

# **Flervariabel analyse med lineær algebra**

av

Tom Lindstrøm og Klara Hveberg

Matematisk institutt  
og  
Senter for matematikk for anvendelser (CMA)  
Universitetet i Oslo  
Revidert versjon for vårsemesteret 2009



## Forord til første versjon, våren 2006

Dette heftet er første del av et planlagt kompendium som skal dekke pensum i lineær algebra og flervariabel analyse i kursene MAT1100 og MAT1110 ved Universitetet i Oslo. I de to kapitlene som nå foreligger, gjennomgår vi først vektor- og matriseregning i  $\mathbb{R}^n$  og deretter den grunnleggende teorien for funksjoner av flere variable (inklusive kjerneregelen, men ikke multippel integrasjon).

Opplegget vårt har to grunntanker. For det første vil vi utnytte samspilllet mellom lineær algebra og flervariabel analyse på en bedre måte enn det som er vanlig i “standardoppdragget” der disse to emnene undervises i separate kurs. Etter vår mening tjener begge emnene på å bli sett i sammenheng — flervariabel analyse blir begrepsmessig enklere og mer oversiktlig når man kan bruke notasjon og terminologi fra lineær algebra, og lineæralgebraen vinner nye anvendelsesområder når den blir knyttet til flervariabel analyse. Vår andre grunntanke er at undervisningen i lineær algebra og flervariabel analyse bør knyttes nærmere til numeriske problemstillinger og bruk av dataverktøy. Siden MAT1100 er et kurs med grafiske lommeregnere som eneste elektroniske hjelpemiddel, er ikke denne tanken så synlig i de kapitlene som nå foreligger, med den vil komme tydligere frem både i stoffutvalg og presentasjonsform i de neste kapitlene. Vi har imidlertid prøvd å legge forholdene til rette allerede i de første kapitlene ved å trekke frem begreper som iterasjon, linearisering og egenverdier/egenvektorer på et tidlig stadium. Så selv om numerikken er så godt som fraværende i dette heftet, ser vi det likevel som en del av fakultetets kampanje for å styrke de numeriske sidene av realfagsutdanningen!

Heftet forutsetter at studentene har en solid forståelse av kontinuitet og deriverbarhet av funksjoner av én variabel, f.eks. tilsvarende de første sju kapitlene i Tom Lindstrøms lærebok *Kalkulus* (siden denne boken brukes i de andre delene av MAT1100, finnes det en del referanser til den i teksten). Siden heftet skal bli del av et større kompendium, finnes det noen få bemerkninger av typen “som vi senere skal se” som refererer til deler som ennå ikke foreligger.

To ord om notasjon: Vi bruker  $\square$  til å markere slutten på et bevis og  $\clubsuit$  til å markere slutten på et eksempel.

Blindern 26/10-2006

Tom Lindstrøm

Klara Hveberg

**Kommentar til den reviderte utgaven for vårsemesteret 2009:** Etter at den første versjonen ble lagt ut på nettet i 2006, har kompendiet gjenomgått en rekke forandringer — feil er rettet opp, nye oppgaver er lagt

til og noen seksjoner er flyttet eller omarbeidet. Til tross for at det ikke gis MATLAB-undervisning i MAT1100, har vi lagt til noen få, korte MATLAB-kommentarer med tanke på dem som har lyst til å sette seg inn i programmet på egen hånd. Vi nevner også at fire nye kapitler lagt ut på nettet som separate filer (disse brukes i oppfølgingskurset MAT1110):

Kapittel 3: “Kurver og flater”

Kapittel 4: “Lineær algebra i  $\mathbb{R}^n$ ”

Kapittel 5: “Iterasjon og optimering”

Kapittel 6: “Multippel integrasjon”

Dersom du finner trykkfeil eller har kommentarer til denne utgaven, så send en e-post til *lindstro@math.uio.no*.

# Innhold

<b>1 Vektorer og matriser</b>	<b>3</b>
1.1 Algebra for $n$ -tupler . . . . .	4
1.2 Geometri for $n$ -tupler . . . . .	9
1.3 Komplekse $n$ -tupler . . . . .	20
1.4 Vektorproduktet . . . . .	23
1.5 Matriser . . . . .	35
1.6 Multiplikasjon av matriser . . . . .	45
1.7 Identitetsmatriser og inverse matriser . . . . .	56
1.8 Determinanter, arealer og volumer . . . . .	62
<b>2 Funksjoner fra <math>\mathbb{R}^n</math> til <math>\mathbb{R}^m</math></b>	<b>75</b>
2.1 Funksjoner av flere variable . . . . .	75
2.2 Kontinuerlige funksjoner . . . . .	79
2.3 Grenseverdier . . . . .	86
2.4 Derivasjon av skalarfelt . . . . .	89
2.5 Partielle deriverte av høyere orden . . . . .	102
2.6 Derivasjon av vektorvaluerte funksjoner . . . . .	106
2.7 Kjerneregelen . . . . .	111
2.8 Lineæravbildninger . . . . .	122
2.9 Affinavbildninger og lineariseringer . . . . .	131
<b>3 Kurver og flater</b>	<b>141</b>
3.1 Parametriserte kurver . . . . .	141
3.2 Kjerneregelen for parametriserte kurver . . . . .	157
3.3 Linjeintegraler for skalarfelt . . . . .	161
3.4 Linjeintegraler for vektorfelt . . . . .	169
3.5 Gradienter og konservative felt . . . . .	178
3.6 Kjeglesnitt . . . . .	185
3.7 Grafisk fremstilling av skalarfelt . . . . .	208
3.8 Grafisk fremstilling av vektorfelt . . . . .	227
3.9 Parametriserte flater . . . . .	232

<b>4 Lineær algebra i <math>\mathbb{R}^n</math></b>	<b>239</b>
4.1 Noen eksempler på Gauss-eliminasjon . . . . .	240
4.2 Trappeform . . . . .	249
4.3 Redusert trappeform . . . . .	261
4.4 Matriseligninger . . . . .	267
4.5 Inverse matriser . . . . .	275
4.6 Lineærkombinasjoner og basiser . . . . .	283
4.7 Underrom . . . . .	300
4.8 Elementære matriser . . . . .	309
4.9 Determinanter . . . . .	313
4.10 Egenvektorer og egenverdier . . . . .	331
4.11 Egenvektorer i praksis . . . . .	347
4.12 Spektralteoremet . . . . .	366
<b>5 Iterasjon og optimering</b>	<b>375</b>
5.1 Litt topologi i $\mathbb{R}^m$ . . . . .	376
5.2 Kompletthet av $\mathbb{R}^m$ . . . . .	385
5.3 Noen konsekvenser av kompletthet . . . . .	390
5.4 Iterasjon av funksjoner . . . . .	402
5.5 Konvergens mot et fikspunkt . . . . .	410
5.6 Newtons metode i flere variable . . . . .	420
5.7 Omvendte og implisitte funksjoner . . . . .	439
5.8 Ekstremalverdisetningen . . . . .	454
5.9 Maksimums- og minimumspunkter . . . . .	456
5.10 Lagranges multiplikatormetode . . . . .	478
5.11 Gradientmetoden . . . . .	502
<b>6 Multippel integrasjon</b>	<b>507</b>
6.1 Dobbeltintegraler over rektangler . . . . .	508
6.2 Dobbeltintegraler over begrensede områder . . . . .	518
6.3 Dobbeltintegraler i polarkoordinater . . . . .	526
6.4 Anwendelser av dobbeltintegraler . . . . .	534
6.5 Greens teorem . . . . .	546
6.6 Jordan-målbare mengder . . . . .	560
6.7 Skifte av variable i dobbeltintegral . . . . .	566
6.8 Uegentlige integraler i planet . . . . .	586
6.9 Trippelintegraler . . . . .	590
6.10 Skifte av variable i trippelintegraler . . . . .	599
6.11 Anwendelser av trippelintegraler . . . . .	607
<b>Fasit</b>	<b>613</b>



# Kapittel 1

## Vektorer og matriser

Tall spiller en sentral rolle i matematikken — så sentral at mange nok vil si at der det er tall, er det matematikk, og der det ikke er tall, er det ikke matematikk! Fullt så enkelt er det ikke — det finnes mange grener av matematikken der tall spiller en underordnet rolle — men det er likevel ikke til å komme forbi at tall er et av fagets aller viktigste bestanddeler.

I din tidligere matematikkutdanning har du lært å regne med mange slags tall: hele tall, desimaltall, brøker, irrasjonale tall og til og med komplekse tall. I dette kapitlet skal vi gå et skritt videre og regne med tupler av tall, dvs. flere tall på en gang. Du har vært borti dette tidligere når du har regnet med vektorer i planet og i rommet — en vektor  $[x, y]$  i planet er et 2-tuppel, mens en vektor  $[x, y, z]$  i rommet er et 3-tuppel. Vi skal nå gå videre og regne med  $n$ -tupler for alle naturlige tall  $n$ . Hvis du tenker geometrisk, kan dette høres skummelt ut — hvordan skal man kunne forestille seg en 4-dimensjonal vektor  $[x, y, z, u]$ ? Tenker du mer algebraisk, er det ikke noe skummelt i det hele tatt; et 4-tuppel  $[x, y, z, u]$  er bare en notasjon for å holde styr på fire tall på en praktisk og kortfattet måte. I dette kapitlet skal vi utvikle både algebra og geometri, for selv om det å regne algebraisk med tupler er trygt og ukomplisert, mister man fort oversikten, og den gjenvinner man først når man lærer å tenke på tupler som geometriske objekter.

Vi skal også gå et skritt videre og arbeide med matriser. Dette er rektangulære oppsett av tall som f.eks.

$$\begin{pmatrix} 2 & -1 & 4 \\ 0 & -3 & 2 \end{pmatrix} \quad \text{og} \quad \begin{pmatrix} 3 & 2 \\ -1 & 12 \\ \frac{3}{2} & -2 \end{pmatrix}$$

Ved hjelp av matriser kan vi “transformere”  $n$ -tupler på en måte som er viktig i svært mange sammenhenger, både regneteknisk og geometrisk. Matriser og tupler kommer til å spille en sentral rolle også i senere kapitler, dels som nyttige verktøy og dels som selvstendige studieobjekter.

## 1.1 Algebra for $n$ -tupler

La oss begynne med den grunnleggende definisjonen. Et  $n$ -tuppel er et uttrykk  $(a_1, a_2, \dots, a_n)$  der  $a_1, a_2, \dots, a_n$  er reelle tall. Vi ser at  $(2, -1, 7, 3)$  er et 4-tuppel, mens  $(0, 1, \pi, \frac{3}{2}, -7, 3)$  er et 6-tuppel. To  $n$ -tupler  $(a_1, a_2, \dots, a_n)$  og  $(b_1, b_2, \dots, b_n)$  regnes som like dersom de inneholder de samme tallene i samme rekkefølge, dvs. hvis  $a_1 = b_1, a_2 = b_2, \dots, a_n = b_n$ . Legg merke til at  $(3, 2, 4) \neq (2, 3, 4)$ ; selv om tallene er de samme, er rekkefølgen forskjellig.

I dette heftet skal vi bruke bokstaver i **fete typer** som navn på  $n$ -tupler, f.eks.  $\mathbf{a} = (-2, 3, 0, -17)$ . Det er vanskelig å bruke fete typer når man skriver for hånd, og man kan da isteden skrive en pil eller en strek over bokstavene; dvs.  $\vec{a} = (-2, 3, 0, -17)$  eller  $\bar{a} = (-2, 3, 0, -17)$ .

Vi skriver  $\mathbf{0}$  for det  $n$ -tuplet som har alle komponenter lik 0, altså  $\mathbf{0} = (0, 0, \dots, 0)$ . Hvis vi har et  $n$ -tuppel  $\mathbf{a} = (a_1, a_2, \dots, a_n)$ , skriver vi  $-\mathbf{a}$  for  $n$ -tuplet  $(-a_1, -a_2, \dots, -a_n)$ .

Det er en naturlig måte å definere addisjon og subtraksjon av  $n$ -tupler på. Dersom  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  og  $\mathbf{b} = (b_1, b_2, \dots, b_n)$ , så er

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

og

$$\mathbf{a} - \mathbf{b} = (a_1 - b_1, a_2 - b_2, \dots, a_n - b_n)$$

Vi sier at vi adderer og subtraherer *komponentvis*. Legg merke til at vi bare kan addere og subtrahere tupler med like mange komponenter — oppskriften ovenfor gir oss ikke noen måte å addere et 3-tuppel og et 7-tuppel på. Før vi ser på et eksempel, tar vi med en regneoperasjon til. Dersom  $s$  er et tall og  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  er et  $n$ -tuppel, definerer vi produktet av  $s$  og  $\mathbf{a}$  til å være

$$s\mathbf{a} = (sa_1, sa_2, \dots, sa_n)$$

Vi ganger altså  $s$  inn i hver komponent i  $\mathbf{a}$ .

**Eksempel 1.** Vi lar  $\mathbf{a} = (-2, 3, 0, -17)$  og  $\mathbf{b} = (4, -1, 3, 17)$ . Da er

$$\mathbf{a} + \mathbf{b} = (-2 + 4, 3 + (-1), 0 + 3, -17 + 17) = (2, 2, 3, 0)$$

og

$$\mathbf{a} - \mathbf{b} = (-2 - 4, 3 - (-1), 0 - 3, -17 - 17) = (-6, 4, -3, -34)$$

Hvis  $s = 3$ , får vi

$$s\mathbf{a} = (3 \cdot (-2), 3 \cdot 3, 3 \cdot 0, 3 \cdot (-17)) = (-6, 9, 0, -51)$$



Vi skal innføre en regneoperasjon til. Dersom  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  og  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  er to  $n$ -tupler, definerer vi *skalarproduktet* (også kalt *prikkproduktet*)  $\mathbf{a} \cdot \mathbf{b}$  ved

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

Legg merke til at  $\mathbf{a} \cdot \mathbf{b}$  ikke er et  $n$ -tuppel, men et tall (eller en *skalar* som man ofte sier når man vil understreke at noe er et tall og ikke et  $n$ -tuppel). Hvis vi lar  $\mathbf{a} = (-2, 3, 0, -17)$  og  $\mathbf{b} = (4, -1, 3, 17)$  som ovenfor, ser vi at

$$\mathbf{a} \cdot \mathbf{b} = (-2) \cdot 4 + 3 \cdot (-1) + 0 \cdot 3 + (-17) \cdot 17 = -8 - 3 + 0 - 289 = -300$$

Vi har nå sett hvordan vi kan regne med  $n$ -tupler, og det er kanskje på tide å ta en kikk på noen eksempler som antyder hvorfor det er et poeng med slike regnestykker. Det første eksemplet viser at  $n$ -tupler er naturlige redskap når vi skal holde styr på mer informasjon enn det som kan rommes i et enkelt tall, og at regneoperasjonene svarer til regnestykker det ofte er naturlig å utføre i slike sammenhenger.

**Eksempel 2.** En forretning har ansatt 7 studenter på timebasis. For å holde styr på hvor mange timer hver student har arbeidet så langt, kan vi bruke et 7-tuppel  $\mathbf{t} = (t_1, t_2, \dots, t_7)$  der  $t_1$  er antall timer den første studenten har arbeidet,  $t_2$  er antall timer den andre studenten har arbeidet osv. Dersom studentene arbeider mer senere, kan vi på samme måte kode tilleggstimene som et 7-tuppel  $\mathbf{s} = (s_1, s_2, \dots, s_7)$ . Det totale antall timer som studentene har arbeidet, er nå gitt ved  $\mathbf{t} + \mathbf{s}$ .

Studentene har ulik erfaring og derfor ulik lønn. Hvis student nummer én har en timelønn på  $p_1$  kroner, student nummer to har en timelønn på  $p_2$  kroner osv., kan vi også representere lønnen som et 7-tuppel  $\mathbf{p} = (p_1, p_2, \dots, p_7)$ . Dersom studentene har arbeidet  $\mathbf{t} = (t_1, t_2, \dots, t_7)$  timer, er den totale lønnen som forretningen skylder, gitt av skalarproduktet  $\mathbf{p} \cdot \mathbf{t} = p_1 t_1 + p_2 t_2 + \dots + p_7 t_7$ . Dersom alle studentene får et lønnstillegg på 7 prosent, får vi det nye lønnstuplet ved å gange det gamle med skalaren 1.07, altså  $1.07\mathbf{p}$ .



Vi tar med noen eksempler til som viser hvordan  $n$ -tupler brukes til å holde styr på tallmessig informasjon i forskjellige sammenhenger.

**Eksempel 3.** Tilstanden til en gassbeholder er bestemt av trykket  $p$ , temperaturen  $T$  og volumet  $V$ . Hvis du får i oppdrag å måle tilstanden til beholderen ved forskjellige tidspunkt, kan det være naturlig å bruke 4-tupler  $\mathbf{a} = (t, p, T, V)$  der  $t$  er tidspunktet for målingen. Forskjellen mellom to målinger  $\mathbf{a}$  og  $\mathbf{b}$  er da gitt ved differensen  $\mathbf{b} - \mathbf{a}$ .



**Eksempel 4.** Et bilde på en fjernsynsskjerm eller en dataskjerm er bygget opp av små lysende punkter (piksler). Et vanlig format er  $1280 \times 1024 = 1310720$  piksler. I hvert punkt må vi angi styrken til hver av de tre grunnfargene rødt, grønt og blått, så totalt har vi  $3 \times 1310720 = 3932160$  tall å holde styr på. En naturlig måte å gjøre dette på er å oppfatte bilder som 3932160-tupler! Dette er ikke noe enestående eksempel — i mange anledninger er man interessert i tupler med svært mange komponenter. ♣

Her er noen enkle regneregler for  $n$ -tupler (det finnes flere). Vær oppmerksom på at vi bruker de samme prioriteringsreglene for vektorer som for tall; dersom det ikke står parenteser, skal multiplikasjoner utføres før addisjoner.

**Setning 1.1.1 (Regneregler for  $n$ -tupler)** *Dersom  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$  er  $n$ -tupler og  $s$  og  $t$  er reelle tall, gjelder følgende regneregler:*

- (a)  $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
- (b)  $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$
- (c)  $s(\mathbf{a} + \mathbf{b}) = s\mathbf{a} + s\mathbf{b}$
- (d)  $(s + t)\mathbf{a} = s\mathbf{a} + t\mathbf{a}$
- (e)  $\mathbf{c} \cdot (\mathbf{a} + \mathbf{b}) = \mathbf{c} \cdot \mathbf{a} + \mathbf{c} \cdot \mathbf{b}$  og  $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$
- (f)  $(s\mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (s\mathbf{b}) = s(\mathbf{a} \cdot \mathbf{b})$
- (g)  $\mathbf{a} \cdot \mathbf{a} \geq 0$  med likhet hvis og bare hvis  $\mathbf{a} = \mathbf{0}$

*Bevis:* Alle disse reglene bevises lett ved å regne ut venstre- og høyresiden og kontrollere at svarene stemmer overens. Vi tar (c) og (g) som eksempler:  
(c) Dersom  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  og  $\mathbf{b} = (b_1, b_2, \dots, b_n)$ , ser vi at venstresiden kan skrives

$$\begin{aligned} s(\mathbf{a} + \mathbf{b}) &= s(a_1 + b_1, a_2 + b_2, \dots, a_n + b_n) \\ &= (s(a_1 + b_1), s(a_2 + b_2), \dots, s(a_n + b_n)) \\ &= (sa_1 + sb_1, sa_2 + sb_2, \dots, sa_n + sb_n) \end{aligned}$$

Tilsvarende kan høyresiden skrives

$$\begin{aligned} s\mathbf{a} + s\mathbf{b} &= (sa_1, sa_2, \dots, sa_n) + (sb_1, sb_2, \dots, sb_n) \\ &= (sa_1 + sb_1, sa_2 + sb_2, \dots, sa_n + sb_n) \end{aligned}$$

Siden de to uttrykkene er like, er (c) bevist.

(g) Vi ser at

$$\mathbf{a} \cdot \mathbf{a} = a_1^2 + a_2^2 + \dots + a_n^2 \geq 0$$

siden kvadrater aldri er negative. Likhet har vi dersom  $a_1^2 = 0, a_2^2 = 0, \dots, a_n^2 = 0$ , dvs. dersom  $a_1 = 0, a_2 = 0, \dots, a_n = 0$ .  $\square$

Vi har hittil skrevet våre n-tupler liggende

$$\mathbf{a} = (a_1, a_2, \dots, a_n),$$

men vi kan også skrive dem stående

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

I det første tilfellet kaller vi **a** en *radvektor*, mens vi i det andre kaller **a** en *søylevektor*. I de fleste situasjoner spiller det ingen rolle om vi skriver *n*-tuplene på den ene eller andre formen, og vi velger da ofte å skrive dem som radvektorer siden det tar minst plass. Det finnes imidlertid tilfeller der det er viktig å skille mellom radvektorer og søylevektorer, men det skal vi komme tilbake til senere — foreløpig kan du skrive dine vektorer på den måten du måtte ønske. Legg for øvrig merke til at det ofte kan være lettere å få øye på strukturen i et regnestykke når du bruker søylevektorer, f.eks. kan

$$s \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + t \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} + r \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} sa_1 + tb_1 + rc_1 \\ sa_2 + tb_2 + rc_2 \\ \vdots \\ sa_n + tb_n + rc_n \end{pmatrix}$$

virke mer oversiktlig enn

$$\begin{aligned} s(a_1, a_2, \dots, a_n) + t(b_1, b_2, \dots, b_n) + r(c_1, c_2, \dots, c_n) &= \\ &= (sa_1 + tb_1 + rc_1, sa_2 + tb_2 + rc_2, \dots, sa_n + tb_n + rc_n) \end{aligned}$$

La oss avslutte denne seksjonen med noen flere ord om notasjon. Mengden av alle *n*-tupler kaller vi  $\mathbb{R}^n$ . Når vi skriver  $\mathbf{a} \in \mathbb{R}^n$ , betyr dette derfor ikke noe annet enn at **a** er et *n*-tuppel. Hittil har vi holdt oss til reelle *n*-tupler, men vi kan selvfølgelig også tenke oss *n*-tupler  $(c_1, c_2, \dots, c_n)$  der komponentene  $c_1, c_2, \dots, c_n$  er *komplekse* tall. Mengden av alle slike *n*-tupler kaller vi  $\mathbb{C}^n$ . Vi skal se nærmere på komplekse *n*-tupler litt senere. Notasjonen kan også gjøres enda mer generell: Dersom  $A$  er en hvilken som helst mengde, betegner  $A^n$  mengden av alle *n*-tupler  $(a_1, a_2, \dots, a_n)$  der  $a_i \in A$  for alle  $i = 1, 2, \dots, n$ .

Helt til slutt legger vi merke til at et 1-tuppel ( $a_1$ ) ikke er noe annet enn et tall inni en parentes. Parentesen spiller ingen rolle (den er bare med for å avgrense uttrykket), og vi skal derfor regne 1-tuplet ( $a_1$ ) og tallet  $a_1$  som det samme objektet. Dette betyr at  $\mathbb{R}^1$  og  $\mathbb{R}$  er den samme mengden.

**MATLAB-kommentar:** For å regne med vektorer i MATLAB, må du første skrive dem inn. Du kan skrive inn radvektorene  $\mathbf{a} = (1, -2, 3, 0, 5)$  og  $\mathbf{b} = (3, -2.4, -2, 0)$  ved hjelp av kommandoene:

```
>> a=[1,-2,3,0,5]
>> b=[3,-2.4,-2,0]
```

Legg merke til at vi bruker hakeparenteser  $[,]$  og ikke runde parenteser  $(, )$  for å beskrive vektorer. Du kan også erstatte kommaene mellom komponentene med mellomrom:

```
>> a=[1 -2 3 0 5]
>> b=[3 -2 4 -2 0]
```

Når vektorene er skrevet inn, kan du regne med dem ved å bruke kommandoer av typen  $>>\mathbf{a}+\mathbf{b}$ ,  $>>\mathbf{a}-\mathbf{b}$ ,  $>>7*\mathbf{a}$ . Skalarproduktet får du ved å skrive  $>>\text{dot}(\mathbf{a}, \mathbf{b})$ . Vil du skrive inn en søylevektor

$$c = \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix},$$

må du bruke semikolon mellom komponentene:

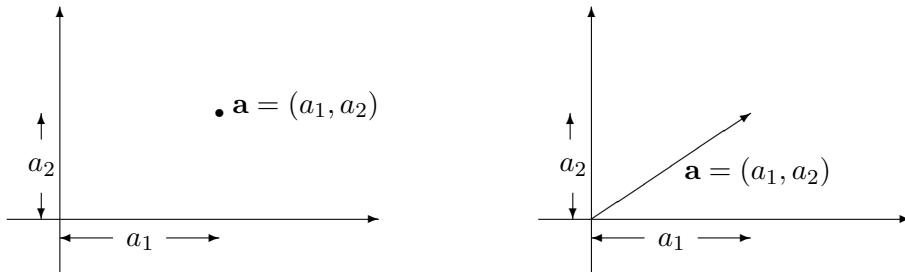
```
c=[-1;3;2]
```

### Oppgaver til seksjon 1.1

1. Finn  $\mathbf{a}+\mathbf{b}$ ,  $\mathbf{a}-\mathbf{b}$ ,  $s\mathbf{a}$  og  $\mathbf{a}\cdot\mathbf{b}$  når  $\mathbf{a} = (1, -2, 4, -5, 1)$ ,  $\mathbf{b} = (-3, 5, 5, 0, -3)$  og  $s = 3$ .
2. Finn  $\mathbf{a} + \mathbf{b}$ ,  $\mathbf{a} - \mathbf{b}$ ,  $s\mathbf{a}$  og  $\mathbf{a} \cdot \mathbf{b}$  når  $\mathbf{a} = (7, 0, 4, -2, -5, 4)$ ,  $\mathbf{b} = (0, 2, 1, -6, 0, -1)$  og  $s = -4$ .
3. Vis at for alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  er:
  - a)  $(\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} + \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} + 2\mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{y}$
  - b)  $(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} - 2\mathbf{x} \cdot \mathbf{y} + \mathbf{y} \cdot \mathbf{y}$
  - c)  $(\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) = \mathbf{x} \cdot \mathbf{x} - \mathbf{y} \cdot \mathbf{y}$ .
4. Bevis punktene d), e), f) i setning 1.1.1.
5. Et grossistfirma har  $n$  vareslag på lager,  $m_1$  enheter av vareslag 1,  $m_2$  enheter av vareslag 2 osv. Verdien av hver enhet er  $p_1$  for vareslag 1,  $p_2$  for vareslag 2 osv. Uttrykk den totale verdien av varelageret som skalarproduktet mellom to  $n$ -tupler.

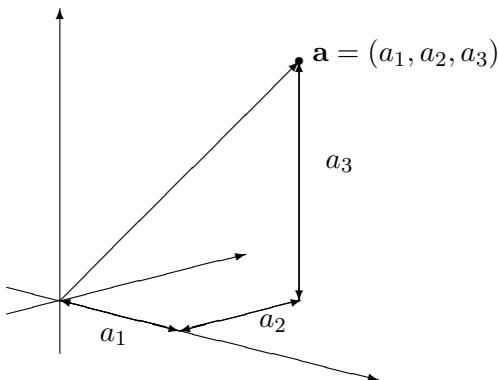
## 1.2 Geometri for $n$ -tupler

Et 2-tuppel er ikke noe annet enn et par  $(a_1, a_2)$ . Geometrisk kan vi tenke på et slikt par på to måter — enten som et *punkt* med koordinater  $a_1$  og  $a_2$ , eller som en *vektor* (pil) som starter i origo og ender i dette punktet (se figur 1).



Figur 1:  $\mathbf{a}$  som et punkt og som en vektor

I skolematematikken bruker man gjerne forskjellig notasjon ettersom man tenker på paret som et punkt eller som en vektor — et punkt  $(a_1, a_2)$  har runde parenteser, mens en vektor  $[a_1, a_2]$  har klammeparenteser. Det er ganske tungvint å bruke to forskjellige notasjoner, og vi vil derfor bruke runde parenteser  $\mathbf{a} = (a_1, a_2)$  uansett om vi tenker på  $\mathbf{a}$  som et punkt eller som en vektor. Hva som er naturlig, fremgår som regel av sammenhengen. Snakker vi om en linje gjennom  $\mathbf{a}$ , er det naturlig å tenke på  $\mathbf{a}$  som et punkt, men snakker vi om en linje parallel med  $\mathbf{a}$ , er det naturlig å tenke på  $\mathbf{a}$  som en vektor. Når vi lager figurer, vil vi noen ganger tegne paret  $(a_1, a_2)$  som en vektor og andre ganger som et punkt, alt etter hva vi synes passer best i hvert enkelt tilfelle (se figur 1).

