner vi at egenverdiene til A er 3 *og* 7*, med tilhørende enhetsegenvektorer $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.*

Sett $P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ og $D = diag(3, 7)$. Variabelskiftet $\mathbf{x} = P\mathbf{y}$ gir da at

$$Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x} = \mathbf{y}^T D\mathbf{y} = 3y_1^2 + 7y_2^2 = Q'(\mathbf{y}).$$

Vi vender tilbake til den generelle teorien, og ser nå nærmere på tilfellet $n = 2$. Betrakt en kvadratisk form på $\mathbb{R}^2$,

$$Q(\mathbf{x}) = a\,x_1^2 + b\,x_1x_2 + c\,x_2^2.$$

Hvordan ser nivåmengdene til Q ut? Vi minner om at nivåmengden til Q svarende til en verdi $d \in \mathbb{R}$ består av alle $\mathbf{x} = (x_1, x_2)$ i $\mathbb{R}^2$ som er slik at $Q(\mathbf{x}) = d$, dvs. som tilfredsstiller likningen

$$a\,x_1^2 + b\,x_1x_2 + c\,x_2^2 = d$$

Vi kan da skifte variabel og gå over til koordinatsystemet angitt i Teorem 5.5.1. Likningen ovenfor forenkles da til likningen

$$d_1y_1^2 + d_2y_2^2 = d$$

der d_1 og d_2 er egenverdiene til den symmetriske matrisen A tilordnet Q. Kurvene bestemt av denne likningen, og dermed nivå kurvene til Q, lar seg lett beskrive.

Anta f.eks. at d_1, d_2 og d alle er forskjellig fra 0. Da har vi at

- hvis d_1, d_2 (og d) alle har samme fortegn så blir kurven en ellipse
- hvis d_1, d_2 har motsatt fortegn så blir kurven en hyperbel.

Eksempel 5.5.3. *Betrakt likningen* $5\,x_1^2 - 4\,x_1x_2 + 5\,x_2^2 = 48$*, mao.* $Q(\mathbf{x}) = 48$ *der* $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$ *med* $A = \begin{bmatrix} 5 & -2 \\ -2 & 5 \end{bmatrix}$.

I koordinatsystemet bestemt av egenvektorbasisen for A *vi fant da, blir likningen omgjort til* $3\,y_1^2 + 7\,y_2^2 = 48$*, altså til*

$$\frac{y_1^2}{4^2} + \frac{y_2^2}{(\sqrt{48/7}\,)^2} = 1$$

som er likningen for en ellipse.

Eksempel 5.5.4. *Eksempel. Betrakt likningen* $3\,x_1^2 + 10\,x_1x_2 + 3\,x_2^2 = 2$*, mao.* $Q(\mathbf{x}) = 2$ *der* $Q(\mathbf{x}) = \mathbf{x}^T A\,\mathbf{x}$ *med* $A = \begin{bmatrix} 3 & 5 \\ 5 & 3 \end{bmatrix}$,

Ved regning finner vi at egenverdiene til A *er* 8 *og* -2*, med tilhørende enhetsegenvektorer* $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Sett $P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$*. Variabelskiftet* $\mathbf{x} = P\mathbf{y}$ *gjør da likningen* $Q(\mathbf{x}) = 2$ *om til likningen*

$$8\,y_1^2 - 2\,y_2^2 = 2, \;\; \textit{dvs.} \;\; \frac{y_1^2}{(1/2)^2} - y_2^2 = 1.$$

Dette er likningen for en hyperbel.

5.6 Oppgaver

1. Betrakt den kvadratiske formen $Q(\mathbf{x}) = x_1^2 - 2x_1x_2 + 2x_2^2$, der $x = (x_1, x_2)$.

   (a) Beregn $Q(1, 2)$.

   Løsning: $Q(1,2) = 1 - 2 \cdot 1 \cdot 2 + 2 \cdot 2^2 = 5$. ◇

   (b) Definer funksjonen (i én variable) $f(y) = Q(1, y)$. Finn funksjonsuttrykket for f og finn minimum av f over alle $y \in \mathbb{R}$.