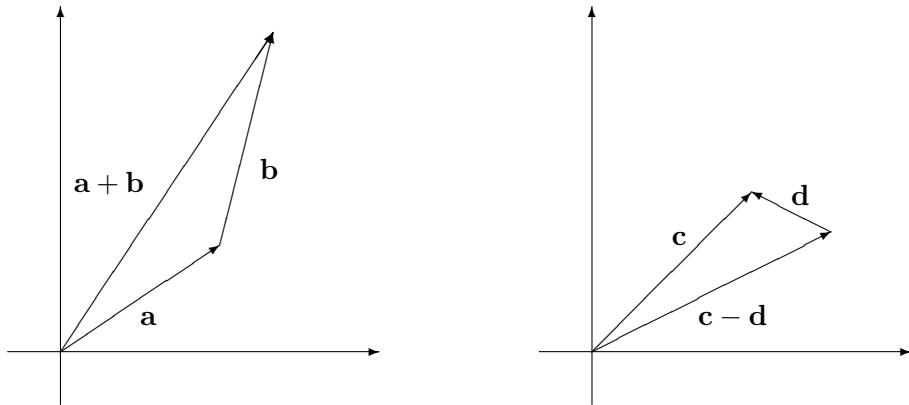


Figur 2: Et 3-tuppel som en vektor i rommet

På tilsvarende vis kan vi oppfatte 3-tupler som punkter og vektorer i

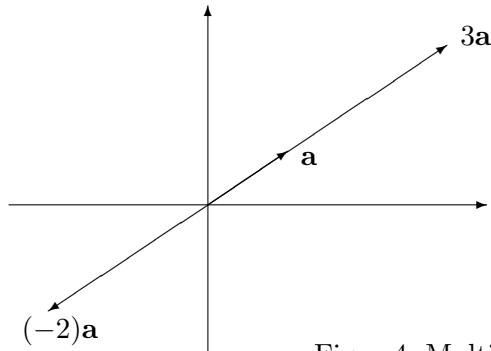
rommet. Figuren ovenfor viser hvordan et 3-tuppel  $\mathbf{a} = (a_1, a_2, a_3)$  kan oppfattes som en vektor i rommet.

Som du vet fra skolematematikken, har de algebraiske operasjonene vi innførte i forrige seksjon, en geometrisk tolkning når vi tenker på tupler som vektorer i planet eller rommet. Figur 3 viser hvordan vi får frem addisjon og subtraksjon ved å sette sammen vektorer:



Figur 3: Addisjon og subtraksjon av vektorer

Multiplikasjon med en skalar har også en geometrisk tolkning. Dersom vi ganger  $\mathbf{a}$  med et *positivt* tall  $s$ , beholder vektoren retningen, men blir  $s$  ganger så lang. Dersom vi ganger  $\mathbf{a}$  med et *negativt* tall  $s$ , snur retningen  $180^\circ$ , og den nye vektoren blir  $|s|$  ganger så lang som den opprinnelige (se figur 4).



Figur 4: Multiplikasjon med et tall

Det er også andre sammenhenger mellom regneoperasjoner og geometri. Fra skolematematikken vet du for eksempel at to (ikke-null) vektorer  $\mathbf{a}, \mathbf{b}$  er parallele dersom det finnes et tall  $s \neq 0$  slik at  $\mathbf{a} = s\mathbf{b}$ , og at de er ortogonale (dvs. står normalt på hverandre) dersom  $\mathbf{a} \cdot \mathbf{b} = 0$ . Du vet også at lengden  $|\mathbf{a}|$  til vektoren  $\mathbf{a} = (a_1, a_2, a_3)$  kan regnes ut fra koordinatene:

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

og at det er en sammenheng mellom lengden og skalarproduktet:

$$|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a}$$

Det er sammenhengen mellom geometri og algebra som gir liv til vektorregning i to og tre dimensjoner, og det hadde vært nyttig om vi kunne bruke vår geometriske intuisjon på samme måte når vi arbeidet med generelle  $n$ -tupler. Dette kan virke som en uoverkommelig oppgave — hvis 2-tupler representerer 2-dimensjonale objekter i planet, og 3-tupler representerer 3-dimensjonale objekter i rommet, så burde 4-tupler representerere 4-dimensjonale objekter i et slags 4-dimensjonalt rom? Og, enda verre, 5-tupler burde representerere 5-dimensjonale objekter i et 5-dimensjonalt rom, 6-tupler burde representerere 6-dimensjonale objekter i et 6-dimensjonalt rom osv? Hvem av oss kan med hånden på hjertet si at de har noen særlig geometrisk intuisjon for det som skjer i 4-, 5- og 6-dimensjonale rom?

Heldigvis behøver vi ikke å ha en slik intuisjon på forhånd, men kan bygge den opp gradvis. Ideen er enkel: vi overfører geometriske begreper fra planet og rommet til det generelle tilfellet ved å bruke de algebraiske beskrivelsene av geometriske egenskaper. Her er et eksempel: At to vektorer  $\mathbf{a}$  og  $\mathbf{b}$  er ortogonale (dvs. at de står normalt på hverandre), er i utgangspunktet en geometrisk egenskap. Denne egenskapen kan vi beskrive algebraisk ved  $\mathbf{a} \cdot \mathbf{b} = 0$ . Vi bruker nå denne algebraiske beskrivelsen til å *definere* at to  $n$ -tupler  $\mathbf{a}$  og  $\mathbf{b}$  er ortogonale dersom  $\mathbf{a} \cdot \mathbf{b} = 0$  (vi sier da at de står *normalt på hverandre*). På denne måten får vi innført det geometriske begrepet orthogonalitet i høyere dimensjoner uten å måtte støtte oss til noen geometrisk intuisjon. Når begrepet først er innført på denne måten, kan vi undersøke i hvilken grad det svarer til våre (geometriske) forestillinger om hva orthogonalitet er. På den måten bygger vi etter hvert opp en intuisjon om orthogonalitet av  $n$ -tupler, og denne intuisjonen tar fort en geometrisk form.

La oss begynne med litt terminologi: Mengden  $\mathbb{R}^n$  av alle  $n$ -tupler kalles det  *$n$ -dimensjonale euklidske rommet*, og et  $n$ -tuppel  $\mathbf{a}$  kalles også en  *$n$ -dimensjonal vektor* eller et  *$n$ -dimensjonalt punkt*. Som i det 2- og 3-dimensjonale tilfellet skal vi ofte bruke ordet “vektor” når det er naturlig å tenke på  $\mathbf{a}$  som et geometrisk objekt med lengde og retning, og vi skal bruke ordet “punkt” når vi er opptatt av noe (f.eks. en linje eller et plan) som går gjennom  $\mathbf{a}$ . Logisk sett er det selvfølgelig unødvendig å ha mer enn ett navn på disse objektene, men pedagogisk er det ofte en fordel å kunne bruke et ord som antyder hvilke egenskaper vi er opptatt av i hver enkelt situasjon. Vi har valgt det nøytrale ordet ” $n$ -tuppel” som utgangspunkt for ikke å binde oss sterkt til den ene eller andre tolkningen.

Hvis  $\mathbf{a} = (a_1, a_2)$  er en to-dimensjonal vektor, er lengden gitt ved

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2}$$

Tilsvarende er lengden til en tre-dimensjonal vektor  $\mathbf{a} = (a_1, a_2, a_3)$  gitt ved

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

For en  $n$ -dimensjonal vektor  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  er det derfor naturlig å definere *lengden* (eller *normen* som den også kalles) ved

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

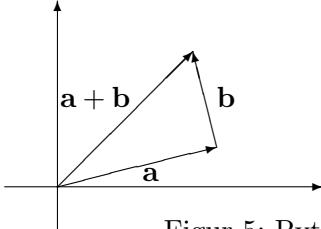
I kapittel 1 definerte vi skalarproduktet av vektorene  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  og  $\mathbf{b} = (b_1, b_2, \dots, b_n)$  til å være

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

og vi ser at vi har den vanlige sammenhengen mellom lengden og skalarproduktet:

$$|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} \quad \text{eller med andre ord} \quad \mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$$

Vi har allerede definert to  $n$ -tupler  $\mathbf{a}$  og  $\mathbf{b}$  til å være *ortogonale* (eller *står normalt på hverandre*) dersom  $\mathbf{a} \cdot \mathbf{b} = 0$ . Ved hjelp av denne definisjonen kan vi formulere en  $n$ -dimensjonal versjon av et meget berømt resultat (figur 5 viser den geometriske motivasjonen).



Figur 5: Pythagoras' setning i planet

**Setning 1.2.1 (Pythagoras' setning for  $n$ -tupler)** Dersom  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  er ortogonale, så er

$$|\mathbf{a} + \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2$$

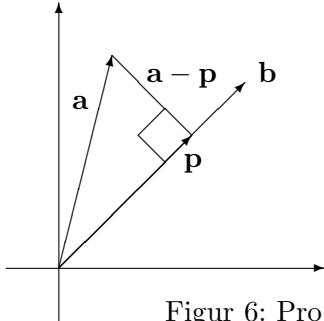
*Bevis:* Dette er bare et enkelt regnestykke (husk regnereglene for  $n$ -tupler fra kapittel 1):

$$\begin{aligned} |\mathbf{a} + \mathbf{b}|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) = \\ &= \mathbf{a} \cdot \mathbf{a} + 2\mathbf{a} \cdot \mathbf{b} + \mathbf{b} \cdot \mathbf{b} = |\mathbf{a}|^2 + 2 \cdot 0 + |\mathbf{b}|^2 = \\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 \end{aligned}$$

□

Setningen ovenfor er vårt første eksempel på et resultat om  $n$ -tupler som er inspirert av en geometrisk observasjon. Vårt neste problem tar utgangspunkt i figur 6. Vi tenker oss at vi er gitt to vektorer  $\mathbf{a}$ ,  $\mathbf{b}$ , og at vi ønsker

å finne *projeksjonen*  $\mathbf{p}$  av  $\mathbf{a}$  ned på  $\mathbf{b}$ . Dette betyr at  $\mathbf{p}$  er vektoren parallell med  $\mathbf{b}$  slik at  $\mathbf{a} - \mathbf{p}$  står normalt på  $\mathbf{b}$ .



Figur 6: Projeksjonen  $\mathbf{p}$  av  $\mathbf{a}$  ned på  $\mathbf{b}$

I utgangspunktet er dette en geometrisk problemstilling som bare gir mening for vektorer i planet og rommet, men vi kan bruke vår oversettingsfilosofi til å gi mening til problemet for generelle  $n$ -tupler  $\mathbf{a}$  og  $\mathbf{b}$ . Siden det  $n$ -tuplet  $\mathbf{p}$  vi er på jakt etter skal være parallelt med  $\mathbf{b}$ , må det finnes et tall  $t$  slik at  $\mathbf{p} = t\mathbf{b}$ , og siden  $\mathbf{a} - \mathbf{p}$  skal stå normalt på  $\mathbf{b}$ , må vi ha

$$0 = (\mathbf{a} - \mathbf{p}) \cdot \mathbf{b} = (\mathbf{a} - t\mathbf{b}) \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b} - t|\mathbf{b}|^2$$

Løser vi denne ligningen med hensyn på  $t$ , får vi

$$t = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2},$$

som betyr at

$$\mathbf{p} = t\mathbf{b} = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \mathbf{b}$$

Vi får dermed dette resultatet:

**Setning 1.2.2** *Anta at  $\mathbf{a}$  og  $\mathbf{b}$  er to ikke-null vektorer i  $\mathbb{R}^n$ . Da er projeksjonen  $\mathbf{p}$  av  $\mathbf{a}$  ned på  $\mathbf{b}$  gitt ved:*

$$\mathbf{p} = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2} \mathbf{b}$$

*Lengden til projeksjonen er  $|\mathbf{p}| = \frac{|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{b}|}$ .*

*Bevis:* Den første formelen har vi allerede utledet. Den andre kan vi for eksempel finne med følgende regnestykke:

$$|\mathbf{p}| = |t\mathbf{b}| = |t||\mathbf{b}| = \frac{|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{b}|^2} |\mathbf{b}| = \frac{|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{b}|}$$

□

La oss kombinere resultatet vi nettopp har bevist med Pythagoras' setning. Siden  $\mathbf{a} - \mathbf{p}$  står normalt på  $\mathbf{b}$ , må den også stå normalt på  $\mathbf{p}$  som er parallell med  $\mathbf{b}$  (sjekk dette!). Det betyr at vi kan bruke Pythagoras' setning på vektorene  $\mathbf{p}$ ,  $\mathbf{a} - \mathbf{p}$  og  $\mathbf{a}$  (se figur 6 for å få intuisjonen):

$$|\mathbf{a}|^2 = |\mathbf{p}|^2 + |\mathbf{a} - \mathbf{p}|^2$$

Siden  $|\mathbf{a} - \mathbf{p}|^2 \geq 0$ , betyr dette at

$$|\mathbf{a}|^2 \geq |\mathbf{p}|^2$$

som medfører at

$$|\mathbf{a}| \geq |\mathbf{p}|$$

(husk at både  $|\mathbf{a}|$  og  $|\mathbf{p}|$  er positive). Ifølge setningen ovenfor er  $|\mathbf{p}| = \frac{|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{b}|}$ , og setter vi dette inn i ulikheten, får vi

$$|\mathbf{a}| \geq \frac{|\mathbf{a} \cdot \mathbf{b}|}{|\mathbf{b}|}$$

Ganger vi med  $|\mathbf{b}|$  på begge sider, sitter vi igjen med

$$|\mathbf{a}||\mathbf{b}| \geq |\mathbf{a} \cdot \mathbf{b}|$$

Vi har kommet frem til en berømt og meget nyttig ulikhet:

**Setning 1.2.3 (Schwarz' ulikhet)** *For alle  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  gjelder*

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|$$

*Vi har likhet (dvs.  $|\mathbf{a} \cdot \mathbf{b}| = |\mathbf{a}||\mathbf{b}|$ ) hvis og bare hvis  $\mathbf{a}$  og  $\mathbf{b}$  er parallelle eller minst én av dem er null.*

*Bevis:* I utledningen av ulikheten har vi strengt tatt gått ut i fra at  $\mathbf{a}, \mathbf{b} \neq \mathbf{0}$ , men ulikheten gjelder åpenbart også om én eller begge vektorer er lik  $\mathbf{0}$  (for da er venstresiden i ulikheten lik 0). Det gjenstår dermed bare å sjekke den siste påstanden. Leser du gjennom utledningen av ulikheten en gang til, vil du se at vi har likhet når  $|\mathbf{a} - \mathbf{p}| = 0$ , dvs. når  $\mathbf{a} = \mathbf{p}$ . Siden  $\mathbf{p}$  er parallel med  $\mathbf{b}$ , skjer dette når  $\mathbf{a}$  og  $\mathbf{b}$  er parallele.  $\square$

Du husker sikkert fra skolematematikken at

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos v$$

der  $v$  er vinkelen mellom vektorene  $\mathbf{a}$  og  $\mathbf{b}$ . I utgangspunktet gir denne formelen bare mening når  $\mathbf{a}$  og  $\mathbf{b}$  er vektorer i planet eller rommet — for generelle  $n$ -tupler vet vi jo ikke hva vinkler er. Ved hjelp av Schwarz' ulikhet kan vi nå snu situasjonen på hodet; vi definerer rett og slett vinkelen mellom

to ikke-null  $n$ -tupler  $\mathbf{a}$  og  $\mathbf{b}$  til å være den vinkelen  $v$  mellom  $0^\circ$  og  $180^\circ$  som er slik at  $\cos v = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}$ . Legg merke til at siden Schwarz' ulikhet garanterer at  $-1 \leq \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} \leq 1$ , så finnes det alltid en slik vinkel  $v$  som definisjonen forutsetter. Vi ser også at vi får  $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos v$ .

Hva er så vitsen med et slikt abstrakt og merkelig vinkelbegrep? Kan disse vinklene brukes til noe, og oppfører de seg som de vinklene vi er vant til fra planet og rommet? Dette er fornuftige spørsmål som bare erfaring kan gi svar på. Erfaringen viser at disse vinklene fungerer utmerket, og at de i det store og hele har de samme egenskapene som vinkler i 2 og 3 dimensjoner. Vi skal ikke komme nærmere inn på dette her, men tar med et eksempel på hvordan man finner en vinkel:

**Eksempel 1:** Finn vinkelen mellom vektorene  $\mathbf{a} = (2, -1, 0, 1, 1)$  og  $\mathbf{b} = (0, 1, 3, -2, 0)$ . Vi har

$$\begin{aligned}\cos v &= \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} = \\ &= \frac{(2, -1, 0, 1, 1) \cdot (0, 1, 3, -2, 0)}{|(2, -1, 0, 1, 1)| |(0, 1, 3, -2, 0)|} = \frac{-3}{\sqrt{7} \sqrt{14}} = -\frac{3\sqrt{2}}{14}\end{aligned}$$

Bruker vi en lommeregner, finner vi at  $-\frac{3\sqrt{2}}{14} \approx -0.3030$ . Dette gir  $v \approx \arccos(-0.3030) \approx 107.6^\circ$ . ♣

Et viktig resultat i planet og rommet er trekantulikheten som sier at  $|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$ . Ved hjelp av Schwarz' ulikhet skal vi nå vise at trekantulikheten også gjelder i  $n$  dimensjoner.

**Setning 1.2.4 (Trekantulikheten)** *For alle  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  gjelder*

$$|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$$

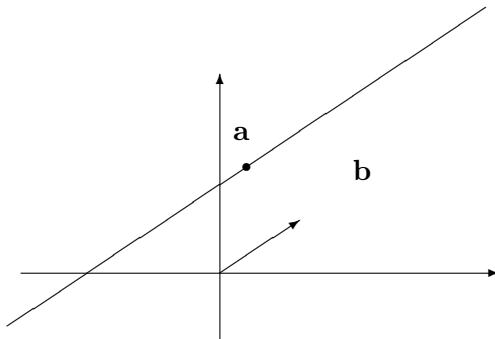
*Bevis:* Vi har

$$\begin{aligned}|\mathbf{a} + \mathbf{b}|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) \\ &= |\mathbf{a}|^2 + 2 \mathbf{a} \cdot \mathbf{b} + |\mathbf{b}|^2 \leq |\mathbf{a}|^2 + 2 |\mathbf{a}| |\mathbf{b}| + |\mathbf{b}|^2 = (|\mathbf{a}| + |\mathbf{b}|)^2\end{aligned}$$

der vi har brukt at ifølge Schwarz' ulikhet er  $\mathbf{a} \cdot \mathbf{b} \leq |\mathbf{a}| |\mathbf{b}|$ . □

Geometrisk sier trekantulikheten at lengden til den ene siden i en trekant alltid er mindre enn summen av de to andre sidene. Resultatet ovenfor forteller oss at dette også gjelder i høyere dimensjoner. Faktisk spiller trekantulikheten en nøkkelrolle i de fleste forsøk på å generalisere avstandsbegrepet til nye sammenhenger. I dette heftet skal vi ha stor glede av trekantulikheten når vi studerer funksjoner av flere variable.

La oss avslutte denne seksjonen med å se på hvordan vi kan generalisere begrepet *linje* til  $\mathbb{R}^n$ . Vi starter i planet. Figur 7 viser en rett linje gjennom punktet  $\mathbf{a}$  parallel med vektoren  $\mathbf{b}$ .

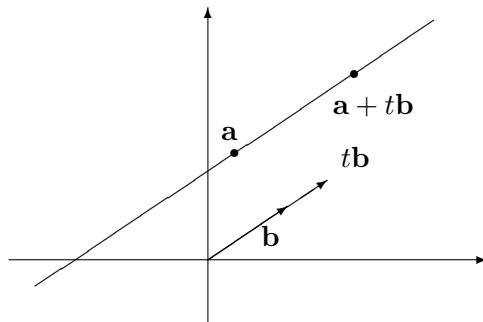
Figur 7: Rett linje gjennom  $\mathbf{a}$  parallel med  $\mathbf{b}$ 

Siden enhver vektor  $t\mathbf{b}$  er parallel med  $\mathbf{b}$ , ser vi at alle punkter av typen  $\mathbf{a} + t\mathbf{b}$  må ligge på linjen (se figur 8). Det er heller ikke så vanskelig å overbevise seg om at ethvert punkt på linjen må være av formen  $\mathbf{a} + t\mathbf{b}$  for ett eller annet tall  $t$ .

Vi har dermed kommet frem til at de punktene som ligger på den rette linjen, er nøyaktig de som er av typen  $\mathbf{a} + t\mathbf{b}$  for et reelt tall  $t$ . Vi skriver gjerne

$$\mathbf{r}(t) = \mathbf{a} + t\mathbf{b}$$

og tenker på  $\mathbf{r}(t)$  som et punkt som beveger seg langs linjen når  $t$  endrer seg.



Figur 8: Parameterfremstilling av en rett linje

Det er nå lett å generalisere begrepet rett linje til  $\mathbb{R}^n$ . Hvis  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ ,  $\mathbf{b} \neq \mathbf{0}$ , så består den rette linjen gjennom punktet  $\mathbf{a}$  og med retningsvektor  $\mathbf{b}$  av alle punkter på formen

$$\mathbf{r}(t) = \mathbf{a} + t\mathbf{b}$$

Bruker vi koordinater, ser vi at hvis  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  og  $\mathbf{b} = (b_1, b_2, \dots, b_n)$ , så blir

$$\mathbf{r}(t) = \mathbf{a} + t\mathbf{b} = (a_1 + tb_1, a_2 + tb_2, \dots, a_n + tb_n)$$

**Eksempel 2:** Finn en parameterfremstilling til linjen gjennom punktet  $\mathbf{a} = (1, 2, 0, -1)$  med retningsvektor  $\mathbf{b} = (-1, 2, -3, 1)$ , og avgjør om punktet  $\mathbf{c} = (2, -1, 1, 4)$  ligger på linjen.

Parameterfremstillingen er

$$\mathbf{r}(t) = \mathbf{a} + t\mathbf{b} = (1 - t, 2 + 2t, -3t, -1 + t)$$

Skal punktet  $\mathbf{c}$  ligge på linjen, må det finnes et tall  $t$  slik at  $\mathbf{r}(t) = \mathbf{c}$ , dvs. at følgende ligninger må være oppfylt:

$$1 - t = 2, \quad 2 + 2t = -1, \quad -3t = 1, \quad -1 + t = 4$$

Siden det ikke finnes noe tall  $t$  som oppfyller alle disse ligningene, ligger ikke  $\mathbf{c}$  på linjen. ♣

La oss avslutte denne seksjonen med et begrep som først vil spille en sentral rolle i senere kapitler, men som det kan være greit å vite om allerede nå. Anta at vi har vektorer  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  i  $\mathbb{R}^n$ . Vi sier at vektoren  $\mathbf{v} \in \mathbb{R}^n$  er en *lineærkombinasjon* av  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  dersom det finnes tall  $s_1, s_2, \dots, s_k$  slik at

$$\mathbf{v} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k$$

**MATLAB-kommentar:** MATLAB har en egen kommando for å regne ut lengden (eller normen) til en vektor  $\mathbf{a}$ . Vi skriver `>>norm(a)`.

## Oppgaver til seksjon 1.2

1. Finn skalarproduktet av  $(-2, 3)$  og  $(4, 1)$ . Finn også vinkelen mellom vektorene.
2.  $|\mathbf{a}| = 4$ ,  $|\mathbf{b}| = 5$  og vinkelen mellom  $\mathbf{a}$  og  $\mathbf{b}$  er  $45^\circ$ . Finn  $\mathbf{a} \cdot \mathbf{b}$ .
3. Finn vinkelen mellom vektorene  $(1, 2, 3)$  og  $(-1, 0, 1)$ .
4. Regn ut vinkelen mellom  $(-1, 2, 6, 2, 4)$  og  $(1, 0, 3, 1, 1)$ .
5. Finn vinkelen mellom vektorene  $\mathbf{a} = (4, 3, 1, 2)$  og  $\mathbf{b} = (-1, 3, 2, 0)$ . Finn også projeksjonen av  $\mathbf{a}$  ned på  $\mathbf{b}$ .
6. Hvor lang er projeksjonen av  $(-3, 4, 2, 5)$  ned på  $(0, 3, 1, 2)$ ?

**7.** Skriv  $\mathbf{a} = (4, 3)$  som en sum av to vektorer  $\mathbf{b}$  og  $\mathbf{c}$  der  $\mathbf{b}$  er parallel med  $\mathbf{d} = (1, 2)$  og  $\mathbf{c}$  står normalt på  $\mathbf{d}$ .

**8.** Skriv  $\mathbf{a} = (2, 2, 1)$  som en sum av to vektorer  $\mathbf{b}$  og  $\mathbf{c}$  der  $\mathbf{b}$  er parallel med  $\mathbf{d} = (1, 0, -1)$  og  $\mathbf{c}$  står normalt på  $\mathbf{d}$

**9.** Finn vinkelen som hver av vektorene  $\mathbf{a} = (\sqrt{3}, 1)$  og  $\mathbf{b} = (1, 1)$  danner med  $x$ -aksen. Regn ut  $\mathbf{a} \cdot \mathbf{b}$  og bruk svaret til å finne et eksakt uttrykk for  $\cos(15^\circ)$ .

**10.** Finn to vektorer som begge står normalt på  $(3, 2, -1)$  og som ikke er parallelle.

**11.** Vis at dersom  $\mathbf{a}$  står normalt på både  $\mathbf{b}$  og  $\mathbf{c}$ , så står  $\mathbf{a}$  normalt på  $\mathbf{b} + \mathbf{c}$ .

**12.** I denne oppgaver er  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ .

a) Vis at  $|\mathbf{a} + \mathbf{b}|^2 = |\mathbf{a}|^2 + 2 \mathbf{a} \cdot \mathbf{b} + |\mathbf{b}|^2$ .

b) Finn  $\mathbf{a} \cdot \mathbf{b}$  når  $|\mathbf{a}| = 6$ ,  $|\mathbf{b}| = 4$  og  $|\mathbf{a} + \mathbf{b}| = 3$ ,

c) Anta at  $|\mathbf{c}| = 3$ ,  $|\mathbf{d}| = 4$  og  $|\mathbf{c} + \mathbf{d}| = 5$ . Finn vinkelen mellom  $\mathbf{c}$  og  $\mathbf{d}$ .

**13.** Per påstår at han har to vektorer  $\mathbf{a}$  og  $\mathbf{b}$  slik at  $|\mathbf{a}| = 3$ ,  $|\mathbf{b}| = 2$  og  $|\mathbf{a} + \mathbf{b}| = 7$ . Hvorfor tror du ikke på ham?

**14.** Kari påstår at hun har to vektorer  $\mathbf{a}$  og  $\mathbf{b}$  slik at  $|\mathbf{a}| = 7$ ,  $|\mathbf{b}| = 2$  og  $\mathbf{a} \cdot \mathbf{b} = -16$ . Hvorfor tror du ikke på henne?

**15.** Vis at for alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  er  $|\mathbf{x}| - |\mathbf{y}| \leq |\mathbf{x} - \mathbf{y}|$ . Vis også at  $|\mathbf{y}| - |\mathbf{x}| \leq |\mathbf{x} - \mathbf{y}|$ , og konkluder med at  $||\mathbf{x}| - |\mathbf{y}|| \leq |\mathbf{x} - \mathbf{y}|$ .

**16.** Avstanden  $d(\mathbf{a}, \mathbf{b})$  mellom to punkter  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$  er lik lengden til vektoren som forbinder dem, dvs.  $d(\mathbf{a}, \mathbf{b}) = |\mathbf{b} - \mathbf{a}|$ . Bevis at  $d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b})$  for alle vektorer  $\mathbf{a}, \mathbf{b}, \mathbf{c}$ . Hva er den geometriske tolkningen av denne ulikheten?

**17.** Vis at for alle vektorer  $\mathbf{x}$  og  $\mathbf{y}$  gjelder  $|\mathbf{x} + \mathbf{y}|^2 + |\mathbf{x} - \mathbf{y}|^2 = 2|\mathbf{x}|^2 + 2|\mathbf{y}|^2$ . Vis at i et parallellogram er summen av kvadratene av sidene lik summen av kvadratene av diagonalene.

**18.** Finn en parameterfremstilling av linjen som går gjennom punktet  $(-1, -1, 2)$  og er parallel med  $(2, 3, 1)$ .

**19.** Finn en parameterfremstilling av linjen gjennom  $(-3, -2, 5, 8)$  parallel med  $(1, -2, -1, 3)$ . Sjekk om punktet  $(1, -6, 3, 14)$  ligger på linjen.

**20.** Finn en parameterfremstilling av linjen som går gjennom punktene  $(2, -1, 3)$  og  $(3, 8, -2)$ .

**21.** Finn en parameterfremstilling av linjen som går gjennom punktene  $(7, -3, 2, 4, -2)$  og  $(2, 1, -1, -1, 5)$ .

**22.** Finn en parameterfremstilling for linjen som går gjennom  $(5, -2)$  og som står normalt på  $(-1, 2)$ .

**23.** Finn en parameterfremstilling for linjen i planetet som har ligning  $2x + 3y = 6$ .

**24.** En linje i planetet har parameterfremstilling  $(-3 + 2t, 2 - t)$ . Finn en ligning av typen  $y = ax + b$  for denne linjen.

**25.** To skip er på kryssende kurs. Ved tiden  $t = 0$  er det ene skipet i punktet  $(0, 4)$ , og det andre skipet i punktet  $(39, 14)$  (alle avstander er målt i nautiske mil.) Det første skipet beveger seg parallelt med vektoren  $(3, 4)$  med en fart av 15 knop (1 knop = 1 nautisk mil per time). Det andre skipet beveger seg parallelt med vektoren  $(-12, 5)$  med en fart av 13 knop.

a) Hvor vil kursene krysse hverandre?

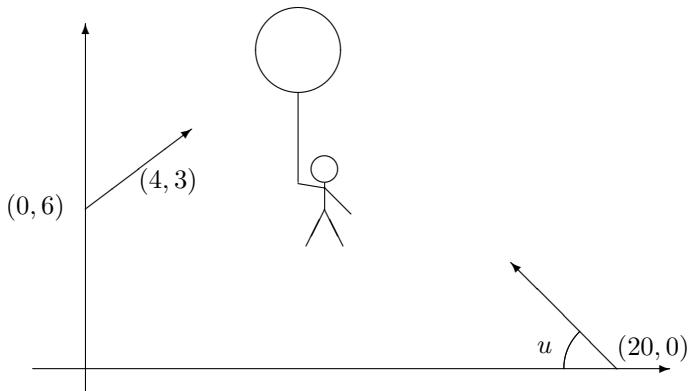
b) Vil skipene kolidere?

**26.** To fly er i det samme området. Ved tiden  $t = 0$  er det ene flyet i punktet  $(0, 0, 2000)$  og flyr med en fart på 150m/s parallelt med vektoren  $(2, 2, 1)$ . Det andre flyet er ved tiden  $t = 0$  i punktet  $(5000, -1000, 4000)$  og 20 sekunder senere i punktet  $(4400, 2000, 4000)$ . Flyet følger en rett linje og holder konstant hastighet.

a) Vil kursene til de to flyene skjære hverandre?

b) Vil flyene kolidere?

**27.** I sin evige jakt etter honning forsøker Ole Brumm å invadere et tre ved hjelp av en ballong. Plutselig blir ballongen tatt av et vindkast og farer av sted med Ole Brumm. Etter å ha tenkt seg om et øyeblikk, innser Kristoffer Robin at hans eneste sjanse til å redde vennen er å skyte istykker ballongen med lekegeværet sitt. Figuren nedenfor viser en skisse av situasjonen.



Når vindkastet kommer ved tiden  $t = 0$ , befinner ballongen seg i punktet  $(0, 6)$ . Den blir ført av gårde med en fart av 5m/s i retningen  $(4, 3)$ . Ved tiden  $t = 2$  skyter Kristoffer Robin mot ballongen fra sin posisjon  $(20, 0)$ . Vinkelen mellom geværet og underlaget er  $u$ , og vi regner med at kulen beveger seg rettlinjet med en fart av 70m/s. Alle avstander er målt i meter og tiden er målt i sekunder.

a) Forklar at ballongens posisjon ved tiden  $t$  er  $(4t, 6 + 3t)$ .

- b) Vis at kulens posisjon ved tiden  $t$  er  $(20 - 70(t-2)\cos u, 70(t-2)\sin u)$ .  
c) Hvilken vinkel  $u$  må Kristoffer Robin holde geværet i for å treffe midt i ballongen? Hvor langt er det ned til bakken når ballongen blir truffet?

**28.** I denne oppgaven skal vi se på et annet bevis for Schwarz' ulikhet.

- a) Vis at for alle  $a, b \in \mathbb{R}$  og alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  er

$$0 \leq |a\mathbf{x} \pm b\mathbf{y}|^2 = a^2|\mathbf{x}|^2 \pm 2ab \mathbf{x} \cdot \mathbf{y} + b^2|\mathbf{y}|^2$$

- b) Velg  $a = |\mathbf{y}|$ ,  $b = |\mathbf{x}|$  i ulikhettene ovenfor og utled Schwarz' ulikhet.

### 1.3 Komplekse $n$ -tupler

Hittil har vi bare sett på  $n$ -tupler  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  der komponentene  $a_1, a_2, \dots, a_n$  er reelle tall. Vi skal nå ta en rask titt på det komplekse tilfellet. Som nevnt tidligere kalles mengden av alle komplekse  $n$ -tupler for  $\mathbb{C}^n$ . Addisjon og subtraksjon av komplekse  $n$ -tupler foregår komponentvis akkurat som i det reelle tilfellet. Også multiplikasjon med skalar (som nå godt kan være kompleks) foregår akkurat som før. La oss se på et eksempel:

**Eksempel 1:** Regn ut  $s\mathbf{a} + t\mathbf{b}$  når  $s = 1 + i$ ,  $t = i$ ,  $\mathbf{a} = \begin{pmatrix} 2 \\ i \end{pmatrix}$ ,  $\mathbf{b} = \begin{pmatrix} 1-i \\ 1+i \end{pmatrix}$ . Vi får (husk at  $i^2 = -1$ ):

$$\begin{aligned} s\mathbf{a} + t\mathbf{b} &= (1+i) \begin{pmatrix} 2 \\ i \end{pmatrix} + i \begin{pmatrix} 1-i \\ 1+i \end{pmatrix} = \begin{pmatrix} (1+i)2 \\ (1+i)i \end{pmatrix} + \begin{pmatrix} i(1-i) \\ i(1+i) \end{pmatrix} = \\ &= \begin{pmatrix} 2+2i \\ i-1 \end{pmatrix} + \begin{pmatrix} i+1 \\ i-1 \end{pmatrix} = \begin{pmatrix} 3+3i \\ -2+2i \end{pmatrix} \end{aligned}$$



Når vi skal definere *normen* (dvs. *lengden*) til et komplekst  $n$ -tuppel, må vi være litt forsiktig. Dersom  $\mathbf{a} = (z_1, z_2, \dots, z_n)$ , setter vi

$$|\mathbf{a}| = \sqrt{|z_1|^2 + |z_2|^2 + \dots + |z_n|^2}$$

Vi tar altså tallverdien til komponentene før vi kvadrerer dem. Grunnen til dette er at kvadratet til et komplekst tall ikke behøver å være positivt og derfor gir et dårlig mål på størrelse.

**Eksempel 2:** Finn normen til vektoren  $\mathbf{a} = (2+i, 4+i, 1-3i)$ . Husk at dersom  $z = a+ib$ , så er  $|z| = \sqrt{a^2+b^2}$  og dermed  $|z|^2 = a^2+b^2$ . Dette gir

$$|\mathbf{a}| = \sqrt{|2+i|^2 + |4+i|^2 + |1-3i|^2} =$$

$$\begin{aligned}
&= \sqrt{(2^2 + 1^2) + (4^2 + 1^2) + (1^2 + (-3)^2)} = \\
&= \sqrt{4 + 1 + 16 + 1 + 1 + 9} = \sqrt{32} = 4\sqrt{2}
\end{aligned}$$

Lengden til vektoren er altså  $4\sqrt{2}$ . ♣

For å få til det riktige samsillet mellom normen og skalarproduktet, må vi også gjøre en liten justering i definisjonen av skalarprodukt. Dersom  $\mathbf{a} = (z_1, z_2, \dots, z_n)$  og  $\mathbf{b} = (w_1, w_2, \dots, w_n)$ , definerer vi

$$\mathbf{a} \cdot \mathbf{b} = z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}$$

Vi komplekskonjugerer altså den andre faktoren i skalarproduktet. Siden  $z\overline{z} = |z|^2$ , ser vi at

$$\mathbf{a} \cdot \mathbf{a} = z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 = |\mathbf{a}|^2$$

Vi har altså den vanlige sammenhengen mellom norm og skalarprodukt:

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

La oss ta med et eksempel på hvordan man regner ut et komplekst skalarprodukt.

**Eksempel 3:** Regn ut  $\mathbf{a} \cdot \mathbf{b}$  når  $\mathbf{a} = (1+i, -2, 1+3i)$  og  $\mathbf{b} = (2+2i, 1-2i, 3+4i)$ . Vi får

$$\begin{aligned}
\mathbf{a} \cdot \mathbf{b} &= (1+i)(2-2i) + (-2)(1+2i) + (1+3i)(3-4i) = \\
&= 2-2i+2i+2-2-4i+3-4i+9i+12 = 17+i
\end{aligned}$$

Skalarproduktet av to komplekse vektorer er altså et komplekst tall. ♣

På grunn av komplekskonjugasjonen i annen faktor, har vi ikke lenger at  $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$  for komplekse vektorer (den *kommutative lov* gjelder altså ikke for komplekse skalarprodukter). Bruker vi regnereglene for konjugasjon (se *Kalkulus*, setning 3.1.5), ser vi imidlertid at

$$\begin{aligned}
\overline{\mathbf{b} \cdot \mathbf{a}} &= \overline{w_1\overline{z_1} + w_2\overline{z_2} + \cdots + w_n\overline{z_n}} = \\
&= \overline{w_1}\overline{\overline{z_1}} + \overline{w_2}\overline{\overline{z_2}} + \cdots + \overline{w_n}\overline{\overline{z_n}} = \\
&= \overline{w_1}z_1 + \overline{w_2}z_2 + \cdots + \overline{w_n}z_n = \\
&= z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n} = \mathbf{a} \cdot \mathbf{b}
\end{aligned}$$

Bytter vi om på faktorenes rekkefølge, så komplekskonjugerer vi altså resultatet!

Siden vi er så vant til at faktorenes rekkefølge ikke spiller noen rolle, er det lett å bli lurt av det komplekse skalarproduktet. Dersom vi skal regne

ut  $(\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b})$ , er det f.eks. fristende å bruke første kvadratsetning til å skrive svaret

$$\mathbf{a} \cdot \mathbf{a} + 2\mathbf{a} \cdot \mathbf{b} + \mathbf{b} \cdot \mathbf{b},$$

men dette blir ikke riktig! Ganger vi nemlig ut parentesene litt forsiktig, ser vi at

$$(\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) = \mathbf{a} \cdot \mathbf{a} + \mathbf{a} \cdot \mathbf{b} + \mathbf{b} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b}$$

og siden  $\mathbf{a} \cdot \mathbf{b}$  og  $\mathbf{b} \cdot \mathbf{a}$  ikke er like, kan vi ikke slå sammen de to midterste leddene på vanlig måte. Siden tallene  $\mathbf{a} \cdot \mathbf{b}$  og  $\mathbf{b} \cdot \mathbf{a}$  er komplekskonjugerte, har vi imidlertid at  $\mathbf{a} \cdot \mathbf{b} + \mathbf{b} \cdot \mathbf{a} = 2 \operatorname{Re}(\mathbf{a} \cdot \mathbf{b})$  (der  $\operatorname{Re}$  står for realdel), og dermed kan vi skrive

$$\begin{aligned} (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) &= \mathbf{a} \cdot \mathbf{a} + 2 \operatorname{Re}(\mathbf{a} \cdot \mathbf{b}) + \mathbf{b} \cdot \mathbf{b} = \\ &= |\mathbf{a}|^2 + 2 \operatorname{Re}(\mathbf{a} \cdot \mathbf{b}) + |\mathbf{b}|^2 \end{aligned}$$

Dette regnestykket viser at vi må være litt forsiktige når vi overfører standard regneprosedyrer til komplekse skalarprodukt.

La oss skrive opp de grunnleggende regnereglene for komplekse  $n$ -tupler (sammenlign med setning 1.1.1 for reelle  $n$ -tupler):

**Setning 1.3.1 (Regneregler for komplekse  $n$ -tupler.)** Dersom  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$  er  $n$ -tupler, og  $s$  og  $t$  er komplekse tall, gjelder følgende regneregler:

- (a)  $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$
- (b)  $\mathbf{a} \cdot \mathbf{b} = \overline{\mathbf{b} \cdot \mathbf{a}}$
- (c)  $s(\mathbf{a} + \mathbf{b}) = s\mathbf{a} + s\mathbf{b}$
- (d)  $(s + t)\mathbf{a} = s\mathbf{a} + t\mathbf{a}$
- (e)  $\mathbf{c} \cdot (\mathbf{a} + \mathbf{b}) = \mathbf{c} \cdot \mathbf{a} + \mathbf{c} \cdot \mathbf{b}$  og  $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$
- (f)  $(s\mathbf{a}) \cdot \mathbf{b} = s(\mathbf{a} \cdot \mathbf{b})$  og  $\mathbf{a} \cdot (s\mathbf{b}) = \overline{s}(\mathbf{a} \cdot \mathbf{b})$
- (g)  $\mathbf{a} \cdot \mathbf{a} \geq 0$  med likhet hvis og bare hvis  $\mathbf{a} = \mathbf{0}$

*Bevis:* Med unntak av b) (som vi nettopp har bevist) og andre del av f), er dette nøyaktig de samme reglene som i setning 1.1.1, og bevisene er også de samme. Vi nøyer oss derfor med å vise andre del av f). Hvis  $\mathbf{a} = (z_1, z_2, \dots, z_n)$  og  $\mathbf{b} = (w_1, w_2, \dots, w_n)$ , ser vi at  $s\mathbf{b} = (sw_1, sw_2, \dots, sw_n)$ , og dermed er

$$\begin{aligned} \mathbf{a} \cdot (s\mathbf{b}) &= z_1 \overline{sw_1} + z_2 \overline{sw_2} + \cdots + z_n \overline{sw_n} = \\ &= \overline{s}(z_1 \overline{w_1} + z_2 \overline{w_2} + \cdots + z_n \overline{w_n}) = \overline{s}(\mathbf{a} \cdot \mathbf{b}) \end{aligned} \quad \square$$

Vi kan innføre geometriske begreper for komplekse  $n$ -tupler akkurat som for reelle selv om visualiseringen blir enda vanskeligere i dette tilfellet. Vi sier f.eks. at to vektorer  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^n$  er *ortogonale* dersom  $\mathbf{a} \cdot \mathbf{b} = 0$ . Legg merke til at dette medfører at  $\mathbf{b} \cdot \mathbf{a} = 0$  siden  $\mathbf{b} \cdot \mathbf{a} = \overline{\mathbf{a} \cdot \mathbf{b}} = \overline{0} = 0$ . Argumentkjeden som ga oss Pythagoras' setning, Schwarz' ulikhet og trekantulikheten i forrige seksjon, fungerer med små justeringer også i det komplekse tilfellet, og vi nøyer oss med å skrive opp resultatet:

**Setning 1.3.2** *For komplekse vektorer  $\mathbf{a}, \mathbf{b}$  gjelder:*

- (i) (*Pythagoras' setning*) Dersom  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^n$  er ortogonale, så er  $|\mathbf{a}|^2 + |\mathbf{b}|^2 = |\mathbf{a} + \mathbf{b}|^2$
- (ii) (*Schwarz' ulikhet*) For alle  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^n$  er  $|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}| |\mathbf{b}|$
- (iii) (*Trekantulikheten*) For alle  $\mathbf{a}, \mathbf{b} \in \mathbb{C}^n$  er  $|\mathbf{a} + \mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}|$

□

### Oppgaver til seksjon 1.3

1. Regn ut  $s\mathbf{x} + t\mathbf{y}$  når  $s = i$ ,  $t = 1 + 2i$ ,  $\mathbf{x} = \begin{pmatrix} -4i \\ 2-i \end{pmatrix}$  og  $\mathbf{y} = \begin{pmatrix} 2+i \\ 2i \end{pmatrix}$ .
2. Finn lengden til vektorene  $\mathbf{a} = (3 + 2i, -1 + i)$  og  $\mathbf{b} = (i, 2 + 3i, -2 - i)$ .
3. Regn ut skalarproduktet  $\mathbf{x} \cdot \mathbf{y}$  når  $\mathbf{x} = (1+3i, -2i, 2+3i)$  og  $\mathbf{y} = (2, 1+2i, -1+i)$ .
4. Vis at for alle  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$  er

$$|\mathbf{x} - \mathbf{y}|^2 = |\mathbf{x}|^2 - 2 \operatorname{Re}(\mathbf{x} \cdot \mathbf{y}) + |\mathbf{y}|^2$$

og

$$(\mathbf{x} + \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) = |\mathbf{x}|^2 - 2 \operatorname{Im}(\mathbf{x} \cdot \mathbf{y}) - |\mathbf{y}|^2$$

hvor  $\operatorname{Re}(z)$  og  $\operatorname{Im}(z)$  betegner hhv. realdelen og imaginærdelen til  $z$ .

5. Bevis setning 1.3.2 ved å gå gjennom beviset for de tilsvarende resultatene i seksjon 1.2 og se hvilke modifikasjoner som må gjøres.

## 1.4 Vektorproduktet

De regneoperasjonene vi hittil har sett på, er definert for vektorer av alle dimensjoner. I denne seksjonen skal vi se studere en operasjon — *vektorproduktet* — som bare er definert for tredimensjonale vektorer. Siden 3 er den fysiske romdimensjonen, brukes vektorproduktet ofte i geometriske problemstillinger. Det brukes også mye i fysikk og mekanikk — tok du fysikk i videregående skole, har du sikkert støtt på "høyrehåndsregler" i en del

sammenhenger, og bak enhver slik høyrehåndsregel skjuler det seg et vektorprodukt.

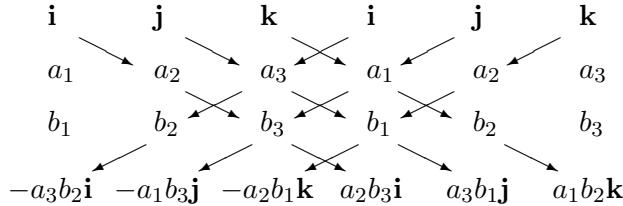
Det er to måter å definere vektorproduktet på, en geometrisk og en algebraisk, og det er samspillet mellom disse to betraktningsmåtene som gir vektorproduktet slagkraft. Vi skal ta utgangspunkt i den algebraiske definisjonen. Før vi begynner, minner vi om at man i tre dimensjoner gjerne skriver enhetsvektorene langs aksene på denne måten:

$$\mathbf{i} = (1, 0, 0), \quad \mathbf{j} = (0, 1, 0), \quad \mathbf{k} = (0, 0, 1).$$

Gitt to vektorer  $\mathbf{a} = (a_1, a_2, a_3)$  og  $\mathbf{b} = (b_1, b_2, b_3)$  i  $\mathbb{R}^3$  definerer vi nå vektorproduktet (også kalt *kryssproduktet*)  $\mathbf{a} \times \mathbf{b}$  ved:

$$\begin{aligned}\mathbf{a} \times \mathbf{b} &= (a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1) \\ &= (a_2 b_3 - a_3 b_2)\mathbf{i} + (a_3 b_1 - a_1 b_3)\mathbf{j} + (a_1 b_2 - a_2 b_1)\mathbf{k}\end{aligned}$$

Denne formelen kan være vanskelig å huske, men det finnes huskeregler. Én slik regel er vist i skjemaet nedenfor. Vi multipliserer langs pilene og gir resultatet positiv verdi dersom pilene går fra venstre mot høyre og negativ verdi dersom de går fra høyre mot venstre (kjennen du en annen huskeregel fra før, kan du trygt bruke den).



Figur 1: Huskeregel for vektorproduktet

La oss regne ut et vektorprodukt:

**Eksempel 1:** Finn vektorproduktet av  $\mathbf{a} = (3, -1, 2)$  og  $\mathbf{b} = (4, -2, 5)$ . Vi får:

$$\begin{aligned}\mathbf{a} \times \mathbf{b} &= ((-1) \cdot 5 - 2 \cdot (-2))\mathbf{i} + (2 \cdot 4 - 3 \cdot 5)\mathbf{j} + (3 \cdot (-2) - (-1) \cdot 4)\mathbf{k} \\ &= (-1, -7, -2)\end{aligned}$$



La oss så se hva som skjer dersom vi regner ut  $\mathbf{b} \times \mathbf{a}$  istedenfor  $\mathbf{a} \times \mathbf{b}$ :

$$\begin{aligned}\mathbf{b} \times \mathbf{a} &= (b_2 a_3 - b_3 a_2, b_3 a_1 - b_1 a_3, b_1 a_2 - b_2 a_1) \\ &= -(a_2 b_3 - a_3 b_2, a_3 b_1 - a_1 b_3, a_1 b_2 - a_2 b_1) = -(\mathbf{a} \times \mathbf{b})\end{aligned}$$

Akkurat som skalarproduktet for komplekse vektorer er altså vektorproduktet ikke-kommutativt, men vi har en formel som gjør at vi lett kan regne ut  $\mathbf{b} \times \mathbf{a}$  når vi kjenner  $\mathbf{a} \times \mathbf{b}$ .

Her er en liste over de grunnleggende egenskapene til vektorproduktet:

**Setning 1.4.1** For vektorer  $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$  gjelder:

- (a)  $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$
- (b)  $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$  og  $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$
- (c)  $\mathbf{a} \times (s\mathbf{b}) = s(\mathbf{a} \times \mathbf{b})$  og  $(s\mathbf{a}) \times \mathbf{b} = s(\mathbf{a} \times \mathbf{b})$  der  $s \in \mathbf{R}$
- (d)  $\mathbf{a} \times \mathbf{b}$  står ortogonalt på både  $\mathbf{a}$  og  $\mathbf{b}$
- (e) (Lagranges identitet)  $|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2$

*Bevis:* Punkt a) har vi allerede bevist og de andre bevisene er av samme type — vi skriver vektorene på koordinatform, regner ut og ser at det stemmer. Vi tar c), d) og e) som eksempler:

c) Hvis  $\mathbf{a} = (a_1, a_2, a_3)$  og  $\mathbf{b} = (b_1, b_2, b_3)$ , så er  $s\mathbf{a} = (sa_1, sa_2, sa_3)$ . Vi får:

$$\begin{aligned} (s\mathbf{a}) \times \mathbf{b} &= ((sa_2)b_3 - (sa_3)b_2, (sa_3)b_1 - (sa_1)b_3, (sa_1)b_2 - (sa_2)b_1) \\ &= s(a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1) = s(\mathbf{a} \times \mathbf{b}) \end{aligned}$$

Den andre likheten i c) går på samme måte.

d) For å vise at  $\mathbf{a}$  står ortogonalt på  $\mathbf{a} \times \mathbf{b}$ , må vi vise at  $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0$ . Vi får:

$$\begin{aligned} \mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) &= a_1(a_2b_3 - a_3b_2) + a_2(a_3b_1 - a_1b_3) + a_3(a_1b_2 - a_2b_1) \\ &= a_1a_2b_3 - a_1a_3b_2 + a_2a_3b_1 - a_2a_1b_3 + a_3a_1b_2 - a_3a_2b_1 = 0 \end{aligned}$$

En helt tilsvarende regning viser at  $\mathbf{b}$  står ortogonalt på  $\mathbf{a} \times \mathbf{b}$ .

e) Vi skriver  $\mathbf{a} = (a_1, a_2, a_3)$ ,  $\mathbf{b} = (b_1, b_2, b_3)$  og regner ut begge sider (du er ikke forpliktet til å føle at dette er spesielt festlig):

$$\begin{aligned} |(\mathbf{a} \times \mathbf{b})|^2 &= (a_2b_3 - a_3b_2)^2 + (a_3b_1 - a_1b_3)^2 + (a_1b_2 - a_2b_1)^2 \\ &= a_2^2b_3^2 - 2a_2a_3b_2b_3 + a_3^2b_2^2 + a_3^2b_1^2 - 2a_1a_3b_1b_3 + a_1^2b_3^2 + a_1^2b_2^2 - 2a_1a_2b_1b_2 + a_2^2b_1^2 \\ \text{og} \\ |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2 &= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\ &= a_1^2b_1^2 + a_1^2b_2^2 + a_1^2b_3^2 + a_2^2b_1^2 + a_2^2b_2^2 + a_2^2b_3^2 + a_3^2b_1^2 + a_3^2b_2^2 + a_3^2b_3^2 \\ &\quad - a_1^2b_1^2 - a_2^2b_2^2 - a_3^2b_3^2 - 2a_1a_2b_1b_2 - 2a_1a_3b_1b_3 - 2a_2a_3b_2b_3 \end{aligned}$$

$$= a_1^2 b_2^2 + a_1^2 b_3^2 + a_2^2 b_1^2 + a_2^2 b_3^2 + a_3^2 b_1^2 + a_3^2 b_2^2 - 2a_1 a_2 b_1 b_2 - 2a_1 a_3 b_1 b_3 - 2a_2 a_3 b_2 b_3$$

Bortsett fra rekkefølgen på leddene er dette det samme uttrykket som vi fikk ovenfor. Dermed er e) bevist.  $\square$

**Bemerkning:** Legg merke til at det ikke finnes noen assosiativ lov i listen ovenfor — generelt er nemlig  $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} \neq \mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ . Som et eksempel lar vi  $\mathbf{a} = (1, 1, 0)$ ,  $\mathbf{b} = (1, 0, 0)$  og  $\mathbf{c} = (0, 0, 1)$ . Da er

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = ((1, 1, 0) \times (1, 0, 0)) \times (0, 0, 1) = (0, 0, -1) \times (0, 0, 1) = \mathbf{0}$$

mens

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (1, 1, 0) \times ((1, 0, 0) \times (0, 0, 1)) = (1, 1, 0) \times (0, -1, 0) = (0, 0, -1)$$

At  $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} \neq \mathbf{a} \times (\mathbf{b} \times \mathbf{c})$  betyr at uttrykket  $\mathbf{a} \times \mathbf{b} \times \mathbf{c}$  ikke gir noen mening — vi må ha med parenteser for å presisere hvilken rekkefølge produktene skal utføres i.

Som allerede nevnt, er det også en geometrisk måte å beskrive vektorproduktet på. For å finne frem til denne geometriske beskrivelsen, tar vi utgangspunkt i punkt e) i setningen ovenfor:

$$|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2$$

Siden vi allerede vet at  $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}| |\mathbf{b}| \cos v$ , der  $v$  er vinkelen mellom  $\mathbf{a}$  og  $\mathbf{b}$ , så er

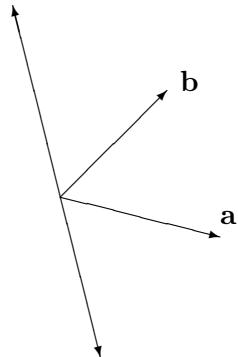
$$|\mathbf{a} \times \mathbf{b}|^2 = |\mathbf{a}|^2 |\mathbf{b}|^2 - |\mathbf{a}|^2 |\mathbf{b}|^2 \cos^2 v = |\mathbf{a}|^2 |\mathbf{b}|^2 \sin^2 v$$

der vi har benyttet at  $1 - \cos^2 v = \sin^2 v$ . Altså er

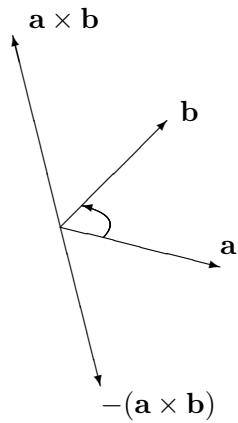
$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| |\mathbf{b}| \sin v$$

(husk at siden  $0^\circ \leq v \leq 180^\circ$ , er  $\sin v$  aldri negativ). Dermed vet vi hvor lang vektoren  $\mathbf{a} \times \mathbf{b}$  er. Legg spesielt merke til at  $\mathbf{a} \times \mathbf{b} = \mathbf{0}$  hvis og bare hvis  $\sin v = 0$ , dvs. dersom  $\mathbf{a}$  og  $\mathbf{b}$  er parallelle.

Fra punkt d) i setning 1.4.1 vet vi også noe om retningen til  $\mathbf{a} \times \mathbf{b}$ , nemlig at  $\mathbf{a} \times \mathbf{b}$  står normalt på både  $\mathbf{a}$  og  $\mathbf{b}$ . Nå finnes det to motsatt rettede vektorer som har lengde  $|\mathbf{a}| |\mathbf{b}| \sin v$  og står normalt på både  $\mathbf{a}$  og  $\mathbf{b}$  (se figur 2). For å vite hvilken av disse to vekatorene som er  $\mathbf{a} \times \mathbf{b}$ , bruker vi *høyrehåndsregelen*:

Figur 2: To like lange vektorer som står normalt på både  $\mathbf{a}$  og  $\mathbf{b}$ 

Vi legger høyre hånd med fingrene pekende den korteste veien fra  $\mathbf{a}$  til  $\mathbf{b}$  mens vi spriker med tommelen. Da er  $\mathbf{a} \times \mathbf{b}$  den av de to normalvektorene som peker i tommelens retning (se figur 3 der den krumme pilen viser den retningen fingrene peker). Legg merke til at når vi regner ut  $\mathbf{b} \times \mathbf{a}$ , skal fingrene spenne over den samme vinkelen, men i motsatt retning (fra  $\mathbf{b}$  mot  $\mathbf{a}$ ), og tommelen kommer derfor til å peke motsatt vei. Dette er den geometriske forklaringen på regelen  $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$ .