Løsning: $f(y) = 1 - 2y + 2y^2$. Derivert: $f'(y) = -2 + 4y$, og $f'(y) = 0$ gir $y = 1/2$, som må være minimumspunkt fordi $f''(1/2) = 4 > 0$. Dette gir $\min f = f(1/2) = 1/2$. ◇

(c) Bruk kvadratsetningen $(x_1 - x_2)^2 = x_1^2 - 2x_1x_2 + x_2^2$ til å vise at $Q(x) \geq 0$ for alle x, dvs. Q er posititiv semidefinit.

Løsning: $Q(x) = x_1^2 - 2x_1x_2 + 2x_2^2 = (x_1^2 - 2x_1x_2 + x_2^2) + x_2^2 = (x_1 - x_2)^2 + x_2^2 \geq 0$. ◇

2. Betrakt igjen den kvadratiske formen $Q(x) = x_1^2 - 2x_1x_2 + 2x_2^2$, der $x = (x_1, x_2)$.

   (a) Beregn gradienten $\nabla Q(x)$.

   Løsning: $\partial Q/\partial x_1 = 2x_1 - 2x_2$ og $\partial Q/\partial x_2 = -2x_1 + 4x_2$. Så

   $$\nabla Q(x) = (2x_1 - 2x_2, -2x_1 + 4x_2)$$

   (a) Anta at du er i punktet $x = (2, 1)$. Hvilken retning vil du gå i for å få en lavere verdi på Q, når du bare får få en "lite" skritt (f.eks. lengde 0.01)?

   Løsning: Den største reduksjonen av Q "lokalt" får vi ved å gå i retningen negativ gradient, altså $-\nabla Q(2, 1) = -(2, 0)$.

   ◇

3. Betrakt den kvadratiske formen $Q(\mathbf{x}) = 3x_1^2$, der $x = (x_1, x_2)$. (Merk: Det er en kvadratisk for i to variable selv om x_2-ledd ikke "synes".)

   (a) Beregn gradienten $\nabla Q(\mathbf{x})$. Når er $\nabla Q(\mathbf{x}) = \mathbf{0}$?

   Løsning: $\partial Q/\partial x_1 = 6x_1$ og $\partial Q/\partial x_2 = 0$. Så

   $$\nabla Q(\mathbf{x}) = (6x_1, 0).$$

   Gradienten er null hvis og bare hvis $x_1 =$, x_2 vilkårlig.

   (b) Bestem en symmetrisk matrise A slik at $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$.

   Løsning:

   $$A = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}.$$

   ◇

4. Betrakt den kvadratiske formen $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ der

$$A = \begin{bmatrix} 2 & -3 \\ -3 & 1 \end{bmatrix}.$$

(a) Finn et uttrykk for $Q(\mathbf{x})$ (dvs. regn ut matriseproduktet).

Løsning: $A\mathbf{x} = \begin{bmatrix} 2x_1 - 3x_2 \\ -3x_1 + x_1 \end{bmatrix}$. Så $\mathbf{x}^T A\mathbf{x} = 2x_1^2 - 6x_1x_2 + x_2^2$. ◇

5. Betrakt den kvadratiske formen

$$Q(\mathbf{x}) = x_1^2 + 2x_2^2 - x_3^2 + x_1x_2 + 4x_1x_3$$

i variable $\mathbf{x} = (x_1, x_2, x_3)$). Finn en symmetrisk 3×3 matrise A slik at $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ for alle $\mathbf{x}$. der

Løsning:

$$A = \begin{bmatrix} 1 & 1/2 & 2 \\ 1/2 & 2 & 0 \\ 2 & 0 & -1 \end{bmatrix}.$$

◇

6. Betrakt den kvadratiske *funksjonen*

$$H(\mathbf{x}) = x_1^2 + 2x_2^2 - x_3^2 + x_1x_2 + 4x_1x_3 + 3x_1 + x_2 + 5$$

i variable $\mathbf{x} = (x_1, x_2, x_3)$). Husk at en kvadratisk funksjon er summen av en kvadratisk form og en lineær funksjon og en konstant.

(a) Skriv $H(x)$ på formen $H(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} + \mathbf{c}^T\mathbf{x} + d$ der A er en $3 \times s$ matrise, $\mathbf{c} \in \mathbb{R}^3$ og d er en konstant. (Bruk forrige oppgave for å finne A.)

Løsning: A er som i forrige oppgave, $\mathbf{c} = (3, 1)$ og $d = 5$.

(b) Beregn $\nabla H(\mathbf{x})$.