Figur 3: Vektorene  $\mathbf{a} \times \mathbf{b}$  og  $-(\mathbf{a} \times \mathbf{b})$ 

La oss oppsummere det vi har kommet frem til:

**Setning 1.4.2** La  $\mathbf{a}$  og  $\mathbf{b}$  være to vektorer i  $\mathbb{R}^3$  og kall vinkelen mellom dem  $v$ . Da har vektorproduktet  $\mathbf{a} \times \mathbf{b}$  lengde  $|\mathbf{a}||\mathbf{b}| \sin v$  og står normalt på både  $\mathbf{a}$  og  $\mathbf{b}$ . Retningen til  $\mathbf{a} \times \mathbf{b}$  er gitt ved høyrehåndsregelen.

\**Bevis:* Vi har bevist alt bortsett fra høyrehåndsregelen. Beviset vi skal gi for denne regelen kan se litt umatematisk og skissemessig (og vanskelig!) ut,

men det kan uten store endringer bygges ut til et fullverdig bevis. Du kan godt hoppe over dette beviset uten å få problemer med det som kommer senere. Legg merke til at hvis  $\mathbf{a}$  og  $\mathbf{b}$  er parallelle, så er det ingen ting å bevise siden  $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ . Vi ser derfor på tilfellet der  $\mathbf{a}$  og  $\mathbf{b}$  ikke er parallelle.

Vi skal først bevise høyrehåndsregelen for det spesialtilfellet der  $\mathbf{a}$  og  $\mathbf{b}$  ligger i  $xy$ -planet, og  $\mathbf{a}$  peker langs den positive  $x$ -aksen. Det betyr at  $\mathbf{a}$  har koordinater  $\mathbf{a} = (a_1, 0, 0)$  der  $a_1 > 0$ , og  $\mathbf{b}$  har koordinater  $\mathbf{b} = (b_1, b_2, 0)$ . I dette tilfellet vil  $\mathbf{a} \times \mathbf{b}$  være parallel med  $z$ -aksen, og vi må undersøke når den peker i positiv retning. Bruker vi høyrehåndsregelen, får vi at  $\mathbf{a} \times \mathbf{b}$  skal peke langs den positive  $z$ -aksen dersom  $\mathbf{b}$  ligger i første eller annen kvadrant av  $xy$ -planet (dvs. når  $b_2 > 0$ ), og langs den negative  $z$ -aksen dersom  $\mathbf{b}$  ligger i tredje eller fjerde kvadrant (dvs. når  $b_2 < 0$ ). Bruker vi isteden formelen for vektorproduktet, ser vi at

$$\mathbf{a} \times \mathbf{b} = 0\mathbf{i} + 0\mathbf{j} + a_1 b_2 \mathbf{k} = a_1 b_2 \mathbf{k}$$

Denne vektoren peker langs den positive eller negative  $z$ -aksen ettersom  $b_2$  er positiv eller negativ (husk at  $a_1 > 0$ ). Dette betyr at formelen og høyrehåndsregelen gir samme resultat, og dermed er høyrehåndsregelen bevist i dette tilfellet.

Vi er nå rede til å se på det generelle tilfellet  $\mathbf{a} = (a_1, a_2, a_3)$ ,  $\mathbf{b} = (b_1, b_2, b_3)$ . Velg et par av vektorer  $\mathbf{a}_0$ ,  $\mathbf{b}_0$  i  $xy$ -planet slik at  $\mathbf{a}$  peker langs den positive  $x$ -aksen, og slik at  $\mathbf{a}_0, \mathbf{b}_0$  er en “kopii” av paret  $\mathbf{a}, \mathbf{b}$ . Med dette mener vi at  $\mathbf{a}_0$  er like lang som  $\mathbf{a}$ ,  $\mathbf{b}_0$  er like lang som  $\mathbf{b}$ , og at vinkelen fra  $\mathbf{a}_0$  til  $\mathbf{b}_0$  er lik vinkelen fra  $\mathbf{a}$  til  $\mathbf{b}$ . La  $\mathbf{c}_0 = \mathbf{a}_0 \times \mathbf{b}_0$ . Etter det vi allerede har vist, gjelder høyrehåndsregelen for  $\mathbf{a}_0$ ,  $\mathbf{b}_0$  og  $\mathbf{c}_0$ .

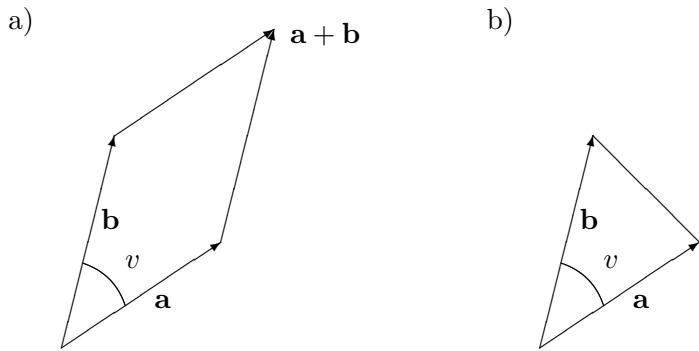
La oss nå tenke på vektortriplet  $(\mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0)$  som en materiell gjenstand, f.eks. tre sammensveisede biter av ståltråd. Vi flytter nå denne gjenstanden med en kontinuerlig bevegelse, uten å deformere den på noen måte, slik at  $\mathbf{a}_0$  ender opp som  $\mathbf{a}$ , og  $\mathbf{b}_0$  ender opp som  $\mathbf{b}$ . La  $\mathbf{a}(t)$  være posisjonen til  $\mathbf{a}_0$  etter  $t$  sekunder av denne bevegelsen, og la  $\mathbf{b}(t)$  og  $\mathbf{c}(t)$  være de tilsvarende posisjonene til  $\mathbf{b}_0$  og  $\mathbf{c}_0$ . Hvis bevegelsen tar  $T$  sekunder, er dermed  $\mathbf{a} = \mathbf{a}(T)$  og  $\mathbf{b} = \mathbf{b}(T)$ .

Dersom  $\mathbf{c}(T) = \mathbf{a}(T) \times \mathbf{b}(T)$ , er høyrehåndsregelen oppfylt for triplet  $\mathbf{a}(T) = \mathbf{a}$ ,  $\mathbf{b}(T) = \mathbf{b}$ ,  $\mathbf{c}(T) = \mathbf{c}$ . Vi skal derfor anta at  $\mathbf{c}(T) = -\mathbf{a}(T) \times \mathbf{b}(T)$  (den eneste andre muligheten) og vise at dette fører til en selvmotsigelse. La  $t_0$  være det første tidspunktet der  $\mathbf{c}(t)$  skifter fra å være lik  $\mathbf{a}(t) \times \mathbf{b}(t)$  til å være lik  $-\mathbf{a}(t) \times \mathbf{b}(t)$  (formelt er  $t_0 = \inf\{t : \mathbf{c}(t) = -\mathbf{a}(t) \times \mathbf{b}(t)\}$ ). Siden  $\mathbf{c}(t)$  beveger seg kontinuerlig, betyr dette at  $\mathbf{a}(t) \times \mathbf{b}(t)$  må gjøre et sprang ved tidspunktet  $t_0$ . Men det er umulig siden  $\mathbf{a}(t) \times \mathbf{b}(t)$  vil bevege seg kontinuerlig når  $\mathbf{a}(t)$  og  $\mathbf{b}(t)$  gjør det (tenk på det algebraiske uttrykket for vektorproduktet). Dermed har vi fått vår selvmotsigelse, og beviset er fullført.  $\square$

Vi skal nå se på noen av de tingene vektorproduktet kan brukes til. Først

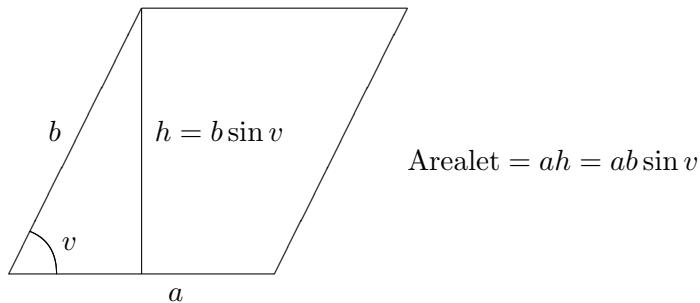
et enkelt eksempel.

**Eksempel 2:** Finn en vektor som står ortogonalt på både  $\mathbf{a} = (1, -2, 3)$  og  $\mathbf{b} = (4, -1, -2)$ . Vi regner rett og slett ut vektorproduktet:  $\mathbf{a} \times \mathbf{b} = (1, -2, 3) \times (4, -1, -2) = (7, 14, 7)$ . Legg merke til at siden  $(7, 14, 7) = 7(1, 2, 1)$ , kan vi forenkle løsningen til  $(1, 2, 1)$ . ♣



Figur 4: Parallellogrammet og trekanten utspent av  $\mathbf{a}$  og  $\mathbf{b}$

Det neste vi skal se på, er hvordan vektorproduktet kan brukes til å regne ut arealer. To vektorer  $\mathbf{a}$  og  $\mathbf{b}$  utspenner på en naturlig måte et parallellogram (se figur 4a). Halvparten av dette parallellogrammet (se figur 4b ovenfor), utgjør trekanten utspent av  $\mathbf{a}$  og  $\mathbf{b}$ .



Figur 5: Arealet til et parallellogram

**Setning 1.4.3** *Arealet til parallellogrammet utspent av vektorene  $\mathbf{a}$  og  $\mathbf{b}$  er lik  $|\mathbf{a} \times \mathbf{b}|$ . Arealet til trekanten utspent av  $\mathbf{a}$  og  $\mathbf{b}$  er  $\frac{1}{2}|\mathbf{a} \times \mathbf{b}|$*

*Bevis:* Enkel geometri forteller oss at arealet til et parallellogram er produktet av de to sidene gange med sinus til den mellomliggende vinkelen (se figur 5). For vårt parallellogram blir dette  $|\mathbf{a}||\mathbf{b}|\sin v$ , som vi vet er lik  $|\mathbf{a} \times \mathbf{b}|$ . Siden arealet av trekanten er halvparten av arealet til parallellogrammet, får vi også formelen for arealet av trekanten. □

**Eksempel 3:** Finn arealet til trekanten med hjørner i punktene  $\mathbf{a} = (2, -7, 3)$ ,  $\mathbf{b} = (-2, 3, 2)$  og  $\mathbf{c} = (2, 2, 2)$ . Denne trekanten har samme areal som den utspent av vektorene  $\mathbf{c} - \mathbf{a}$  og  $\mathbf{b} - \mathbf{a}$  (hvorfor?) Siden  $\mathbf{c} - \mathbf{a} = (0, 9, -1)$  og  $\mathbf{b} - \mathbf{a} = (-4, 10, -1)$ , får vi:

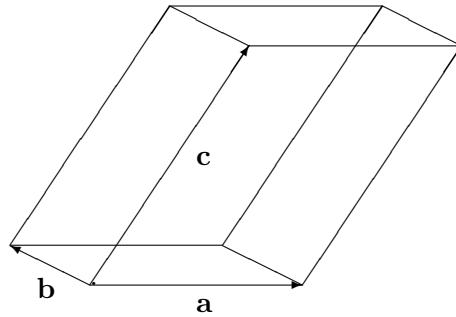
$$(\mathbf{c} - \mathbf{a}) \times (\mathbf{b} - \mathbf{a}) = (0, 9, -1) \times (-4, 10, -1) = (1, 4, 36)$$

Dermed er arealet til trekanten lik

$$\frac{1}{2}|(1, 4, 36)| = \frac{1}{2}\sqrt{1^2 + 4^2 + 36^2} = \frac{1}{2}\sqrt{1313} \quad \clubsuit$$

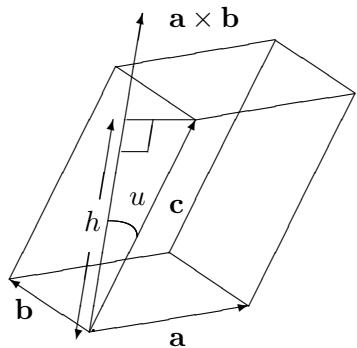
Eksemplet ovenfor viser noe av styrken ved å bruke vektorregning til å regne ut geometriske størrelser; det spiller ikke noen rolle hvor komplisert punktene ligger i forhold til hverandre — vi bare kopler inn den generelle formelen og ut faller svaret. Hadde vi prøvd å finne arealet med tradisjonelle geometriske metoder, hadde vi fort druknet i finurlige tegninger og kompliserte beregninger. Ulempen ved å bruke vektorregning er at vi ofte mister kontakten med det geometriske bildet — regningen viser oss at noe er riktig, men vi skjønner ikke riktig hvorfor.

Vektorproduktet kan også brukes til å regne ut volumer. Tre vektorer  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  i rommet definerer på en naturlig måte et romlegeme, et *parallellepiped*, som vist på figur 6.



Figur 6: Parallellepipedet utspent av  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$

Fra skolen vet vi at volumet til et parallellepiped er arealet av grunnflaten ganget med høyden. Sier vi at grunnflaten er parallelogrammet utspent av  $\mathbf{a}$  og  $\mathbf{b}$ , vet vi at arealet til grunnflaten er  $|\mathbf{a} \times \mathbf{b}|$ . På figur 7 har vi kalt vinkelen mellom  $\mathbf{a} \times \mathbf{b}$  og den tredje vektoren  $\mathbf{c}$  for  $u$ .



Figur 7: Volumet til et parallellepiped

Siden  $\mathbf{a} \times \mathbf{b}$  står normalt på grunnflaten, blir høyden  $h$  lik  $|\mathbf{c}| |\cos u|$  (vi må ha med tallverdien rundt cosinus i tilfelle  $u$  er større en  $90^\circ$ ). Volumet til parallelepipedet er derfor  $|\mathbf{a} \times \mathbf{b}| |\mathbf{c}| |\cos u|$ . Men dette uttrykket er jo lik  $|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$  (husk den geometriske beskrivelsen av skalarproduktet). Dermed har vi vist:

**Setning 1.4.4** *Volumet til parallelepipedet utspent av vektorene  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  er  $|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$*

**Bemerkning:** Når vi skal regne ut volumet til parallelepipedet utspent av  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$ , spiller selvfølgelig ikke rekkefølgen av de tre vektorene noen rolle. Volumet kan derfor skrives som både  $|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$ ,  $|(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b}|$ ,  $|(\mathbf{b} \times \mathbf{a}) \cdot \mathbf{c}|$ ,  $|(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{a}|$ ,  $|(\mathbf{c} \times \mathbf{a}) \cdot \mathbf{b}|$ , og  $|(\mathbf{c} \times \mathbf{b}) \cdot \mathbf{a}|$ . Disse seks uttrykkene må derfor være like. Hvis du ønsker, kan du sjekke dette ved direkte utregning.

**Eksempel 4:** Finn volumet til parallelepipedet som utspennes av vektorene  $(4, 0, 3)$ ,  $(-1, 2, -3)$  og  $(0, 2, 1)$ . Ifølge setningen er dette volumet gitt ved  $|((4, 0, 3) \times (-1, 2, -3)) \cdot (0, 2, 1)|$ . Vi regner først ut

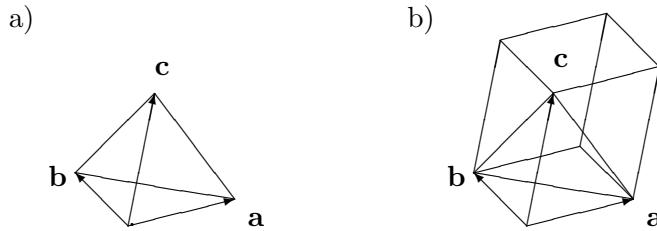
$$(4, 0, 3) \times (-1, 2, -3) = (-6, 9, 8)$$

Deretter tar vi

$$|(-6, 9, 8) \cdot (0, 2, 1)| = |0 + 18 + 8| = 26$$

Volumet er altså 26. ♣

Tre vektorer  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  utspenner også en pyramide (se figur 8a). For å finne volumet til denne pyramiden husker vi at volumet til en generell pyramide er  $\frac{1}{3}gh$ , der  $h$  er høyden og  $g$  er arealet til grunnflaten.

Figur 8: Pyramiden og parallellepipedet utspent av  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$ 

Sammenligner vi pyramiden utspent av  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  med parallellepipedet utspent av de samme vektorene (figur 8b), ser vi at høydene er like, men at grunnflaten til pyramiden er halvparten av grunnflaten til parallellepipedet. Det betyr at volumet til pyramiden må være en seksdel av volumet til parallellepipedet. Dermed har vi:

**Korollar 1.4.5** *Volumet av pyramiden utspent av de tre vektorene  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  er  $\frac{1}{6}|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$*

**Eksempel 5:** Finn volumet til pyramiden med hjørner i punktene  $\mathbf{a} = (-1, 2, -3)$ ,  $\mathbf{b} = (1, 4, 1)$ ,  $\mathbf{c} = (0, 4, 7)$  og  $\mathbf{d} = (3, 0, 5)$ . Denne pyramiden har samme volum som pyramiden utspent av vektorene  $\mathbf{b} - \mathbf{a}$ ,  $\mathbf{c} - \mathbf{a}$  og  $\mathbf{d} - \mathbf{a}$  (forklar hvorfor!) Siden

$$\mathbf{b} - \mathbf{a} = (2, 2, 4) \quad \mathbf{c} - \mathbf{a} = (1, 2, 10) \quad \mathbf{d} - \mathbf{a} = (4, -2, 8)$$

får vi:

$$\begin{aligned} \text{Volum} &= \frac{1}{6}|((2, 2, 4) \times (1, 2, 10)) \cdot (4, -2, 8)| \\ &= \frac{1}{6}|(12, -16, 2) \cdot (4, -2, 8)| = \frac{96}{6} = 16 \end{aligned}$$

♣

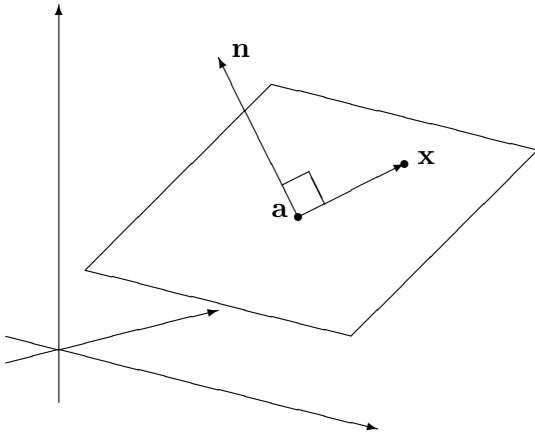
Helt til slutt i dette avsnittet skal vi se hvordan vi kan bruke vektorproduktet til å finne ligningen til et plan. Den enkleste måten å beskrive et plan på, er som regel å angi en normalvektor  $\mathbf{n} = (n_1, n_2, n_3)$  pluss et punkt  $\mathbf{a} = (a_1, a_2, a_3)$  som planet går gjennom. Planet består da av alle de punkter  $\mathbf{x} = (x, y, z)$  slik at  $\mathbf{x} - \mathbf{a}$  står normalt på  $\mathbf{n}$  (se figur 9). At  $\mathbf{x} - \mathbf{a}$  står normalt på  $\mathbf{n}$  er ekvivalent med at

$$0 = \mathbf{n} \cdot (\mathbf{x} - \mathbf{a}) = n_1x + n_2y + n_3z - n_1a_1 - n_2a_2 - n_3a_3$$

eller med andre ord

$$n_1x + n_2y + n_3z = n_1a_1 + n_2a_2 + n_3a_3$$

Dette kaller vi *ligningen til planeten*. Legg merke til at koeffisientene til  $x$ ,  $y$  og  $z$  rett og slett er koordinatene til  $\mathbf{n}$



Figur 9:  $\mathbf{x}$  ligger i planet gjennom  $\mathbf{a}$  normalt på  $\mathbf{n}$

**Eksempel 6:** Finn ligningen til planeten som går gjennom  $\mathbf{a} = (-3, 1, 2)$  og står normalt på  $\mathbf{n} = (-4, 2, -1)$ . Undersøk om punktet  $(-2, 4, 3)$  ligger i dette planet.

Vi skal finne alle vektorer  $\mathbf{x} = (x, y, z)$  slik at  $0 = \mathbf{n} \cdot (\mathbf{x} - \mathbf{a})$ , dvs.

$$\begin{aligned} 0 &= (-4, 2, -1) \cdot (x - (-3), y - 1, z - 2) \\ &= (-4)(x + 3) + 2(y - 1) - 1(z - 2) = -4x + 2y - z - 12 \end{aligned}$$

Altså blir ligningen

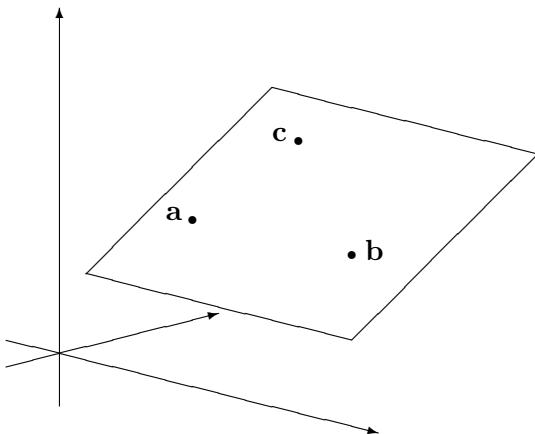
$$-4x + 2y - z = 12$$

For å undersøke om punktet  $(-2, 4, 3)$  ligger i planet, sjekker vi om det passer i ligningen. Setter vi inn på venstre side, får vi:

$$-4 \cdot (-2) + 2 \cdot 4 - 3 = 13 \neq 12$$

som viser at punktet ikke ligger i planet. ♣

En annen måte å beskrive et plan på, er å spesifisere tre punkter  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  som planeten går gjennom, se figur 10 (sørg for at punktene ikke ligger på samme rette linje!). For å finne ligningen til planeten, må vi da først finne en normalvektor. Det er ikke så vanskelig; en normalvektor til planeten må stå normalt på begge vektorene  $\mathbf{b} - \mathbf{a}$  og  $\mathbf{c} - \mathbf{a}$  (siden begge disse vektorene i sin helhet ligger i planeten), og kryssproduktet  $\mathbf{n} = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})$  er derfor et naturlig valg.

Figur 10: Planet gjennom  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$ 

**Eksempel 7:** Finn ligningen til planet som går gjennom  $\mathbf{a} = (1, -2, 1)$ ,  $\mathbf{b} = (3, -2, 2)$  og  $\mathbf{c} = (2, 0, 1)$ .

Siden  $\mathbf{b} - \mathbf{a} = (2, 0, 1)$  og  $\mathbf{c} - \mathbf{a} = (1, 2, 0)$ , så er en normalvektor gitt ved

$$\mathbf{n} = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a}) = (2, 0, 1) \times (1, 2, 0) = (-2, 1, 4)$$

Siden normalvektoren er  $(-2, 1, 4)$ , vet vi at planligningen har formen

$$-2x + y + 4z = d$$

for et tall  $d$ . For å finne  $d$ , kan vi f.eks. sette inn koordinatene til  $\mathbf{a}$  i ligningene:

$$d = -2 \cdot 1 + (-2) + 4 \cdot 1 = 0$$

Altså er ligningen til planet

$$-2x + y + 4z = 0$$

♣

**MATLAB-kommentar:** Du kan bruke MATLAB til å regne ut kryssproduktet til to vektorer  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$  ved å skrive `cross(a, b)`.

### Oppgaver til seksjon 1.4

1. Regn ut  $\mathbf{a} \times \mathbf{b}$  når

a)  $\mathbf{a} = (-1, 3, 2)$      $\mathbf{b} = (-2, 1, 7)$       b)     $\mathbf{a} = (4, -3, 1)$      $\mathbf{b} = (-6, 1, 0)$

2. Finn arealet til parallellogrammet utspent av  $\mathbf{a} = (-2, 3, 1)$  og  $\mathbf{b} = (4, 0, -2)$ .

3. En trekant har hjørner i punktene  $(0, -1, 2)$ ,  $(2, -1, 4)$  og  $(3, 0, 4)$ . Finn arealet.

4. Finn en vektor som står normalt på både  $(2, 0, -3)$  og  $(-1, 3, 4)$ .
5. Regn ut  $\mathbf{i} \times \mathbf{j}$ ,  $\mathbf{j} \times \mathbf{k}$ ,  $\mathbf{k} \times \mathbf{i}$ .
6. Finn volumet til parallellepipedet utspent av  $(3, -2, -2)$ ,  $(0, 0, 4)$  og  $(-3, 2, 1)$ .
7. En pyramide har hjørner i punktene  $(2, -1, 2)$ ,  $(0, 5, -3)$ ,  $(2, 4, 6)$  og  $(3, -2, 4)$ . Finn volumet.
8. Finn en ligning for planet som går gjennom punktene  $\mathbf{a} = (1, 1, -1)$ ,  $\mathbf{b} = (0, -2, -6)$ ,  $\mathbf{c} = (2, 3, 3)$ .
9. Finn en ligning for planet som går gjennom punktene  $\mathbf{a} = (1, 2, 1)$ ,  $\mathbf{b} = (2, 3, 0)$ ,  $\mathbf{c} = (2, 1, -1)$ .
10. Anta at alle hjørnene i et parallellepiped har heltallige koeffisienter. Vis at volumet er et helt tall.
11. Bevis setning 1.4.1b).

## 1.5 Matriser

Hittil har vi bare studert  $n$ -tupler, men tiden er nå inne for å introdusere matriser. En  $m \times n$ -matrise

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

er et rektangulært oppsett av tall med  $m$  rader (linjer) og  $n$  søyler. To eksempler er  $2 \times 3$ -matrisen

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

og  $3 \times 3$ -matrisen

$$C = \begin{pmatrix} 2 & -3 & 1 \\ \frac{1}{2} & -2 & \pi \\ 1 & -\frac{3}{4} & 0 \end{pmatrix}$$

Tallene  $a_{11}, a_{12}, \dots, a_{mn}$  som inngår i matrisen  $A$ , kaller vi *elementene* i  $A$ . Legg merke til hvordan vi nummererer elementene:  $a_{34}$  er elementet i rad 3 og søyle 4 (strengt tatt burde vi ha skrevet et komma mellom 3-tallet og 4-tallet for å gjøre det klart at det ikke er element nummer 34 det er snakk om, men man blir fort lei av å skrive alle kommaene, og det er derfor vanlig å droppe dem dersom det ikke kan oppstå misforståelser). Legg også merke til at vi kan oppfatte vektorer som matriser. En radvektor

$$\mathbf{a} = (a_1, a_2, \dots, a_n)$$

er en  $1 \times n$ -matrise, mens en søylevektor

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

er en  $n \times 1$ -matrise. Som matriser betraktet er altså radvektoren og søylevektoren av forskjellig type, og når vi blander vektorer og matriser er det derfor viktig å holde styr på om  $\mathbf{a}$  skal oppfattes som en rad- eller en søylevektor. Vi skal komme tilbake til dette etter hvert. Legg forøvrig merke til at en  $1 \times 1$ -matrise ( $a_{11}$ ) bare er et tall med en parentes rundt. Parentesen spiller ingen rolle, og vi skal derfor si at en  $1 \times 1$ -matrise bare er et tall  $a_{11}$ .

Akkurat som for  $n$ -tupler kan vi definere addisjon og subtraksjon komponentvis. Dersom vi har to  $m \times n$ -matriser

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

så definerer vi

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

og

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

Siden alle regneoperasjoner foregår komponentvis, vil de vanlige regnereglene for addisjon og subtraksjon også gjelde for matriser.

Vi kan også multiplisere en matrise med et tall  $s$  ved å gange tallet inn i hver komponent:

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

La oss ta et eksempel der vi kombinerer flere av regneoperasjonene:

**Eksempel 1:** Regn ut  $3A - 2B$  når

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 4 & 0 & -1 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} 2 & -3 & 0 \\ 6 & -2 & 3 \end{pmatrix}$$

Vi får

$$3A - 2B = \begin{pmatrix} 3 & -3 & 6 \\ 12 & 0 & -3 \end{pmatrix} - \begin{pmatrix} 4 & -6 & 0 \\ 12 & -4 & 6 \end{pmatrix} = \begin{pmatrix} -1 & 3 & 6 \\ 0 & 4 & -9 \end{pmatrix}$$



*Transponering* er en viktig operasjon for matriser som ikke finnes for tall. Vi transponerer en matrise ved å bytte om rader og søyler. Den *transponerte* til matrisen  $A$  ovenfor er derfor

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

Legg merke til at når  $A$  er en  $m \times n$ -matrise, så er  $A^T$  en  $n \times m$ -matrise.

**Eksempel 2:** Den transponerte til  $2 \times 3$ -matrisen

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

er

$$A^T = \begin{pmatrix} 2 & 5 \\ -1 & -2 \\ 3 & 0 \end{pmatrix}$$

som er en  $3 \times 2$ -matrise.



Transponering kan lett kombineres med addisjon, subtraksjon og multiplikasjon med skalar. Du kan sjekke at

$$(A + B)^T = A^T + B^T, \quad (A - B)^T = A^T - B^T, \quad (cA)^T = cA^T$$

Hvis vi transponerer den transponerte, kommer vi tilbake til utgangspunktet:

$$(A^T)^T = A$$

Legg også merke til at rad- og søylevektorer er transponerte av hverandre. Oppfatter vi søylevektoren

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

som en matrise, blir den transponerte en radvektor

$$\mathbf{a}^T = (a_1, a_2, \dots, a_n)$$

De grunnleggende regneoperasjonene for matriser som vi nå har sett på, er ganske enkle, men man kan jo lure på hva det hele er godt for — hvorfor innfører vi egentlig matriser når vi allerede har  $n$ -tupler til å holde styr på lister av tall? Ett svar er at en del lister kommer naturlig i rektangelform, og at det er greit å beholde denne formen for ikke å miste oversikt over informasjonen. Opplysningene vi legger inn i et regnark vil for eksempel ofte ha matriseform. Det neste eksemplet viser en situasjon som ofte forekommer, og der elementene i matrisen viser fordelingen mellom forskjellige muligheter.

**Eksempel 3:** En fruktpresse mottar epler fra fire forskjellige produsenter. Eplene blir sortert i tre kategorier: god, middels, dårlig. Erfaringene viser at produsentene har forskjellig kvalitet på sine produkter. Produsent 1 leverer 50% av god kvalitet, 30% av middels kvalitet og 20% av dårlig kvalitet. Tallene for de andre produsentene er: Produsent 2: 30%, 40%, 30%; produsent 3: 25%, 40%, 35%, produsent 4: 20%, 60%, 20%. Dersom vi gir hver produsent en søyle, kan vi sette opp denne informasjonen som en  $3 \times 4$ -matrise (husk at prosent betyr “hundredel”):

$$A = \begin{pmatrix} 0.5 & 0.3 & 0.25 & 0.2 \\ 0.3 & 0.4 & 0.4 & 0.6 \\ 0.2 & 0.3 & 0.35 & 0.2 \end{pmatrix}$$

Anta nå at fruktpressen mottar leveringer fra hver av produsentene: 4 tonn fra produsent 1, 5 tonn fra produsent 2, 3 tonn fra produsent 3 og 6 tonn fra produsent 4. Vi setter opp denne leveransen som en søylevektor

$$\mathbf{b} = \begin{pmatrix} 4 \\ 5 \\ 3 \\ 6 \end{pmatrix}$$

Vi ønsker å finne ut hvor mange tonn vi har fått av hver kvalitet. Observer at dersom vi ganger hvert av tallene i første rad i  $A$  med tilsvarende tall i  $\mathbf{b}$ , og så legger sammen, får vi antall tonn av beste kvalitet:

$$0.5 \cdot 4 + 0.3 \cdot 5 + 0.25 \cdot 3 + 0.2 \cdot 6 = 2 + 1.5 + 0.75 + 1.2 = 5.45$$

Tilsvarende får vi antall tonn av nest beste kvalitet ved å gange tallene i annen rad i  $A$  med tilsvarende tall i  $\mathbf{b}$ , og så legge sammen:

$$0.3 \cdot 4 + 0.4 \cdot 5 + 0.4 \cdot 3 + 0.6 \cdot 6 = 1.2 + 2 + 1.2 + 3.6 = 8$$

Til slutt får vi antall tonn av dårligste kvalitet ved å gange tallene i nederste rad i  $A$  med tilsvarende tall i  $\mathbf{b}$  og så legge sammen:

$$0.2 \cdot 4 + 0.3 \cdot 5 + 0.35 \cdot 3 + 0.2 \cdot 6 = 0.8 + 1.5 + 1.05 + 1.2 = 4.55$$

Legg merke til at i hvert av disse regnestykkene har vi regnet ut et slags skalarprodukt mellom en rad i matrisen  $A$  og vektoren  $\mathbf{b}$  (vi sier et “slags” skalarprodukt siden det er et litt uortodokst produkt mellom en radvektor og en søylevektor). Legg også merke til at vi kan tenke på resultatet av regnestykkene som en ny vektor

$$\mathbf{c} = \begin{pmatrix} 5.45 \\ 8 \\ 4.55 \end{pmatrix}$$

som forteller oss hvor mange tonn vi har av hver kvalitet.



Vi skal se på et eksempel til av lignende type.

**Eksempel 4:** Et kjøpesenter har tre stativ  $X$ ,  $Y$  og  $Z$  hvor du kan hente og avlevere handlevogner. Av de vognene som starter dagen i stativ  $X$ , vil 70% avslutte den på samme sted, 10% vil ha endt opp i  $Y$ , og 20% i  $Z$ . Av de vognene som startet dagen i stativ  $Y$ , vil 30% avslutte dagen i stativ  $X$ , mens henholdsvis 50% og 20% vil havne i stativene  $Y$  og  $Z$ . De tilsvarende tallene for vogner som starter i  $Z$ , er at 40% ender dagen i  $X$ , 20% i  $Y$  og 40% i  $Z$ . Vi kan ordne disse tallene i en matrise  $A$  der første søyle gir fordelingen av de vognene som startet i  $X$ , andre søyle gir fordelingen av de vognene som startet i  $Y$  og tredje søyle gir fordelingen av vognene som startet i  $Z$ :

$$A = \begin{pmatrix} 0.7 & 0.3 & 0.4 \\ 0.1 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.4 \end{pmatrix}$$

Anta nå at vi startet dagen med 100 vogner i  $X$ , 70 vogner i  $Y$  og 30 vogner i  $Z$ , og la oss skrive dette som en søylevektor

$$\mathbf{b} = \begin{pmatrix} 100 \\ 70 \\ 30 \end{pmatrix}$$

Vi ønsker å finne ut hvor mange vogner som befinner seg på hvert sted ved slutten av dagen. Antall vogner i  $X$  får vi ved å gange tallene i første rad i  $A$  med tilsvarende komponent i vektoren  $\mathbf{b}$ , og så legge sammen:

$$0.7 \cdot 100 + 0.3 \cdot 70 + 0.4 \cdot 30 = 70 + 21 + 12 = 103$$

Antall vogner i  $Y$  får vi tilsvarende ved å gange hvert tall i annen rad i  $A$  med tilsvarende tall i  $\mathbf{b}$ , og så legge sammen:

$$0.1 \cdot 100 + 0.5 \cdot 70 + 0.2 \cdot 30 = 10 + 35 + 6 = 51$$

Antall vogner i  $Z$  ved slutten av dagen, får vi så ved å gange hvert tall i tredje rad i  $A$  med tilsvarende tall i  $\mathbf{b}$ , og legge sammen:

$$0.2 \cdot 100 + 0.2 \cdot 70 + 0.4 \cdot 20 = 10 + 14 + 12 = 46$$

Vi kan skrive opp resultatet som en ny vektor

$$\mathbf{c} = \begin{pmatrix} 103 \\ 51 \\ 46 \end{pmatrix}$$

som gir oss fordelingen av handlevogner ved slutten av dagen. Legg merke til at vi i dette eksemplet har gjort akkurat samme operasjoner som i det forrige; vi har tatt skalarproduktene mellom radene i  $A$  og søylevektoren  $\mathbf{b}$ .



I begge eksemplene ovenfor gjennomførte vi samme type regneoperasjoner. Vi startet med en matrise  $A$  og en søylevektor  $\mathbf{b}$ , og laget en ny søylevektor  $\mathbf{c}$  der komponentene fremkom som skalarprodukt av radene i  $A$  og vektoren  $\mathbf{b}$ . Denne regneoperasjonen er så vanlig at det er greit å ha et eget navn på den.

**Definisjon 1.5.1 (Multiplikasjon av matrise og søylevektor)** *Anta at  $A$  er en  $m \times n$ -matrise og at  $\mathbf{b}$  er en  $n$ -dimensjonal søylevektor:*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{og} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

Produktet av  $A$  og  $\mathbf{b}$  er da den  $m$ -dimensjonale søylevektoren  $\mathbf{c} = A\mathbf{b}$  gitt ved

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \begin{pmatrix} a_{11}b_1 + a_{12}b_2 + \cdots + a_{1n}b_n \\ a_{21}b_1 + a_{22}b_2 + \cdots + a_{2n}b_n \\ \vdots \\ a_{m1}b_1 + a_{m2}b_2 + \cdots + a_{mn}b_n \end{pmatrix}$$

Den  $i$ -te komponentene i  $\mathbf{c} = A\mathbf{b}$  fremkommer altså ved at vi tar skalarproduktet av den  $i$ -te raden i  $A$  med vektoren  $\mathbf{b}$ .

Legg merke til at produktet  $A\mathbf{b}$  bare er definert når  $A$  og  $\mathbf{b}$  passer sammen størrelsesmessig;  $\mathbf{b}$  må ha like mange komponenter som  $A$  har søyler. Observer også at produktet  $A\mathbf{b}$  er en søylevektor med like mange rader som  $A$ .

**Eksempel 5:** Finn  $A\mathbf{b}$  når

$$A = \begin{pmatrix} 2 & 1 \\ 4 & -1 \\ -2 & 5 \end{pmatrix} \quad \text{og} \quad \mathbf{b} = \begin{pmatrix} 3 \\ 6 \end{pmatrix}$$

Vi får

$$A\mathbf{b} = \begin{pmatrix} 2 \cdot 3 + 1 \cdot 6 \\ 4 \cdot 3 + (-1) \cdot 6 \\ (-2) \cdot 3 + 5 \cdot 6 \end{pmatrix} = \begin{pmatrix} 12 \\ 6 \\ 24 \end{pmatrix}$$



Vi skal komme tilbake til regneregler for produkter senere, men skriver opp noen av de enkleste og vanligste (i formlene nedenfor er  $s$  et tall):

$$(A + B)\mathbf{b} = A\mathbf{b} + B\mathbf{b}, \quad (sA)\mathbf{b} = s(A\mathbf{b})$$

$$A(\mathbf{b} + \mathbf{c}) = A\mathbf{b} + A\mathbf{c}, \quad A(s\mathbf{b}) = s(A\mathbf{b})$$

Det er god trening å sjekke at disse reglene holder. Vær oppmerksom på at det også er en del ting du kan gjøre med vanlige produkter, som du *ikke* kan gjøre med produktet ovenfor, f.eks. kan du ikke bytte om på rekkefølgen av faktorene ( $\mathbf{b}A$  gir ikke mening). Du kan heller ikke forkorte  $A$  i uttrykk av typen  $A\mathbf{b} = A\mathbf{c}$  eller  $\mathbf{b}$  i uttrykk av typen  $A\mathbf{b} = B\mathbf{b}$  — å forkorte betyr egentlig å gange med et inverst element på begge sider av ligningen, og vi vet foreløpig ikke noe om inverse vektorer og matriser (det viser seg at vektorer ikke har inverser, men at noen matriser har det!).

La oss ta en ny kikk på eksempel 3 og 4 i lys av de begrepene vi nå har innført:

**Eksempel 3 og 4 på nytt:** I eksempel 3 (fruktpressen) ble leveransene fra de fire produsentene kodet som et 4-tuppel  $\mathbf{b}$ , mens 3-tuplet  $\mathbf{c}$  fortalte oss hvor mange tonn vi mottok av hver kvalitet. Vi regnet ut elementene i  $\mathbf{c}$  ved å ta skalarproduktet av radene i matrisen  $A$  med vektoren  $\mathbf{b}$  — med andre ord:

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

Som vi skal komme tilbake til senere, kan vi tenke på multiplikasjon med  $A$  som en “transformasjon” — vektoren  $\mathbf{b}$  forteller oss hvor mye frukt vi får fra hver produsent, og multiplikasjon med  $A$  “transformerer” denne kunnskapen til kunnskap om hvor mye vi mottar av hver kvalitet (representert ved vektoren  $\mathbf{c}$ ).

Noe lignende skjer i eksempel 4 (handlevognene). Her representerer vektor **b** fordelingen av handlevogner ved begynnelsen av dagen, mens vektor **c** representerer fordelingen ved slutten av dagen. Vi regnet ut elementene i **c** ved å ta skalarproduktet av radene i matrisen  $A$  med vektoren **b** — med andre ord:

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

Igjen kan vi tenke på multiplikasjon med  $A$  som en transformasjon som transformerer kunnskap om hvor handlevognene er ved begynnelsen av dagen til kunnskap om hvor de er ved slutten av dagen. ♣

La oss avsluttet med en liten observasjon som vi kommer til å få glede av senere. Anta at

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{og} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

er to søylevektorer. Transponerer vi **a**, får vi radvektoren

$$\mathbf{a}^T = (a_1, a_2, \dots, a_n)$$

som vi kan oppfatte som en  $1 \times n$ -matrise. Vi kan nå ta produktet av matrisen  $\mathbf{a}^T$  og vektoren **b**. Dette gir oss en  $1 \times 1$ -matrise (dvs. et tall):

$$\mathbf{a}^T \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \mathbf{a} \cdot \mathbf{b}$$

Matriseproduktet  $\mathbf{a}^T \mathbf{b}$  er altså lik skalarproduktet  $\mathbf{a} \cdot \mathbf{b}$ . Dette kan virke som en kuriositet, men viser seg å være viktig i en del sammenhenger.

**MATLAB-kommentar:** Du kan skrive inn en matrise

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

i MATLAB ved med kommandoen

```
>>A=[2 -1 3
    1 0 -3
    4 0 1]
```

Du kan også markere linjeskiftene med semikolon:

```
>>A=[2 -1 3;1 0 -3;4 0 1]
```

Når matrisene er lastet inn, kan du utføre operasjoner med kommandoer som  $\gg A+B$  og  $\gg 3*A-2*B$ . Den transponerte til  $A$  får du ved å skrive  $\gg A'$ , og produktet mellom en matrise  $A$  og en vektor  $\mathbf{b}$  er gitt ved kommandoen  $\gg A*\mathbf{b}$ . Alle disse kommandoene forutsetter at matrisene og vektorene har riktige dimensjoner slik at regneoperasjonene gir mening.

### Oppgaver til seksjon 1.5

1. Regn ut  $2A$ ,  $(-3)B$ ,  $A + B$  og  $A - B$  når

$$A = \begin{pmatrix} 3 & -2 & 4 \\ -1 & 0 & 6 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} 2 & 2 & -1 \\ 4 & 0 & -6 \end{pmatrix}$$

2. Regn ut  $4A - 3B$  når

$$A = \begin{pmatrix} 1 & -3 \\ 2 & -1 \\ 3 & -1 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} 3 & 2 \\ -2 & -1 \\ 3 & 0 \end{pmatrix}$$

3. Finn de transponerte:

$$A = \begin{pmatrix} 1 & -2 & 6 \\ -1 & 8 & 6 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} 2 & 2 & -1 \\ 4 & 0 & -6 \\ 1 & 7 & -3 \end{pmatrix}$$

4. Finn de transponerte til  $A$ ,  $B$  og  $4A - B$  når

$$A = \begin{pmatrix} 1 & 0 \\ -5 & 7 \\ 2 & -1 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} 4 & -1 \\ -1 & 3 \\ 2 & -6 \end{pmatrix}$$

5. Regn ut  $A\mathbf{x}$  når:

a)  $A = \begin{pmatrix} 1 & 0 & -3 \\ -2 & -3 & 2 \end{pmatrix}$  og  $\mathbf{x} = \begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix}$ .

b)  $A = \begin{pmatrix} 2 & 0 \\ 3 & 1 \\ 6 & -2 \end{pmatrix}$  og  $\mathbf{x} = \begin{pmatrix} 3 \\ -2 \end{pmatrix}$ .

c)  $A = \begin{pmatrix} 2 & 1 & 0 \\ -3 & 4 & -2 \\ 1 & -3 & 2 \end{pmatrix}$  og  $\mathbf{x} = \begin{pmatrix} 4 \\ 0 \\ 3 \end{pmatrix}$ .

6. Regn ut hvor mange tonn vi får av hver kvalitet i eksempel 3 dersom produsent 1 leverer 10 tonn, produsent 2 leverer 5 tonn, produsent 3 leverer 8 tonn og produsent 4 leverer 6 tonn.

7. Hvordan vil handlevognene i eksempel 4 fordele seg ved slutten av dagen dersom vi begynner med 50 vogner i stativ  $X$ , 70 i stativ  $Y$  og 80 i stativ  $Z$ ?

8. Bevis regnereglene for transponering som står rett etter eksempel 2.

**9.** Bevis regneregelen for multiplikasjon av matrise og vektor som står rett etter eksempel 5.

**10.** Et oljemottak får olje fra 3 oljefelt. Oljen inndeles i fem kvaliteter A, B, C, D, E. Fra oljefelt 1 er fordelingen: A: 10%, B: 20%, C: 30%, D: 30%, E: 10%. Fra oljefelt 2 er fordelingen: A: 0%, B: 30%, C: 30%, D: 30%, E: 10%. Fra oljefelt 3 er fordelingen: A: 20%, B: 20%, C: 10%, D: 10%, E: 40%. Anta at mottaket får inn  $x$  enheter fra felt 1,  $y$  enheter fra felt 2 og  $z$  enheter fra felt 3. Finn en matrise som kan brukes til å regne ut hvor mange enheter man får av hver kvalitet. Hva blir resultatet når  $x = 10$ ,  $y = 12$  og  $z = 8$ ?

**11.** En smittsom sykdom sprer seg i et land. Helsemyndighetene deler befolkningen i tre grupper: smitteutsatte, syke, immune. Fra en uke til den neste regner man at 5% av de smitteutsatte blir syke, mens 1% av dem blir immune uten å ha vært syke. Av de syke blir 80% immune, mens resten fortsatt er syke. Én prosent av de immune mister immunitetten og blir smitteutsatte, mens resten fortsatt er immune. La  $x_n$ ,  $y_n$ ,  $z_n$  være den andelen av befolkningen som er hhv. smitteutsatt, syk og immun etter  $n$  uker. La  $\mathbf{v}_n = \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}$ .

- a) Finn en matrise  $A$  slik at  $A\mathbf{v}_n = \mathbf{v}_{n+1}$ .
- b) I uke 0 er 10% av befolkningen syke, mens resten er smitteutsatte. Finn fordelingen av smitteutsatte, syke og immune i uke 1 og 2.

**12.** En forskergruppe som studerer et bestemt dyreslag, deler bestanden inn i fire grupper:

1. unge, dvs. dyr født samme vår
2. unge voksne, dvs. dyr født året før
3. voksne, dvs. dyr født to år før
4. gamle, dvs. dyr født mer enn to år før

Statistikken viser at bare 5% av de unge overlever til året etter. Av de unge voksne overlever 50% til året etter. I tillegg gir en ung voksen i gjennomsnitt opphav til 20 unger som blir født året etter. Blant de voksne overlever 30% til året etter. I tillegg gir en voksen i gjennomsnitt opphav til 50 unger som blir født året etter. Av de gamle overlever 10% til året etter. I tillegg gir en gammel i gjennomsnitt opphav til 10 unger som blir født året etter. La  $x_n$ ,  $y_n$ ,  $z_n$ ,  $u_n$  være hhv. antall unge, antall unge voksne, antall voksne og antall gamle etter  $n$  år, og skriv

$$\mathbf{v}_n = \begin{pmatrix} x_n \\ y_n \\ z_n \\ u_n \end{pmatrix}$$

- a) Finn en matrise  $A$  slik at  $\mathbf{v}_{n+1} = A\mathbf{v}_n$ .
- b) Anta at det et år settes ut 100 voksne dyr i et terrenget der det ikke er noen dyr fra før av. Hvor mange dyr av hver kategori vil det være i dette terrenget to år senere?

**13.** Per, Pål og Espen leker sisten. Hver gang Per eller Pål har sisten, er det 90% sjanse for at Espen får sisten neste gang, mens hver gang Espen har sisten, har Per og Pål 50% sjanse hver for å få sisten neste gang. Anta at  $x_n$  er sannsynligheten for at Per er den  $n$ -te som har sisten, at  $y_n$  er sannsynligheten for at Pål er den  $n$ -te som har sisten, og at  $z_n$  er sannsynligheten for at Espen er den  $n$ -te som har sisten. La  $\mathbf{v}_n = \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}$ .

- a) Finn en matrise  $A$  slik at  $\mathbf{v}_{n+1} = A\mathbf{v}_n$ .
- b) Anta at Pål har sisten først. Hva er sannsynligheten for at henholdsvis Per, Pål og Espen er den tredje som har sisten?

## 1.6 Multiplikasjon av matriser

I forrige seksjon så vi at addisjon og subtraksjon av matriser er lett; vi bare adderer og subtraherer komponentvis. Det er fristende å definere multiplikasjon på samme måte; hvis

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}$$

så bør kanskje produktet være

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

Det er ikke noe galt med dette produktet (det kalles *Hadamard-produktet* og er nyttig for noen formål), men det er en annen type matriseprodukt som er langt viktigere. Ved første øyekast ser dette produktet litt merkelig ut, og for å motivere det skal vi gå tilbake til produktet  $A\mathbf{b}$  fra slutten av forrige seksjon. (Dersom du synes denne motivasjonen er vanskelig, kan du trøste deg med at det kommer en oppsummering i definisjon 1.6.1.)

La oss se litt nærmere på produktet  $\mathbf{c} = A\mathbf{b}$  mellom en matrise  $A$  og en vektor  $\mathbf{b}$ . Som allerede nevnt, er det ofte smart å tenke på dette som en *transformasjon*; vi starter med en vektor  $\mathbf{b}$ , og matrisen  $A$  transformerer  $\mathbf{b}$  til en ny vektor  $\mathbf{c}$ . I eksempel 3 i forrige seksjon (fruktpressen), så vi f.eks. hvordan  $A$  transformerer kunnskap om hvor mange tonn epler vi får fra hver produsent (kodet opp i vektoren  $\mathbf{b}$ ) til kunnskap om hvor mange tonn epler vi har av hver kvalitet (kodet opp i vektoren  $\mathbf{c} = A\mathbf{b}$ ). Noe tilsvarende skjedde i eksempel 4 (handlevognene); i dette tilfellet transformerer  $A$  kunnskap om

hvor vognene er ved begynnelsen av dagen (gitt ved vektoren  $\mathbf{b}$ ), til kunnskap om hvor de er ved slutten av dagen (gitt ved vektoren  $\mathbf{c} = A\mathbf{b}$ ). Legg merke til at i begge eksemplene bruker vi den samme matrisen  $A$  uansett hvilken input-vektor  $\mathbf{b}$  vi har — det er slik gjenbruk som ofte gjør matriser nyttige.

Vi kan illustrere situasjonen med diagrammet nedenfor; en  $m \times n$ -matrise  $A$  transformerer vektorer  $\mathbf{x} \in \mathbb{R}^n$  til vektorer  $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ :

$$\begin{array}{ccc} & A & \\ \mathbf{x} \in \mathbb{R}^n & \xrightarrow{\hspace{1cm}} & \mathbf{y} = A\mathbf{x} \in \mathbb{R}^m \end{array}$$

I mange situasjoner må vi foreta flere transformasjoner etter hverandre. Neste diagram viser en slik situasjon; først transformeres  $\mathbf{x} \in \mathbb{R}^k$  til  $\mathbf{y} = B\mathbf{x} \in \mathbb{R}^n$  ved hjelp av matrisen  $B$ , og deretter transformeres  $\mathbf{y} \in \mathbb{R}^n$  til  $\mathbf{z} = A\mathbf{y} = A(B\mathbf{x}) \in \mathbb{R}^m$  ved hjelp av matrisen  $A$ :

$$\begin{array}{ccccc} & B & & A & \\ \mathbf{x} \in \mathbb{R}^k & \xrightarrow{\hspace{1cm}} & \mathbf{y} = B\mathbf{x} \in \mathbb{R}^n & \xrightarrow{\hspace{1cm}} & \mathbf{z} = A\mathbf{y} = A(B\mathbf{x}) \in \mathbb{R}^m \end{array}$$

Det er naturlig å spørre om det finnes en matrise  $C$  som tar oss direkte fra  $\mathbf{x}$  til  $\mathbf{z}$  uten å gå veien om  $\mathbf{y}$ , dvs. slik at  $\mathbf{z} = C\mathbf{x}$ . Neste diagram viser hvordan en slik  $C$  vil fungere:

$$\begin{array}{ccccc} & B & & A & \\ \mathbf{x} \in \mathbb{R}^k & \xrightarrow{\hspace{1cm}} & \mathbf{y} = B\mathbf{x} \in \mathbb{R}^n & \xrightarrow{\hspace{1cm}} & \mathbf{z} = A\mathbf{y} = A(B\mathbf{x}) \in \mathbb{R}^m \\ & \searrow C & & \nearrow \parallel & \\ & & \mathbf{z} = C\mathbf{x} & & \end{array}$$

Det viser seg at det eksisterer en slik matrise  $C$ , og vi skal nå finne den. Før vi begynner kan det være greit å bestemme dimensjonene til de tre matrisene våre  $A$ ,  $B$  og  $C$ . Vi ser at  $A$  transformerer vektorer i  $\mathbb{R}^n$  til vektorer i  $\mathbb{R}^m$ , og det betyr at  $A$  er en  $m \times n$ -matrise. Tilsvarende transformerer  $B$  vektorer i  $\mathbb{R}^k$  til vektorer i  $\mathbb{R}^n$ , og det betyr at  $B$  er en  $n \times k$ -matrise. Den ukjente matrisen  $C$  skal transformere vektorer i  $\mathbb{R}^k$  til vektorer i  $\mathbb{R}^m$ , og må derfor være en  $m \times k$ -matrise. Vi har derfor

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{pmatrix}$$

som våre kjente matriser, og

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mk} \end{pmatrix}$$

som vår ukjente matrise. Hvis  $\mathbf{x}$  er vektoren

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

ser vi at

$$\mathbf{y} = B\mathbf{x} = \begin{pmatrix} b_{11}x_1 + b_{12}x_2 + \cdots + b_{1k}x_k \\ b_{21}x_1 + b_{22}x_2 + \cdots + b_{2k}x_k \\ \vdots \\ b_{n1}x_1 + b_{n2}x_2 + \cdots + b_{nk}x_k \end{pmatrix}$$

Ser vi på den  $j$ -te raden i denne ligningen, får vi  $y_j = b_{j1}x_1 + b_{j2}x_2 + \cdots + b_{jk}x_k$ . Vi har også

$$\mathbf{z} = A\mathbf{y} = \begin{pmatrix} a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ \vdots \\ a_{m1}y_1 + a_{m2}y_2 + \cdots + a_{mn}y_n \end{pmatrix}$$

og setter vi  $y_j = b_{j1}x_1 + b_{j2}x_2 + \cdots + b_{jk}x_k$  inn i dette uttrykket, ser vi at den  $i$ -te komponenten til  $\mathbf{z}$  er gitt ved

$$\begin{aligned} z_i &= a_{i1}y_1 + a_{i2}y_2 + \cdots + a_{in}y_n = a_{i1}(b_{11}x_1 + b_{12}x_2 + \cdots + b_{1k}x_k) + \\ &\quad + a_{i2}(b_{21}x_1 + b_{22}x_2 + \cdots + b_{2k}x_k) + \cdots + a_{in}(b_{n1}x_1 + b_{n2}x_2 + \cdots + b_{nk}x_k) \end{aligned}$$

Samler vi alle  $x_1$ -ledd for seg, alle  $x_2$ -ledd for seg osv, får vi:

$$\begin{aligned} z_i &= (a_{i1}b_{11} + a_{i2}b_{21} + \cdots + a_{in}b_{n1})x_1 + \\ &\quad + (a_{i1}b_{12} + a_{i2}b_{22} + \cdots + a_{in}b_{n2})x_2 + \cdots \\ &\quad + (a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk})x_k \end{aligned}$$

På den annen side: hvis vi tenker oss at  $\mathbf{z}$  fremkommer direkte fra  $\mathbf{x}$  ved bruk av matrisen  $C$ , har vi

$$\mathbf{z} = C\mathbf{x} = \begin{pmatrix} c_{11}x_1 + c_{12}x_2 + \cdots + c_{1k}x_k \\ c_{21}x_1 + c_{22}x_2 + \cdots + c_{2k}x_k \\ \vdots \\ c_{m1}x_1 + c_{m2}x_2 + \cdots + c_{mk}x_k \end{pmatrix}$$

dvs. at den  $i$ -te komponenten til  $z$  er gitt ved

$$z_i = c_{i1}x_1 + c_{i2}x_2 + \cdots + c_{ik}x_k$$

Skal de to uttrykkene for  $z_i$  være like, må vi ha

$$c_{i1} = a_{i1}b_{11} + a_{i2}b_{21} + \cdots + a_{in}b_{n1}$$

$$c_{i2} = a_{i1}b_{12} + a_{i2}b_{22} + \cdots + a_{in}b_{n2}$$

osv. ned til

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk}$$

Generelt har vi altså

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

Legg merke til at dette er skalarproduktet av den  $i$ -te raden i  $A$  med den  $j$ -te søylen i  $B$ .