Løsning:

$$\begin{aligned} \partial H/\partial x_1 &= 2x_1 + x_2 + 4x_3 + 3, \\ \partial H/\partial x_2 &= 4x_2 + x_1 + 1, \\ \partial H/\partial x_3 &= -2x_3 + 4x_1. \end{aligned}$$

så $\nabla H(\mathbf{x}) = (2x_1 + x_2 + 4x_3 + 3, 4x_2 + x_1 + 1, -2x_3 + 4x_1)$.

(c) Finn alle stasjonære punkt H, dvs $\mathbf{x}$ slik at $\nabla H(\mathbf{x}) = \mathbf{0}$. (Dette gir et lineært likningssystem! Beklager, litt stygge tall!)

Løsning: Løser det lineære likningssystemet ved eliminasjon. (Tips: siste likning gir uttrykk for x_1 som vi setter inn i de to første, etc.) Dette gir $x_1 = -11/39$, $x_2 = -7/39$, $x_3 = -22/39$, som er entydig løsning. ◇

7. Hvilke av følgende 2×2 diagonalmatriser er positiv semidefinit (PSD)?

$$A_1 = \begin{bmatrix} 2 & 0 \\ 0 & \pi \end{bmatrix}, \; A_2 = \begin{bmatrix} -1 & 0 \\ 0 & 7 \end{bmatrix}, \; A_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Løsning: A_1 og A_3 er PSD fordi diagonalelementene er ≥ 0. A_2 er ikke PSD fordi et diagonalelement er negativt. ◇

8. Betrakt matrisen

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Finn en $\mathbf{x}$ slik at $\mathbf{x}^T A \mathbf{x} < 0$.

Løsning: Kan f.eks. bruke $\mathbf{x} = (0, 1)$ for da er $x^T A x = -1$. ◇

9. La A og B være to $n \times n$ matriser som begger er positiv semidefinit. Vis at da er $A + B$ også positiv semidefinit.

Løsning: La $\mathbf{x} \in \mathbb{R}^n$. Da er $\mathbf{x}^T(A + B)\mathbf{x} = \mathbf{x}^T A \mathbf{x}^T + \mathbf{x}^T B \mathbf{x} \geq 0$ fordi hvert ledd er ikkenegativt. ◇

10. Betrakt følgende matriser

$$A = \begin{bmatrix} 1.64 & 0.48 \\ 0.48 & 1.36 \end{bmatrix}, \; V = \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix}, \; D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}.$$

(a) Er A symmetrisk?

Løsning: Ja, fordi $a_{12} = a_{21}$ (så $A^T = A$). ◇

(b) Regn ut DV^T.

Løsning:

$$DV^T = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix} = \begin{bmatrix} 1.6 & 1.2 \\ 0.6 & -0.8 \end{bmatrix}$$

◇

(c) Vis at $A = VDV^T$ og at V er ortogonal (dvs. $V^TV = I$).

Løsning:

$$VDV^T = \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix} \begin{bmatrix} 1.6 & 1.2 \\ 0.6 & -0.8 \end{bmatrix} = \begin{bmatrix} 1.64 & 0.48 \\ 0.48 & 1.36 \end{bmatrix} = A.$$

Videre

$$V^TV = \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

◇

(d) Finn egenverdiene og tilhørende egenvektorer for A.

Løsning: La V ha kolonner $\mathbf{v}_1$ og $\mathbf{v}_2$, så (som tupler) er $v_1 = (0.8, 0.6)$ og $v_2 = (0.6, -0.8)$. Over fant vi spektraldekomposisjonen til A og fra $A = VDV^T$ får vi $Av_1 = 2v_1$ og $Av_2 = v_2$. Så da har A egenverdier 2 og 1 med tilhørende egenvektorer v_1 og v_2, henholdsvis. ◇

11. Betrakt samme matriser som i forrige oppgave.

(a) Forklar hvorfor A er positiv definit.