La oss oppsummere. Vi har vist at den matrisen

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mk} \end{pmatrix}$$

som i én operasjon utfører den samme transformasjonen som  $B$  etterfulgt av  $A$ , er gitt ved

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

Siden det å transformere vektorer er det viktigste matriser gjør, tar vi denne formelen som utgangspunkt for vår definisjon av matriseprodukt.

**Definisjon 1.6.1** *Anta at  $A$  er en  $m \times n$ -matrise og at  $B$  er en  $n \times k$ -matrise. Da er matriseproduktet  $C = AB$  definert som  $m \times k$ -matrisen  $C$  med komponenter*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

*Vi får altså den  $ij$ -te komponenten i  $C$  ved å ta skalarproduktet av den  $i$ -te raden i  $A$  med den  $j$ -te søylen i  $B$ .*

Figuren viser grafisk hvordan vi finner det  $ij$ -te elementet i produktmatrisen  $C$ : Vi tar skalarproduktet av den  $i$ -te raden i  $A$  med den  $j$ -te søylen i  $B$ :

$$\left( \begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \boxed{\begin{array}{cccc} a_{i1} & a_{i2} & \cdots & a_{in} \end{array}} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array} \right) \left( \begin{array}{cccc} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} \\ \boxed{\begin{array}{c} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{array}} \end{array} \right)$$

Legg merke til at matriseproduktet  $AB$  bare er definert når  $A$  og  $B$  passer sammen i størrelse: radene i  $A$  må være like lange som søylene i  $B$ . Dette betyr at den “siste” dimensjonen  $n$  i  $m \times n$ -matrisen  $A$  er lik den “første” dimensjonen  $n$  i  $n \times k$ -matrisen  $B$ . Legg merke til at hvis vi stryker de to  $n$ -ene i  $m \times n$  og  $n \times k$ , sitter vi igjen med størrelsen  $m \times k$  til produktmatrisen  $C$ .

**Eksempel 1:** Regn ut  $AB$  når

$$A = \begin{pmatrix} 2 & 3 \\ 1 & -1 \\ 4 & -2 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} 4 & -3 & 2 \\ -1 & 4 & 5 \end{pmatrix}$$

Siden  $A$  er en  $3 \times 2$  og  $B$  er  $2 \times 3$ -matrise, eksisterer produktet og er en  $3 \times 3$ -matrise. Vi får

$$\begin{aligned} AB &= \begin{pmatrix} 2 \cdot 4 + 3 \cdot (-1) & 2 \cdot (-3) + 3 \cdot 4 & 2 \cdot 2 + 3 \cdot 5 \\ 1 \cdot 4 + (-1) \cdot (-1) & 1 \cdot (-3) + (-1) \cdot 4 & 1 \cdot 2 + (-1) \cdot 5 \\ 4 \cdot 4 + (-2) \cdot (-1) & 4 \cdot (-3) + (-2) \cdot 4 & 4 \cdot 2 + (-2) \cdot 5 \end{pmatrix} = \\ &= \begin{pmatrix} 5 & 6 & 19 \\ 5 & -7 & -3 \\ 18 & -20 & -2 \end{pmatrix} \end{aligned}$$



Legg merke til at selv i de tilfellene hvor  $AB$  er definert, kan vi ikke nødvendigvis regne ut  $BA$ . Her prøver vi nemlig å multiplisere en  $n \times k$ -matrise med en  $m \times n$ -matrise, og det er bare mulig når  $k = m$ . Selv i de tilfellene hvor både  $AB$  og  $BA$  er definert, er de som regel forskjellige. Her er et eksempel.

**Eksempel 2:** Regn ut  $AB$  og  $BA$  når

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix} \quad \text{og} \quad B = \begin{pmatrix} -3 & 2 \\ 4 & 5 \end{pmatrix}$$

Vi får

$$AB = \begin{pmatrix} 2 \cdot (-3) + 1 \cdot 4 & 2 \cdot 2 + 1 \cdot 5 \\ (-1) \cdot (-3) + 4 \cdot 4 & (-1) \cdot 2 + 4 \cdot 5 \end{pmatrix} = \begin{pmatrix} -2 & 9 \\ 19 & 18 \end{pmatrix}$$

og

$$BA = \begin{pmatrix} (-3) \cdot 2 + 2 \cdot (-1) & (-3) \cdot 1 + 2 \cdot 4 \\ 4 \cdot 2 + 5 \cdot (-1) & 4 \cdot 1 + 5 \cdot 4 \end{pmatrix} = \begin{pmatrix} -8 & 5 \\ 3 & 24 \end{pmatrix}$$

Vi har altså  $AB \neq BA$ .



Regnereglen  $ab = ba$  (som altså ikke gjelder for matriser!) kalles for *den kommutative lov*. Siden matrisemultiplikasjon ikke oppfyller denne regelen, sier vi at matriseproduktet er *ikke-kommutativt*. Vi har tidligere sett at vektorproduktet ikke er kommutativt (det er heller ikke skalarproduktet for komplekse  $n$ -tupler), men der har vi i hvert fall en grei regel for sammenhengen mellom  $\mathbf{a} \times \mathbf{b}$  og  $\mathbf{b} \times \mathbf{a}$ , nemlig at  $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$ . Det har vi ikke for multiplikasjon av matriser; kjenner vi  $AB$ , har vi fortsatt ingen anelse om hva  $BA$  er!

Det tar tid å vende seg til ikke-kommulative operasjoner, og det eneste rådet vi kan gi er: Vær forsiktig og begrunn hvert eneste skritt når du regner med slike operasjoner! Heldigvis er det mange regneregler som fortsatt gjelder for matriseprodukter:

**Setning 1.6.2 (Regneregler for matrisemultiplikasjon)** *I hvert av punktene nedenfor antar vi at  $A$ ,  $B$  og  $C$  er matriser slik at regneoperasjonene er definert.*

- (i)  $(AB)C = A(BC)$
- (ii)  $A(B + C) = AB + AC$
- (iii)  $(B + C)A = BA + CA$
- (iv)  $(sA)B = A(sB) = s(AB)$  for alle tall  $s$
- (v)  $(AB)^T = B^T A^T$

*Bevis:* Bevisene går rett og slett ut på å gange ut venstre og høyre side av likhetene og sjekke at svarene er like. Vi tar (i) og (v) som eksempler — de andre punktene er atskillig enklere enn (i):

\* (i) (Dette punktet er kronglete og du mister ikke noe i fortsettelsen hvis du hopper over beviset.) Vi skal bevise at de to matrisene  $X = (AB)C$  og  $Y = A(BC)$  er like. Før man begynner, er det lurt å være helt sikker på hva som er forskjellen på de to uttrykkene. Poenget er at vi alltid gjennomfører operasjonene inni parentesene først. Når vi regner ut  $X = (AB)C$ , multipliserer vi altså først sammen  $A$  og  $B$  til en matrise  $D = AB$ , og så multipliserer vi  $D$  fra høyre med  $C$  slik at vi får  $X = DC$ . Når vi regner ut  $Y = A(BC)$ , ganger vi først sammen  $B$  og  $C$  til en matrise  $E = BC$ , og så ganger vi denne fra venstre med  $A$  slik at vi får  $Y = AE$ . Vår oppgave er altså å vise at  $DC = AE$ . Formulert på denne måten er ikke likheten så selvfølgelig som den kan se ut ved første blikk.

Før vi begynner på regningene, er det greit å se litt på dimensjonene til de involverte matrisene. For at produktene  $(AB)C$  og  $A(BC)$  skal gi mening, må det finnes tall  $m, n, k, l$  slik at  $A$  er en  $m \times n$ -matrise,  $B$  en  $n \times k$ -matrise og  $C$  en  $k \times l$ -matrise. Dermed blir  $D = AB$  en  $m \times k$ -matrise og  $E = BC$  en

$n \times l$ -matrise. Legg merke til at både  $X = DC$  og  $Y = AE$  er  $m \times l$ -matriser, så dimensjonsmessig er i hvert fall de to matrisene like.

Vi er nå klare til å gjøre regningene våre. Vi observerer først at det  $ij$ -te elementet i  $X = DC$  er lik

$$x_{ij} = d_{i1}c_{1j} + d_{i2}c_{2j} + \cdots + d_{ik}c_{kj} = \sum_{p=1}^k d_{ip}c_{pj}$$

Siden  $D = AB$  er videre

$$d_{ip} = a_{i1}b_{1p} + a_{i2}b_{2p} + \cdots + a_{in}b_{np} = \sum_{q=1}^n a_{iq}b_{qp}$$

Setter vi dette inn i uttrykket ovenfor, får vi

$$x_{ij} = \sum_{p=1}^k d_{ip}c_{pj} = \sum_{p=1}^k \left( \sum_{q=1}^n a_{iq}b_{qp} \right) c_{pj} = \sum_{p=1}^k \sum_{q=1}^n a_{iq}b_{qp}c_{pj}$$

dvs. at  $x_{ij}$  er summen av alle mulige kombinasjoner  $a_{iq}b_{qp}c_{pj}$ .

På tilsvarende måte regner vi ut  $Y = AE$ . Det  $ij$ -te elementet i  $Y$  er

$$y_{ij} = a_{i1}e_{1j} + a_{i2}e_{2j} + \cdots + a_{in}e_{nj} = \sum_{q=1}^n a_{iq}e_{qj}$$

Siden  $E = BC$  er videre

$$e_{qj} = b_{q1}c_{1j} + b_{q2}c_{2j} + \cdots + b_{qk}c_{kj} = \sum_{p=1}^k b_{qp}c_{pj}$$

Setter vi dette inn i uttrykket ovenfor, får vi

$$y_{ij} = \sum_{q=1}^n a_{iq}e_{qj} = \sum_{q=1}^n a_{iq} \left( \sum_{p=1}^k b_{qp}c_{pj} \right) = \sum_{q=1}^n \sum_{p=1}^k a_{iq}b_{qp}c_{pj}$$

Dette viser at  $y_{ij}$  er lik summen av alle mulige kombinasjoner  $a_{iq}b_{qp}c_{pj}$ , akkurat som  $x_{ij}$  er. Altså er  $x_{ij} = y_{ij}$ , og beviset for (i) er ferdig.

(v) Anta at  $A$  er en  $m \times n$ -matrise og at  $B$  er en  $n \times k$ -matrise. Da er  $A^T$  en  $n \times m$ -matrise og  $B^T$  en  $k \times n$ -matrise, så både  $(AB)^T$  og  $B^T A^T$  er  $k \times m$ -matriser. Det  $ij$ -te elementer i  $C = (AB)^T$  er lik det  $ji$ -te elementet i  $AB$ , dvs:

$$c_{ij} = a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni}$$

På den annen side fremkommer det  $ij$ -te elementet i  $D = B^T A^T$  ved at vi tar skalarproduktet av den  $i$ -te raden i  $B^T$  med den  $j$ -te søylen i  $A^T$ . Men

den  $i$ -te raden i  $B^T$  er den  $i$ -te søylen i  $B$ , og den  $j$ -te søylen i  $A^T$ , er den  $j$ -te raden i  $A$ , så dette blir skalarproduktet mellom den  $j$ -te raden i  $A$  og den  $i$ -søylen i  $B$ , dvs.

$$d_{ij} = a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni}$$

Altså er  $c_{ij} = d_{ij}$ , og beviset er fullført.  $\square$

Regnereglene ovenfor gjør det enklere å regne med matriser, men vi må som sagt være forsiktige med en del ting — spesielt at vi ikke bytter om på rekkefølgen av matrisene (husk at vanligvis er  $AB$  ulik  $BA$ ). Vi kan heller ikke forkorte matriser på vanlig måte — å forkorte betyr egentlig å multiplisere med et inverst element, og foreløpig vet vi ingen ting om inverse matriser.

La oss ta med oss en liten observasjon. Dersom  $A$  er en  $m \times n$ -matrise og  $\mathbf{b}$  er en søylevektor med  $n$ -komponenter, kan vi danne produktet  $A\mathbf{b}$  på to måter — vi kan enten tenke på det som produktet av matrisen  $A$  og vektoren  $\mathbf{b}$  slik vi definerte det i forrige seksjon, eller vi kan tenke på det som produktet av  $m \times n$ -matrisen  $A$  med  $n \times 1$ -matrisen

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

slik vi har definert det i denne seksjonen. Det er lett å se at disse to måtene gir akkurat samme svar. Dette betyr at vi kan bruke regnereglene ovenfor også når noen av matrisene er vektorer.

**Eksempel 3:** Vi har en  $m \times n$ -matrise  $A$  og en  $n \times k$ -matrise  $B$ . Vi lar  $B$  transformere en søylevektor  $\mathbf{x} \in \mathbb{R}^k$  ved  $\mathbf{y} = B\mathbf{x}$ . Så lar vi  $A$  transformere  $\mathbf{y}$  slik at  $\mathbf{z} = A\mathbf{y}$ . Det samlede resultatet er:

$$\mathbf{z} = A\mathbf{y} = A(B\mathbf{x})$$

Ved hjelp av regneregel (i) ovenfor kan vi flytte parentesene:

$$\mathbf{z} = A\mathbf{y} = A(B\mathbf{x}) = (AB)\mathbf{x}$$

Dette viser noe vi allerede vet, nemlig at når vi transformerer en vektor to ganger ved først å bruke  $B$  og deretter  $A$ , så er det det samme som å transformere én gang ved hjelp av produktet  $AB$ .  $\clubsuit$

Vi avslutter denne seksjonen med en ulikhet som vi vil få bruk for i neste kapittel. Hvis

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

er en  $m \times n$ -matrise, definerer vi *normen til A* til å være

$$\|A\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$$

(vi legger altså sammen kvadratene  $a_{ij}^2$  av alle leddene og tar kvadratroten til summen).

**Setning 1.6.3** *Hvis A er en  $m \times n$ -matrise og  $\mathbf{x} \in \mathbb{R}^n$ , så er*

$$|A\mathbf{x}| \leq \|A\| |\mathbf{x}|$$

*Med andre ord: Når vi ganger vektoren  $\mathbf{x}$  med matrisen A, så øker lengden maksimalt med en faktor  $\|A\|$ .*

*Bevis:* La  $\mathbf{y} = A\mathbf{x}$ . Da er

$$|A\mathbf{x}| = \sqrt{y_1^2 + y_2^2 + \cdots + y_m^2}$$

der  $y_1, y_2, \dots, y_m$  er komponentene til  $\mathbf{y}$ . Dersom  $\mathbf{a}_i = (a_{i1}, a_{i2}, \dots, a_{in})$  er den i-te linjen i A, vet vi at  $y_i = \mathbf{a}_i \cdot \mathbf{x}$ . Ifølge Schwarz' ulikhet er dermed  $|y_i| \leq |\mathbf{a}_i| |\mathbf{x}|$ . Kvadrerer vi, får vi  $y_i^2 \leq |\mathbf{a}_i|^2 |\mathbf{x}|^2$ . Setter vi dette inn i uttrykket for  $A\mathbf{x}$ , ser vi at

$$\begin{aligned} |A\mathbf{x}| &= \sqrt{y_1^2 + y_2^2 + \cdots + y_m^2} \leq \\ &\leq \sqrt{|\mathbf{a}_1|^2 |\mathbf{x}|^2 + |\mathbf{a}_2|^2 |\mathbf{x}|^2 + \cdots + |\mathbf{a}_m|^2 |\mathbf{x}|^2} = \\ &= \sqrt{|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2 + \cdots + |\mathbf{a}_m|^2} |\mathbf{x}| = \|A\| |\mathbf{x}| \end{aligned}$$

der vi har brukt at

$$\sqrt{|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2 + \cdots + |\mathbf{a}_m|^2} = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2} = \|A\|$$

□

**Bemerkning:**  $|A|$  er normalt ikke det *minste* tallet slik at  $|A\mathbf{x}| \leq |A||\mathbf{x}|$  for alle  $\mathbf{x}$ . Vi skal komme tilbake til denne problemstillingen i seksjon 5.2.

**MATLAB-kommentar:** Har du skrevet inn to matriser  $A$  og  $B$  med passende dimensjoner, får du MATLAB til å regne ut produktet ved å skrive  $A*B$ .

### Oppgaver til seksjon 1.6

1. Regn ut  $AB$  og  $BA$  når

$$\begin{aligned} \text{a) } A &= \begin{pmatrix} 1 & -2 \\ 3 & 1 \end{pmatrix} \text{ og } B = \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \\ \text{b) } A &= \begin{pmatrix} 1 & -1 & 0 \\ -2 & 0 & 1 \\ -1 & 2 & 1 \end{pmatrix} \text{ og } B = \begin{pmatrix} 0 & 2 & 1 \\ -1 & -2 & 0 \\ 3 & -1 & 2 \end{pmatrix} \end{aligned}$$

2. Regn ut  $AB$  når

$$A = \begin{pmatrix} 1 & -2 & 3 \\ 0 & -1 & 2 \end{pmatrix} \text{ og } B = \begin{pmatrix} 2 & 1 \\ 0 & -3 \\ 1 & 0 \end{pmatrix}.$$

3. Regn ut  $AB$  når

$$A = \begin{pmatrix} 1 & -2 \\ 3 & 0 \\ -1 & 2 \end{pmatrix} \text{ og } B = \begin{pmatrix} 2 & 1 & 0 \\ -3 & 1 & 1 \end{pmatrix}.$$

4. a) Hvilken dimensjon har produktmatrisen  $AB$  hvis  $A$  er en  $8 \times 6$ -matrise og  $B$  er en  $6 \times 9$ -matrise?

b) Hvilken dimensjon har matrisen  $B$  hvis  $A$  er en  $4 \times 3$ -matrise og produktmatrisen  $AB$  er en  $4 \times 5$ -matrise?

c) Hvor mange søyler har matrisen  $B$  hvis produktet  $AB$  er en  $5 \times 7$ -matrise?

5. Gitt følgende matriser:

$$A = \begin{pmatrix} 1 & 9 \\ 0 & 8 \\ 7 & 4 \end{pmatrix}, B = \begin{pmatrix} 2 & 8 \\ 9 & 6 \end{pmatrix} \text{ og } C = \begin{pmatrix} 3 & 0 & 5 \\ 0 & 9 & 4 \end{pmatrix}$$

Regn ut følgende uttrykk hvis det er definert (hvis uttrykket ikke er definert, skal du begrunne hvorfor):

- a)  $AB$
- b)  $AC$
- c)  $A(B + C)$
- d)  $(BC)^T$
- e)  $B^T C^T$
- f)  $(A + C^T)B$
- g)  $B(A^T - 2C)$

**6.** Gitt matrisene

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}, B = \begin{pmatrix} 1 & 1 \\ 3 & 4 \end{pmatrix}, C = \begin{pmatrix} 2 & 5 \\ 3 & 4 \end{pmatrix} \text{ og } D = \begin{pmatrix} 3 & 7 \\ 0 & 0 \end{pmatrix}$$

Sjekk at  $AB = AC$  selv om  $B \neq C$ . Sjekk også at  $AD = 0$  selv om  $A \neq 0$  og  $D \neq 0$ .

**7.** Gitt matrisene  $D = \begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix}$  og  $E = \begin{pmatrix} 2 & 7 & 1 \\ 8 & 2 & 8 \\ 1 & 8 & 2 \end{pmatrix}$ .

Finn første rad i produktmatrisen  $DE$ . Finn deretter andre søyle i produktmatrisen  $DE$ .

**8.** a) Hvis første og andre søyle i matrisen  $B$  er like, hva kan du si om første og andre søyle i produktmatrisen  $AB$  (dersom  $AB$  er definert)?

b) Hvis andre søyle i matrisen  $B$  bare består av nuller, hva kan du si om andre søyle i produktmatrisen  $AB$ ?

**9.** Finn normen  $\|A\|$  til matrisen  $A = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 1 & 0 \end{pmatrix}$ .

**10.** Dersom  $A$  er en  $n \times n$ -matrise (dvs. har like mange rader og søyler) definerer vi potensene  $A^k$  på vanlig måte:  $A^2 = AA$ ,  $A^3 = AA^2$ ,  $A^4 = AA^3$  osv. Regn ut  $A^2$  og  $A^3$  når  $A = \begin{pmatrix} 1 & -2 \\ -1 & 1 \end{pmatrix}$ .

**11.** La  $A = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & 3 \end{pmatrix}$ ,  $B = \begin{pmatrix} 2 & -1 \\ -1 & 0 \\ 3 & -1 \end{pmatrix}$ ,  $\mathbf{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ .

a) Finn  $\mathbf{y} = B\mathbf{x}$ .

b) Regn ut  $C = AB$ .

c) Regn ut både  $A\mathbf{y}$  og  $C\mathbf{x}$ . Sammenlign resultatene.

**12.** I en barnehage kjører barna sanden frem og tilbake i trillebår mellom tre sandkasser  $A$ ,  $B$  og  $C$  (og mister noe på veien!). Av den sanden som er i sandkasse  $A$  ved begynnelsen av dagen, vil 70% være i  $A$  ved slutten av dagen, 15% vil være i  $B$ , 10% i  $C$  (og resten vil være mistet). Av den sanden som starter dagen i  $B$ , vil 75% være i  $B$  ved slutten av dagen, 5% vil være i  $A$  og 10% i  $C$ . Av den sanden som startet i  $C$ , vil 70% være i  $C$  ved slutten av dagen, 10% vil være i  $A$  og 15% i  $B$ .

a) Finn en matrise  $M$  slik at hvis  $x, y, z$  er henholdsvis antall liter sand i  $A$ ,  $B$  og  $C$  ved begynnelsen av dagen, og  $\mathbf{u} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ , så angir komponentene til  $\mathbf{v} = M\mathbf{u}$  hvor mange liter sand det er i hver kasse ved slutten av dagen.

b) Etter at barna har gått hjem, forsøker personalet å fordele sanden på nytt. De flytter 20% av sanden i  $B$  til  $A$  og 5% av sanden i  $C$  til  $B$ . Finn en matrise  $N$  slik at komponentene til  $\mathbf{w} = N\mathbf{v}$  angir hvor mye sand det nå er i hver sandkasse.

- c) Regn ut  $K = NM$ . Anta at fordeling ved begynnelsen av dagen var  $x = 200$ ,  $y = 300$ ,  $z = 400$ . Finn fordelingen på slutten av dagen etter at personalet har omfordelt sanden.
- d) Hvor mye sand vil det være i hver kasse på slutten av neste dag dersom den forløper på tilsvarende måte?

**13.** Bevis punktene (ii), (iii) og (iv) i setning 1.6.2.

## 1.7 Identitetsmatriser og inverse matriser

I denne seksjonen skal vi bare arbeide med *kvadratiske* matriser, dvs. matriser med like mange rader som søyler. Matrisene vil altså være  $n \times n$ -matriser for et helt tall  $n$ . En spesiell slik matrise er *identitetsmatrisen*

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

som har 1-ere på diagonalen og 0-er overalt ellers. Ganger du en annen  $n \times n$ -matrise  $A$  med  $I_n$  (gjør det!), ser du at

$$AI_n = A \quad \text{og} \quad I_n A = A$$

Uansett om vi multipliserer  $A$  med  $I_n$  fra høyre eller venstre, får vi altså  $A$  tilbake. Blant tall er det bare 1 som har en tilsvarende egenskap; ganger vi et tall med 1, får vi tallet tilbake. Identitetsmatrisen  $I_n$  spiller derfor mye av den samme rollen for matrisemultiplikasjon som 1 gjør for vanlig multiplikasjon.

Et tall  $a \neq 0$  har alltid et inverst tall  $a^{-1}$ . Dette tallet er definert ved at produktet av  $a$  og  $a^{-1}$  er lik 1, dvs.  $aa^{-1} = 1$ . Vi kan på tilsvarende måte definere inverse matriser:

**Definisjon 1.7.1** *Anta at  $A$  er en  $n \times n$ -matrise. En  $n \times n$ -matrise  $X$  kalles en invers matrise til  $A$  dersom*

$$AX = XA = I_n$$

(siden matrisemultiplikasjon ikke er kommutativ, krever vi å få  $I_n$  som svar uansett hvilken side vi multipliserer fra).

Som allerede nevnt har alle tall bortsett fra 0 en invers. For matriser er det mer komplisert; det finnes mange matriser som ikke har invers! Vi skal komme tilbake til spørsmålet om når en matrise har en invers etter hvert, men foreløpig nøyer vi oss med å vise noe enklere — nemlig at ingen matrise har mer enn én invers.

**Setning 1.7.2** En  $n \times n$ -matrise har høyst én invers.

*Bevis:* Anta at både  $X$  og  $Y$  er inverser til  $A$ . Da har vi

$$X = I_n X = (YA)X = Y(AX) = YI_n = Y$$

Altså er de to inversene like. ♣

Nå som vi vet at det finnes høyst én invers, kan vi være mer konkrete i språkbruken.

**Definisjon 1.7.3** En  $n \times n$ -matrise  $A$  kalles inverterbar dersom den har en invers, og den inverse matrisen betegnes med  $A^{-1}$ . En matrise som ikke er inverterbar, kalles singulær.

Selv om vi ikke skal utvikle noen teori for hvordan man finner inverse matriser på det nåværende tidspunkt (det får vente til kapittel 4) er det instruktivt å se på noen enkle eksempler.

**Eksempel 1:** Finn den inverse matrisen til

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

Vi må finne en  $2 \times 2$ -matrise  $X = \begin{pmatrix} x & y \\ z & u \end{pmatrix}$  slik at  $AX = I_2$  og  $XA = I_2$ .  
Siden

$$AX = \begin{pmatrix} 3x - z & 3y - u \\ x + 2z & y + 2u \end{pmatrix}$$

kan ligningen  $AX = I_2$  skrives

$$\begin{pmatrix} 3x - z & 3y - u \\ x + 2z & y + 2u \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Dette gir oss fire ligninger med fire ukjente

$$\begin{aligned} 3x - z &= 1 \\ 3y - u &= 0 \\ x + 2z &= 0 \\ y + 2u &= 1 \end{aligned}$$

Dette ligningssystemet er lett å løse (legg merke til at vi har to ligninger som bare inneholder  $x$  og  $z$ , og to som bare inneholder  $y$  og  $u$ ), og vi får  $x = \frac{2}{7}$ ,  $y = \frac{1}{7}$ ,  $z = -\frac{1}{7}$  og  $u = \frac{3}{7}$ . Dette betyr at

$$X = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix}$$

tilfredsstiller ligningen  $AX = I_2$ . Vi må også sjekke at  $X$  tilfredsstiller det andre kravet til en invers matrise, nemlig at  $XA = I_2$ :

$$\begin{aligned} XA &= \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix} \begin{pmatrix} 3 & -1 \\ 1 & 2 \end{pmatrix} = \\ &= \begin{pmatrix} \frac{2}{7} \cdot 3 + \frac{1}{7} \cdot 1 & \frac{2}{7} \cdot (-1) + \frac{1}{7} \cdot 2 \\ -\frac{1}{7} \cdot 3 + \frac{3}{7} \cdot 1 & -\frac{1}{7} \cdot (-1) + \frac{3}{7} \cdot 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Siden vi nå har vist at  $A$  er inverterbar og at

$$A^{-1} = \begin{pmatrix} \frac{2}{7} & \frac{1}{7} \\ -\frac{1}{7} & \frac{3}{7} \end{pmatrix}$$

er oppgaven vår fullført. ♣

**Bemerkning:** Det kan virke som vi har flaks på slutten av eksemplet ovenfor når vi sjekker at  $XA = I_2$  — vi har laget  $X$  slik at  $AX = I_2$ , men siden matrisemultiplikasjon ikke er kommutativ, er det ingen grunn til å tro at dette vil medføre at  $XA = I_2$ . Det viser seg imidlertid at det alltid er slik; dersom  $n \times n$ -matrisen  $X$  er en *ensidig invers* til  $n \times n$ -matrisen  $A$  (dvs. at vi enten har  $AX = I_n$  eller  $XA = I_n$ ), så er  $A$  inverterbar og  $X = A^{-1}$ . Det er imidlertid forbausende vanskelig å bevise dette, og vi må utsette det til en senere anledning.

La oss nå se på en matrise som *ikke* har en invers:

**Eksempel 2:** La

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$$

Vi forsøker å finne en invers matrise

$$X = \begin{pmatrix} x & y \\ z & u \end{pmatrix}$$

på samme måte som i forrige eksempel, nemlig ved å kreve at  $AX = I_2$ . Siden

$$AX = \begin{pmatrix} x + 2z & y + 2u \\ 2x + 4z & 2y + 4u \end{pmatrix}$$

gir dette ligningssystemet

$$\begin{aligned} x + 2z &= 1 \\ y + 2u &= 0 \\ 2x + 4z &= 0 \\ 2y + 4u &= 1 \end{aligned}$$

Det er lett å se at dette ligningssystemet ikke har løsninger; ganger vi den første ligningen med 2, får vi  $2x + 4z = 2$  som er i åpenbar strid med den tredje ligningen.



De to eksemplene ovenfor viser at det er en nærliggende sammenheng mellom invertering av matriser og lineære ligningssystemer. Vi skal komme tilbake til denne sammenhengen senere når vi skal utvikle effektive metoder for å finne inverse matriser. Foreløpig nøyser vi oss med å se på noen enkle regneregler for inverse matriser:

**Setning 1.7.4** *Anta at  $A$  og  $B$  er inverterbare  $n \times n$ -matriser. Da er*

- (i)  $sA$  inverterbar for alle tall  $s \neq 0$ , og  $(sA)^{-1} = s^{-1}A^{-1}$
- (ii)  $AB$  inverterbar, og  $(AB)^{-1} = B^{-1}A^{-1}$
- (iii)  $A^T$  inverterbar, og  $(A^T)^{-1} = (A^{-1})^T$ .
- (iv)  $A^{-1}$  er inverterbar og  $(A^{-1})^{-1} = A$ .

*Bevis:* Vi konsentrerer oss om (ii), og overlater de (enkeltre!) (i), (iii) og (iv) til leserne. For å bevise (ii) må vi sjekke at

$$(AB)(B^{-1}A^{-1}) = I_n \quad \text{og} \quad (B^{-1}A^{-1})(AB) = I_n$$

Dette er en liten herjing i parentesflytting (husk setning 1.6.2(i)):

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= ((AB)B^{-1})A^{-1} = (A(BB^{-1}))A^{-1} = \\ &= (AI_n)A^{-1} = AA^{-1} = I_n \end{aligned}$$

Helt tilsvarende får vi:

$$\begin{aligned} (B^{-1}A^{-1})(AB) &= B^{-1}(A^{-1}(AB)) = B^{-1}((A^{-1}A)B) = \\ &= B^{-1}(I_nB) = B^{-1}B = I_n \end{aligned}$$

□

**Bemerkning:** Legg merke til at vi ikke har noen regler for de inverse til  $A + B$  og  $A - B$ . Disse matrisene behøver ikke være inverterbare selv om  $A$  og  $B$  er det, og selv i de tilfellene hvor de er inverterbare, finnes det ikke noen enkel måte å finne den inverse på.

Dersom en matrise  $A$  er inverterbar, kan vi forkorte den bort i ligninger av typen  $AX = AB$  og få  $X = B$ . Grunnen er at vi kan multiplisere fra venstre med  $A^{-1}$  på begge sider av ligningen:

$$\begin{aligned} AX = AB &\implies A^{-1}(AX) = A^{-1}(AB) \implies (A^{-1}A)X = (A^{-1}A)B \implies \\ &\implies I_n X = I_n B \implies X = B \end{aligned}$$

På tilsvarende måte kan vi forkorte med  $A$  i ligningen  $XA = BA$  og få  $X = B$  (i dette tilfellet må vi gange ligningen fra *høyre* med  $A^{-1}$ ). Vi kan imidlertid *ikke* forkorte med  $A$  i ligningen  $AX = BA$  — uansett om vi ganger med  $A^{-1}$  fra venstre eller høyre, vil det være en  $A$  vi ikke greier å forkorte bort.

I forrige seksjon tenkte vi på matriser som transformasjoner; når vi ganger en vektor  $\mathbf{x}$  med en matrise  $A$ , transformerer vi  $\mathbf{x}$  til en ny vektor  $\mathbf{y} = A\mathbf{x}$ . Dersom  $A$  er inverterbar, kan vi gange den siste ligningen fra venstre med den inverse matrisen  $A^{-1}$ . Vi får da

$$A^{-1}\mathbf{y} = A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}$$

Dette viser at  $A^{-1}$  er den omvendte transformasjonen til  $A$  — hvis  $A$  transformerer  $\mathbf{x}$  til  $\mathbf{y}$ , vil  $A^{-1}$  transformere  $\mathbf{y}$  til  $\mathbf{x}$ . Denne ideen om  $A^{-1}$  som den omvendte transformasjonen til  $A$  er viktig når man bruker matriser i praksis.

$$\begin{array}{ccc} & A & \\ \mathbf{x} = A^{-1}\mathbf{y} & \xrightarrow{\hspace{2cm}} & \mathbf{y} = A\mathbf{x} \\ & \xleftarrow{\hspace{2cm}} A^{-1} & \end{array}$$

Siden det finnes mange transformasjoner vi ikke kan vende om på, gir dette bildet oss en bedre forståelse av hvorfor det finnes matriser som ikke er inverterbare. Hvis det f.eks. finnes to forskjellige vektorer  $\mathbf{x}$  og  $\mathbf{x}'$  som transformeres til den samme vektoren  $\mathbf{y}$  (dvs. vi har både  $\mathbf{y} = A\mathbf{x}$  og  $\mathbf{y} = A\mathbf{x}'$ ), så kan ikke transformasjonen vendes om; vi kan ikke starte med  $\mathbf{y}$  og transformere den til både  $\mathbf{x}$  og  $\mathbf{x}'$ !

**MATLAB-kommentar:** Det er en egen kommando for identitetsmatrisen  $I_n$  i MATLAB — du skriver `>>eye(n)` (der  $n$  er tallet du ønsker). Den inverse matrisen til  $A$  får du ved å skrive `>>inv(A)`.

### Oppgaver til seksjon 1.7

**1.** La  $A = \begin{pmatrix} 1 & 3 \\ -2 & 1 \end{pmatrix}$ . Kontroller at  $AI_2 = A$  og  $I_2A = A$  ved å gjennomføre utregningene.

**2.** La  $A = \begin{pmatrix} 1 & 0 & -2 \\ -3 & 1 & 4 \\ 2 & -3 & 4 \end{pmatrix}$ . Vis at  $B = \begin{pmatrix} 8 & 3 & 1 \\ 10 & 4 & 1 \\ \frac{7}{2} & \frac{3}{2} & \frac{1}{2} \end{pmatrix}$  er den inverse matrisen til  $A$  ved å regne ut  $AB$  og  $BA$ .

**3.** Avgjør om følgende matriser er inverterbare eller singulære:

$$A = \begin{pmatrix} 3 & -7 \\ 1 & -2 \end{pmatrix}, B = \begin{pmatrix} 5 & 0 \\ 0 & 4 \end{pmatrix}, C = \begin{pmatrix} 2 & 0 \\ 6 & 0 \end{pmatrix} \text{ og } D = \begin{pmatrix} 2 & 4 \\ 3 & 4 \end{pmatrix}$$

**4.** En inverterbar matrise  $A$  har en invers matrise som er gitt ved  $A^{-1} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$ . Finn matrisen  $A$ .

**5.** Gitt to inverterbare matriser  $A$  og  $B$ , hvor  $A^{-1} = \begin{pmatrix} 1 & 4 \\ 2 & 9 \end{pmatrix}$  og  $B^{-1} = \begin{pmatrix} 8 & 3 \\ 2 & 1 \end{pmatrix}$ . Finn  $(AB)^{-1}$ .

**6. a)** La  $\mathbf{a} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ . Vis at  $I_2\mathbf{a} = \mathbf{a}$ .

**b)** La  $\mathbf{b} = \begin{pmatrix} -3 \\ 2 \\ -1 \end{pmatrix}$ . Vis at  $I_3\mathbf{b} = \mathbf{b}$ .

**c)** Vis at  $I_n\mathbf{c} = \mathbf{c}$  for alle  $\mathbf{c} \in \mathbb{R}^n$ .

**7. a)** Anta at  $A = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$  der  $a, b \neq 0$ . Vis at  $A$  er inverterbar og at  $A^{-1} = \begin{pmatrix} a^{-1} & 0 \\ 0 & b^{-1} \end{pmatrix}$ .

**b)** Anta at  $B = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$  der  $a, b, c \neq 0$ . Vis at  $B$  er inverterbar og at  $B^{-1} = \begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & b^{-1} & 0 \\ 0 & 0 & c^{-1} \end{pmatrix}$ .

**c)** Formuler et tilsvarende resultat for  $n \times n$ -matriser.

**8.** Vis at dersom  $A$  og  $B$  er inverterbare, så er den inverse til  $(AB)^T$  lik  $(A^{-1})^T(B^{-1})^T$ .

**9.** Bevis punktene (i), (iii) og (iv) i setning 1.6.4.

**10. a)** Vis at  $2 \times 2$ -matrisen  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$  er inverterbar hvis og bare hvis  $ad - bc \neq 0$ .

- 0, og at den inverse matrisen i så fall er gitt ved  $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ .
- b) Bruk formelen fra punkt a) til å finne den inverse til matrisen  $A = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}$ .
- c) Bruk matrisen du fant i punkt b) til å løse ligningssystemet

$$2x - 5y = 3$$

$$-x + 3y = 2$$

*Hint:* Systemet kan skrives  $\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ .

## 1.8 Determinanter, arealer og volumer

Til enhver  $n \times n$ -matrise

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

hører det et tall som kalles *determinanten* til  $A$ , og som betegnes med

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

Generelle  $n \times n$ -determinanter kan være litt vanskelige å forstå seg på, så vi skal derfor begynne med  $2 \times 2$ - og  $3 \times 3$ -determinanter og deres geometriske egenskaper.

Hvis  $A$  er en  $2 \times 2$ -matrise

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

så er *determinanten* til  $A$  definert ved

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Legg merke til at uttrykket  $ad - bc$  fremkommer fra diagonalene i matrisen  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ; ganger vi sammen tallene i den ene diagonalen, får vi  $ad$ , og ganger vi sammen tallene i den andre diagonalen, får vi  $bc$ .

La oss regne ut en determinant.

**Eksempel 1:** Vi ser at

$$\begin{vmatrix} -3 & -5 \\ 4 & 2 \end{vmatrix} = (-3) \cdot 2 - (-5) \cdot 4 = -6 + 20 = 14 \quad \clubsuit$$

Dersom vi har to vektorer  $\mathbf{a} = (a_1, a_2)$ ,  $\mathbf{b} = (b_1, b_2)$ , kan vi lage en  $2 \times 2$ -determinant  $\det(\mathbf{a}, \mathbf{b})$  ved å legge inn vektorene som radene i en matrise på denne måten:

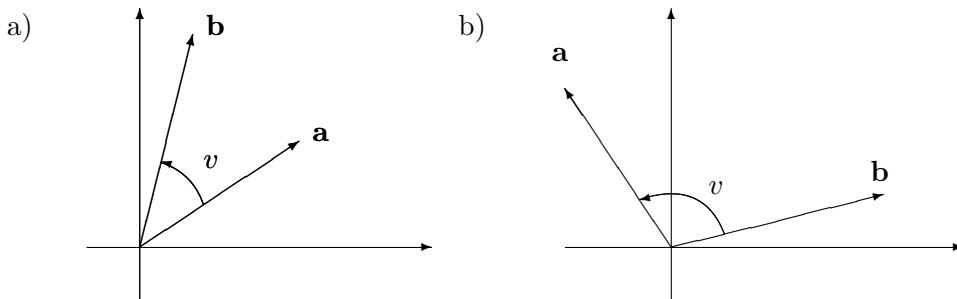
$$\det(\mathbf{a}, \mathbf{b}) = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1$$

Legg merke til at dersom vi bytter om rekkefølgen på vektorene  $\mathbf{a}$  og  $\mathbf{b}$ , så skifter determinanten fortegn:

$$\det(\mathbf{b}, \mathbf{a}) = \begin{vmatrix} b_1 & b_2 \\ a_1 & a_2 \end{vmatrix} = b_1 a_2 - b_2 a_1 = -(a_1 b_2 - a_2 b_1) = -\det(\mathbf{a}, \mathbf{b})$$

Som vi snart skal se, har både fortegnet og størrelsen til  $\det(\mathbf{a}, \mathbf{b})$  en geometrisk betydning.

**Bemerkning:** Man kan lure på hvorfor vi legger inn vektorene  $\mathbf{a}$  og  $\mathbf{b}$  som rader og ikke som søyler. Det viser seg at dette ikke spiller noen rolle siden  $\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ , men vi har valgt å ta utgangspunkt i rader siden det passer best til anvendelsene i senere kapitler (for andre anvendelser hadde det passet vel så bra å bruke søyler!).



Figur 1: Vinkelen  $v$  mellom  $\mathbf{a}$  og  $\mathbf{b}$

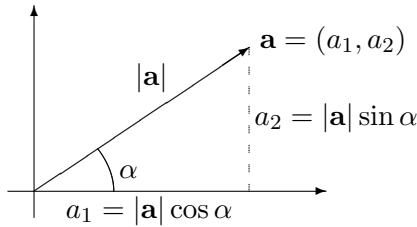
For å forstå den geometriske tolkningen av  $2 \times 2$ -determinanter, er det nyttig å vite litt om orientering av vektorpar. To vektorer  $\mathbf{a}$  og  $\mathbf{b}$  bestemmer en vinkel  $v$  mellom  $0^\circ$  og  $180^\circ$  som vist på figur 1. Vi kaller dette *vinkelen mellom  $\mathbf{a}$  og  $\mathbf{b}$* . Legg merke til at dersom vi beveger oss i positiv omløpsretning, vil denne vinkelen noen ganger starte i  $\mathbf{a}$  og ende i  $\mathbf{b}$  (se figur 1a) og andre ganger starte i  $\mathbf{b}$  og ende i  $\mathbf{a}$  (se figur 1b). I det første

tilfellet sier vi at paret  $(\mathbf{a}, \mathbf{b})$  er *positivt orientert*, i det andre tilfellet at det er *negativt orientert*. Her er åpenbart rekkefølgen til vektorene viktig —  $\mathbf{a}$  er første vektor og  $\mathbf{b}$  er andre vektor. Bytter vi om rekkefølgen av vektorene, bytter vi også orientering.

For å studere den geometriske tolkningen av determinanten

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

lønner det seg å skrive vektorene  $\mathbf{a} = (a_1, a_2)$  og  $\mathbf{b} = (b_1, b_2)$  på polarform.



Figur 2: Vektoren  $\mathbf{a}$  på polarform

Lar vi  $\alpha$  være vinkelen fra den positive  $x$ -aksen til vektoren  $\mathbf{a}$ , så er

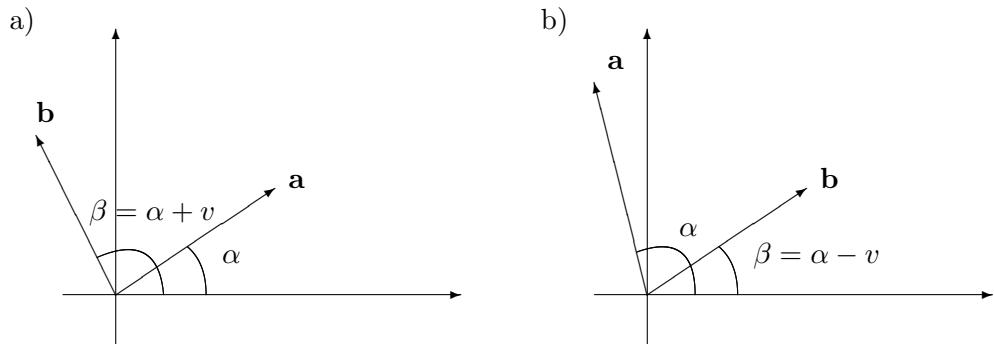
$$\mathbf{a} = (|\mathbf{a}| \cos \alpha, |\mathbf{a}| \sin \alpha)$$

(se figur 2 og husk det du har lært om polarform til komplekse tall). På tilsvarende måte er

$$\mathbf{b} = (|\mathbf{b}| \cos \beta, |\mathbf{b}| \sin \beta)$$

der  $\beta$  er vinkelen fra den positive  $x$ -aksen til vektoren  $\mathbf{b}$ .

Vi lar fortsatt  $v$  være vinkelen mellom vektorene  $\mathbf{a}$  og  $\mathbf{b}$  som på figur 1. Dersom paret  $(\mathbf{a}, \mathbf{b})$  er positivt orientert, ser vi fra figur 3a at  $\beta = \alpha + v$ . Dersom paret  $(\mathbf{a}, \mathbf{b})$  er negativt orientert, ser vi fra figur 3b at vi  $\beta = \alpha - v$ .



Figur 3: Sammenhengen  $\beta = \alpha \pm v$

Vi har altså

$$\beta = \alpha \pm v$$

der fortegnet er pluss eller minus ettersom paret  $(\mathbf{a}, \mathbf{b})$  er positivt eller negativt orientert.

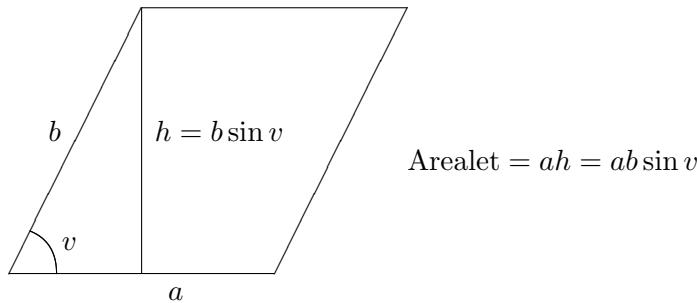
La oss nå finne determinanten uttrykt ved  $\alpha$  og  $\beta$ :

$$\begin{aligned}\det(\mathbf{a}, \mathbf{b}) &= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = \begin{vmatrix} |\mathbf{a}| \cos \alpha & |\mathbf{a}| \sin \alpha \\ |\mathbf{b}| \cos \beta & |\mathbf{b}| \sin \beta \end{vmatrix} \\ &= |\mathbf{a}| \cos \alpha |\mathbf{b}| \sin \beta - |\mathbf{a}| \sin \alpha |\mathbf{b}| \cos \beta \\ &= |\mathbf{a}| |\mathbf{b}| (\sin \beta \cos \alpha - \cos \beta \sin \alpha) \\ &= |\mathbf{a}| |\mathbf{b}| \sin(\beta - \alpha)\end{aligned}$$

der vi i den siste overgangen har brukt formelen for sinus til en differens. Siden  $\beta = \alpha \pm v$ , får vi

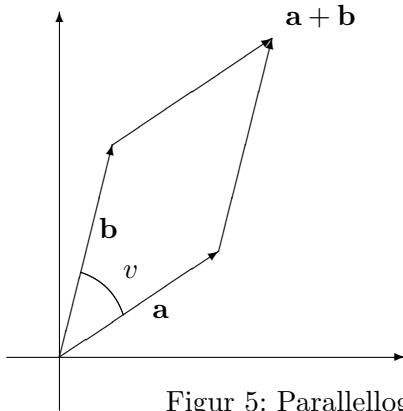
$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = |\mathbf{a}| |\mathbf{b}| \sin(\pm v) = \pm |\mathbf{a}| |\mathbf{b}| \sin v$$

Siden  $\sin v$  aldri er negativ ( $v$  ligger per definisjon i intervallet  $[0^\circ, 180^\circ]$  der sinus er positiv), vil  $\det(\mathbf{a}, \mathbf{b})$  altså være positiv dersom paret  $(\mathbf{a}, \mathbf{b})$  er positivt orientert, og negativ dersom dette paret er negativt orientert. Fortegnet til determinanten  $\det(\mathbf{a}, \mathbf{b})$  gjenspeiler altså orienteringen til paret  $(\mathbf{a}, \mathbf{b})$ . Legg forøvrig merke til at  $\det(\mathbf{a}, \mathbf{b})$  er 0 dersom  $v$  er  $0^\circ$  eller  $180^\circ$ , det vil si når  $\mathbf{a}$  og  $\mathbf{b}$  er parallelle.



Figur 4: Arealet til et parallellogram

Etter at vi nå har funnet ut hva fortegnet til determinanten betyr, er det på tide å se på absoluttverdien. Aller først repete vi formelen for arealet til et parallellogram. Som det fremgår fra figur 4, er dette arealet gitt ved  $A = ab \sin v$ , der  $a$  og  $b$  er lengdene til sidene, og der  $v$  er vinkelen mellom dem (på figuren er vinkel  $v$  spiss, men det er lett å se at resultatet også holder dersom vinkelen er stump).

Figur 5: Parallellogrammet utspent av  $\mathbf{a}$  og  $\mathbf{b}$ 

Arealet til parallellogrammet utspent av vektorene  $\mathbf{a} = (a_1, a_2)$  og  $\mathbf{b} = (b_1, b_2)$  (se figur 5) er derfor lik  $|\mathbf{a}||\mathbf{b}| \sin v = \pm \det(\mathbf{a}, \mathbf{b})$  der fortegnet er pluss eller minus ettersom  $(\mathbf{a}, \mathbf{b})$  er positivt eller negativt orientert. Det betyr at arealet er lik tallverdien til determinanten. La oss oppsummere resultatene.

### Setning 1.8.1 Determinanten

$$\det(\mathbf{a}, \mathbf{b}) = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

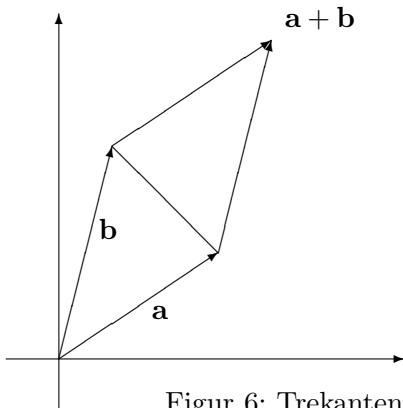
er positiv dersom vektorparet  $(\mathbf{a}, \mathbf{b})$  er positivt orientert og negativ dersom paret er negativt orientert. Arealet til parallellogrammet utspent av  $\mathbf{a}$  og  $\mathbf{b}$  er lik tallverdien til determinanten.

**Bemerkning:** Matematikere sier at determinanten gir oss arealet med fortegn (eller orientering). Det kan virke merkelig å knytte fortegn til areal, men spesielt når man skal studere arealet til flater, viser det seg viktig å holde styr på retningen — det er i mange sammenhenger viktig å vite hva man skal regne som flatens “overside/underside” eller “utside/innde”. Som vi skal se senere i denne seksjonen, kan sammenhengen mellom determinant og “areal med fortegn” generaliseres til tre dimensjoner.

**Eksempel 2:** Finn arealet utspent av vektorene  $\mathbf{a} = (3, -7)$  og  $\mathbf{b} = (-4, 5)$ . Vi får

$$\det(\mathbf{a}, \mathbf{b}) = \begin{vmatrix} 3 & -7 \\ -4 & 5 \end{vmatrix} = 3 \cdot 5 - (-7) \cdot (-4) = 15 - 28 = -13$$

Arealet er dermed  $|-13| = 13$ . Siden  $\det(\mathbf{a}, \mathbf{b})$  er negativ, er paret  $(\mathbf{a}, \mathbf{b})$  negativt orientert, dvs. at vinkelen fra  $\mathbf{a}$  til  $\mathbf{b}$  er større enn  $180^\circ$ . ♣

Figur 6: Trekanten med sider  $\mathbf{a}$  og  $\mathbf{b}$ 

Determinanter kan også brukes til å regne ut arealet til trekanner. Arealet til trekanten med sider  $\mathbf{a}$  og  $\mathbf{b}$  er halvparten av arealet til parallellogrammet utspent av disse vektorene (se figur 6).

Vi har derfor følgende resultat:

**Korollar 1.8.2** *Arealet til trekanten med sider  $\mathbf{a}$  og  $\mathbf{b}$  er  $\frac{1}{2} \cdot |\det(\mathbf{a}, \mathbf{b})|$*

**Eksempel 3:** Finn arealet til trekanten med hjørner i punktene  $\mathbf{c} = (-1, 2)$ ,  $\mathbf{d} = (4, 8)$  og  $\mathbf{e} = (2, -3)$ . Vi regner ut

$$\mathbf{a} = \mathbf{d} - \mathbf{c} = (4, 8) - (-1, 2) = (5, 6)$$

$$\mathbf{b} = \mathbf{e} - \mathbf{c} = (2, -3) - (-1, 2) = (3, -5)$$

Trekanten vi er på jakt etter, har samme areal som trekanten med sider  $\mathbf{a}$  og  $\mathbf{b}$  (hvorfor?). Dermed er

$$\text{Areal} = \frac{1}{2} \cdot |\det(\mathbf{a}, \mathbf{b})| = \frac{1}{2} \cdot \left| \begin{array}{cc} 5 & 6 \\ 3 & -5 \end{array} \right| = \frac{1}{2} \cdot |-25 - 18| = \frac{43}{2} \quad \clubsuit$$

Før vi går over til  $3 \times 3$ -determinanter, tar vi med et resultat til. Dette resultatet kan virke litt underlig på det nåværende tidspunkt, men det skal bli en viktig inspirasjonskilde når vi studerer generelle  $n \times n$ -determinanter i et senere kapittel.

**Setning 1.8.3** *For  $2 \times 2$ -matriser gjelder:*

- (i)  $\det(I_2) = 1$  (husk at  $I_2$  er identitetsmatrisen  $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ).
- (ii) *Dersom vi bytter om to rader, så bytter determinanten fortegn (dvs.  $\det(\mathbf{a}, \mathbf{b}) = -\det(\mathbf{b}, \mathbf{a})$ ).*

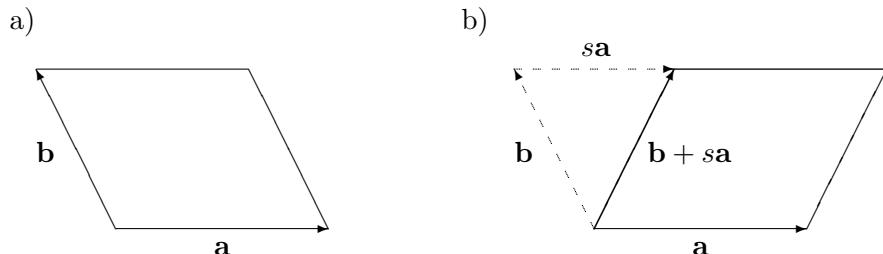
- (iii) Dersom vi ganger alle elementene i en rad med et tall  $s$ , så forandrer også matrisen seg med en faktor  $s$  (dvs.  $\det(s\mathbf{a}, \mathbf{b}) = s \det(\mathbf{a}, \mathbf{b})$  og  $\det(\mathbf{a}, s\mathbf{b}) = s \det(\mathbf{a}, \mathbf{b})$ ).
- (iv) Dersom vi adderer et tall ganger en rad til en av de andre radene, endrer ikke determinanten verdi (dvs.  $\det(\mathbf{a} + s\mathbf{b}, \mathbf{b}) = \det(\mathbf{a}, \mathbf{b})$  og  $\det(\mathbf{a}, \mathbf{b} + s\mathbf{a}) = \det(\mathbf{a}, \mathbf{b})$ ).

*Bevis:* Alle punktene kan vises ved direkte utregning, men vi vil gjerne forstå dem geometrisk selv om det i noen tilfeller er litt mer omstendelig:

(i) Parallellogrammet utspent av radvektorene  $\mathbf{a} = (1, 0)$  og  $\mathbf{b} = (0, 1)$  er et kvadrat med side 1. Arealet er opplagt 1, og siden paret  $(\mathbf{a}, \mathbf{b})$  er positivt orientert, er  $\det(I_2) = 1$ .

(ii) Bytter vi om på radene, bytter vi orientering på paret  $(\mathbf{a}, \mathbf{b})$ , og determinanten bytter dermed fortegn.

(iii) Ganger vi den ene vektoren med et positivt tall  $s$ , blir enten grunnlinjen eller høyden i det utspente parallellogrammet ganget med  $s$ , og arealet øker/avtar derfor med en faktor  $s$ . Ganger vi med et negativt tall, endres høyden eller grunnlinjen med en faktor  $|s|$ , men i tillegg bytter vektorparet orientering slik at også i dette tilfellet endrer determinanten seg med en faktor  $s$ .