Løsning: A har egenverdier 2 og 1, begge positive, så da er A positiv definit. ◇

(b) Forklar hvorfor den kvadratiske formen $Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x}$, ved variableskifte, kan skrives som

$$Q(\mathbf{x}) = 2y_1^2 + y_2^2.$$

Løsning: Vi gjr variabelskiftet $x = Vy$, så de nye variablene er $y = (y_1, y_2)$ som gir koordinatene til x i basisen bestående av egenvektorene v_1 og v_2. Da er

$$Q(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = (V\mathbf{y})^TA(V\mathbf{y}) = \mathbf{y}^T(V^TAV)\mathbf{y} = \mathbf{y}^TD\mathbf{y} = 2y_1^2 + y_2^2.$$

◇

12. La f være en funksjon fra $\mathbb{R}^2$ til $\mathbb{R}$, og la $a, b \in \mathbb{R}$ der $a \neq b$. Hva kan du ha si om snittet av de to nivåmengdene $N_f(a)$ og $N_f(b)$, altså $N_f(a) \cap N_f(a)$?

    Løsning: Snittet må være tomt! Fordi $f(x)$ kan ikke vær lik både a og b når disse tallene er forskjellige. ◇

13. La A og B være to konvekse mengder i planet.

    (a) Vis at snittet av mengdene, $A \cap B$ også er en konveks mengde. (Tegn et eksempel, og studer dette som hjelp.)

    Løsning: Let $\mathbf{x}$ og $\mathbf{y}$ være to punkter i $A \cap B$ og se på en konveks kombinasjon $\mathbf{z} = (1-\lambda)\mathbf{x} + \lambda\mathbf{y}$ der $0 \leq \lambda \leq 1$. Klart at $\mathbf{x}$ og $\mathbf{y}$ ligger i A, så siden A er konveks, vil $\mathbf{z} \in A$. Tilsvarende er $\mathbf{x}, \mathbf{y} \in B$, så siden B er konveks ligger $\mathbf{z} \in B$. Men da er $\mathbf{z} \in A \cap B$. ◇

14. La funksjonen $f : \mathbb{R} \to \mathbb{R}$ være gitt ved $f(x) = x^3$. Vis at f *ikke* er konveks. (Hint: tegn grafen, da kan du gjette på to punkter å se på.)

    Løsning: Se f.eks. på $x = 1$ og $y = -3$ og $z = (1/2)x + (1/2)y = -1$. Da er

    $$f(x) = 1,\ f(y) = (-3)^3 = -27,\ f(z) = -1,$$

    så $f(z) = -1 > -13 = (1/2)f(x) + (1/2)f(y)$ og derfor er f ikke konveks. ◇

15. Vis at eksponentialfunsjonen $f(x) = e^x$ (fra $\mathbb{R}$ til $\mathbb{R}$) er konveks.

    Løsning: Her er det lurt å bruke at hvis $f''(x) \geq 0$ for alle x, så er f konveks. Men det er lett: $f'(x) = e^x$ og $f''(x) = e^x > 0$. ◇

16. La funksjonen $f : \mathbb{R} \to \mathbb{R}$ være gitt ved $f(x) = x^2$. Vis at f er konveks ut fra definisjonen (selv om det er klart ut fra at $f'' > 0$). Du kan gjøre det ved følgende skritt:

    (a) La $x, y \in \mathbb{R}$ og la $0 \leq \lambda \leq 1$. Regn ut LHS=$((1-\lambda)x + \lambda y)^2$ og la RHS=$(1-\lambda)x^2 + \lambda y^2$. Regn ut RHS-LHS.

    Løsning: Ved kvadratsetning er

    $$\text{LHS} = ((1-\lambda)x + \lambda y)^2 = (1-\lambda)^2x^2 + 2(1-\lambda)\lambda xy + \lambda^2 y^2$$

    (b) Forenkle RHS-LHS ved å sette $\lambda(1-\lambda)$ utenfor som felles faktor, og vis at RHS-LHS≥ 0. Forklar hvorfor dette viser at f er konveks.

Løsning: Ved litt regning

$$\text{RHS-LHS} = \lambda(1-\lambda)(x^2 - 2xy + y^2) = \lambda(1-\lambda)(x-y)^2 \geq 0$$

fordi $0 \leq \lambda \leq 1$ (og kvadrat et er ≥ 0). LHS$\leq$RHS er akkurat ulikheten som viser konveksitet. $\diamond$

17. La $f : C \to \mathbb{R}$ være gitt ved $f(x) = x^n$. Her er n et positivt heltall og $C = \{x \in \mathbb{R} : x > 0\}$.

    (a) La $n = 3$. Vis at f er konveks ved å vise at $f''(x) \geq 0$ for alle $x \in C$.

    Løsning: $f'(x) = 3x^2$ og $f''(x) = 6x > 0$ fordi $x > 0$.

    (b) Samme oppgave som (a), men for generell n

    Løsning: $f'(x) = nx^{n-1}$ og $f''(x) = n(n-1)6x^{n-2} > 0$ fordi $x > 0$. $\diamond$

18. La $f : C \to \mathbb{R}$ være gitt ved $f(x) = 1/x$. Her er $C = \{x \in \mathbb{R} : x > 0\}$. Vis at f er konveks.

    Løsning: $f(x) = 1/x = x^{-1}$, $f'(x) = -x^{-2}$ og $f''(x) = 2x^{-3} > 0$ fordi $x > 0$. $\diamond$

19. La $f : C \to \mathbb{R}$ være gitt ved $f(x) = x \ln x$ der ln er den naturlige logaritmefunksjonen og $C = \{x \in \mathbb{R} : x > 0\}$. Vis at f er konveks.

    Løsning: Bruker regel for derivasjon av produkt og får

    $$f'(x) = 1 \cdot \ln x + x \cdot (1/x) = \ln x + 1, \quad f''(x) = 1/x > 0$$

    så f er konveks. $\diamond$

20. La $f : \mathbb{R} \to \mathbb{R}$ være gitt ved $f(x) = |x|$. (Husk: $|x| = x$ når $x \geq 0$, og $|x| = -x$ når $x < 0$.) Vis at f er konveks. OBS: f er ikke deriverbar overalt (den er ikke deriverbar i 0), så du kan ikke bruke derivasjonsmetoden.

    Løsning: Vi kan bruke definisjonen, eller f.eks. at epi(f) er en konveks mengde (det er et polyeder i $\mathbb{R}^2$ gitt ved $y \geq x$, $y \geq -x$). Enda et bevis er å se at f er punktvis maksim av to lineære funksjoner: $g(x) = x$ og $h(x) - x$. $\diamond$

21. (Prøveeksamen UiO V2012) Anta at f, g er konvekse, positive og voksende funksjoner fra $\mathbb{R}$ til $\mathbb{R}$. Vi antar at begge funksjonene er to ganger deriverbare. Vis at $h : \mathbb{R} \to \mathbb{R}$ gitt ved $h(x) = f(x)g(x)$ er konveks. Hint: Se på den andrederiverte til h.

    Løsning: Har: $h'(x) = f'(x)g(x) + f(x)g'(x)$, og $h''(x) = f''(x)g(x) + f(x)g''(x) + 2f'(x)g'(x)$. Siden f og g er konvekse har vi $f''(x) \geq 0$ og $g''(x) \geq 0$. Siden funksjonene er voksende har vi $f'(x) \geq 0$ and $g'(x) \geq 0$. Dessuten, siden funsksjonene også er positive, ser vi at alle de tre leddene i summen over er ≥ 0, så $h''(x) \geq 0$. Da følger det at h også er konveks. ◇

22. La $f : [a, b] \to \mathbb{R}$ være en konveks funksjon. Vis at

$$\max\{f(x) : x \in [a, b]\} = \max\{f(a), f(b)\},$$

    altså at en konveks funksjon på et lukket intervall oppnår sitt maksimum i et av endepunktene.

    Løsning: La $x \in [a, b]$. La $M = \max\{f(a), f(b)\}$. Da finnes $0 \leq \lambda \leq 1$ slik st $x = (1 - \lambda)a + \lambda b$, og ved konveksitet er

$$f(x) = f((1-\lambda)a+\lambda b) \leq (1-\lambda)f(a)+\lambda f(b) \leq (1-\lambda)M+\lambda M = M.$$

    ◇

Kapittel 6

Bibliografi

[1] G. Dahl, An Introduction to Convexity, Lecture notes, Dept. of Mathematics, University of Oslo (2010).

[2] T.L. Lindstrøm, *Kalkulus*, 4. utgave, Universitetsforlaget, Oslo, 2016.

[3] A.B. Sletsjøe, Kompendium, MAT 1050, Matematisk inst., Universitet i Oslo (2024).

[4] A.B. Sletsjøe, Kompendium, MAT 1060, Matematisk inst., Universitet i Oslo (2024).

[5] R. Webster, *Convexity*, Oxford University Press, Oxford, 1994.