Figur 7: Parallellogrammene utspent av henholdsvis  $\mathbf{a}, \mathbf{b}$  og  $\mathbf{a}, \mathbf{b} + s\mathbf{a}$

(iv) Her trenger vi en liten figur. Figur 7a viser parallellogrammet utspent av  $\mathbf{a}$  og  $\mathbf{b}$ , mens figur 7b viser parallellogrammet utspent av  $\mathbf{a}$  og  $\mathbf{b} + s\mathbf{a}$ . De to parallellogrammene har samme grunnlinje og høyde, og derfor samme areal. Det er også lett å se at uansett hvor stor  $s$  er, så har parene  $(\mathbf{a}, \mathbf{b})$  og  $(\mathbf{a}, \mathbf{b} + s\mathbf{a})$  samme orientering. Altså er  $\det(\mathbf{a}, \mathbf{b}) = \det(\mathbf{a}, \mathbf{b} + s\mathbf{a})$ .  $\square$

### 3 × 3-determinanter

Determinanten til en  $3 \times 3$ -matrise

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

er definert ved

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

der  $2 \times 2$ -determinantene på høyre side regnes ut på vanlig måte. Legg merke til hvordan disse  $2 \times 2$ -determinantene fremkommer fra den opprinnelige determinanten — for å finne den  $2 \times 2$ -determinanten som ganges med  $a_{11}$ , stryker vi den linjen og den søylen som går gjennom  $a_{11}$  (se figur 8), for å finne den  $2 \times 2$ -determinanten som ganges med  $a_{12}$ , stryker vi den linjen og den søylen som går gjennom  $a_{12}$ , osv. Legg også merke til at fortegnene til leddene på høyre side veksler mellom pluss og minus.

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

Figur 8:  $2 \times 2$ -determinanten som skal ganges med  $a_1$

La oss regne ut en  $3 \times 3$ -determinant.

**Eksempel 4:** Regn ut

$$\begin{aligned} \begin{vmatrix} 2 & 3 & -1 \\ 5 & -4 & 0 \\ -3 & 1 & 2 \end{vmatrix} &= 2 \begin{vmatrix} -4 & 0 \\ 1 & 2 \end{vmatrix} - 3 \begin{vmatrix} 5 & 0 \\ -3 & 2 \end{vmatrix} + (-1) \begin{vmatrix} 5 & -4 \\ -3 & 1 \end{vmatrix} \\ &= 2((-4) \cdot 2 - 0 \cdot 1) - 3(5 \cdot 2 - 0 \cdot (-3)) + (-1)(5 \cdot 1 - (-4)(-3)) \\ &= 2(-8) - 3 \cdot 10 + (-1)(-7) = -16 - 30 + 7 = -39 \end{aligned}$$



Det er en nær sammenheng mellom  $3 \times 3$ -determinanter og kryssproduktet. Som et første eksempel har vi følgende huskeregel for kryssproduktet:

$$\begin{aligned} \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} &= \\ &= (a_2 b_3 - a_3 b_2) \mathbf{i} + (a_3 b_1 - a_1 b_3) \mathbf{j} + (a_1 b_2 - a_2 b_1) \mathbf{k} = \mathbf{a} \times \mathbf{b} \end{aligned}$$

Siden vi bare har definert determinanten når elementene i første rad er tall (og ikke vektorer), gir det første skrittet i denne utregningen egentlig ikke mening, men resultatet er likevel en grei huskeregel.

Vi har tidligere sett at  $2 \times 2$ -determinanter kan brukes til å regne ut arealer og til å bestemme orienteringen til vektorpar  $(\mathbf{a}, \mathbf{b})$ . På tilsvarende måte kan vi bruke  $3 \times 3$ -determinanter til å regne ut volumer og til å bestemme orienteringen til vektortripler  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ . Før vi begynner, kan det være greit å bli enig om notasjonen. Dersom  $\mathbf{a} = (a_1, a_2, a_3)$ ,  $\mathbf{b} = (b_1, b_2, b_3)$ ,  $\mathbf{c} = (c_1, c_2, c_3)$ , skriver vi

$$\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}$$

Vi observerer så at

$$\begin{aligned} \det(\mathbf{a}, \mathbf{b}, \mathbf{c}) &= \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \\ &= a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \end{aligned}$$

Sammenholder vi dette med setning 1.4.4 og korollar 1.4.5, får vi:

**Setning 1.8.4** *Volumet av parallellepipedet utspent av vektorene  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  er  $|\det(\mathbf{a}, \mathbf{b}, \mathbf{c})|$ . Volumet av pyramiden utspent av  $\mathbf{a}, \mathbf{b}, \mathbf{c}$  er  $\frac{1}{6}|\det(\mathbf{a}, \mathbf{b}, \mathbf{c})|$ .*

Legg merke til at  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = 0$  hvis volumet til parallellepipedet er 0. Det skjer hvis vektorene  $\mathbf{a}, \mathbf{b}$  og  $\mathbf{c}$  ligger i samme plan gjennom origo. På det nåværende tidspunkt kan dette virke som en uvesentlig observasjon, men det viser seg faktisk å være en av hovedårsakene til determinantenes betydning.

Hva så med orienteringen? Først må vi definere når et trippel  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  er positivt og negativt orientert: To ikke-parallelle vektorer  $\mathbf{a}$  og  $\mathbf{b}$  definerer sammen med origo et plan (planet) gjennom punktene  $\mathbf{0}, \mathbf{a}$  og  $\mathbf{b}$ . Dette planet deler rommet i to halvdeler. Dersom  $\mathbf{c}$  ligger på samme side av planet som kryssproduktet  $\mathbf{a} \times \mathbf{b}$ , sier vi at triplet  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  er *positivt orientert*. Dersom  $\mathbf{c}$  ligger på den andre siden av planet, sier vi at triplet er *negativt orientert*. Bruker vi den geometriske tolkningen av skalarproduktet, ser vi at triplet  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  er positivt orientert hvis og bare hvis  $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$  er positiv (for da er vinkelen mellom  $\mathbf{c}$  og  $\mathbf{a} \times \mathbf{b}$  mindre enn  $90^\circ$ ). Det er lett (men ikke særlig spennende) å sjekke at  $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \det(\mathbf{a}, \mathbf{b}, \mathbf{c})$  (det er ikke noe mystisk i dette — både  $|(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}|$  og  $|\det(\mathbf{a}, \mathbf{b}, \mathbf{c})|$  er lik volumet til parallellepipedet utspent av  $\mathbf{a}, \mathbf{b}$  og  $\mathbf{c}$ , så alt vi sjekker er at fortegnet er det samme). Dette betyr at  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  er positivt orientert hvis og bare hvis  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c})$  er positiv. Vi har altså den samme forbindelsen mellom positiv orientering og positiv determinant som i det to-dimensjonale tilfellet.

Vi tar med en tredimensjonal versjon av setning 1.8.3.

**Setning 1.8.5** *For  $3 \times 3$ -matriser gjelder:*

- (i)  $\det(I_3) = 1$  (husk at  $I_3$  er identitetsmatrisen  $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ ).
- (ii) Dersom vi bytter om to rader, så bytter determinanten fortegn (det vil f.eks. si at  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = -\det(\mathbf{c}, \mathbf{b}, \mathbf{a})$ ).
- (iii) Dersom vi ganger alle elementene i en rad med et tall  $s$ , så forandrer også matrisen seg med en faktor  $s$  (det vil f.eks. si at  $\det(\mathbf{a}, s\mathbf{b}, \mathbf{c}) = s\det(\mathbf{a}, \mathbf{b}, \mathbf{c})$ ).
- (iv) Dersom vi adderer et tall ganger en rad til en av de andre radene, endrer ikke determinanten verdi (det vil f.eks. si at  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c} + s\mathbf{a}) = \det(\mathbf{a}, \mathbf{b}, \mathbf{c})$ ).

*Bevis:* Vi skal ikke gjennomgå punktene i detalj, bare se på hovedideene. Punkt (i) kan du bevise enten ved direkte utregning eller ved å observere at parallellogrammet utspent av vektorene  $\mathbf{a} = (1, 0, 0)$ ,  $\mathbf{b} = (0, 1, 0)$  og  $\mathbf{c} = (0, 0, 1)$  er en terning med side 1. I punkt (ii) vet vi allerede at tallverdien til determinanten er uforandret om vi bytter om på radene (fordi begge er lik volumet til det samme parallelepipedet), og alt du behøver å sjekke er at orienteringen snur når du bytter om to vektorer (bruk høyrehåndsregelen). Punkt (iii) følger på samme måte som i det todimensjonale tilfellet; ganger du en av sidekantene i et parallelepiped med  $s > 0$ , øker også volumet med en faktor  $s$ , men ganger du med en faktor  $s < 0$ , må du også ta hensyn til at orienteringen snur. For å bevise punkt (iv) kan du bruke akkurat samme figur som i det todimensjonale tilfellet (figur 7) — den eneste forskjellen er at det nå inngår en vektor til. Denne vektoren stikker på skrå ut av (eller inn i) papiret og inngår ikke i beregningene på noen forstyrrende måte (det eneste den bidrar med er den felles høyden i parallellogrammene).

### $n \times n$ -determinanter

Vi skal ikke studere generelle  $n \times n$ -determinanter for alvor i dette kapitlet, men det kan være morsomt å vite hvordan de regnes ut. Gitt en  $4 \times 4$ -matrise

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}$$

definerer vi determinanten  $\det(A)$  ved

$$\det(A) = a_{11} \left| \begin{array}{ccc} a_{22} & a_{23} & a_{24} \\ a_{32} & a_{33} & a_{34} \\ a_{42} & a_{43} & a_{44} \end{array} \right| - a_{12} \left| \begin{array}{ccc} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{array} \right| +$$

$$+a_{13} \left| \begin{array}{ccc} a_{21} & a_{22} & a_{24} \\ a_{31} & a_{32} & a_{34} \\ a_{41} & a_{42} & a_{44} \end{array} \right| - a_{14} \left| \begin{array}{ccc} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{array} \right|$$

Sammenligner du denne definisjonen med definisjonen av  $3 \times 3$ -determinanter, vil du oppdage det generelle mønsteret som gjør at vi kan gå videre og definere  $5 \times 5$ -determinanter,  $6 \times 6$ -determinanter osv.

Vi har tidligere sett at *tallverdien* til  $2 \times 2$ -determinanten  $\det(\mathbf{a}, \mathbf{b})$  gir oss arealet utspent av vektorene  $\mathbf{a}$  og  $\mathbf{b}$ , mens *fortegnet* til determinanten forteller oss om orienteringen til paret  $(\mathbf{a}, \mathbf{b})$ . På tilsvarende vis vet vi at tallverdien til en  $3 \times 3$ -determinant  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c})$  gir oss volumet utspent av vektorene  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$ , mens fortegnet forteller oss om orienteringen til triplet  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ . Vi kan bruke disse observasjonene til å definere volum og orientering i høyere dimensjoner. Gitt  $n$  vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  definerer vi *volumet utspent av disse vektorene* til å være tallverdien til  $n \times n$ -determinanten  $\det(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ . Vi sier at  $n$ -tuplet  $(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  (legg merke til at dette er et  $n$ -tuppel av vektorer) er *positivt* (henholdsvis *negativt*) orientert dersom determinanten er positiv (henholdsvis negativ). Vi skal ikke gå nærmere inn på volum og orientering i dette heftet, men vi skal komme tilbake til generelle determinanter i kapittel 4.

**Bemerkning:** Som vi skal se i kapittel 4, er det en nær sammenheng mellom determinanter og inverterbarhet — det viser seg at en kvadratisk matrise  $A$  er inverterbar hvis og bare hvis  $\det(A) \neq 0$ . Dette er kanskje ikke så overraskende siden begge begrepene har med degenerasjon å gjøre: En kvadratisk matrise  $A$  er singulær (ikke inverterbar), dersom det finnes to forskjellige vektorer  $\mathbf{x}$  og  $\mathbf{x}'$  slik at  $A\mathbf{x} = A\mathbf{x}'$ , og den har determinant lik 0 dersom radene ikke spenner ut et volum. Det viser seg at disse to formene for degenerasjon er ekvivalente.

**MATLAB-kommentar:** MATLAB regner ut determinanten til  $A$  når du skriver `>>det(A)`.

### Oppgaver til seksjon 1.8

1. Regn ut determinantene

$$\text{a)} \quad \left| \begin{array}{cc} 4 & 3 \\ -2 & 2 \end{array} \right| \quad \text{b)} \quad \left| \begin{array}{cc} 8 & -2 \\ 3 & 4 \end{array} \right| \quad \text{c)} \quad \left| \begin{array}{cc} 6 & -3 \\ -2 & 1 \end{array} \right|$$

2. Finn arealet til parallellogrammet utspent av  $\mathbf{a} = (1, 3)$  og  $\mathbf{b} = (4, 1)$ .

3. En trekant har hjørner i punktene  $(-1, 2)$ ,  $(4, 3)$ ,  $(1, 7)$ . Finn arealet.

4. En firkant har hjørner i punktene  $(0, 1)$ ,  $(5, 1)$ ,  $(1, 7)$  og  $(7, 4)$ . Finn arealet.

5. Avgjør om parene  $(\mathbf{a}, \mathbf{b})$  er positivt eller negativt orientert:

$$\text{a)} \quad \mathbf{a} = (3, -1) \quad \mathbf{b} = (-7, 2) \quad \text{b)} \quad \mathbf{a} = (-1, 5) \quad \mathbf{b} = (3, 2)$$

**6.** Vis at  $\det(\mathbf{a}, \mathbf{b}) = 0$  hvis og bare hvis vektorene  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$  er parallele eller (minst) én av dem er  $\mathbf{0}$ .

**7.** Vis at  $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix}$ , dvs. at vi får den samme determinanten om vi bytter om linjer og søyler.

**8.** Alle hjørnene til et parallellogram har heltallige koordinater. Vis at arealet er et helt tall.

**9.** Anta at  $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0$

a) Vis at ligningssystemet  $a_1x + b_1y = c_1$ ,  $a_2x + b_2y = c_2$  har løsningen

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \quad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}}$$

b) Hva skjer med ligningssystemet når  $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = 0$ ?

**10.** Regn ut determinantene:

$$\text{a)} \begin{vmatrix} 3 & -2 & -1 \\ 1 & 4 & 3 \\ 2 & 1 & 7 \end{vmatrix} \quad \text{b)} \begin{vmatrix} -2 & 4 & 0 \\ -2 & 3 & 3 \\ 1 & 0 & 4 \end{vmatrix} \quad \text{c)} \begin{vmatrix} 1 & 2 & 3 \\ -2 & 5 & 4 \\ 3 & -3 & -1 \end{vmatrix}$$

**11.** Finn volumet til parallelepipedet utspent av  $(-1, 0, 2)$ ,  $(3, -1, 3)$  og  $(4, 0, -1)$ .

**12.** Finn volumet til pyramiden med hjørner i punktene  $(2, 2, 2)$ ,  $(-1, 2, 3)$ ,  $(3, 4, 2)$  og  $(7, 2, 2)$ .

**13.** Avgjør om triplet  $(\mathbf{a}, \mathbf{b}, \mathbf{c})$  er positivt eller negativt orientert når  $\mathbf{a} = (-1, 2, 3)$ ,  $\mathbf{b} = (0, 2, 4)$  og  $\mathbf{c} = (7, -1, 2)$ .

**14.** Vis at  $\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$ , dvs. at determinanten er den samme om vi bytter om søyler og linjer.

**15.** Vis at dersom  $\mathbf{a}$ ,  $\mathbf{b}$  og  $\mathbf{c}$  er ortogonale, så er  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = |\mathbf{a}||\mathbf{b}||\mathbf{c}|$ .

**16.** Regn ut determinanten til  $4 \times 4$ -matrisen

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

**17.** I denne oppgaven er  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  og  $\mathbf{d}$  tredimensjonale vektorer.

a) Vis at dersom to av vektorene  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  er like, så er  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = 0$

- b) Vis at for alle vektorer  $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$  og alle skalarer  $s, t$  gjelder

$$\det(s\mathbf{a} + t\mathbf{d}, \mathbf{b}, \mathbf{c}) = s \det(\mathbf{a}, \mathbf{b}, \mathbf{c}) + t \det(\mathbf{d}, \mathbf{b}, \mathbf{c})$$

- c) Vi sier at en vektor  $\mathbf{a}$  er en *lineærkombinasjon* av vektorene  $\mathbf{b}, \mathbf{c}$  dersom det finnes skalarer  $s, t$  slik at  $\mathbf{a} = s\mathbf{b} + t\mathbf{c}$ . Bruk a) og b) til å vise at dersom  $\mathbf{a}$  er en lineærkombinasjon av  $\mathbf{b}$  og  $\mathbf{c}$ , så er  $\det(\mathbf{a}, \mathbf{b}, \mathbf{c}) = 0$ .
- d) Gi en geometrisk forklaring på resultatet i c).

**18.** Bevis setning 1.8.3 ved regning (dvs. regn ut begge sider av likhetene og se at de stemmer).

**19.** Bevis setning 1.8.5 ved regning (dvs. regn ut begge sider av likhetene og se at de stemmer).

**20.** Vis at en  $2 \times 2$ -matrise  $A$  er inverterbar hvis og bare hvis  $\det(A) \neq 0$ . (*Hint:* Mesteparten av jobben er gjort i oppgave 1.7.10.)

# Kapittel 2

## Funksjoner fra $\mathbb{R}^n$ til $\mathbb{R}^m$

I dette kapitlet skal vi studere funksjoner  $\mathbf{F}$  fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ . En slik funksjon er bare en regel som til hvert  $n$ -tuppel  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  tilordner et  $m$ -tuppel  $\mathbf{y} = \mathbf{F}(\mathbf{x})$ . Dersom  $n > 1$  kaller vi  $\mathbf{F}$  en *funksjon av flere variable*. Funksjoner av flere variable har mange likhetstrekk med de funksjonene av én variabel som du kjenner fra før, men de har også en del nye og litt uvante egenskaper. Spesielt blir geometrien mer komplisert når vi får mer enn én variabel å arbeide med.

### 2.1 Funksjoner av flere variable

Som allerede nevnt er en funksjon  $\mathbf{F}$  fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$  bare en regel som til hver  $\mathbf{x} \in \mathbb{R}^n$  gir oss en  $\mathbf{y} = \mathbf{F}(\mathbf{x})$  i  $\mathbb{R}^m$ . Ofte er disse reglene gitt ved en formler, f.eks. kan en funksjon fra  $\mathbb{R}^4$  til  $\mathbb{R}^3$  være gitt ved

$$\mathbf{F}(x, y, z, u) = (2x^2z + u, xy^2zu^3, ye^{x^2y+u \sin y})$$

Gitt et 4-tuppel  $\mathbf{x} = (x, y, z, u)$  forteller denne formelen oss hvordan vi kan regne ut 3-tuplet  $\mathbf{F}(\mathbf{x})$ . Som du ser, veksler vi mellom skrivemåtene  $\mathbf{F}(\mathbf{x})$  og  $\mathbf{F}(x_1, x_2, \dots, x_n)$ ; den første er ofte mest praktisk når vi snakker om en generell, uspesifisert funksjon, mens den siste som regel er greiest når vi snakker om en bestemt funksjon gitt ved en formel. Husk at  $\mathbb{R}^1 = \mathbb{R}$  slik at teorien vår også dekker funksjoner  $F : \mathbb{R}^n \rightarrow \mathbb{R}$  der  $\mathbf{x}$  er en vektor, men  $F(\mathbf{x})$  er et tall (i prinsippet dekker teorien også funksjoner av typen  $\mathbf{F} : \mathbb{R} \rightarrow \mathbb{R}^m$ , men for øyeblikket er vi ikke så interessert i dem).

Vi husker at vanlige funksjoner  $f(x)$  ikke alltid er definert for alle reelle tall  $x$ , og på samme måte vil heller ikke disse nye funksjonene  $\mathbf{F}(\mathbf{x})$  nødvendigvis være definert for alle vektorer  $\mathbf{x} \in \mathbb{R}^n$ . Den mengden  $A \subset \mathbb{R}^n$  av vektorer  $\mathbf{x}$  som  $\mathbf{F}(\mathbf{x})$  er definert for, kaller vi *definisjonsmengden* eller *definisjonsområdet* til  $\mathbf{F}$ . Vi skal av og til bruke symbolet  $D_{\mathbf{F}}$  for definisjonsområdet til  $\mathbf{F}$ . Når vi skriver

$$\mathbf{F} : A \rightarrow \mathbb{R}^m$$

mener vi at  $\mathbf{F}$  er en funksjon definert på mengden  $A$  med verdier i  $\mathbb{R}^m$ . Hvis en funksjon er gitt ved formler, og definisjonsområdet ikke er spesifisert, regner vi med at funksjonen er definert der alle formlene gir mening.

En funksjon definert på en delmengde av  $\mathbb{R}^n$  kaller vi en funksjon av  $n$ -variable; f.eks. er

$$\mathbf{G}(x, y) = (x^y, \ln(1 - xy), \sin(x^2y))$$

en funksjon av to variable med verdier i  $\mathbb{R}^3$ , mens

$$g(x, y, z) = x^2 + ye^{z+y^2}$$

er en funksjon av tre variable med verdier i  $\mathbb{R}$  (når funksjonene tar verdier i  $\mathbb{R}$  skriver vi dem gjerne uten fete typer og ofte med små bokstaver).

En funksjon som tar verdier i  $\mathbb{R}$  kalles gjerne et *skalarfelt* når vi vil understreke at verdiene er tall og ikke vektorer. Legg merke til at en funksjon  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  kan skrives

$$\mathbf{F}(x_1, x_2, \dots, x_n) = (F_1(x_1, x_2, \dots, x_n), \dots, F_m(x_1, x_2, \dots, x_n))$$

der  $F_1, \dots, F_m$  er skalarfelt. Vi kaller  $F_1, \dots, F_m$  for *komponentene* til  $\mathbf{F}$ . Funksjonen  $\mathbf{G}$  ovenfor har altså komponentene.

$$G_1(x, y) = x^y, \quad G_2(x, y) = \ln(1 - xy), \quad G_3(x, y) = \sin(x^2y)$$

Vi skal ofte gjøre bruk av at funksjoner med verdier i  $\mathbb{R}^m$  er bygget opp av skalarfelt på denne måten, for eksempel ved at vi først innfører begreper og beviser resultater for skalarfelt og så utvider til funksjoner med verdier i  $\mathbb{R}^m$ .

## Eksempler

Det er naturlig å spørre seg selv om *hvorfor* man må studere funksjoner av flere variable — greier det seg ikke med de funksjonene av bare én variabel som man kjenner fra før? Det viser seg imidlertid at funksjoner av flere variable dukker naturlig opp i svært mange sammenhenger. Her er et lite utvalg:

- BMI (body mass index) brukes ofte som en indikator på overvekt og undervekt. For å finne din BMI, tar du vekten din  $v$  (målt i kilo) og deler på kvadratet av høyden din  $h$  (målt i meter). Vi kan tenke på BMI som en funksjon  $f$  av to variable med verdier i  $\mathbb{R}$ :

$$f(v, h) = \frac{v}{h^2}$$

- Når en gjenstand varmes opp, vil temperaturen avhenge av når og hvor vi måler. Det er naturlig å angi temperaturen i punktet  $(x, y, z)$  ved tiden  $t$  som  $T(x, y, z, t)$ . Dette er en funksjon av fire variable med verdier i  $\mathbb{R}$ .
- Gravitasjonskraften fra jorden på en gjenstand i verdensrommet avhenger av posisjonen til gjenstanden. Tenker vi på kraften som en vektor  $\mathbf{F}(\mathbf{x})$  (vi er interessert i både størrelse og retning) som er avhengig av posisjonen  $\mathbf{x} = (x, y, z)$ , får vi en funksjon av tre variable med verdier i  $\mathbb{R}^3$ . Dersom vi plasserer jorden i origo, kan  $\mathbf{F}$  i dette tilfellet skrives

$$\mathbf{F}(\mathbf{x}) = -\gamma \frac{Mm}{|\mathbf{x}|^3} \mathbf{x}$$

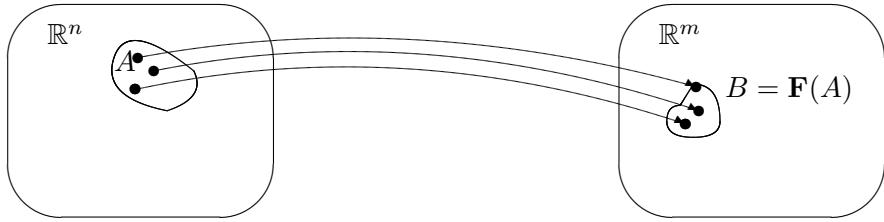
der  $\gamma$  er en naturkonstant (gravitasjonskonstanten),  $M$  er massen til jorden og  $m$  massen til gjenstanden. I dette eksemplet er  $\mathbf{F}$  en funksjon fra  $\mathbb{R}^3$  til  $\mathbb{R}^3$

- Meteorologer arbeider med vind i atmosfæren. Vinden i et punkt med koordinater  $(x, y, z)$  ved tiden  $t$  vil være en vektor  $\mathbf{v}(x, y, z, t)$  (vi er interessert i både vindstyrken og retningen). Vi kan tenke på dette som en funksjon av fire variable med verdier i  $\mathbb{R}^3$ .
- I eksempel 3 i seksjon 1.5 så vi på fire produsenter som leverte epler av tre ulike kvaliteter til en fruktpresse. Dersom produsentene leverte hhv.  $x, y, z$  og  $u$  tonn epler, kunne vi regne ut hvor mange tonn vi ville få av hver kvalitet. Dette gir oss en funksjon av fire variable med verdier i  $\mathbb{R}^3$ . Vi kan skrive funksjonen som  $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ , eller med koordinater

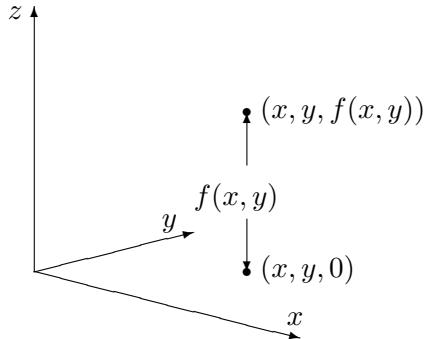
$$\mathbf{F}(x, y, z, u) = \begin{pmatrix} 0.5x + 0.3y + 0.25z + 0.2u \\ 0.3x + 0.4y + 0.4z + 0.6u \\ 0.2x + 0.3y + 0.35z + 0.2u \end{pmatrix}$$

## Grafisk fremstilling

Et av de viktigste verktøyene når man studerer funksjoner av én variabel er grafisk fremstilling. Dessverre er det ikke så lett å gi realistiske grafiske fremstillinger av funksjoner fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ . Noen ganger kan vi imidlertid ha stor glede av mer stiliserte fremstillinger som på figuren nedenfor. Den illustrerer hvordan en funksjon  $\mathbf{F}$  fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$  sender alle punkter i en mengde  $A \subset \mathbb{R}^n$  på en mengde  $B \subset \mathbb{R}^m$ . Mengden  $B$  kalles *bildet av A under F* og betegnes ofte med  $B = \mathbf{F}(A)$ . Du vil se mange figurer av denne typen utover i kapitlet.

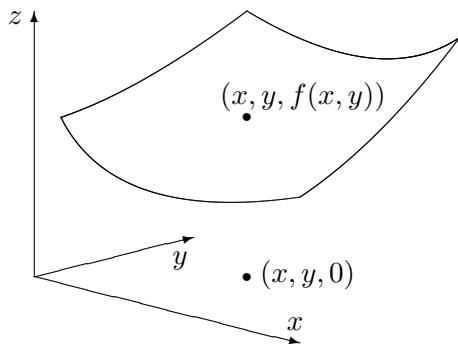
Figur 1: Funksjon fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ 

Det er imidlertid ett tilfelle der det går an å lage gode, realistiske fremstillinger av en funksjon av flere variable, og det er når funksjonen går fra  $\mathbb{R}^2$  til  $\mathbb{R}$ . Vi ser altså på en funksjon  $z = f(x, y)$ . For å tegne funksjonsgrafen lager vi først et tre-dimensjonalt koordinatsystem som vist på figur 2.



Figur 2: Plotting av skalarfelt

Gitt variabelverdier  $x$  og  $y$ , finner vi punktet  $(x, y, 0)$  i  $xy$ -planet. Vi flytter oss nå loddrett (dvs. parallelt med  $z$ -aksen) til vi finner punktet  $(x, y, f(x, y))$ . Dette er det første punktet på funksjonsgrafen vår.



Figur 3: Grafisk fremstilling av skalarfelt

Gjentar vi denne prosedyren for stadig flere variabelverdier  $(x, y)$ , vokser

grafen etterhvert frem som en flate i rommet (se figur 3).

Selv om denne prosedyren på en grei måte forklarer hva grafen til et skalarfelt er, så er den i praksis ubruklig som en oppskrift på hvordan man tegner grafen. Prøver du den, selv på en enkel funksjon, oppdager du fort at du helt mister romfølelsen i bildet. I neste kapittel skal vi se på mer praktiske metoder for å tegne slike funksjonsgrafer — det som er viktig i dette kapittelet, er at du vet hvordan du kan tenke på grafen til  $f$  som en flate.

**MATLAB-kommentar:** MATLAB er et utmerket verktøy for grafisk fremstilling av funksjoner av flere variable, men vi utsetter dette temaet til seksjon 3.7 og 3.8.

### Oppgaver til seksjon 2.1

1. Finn definisjonsområdet til funksjonen:

- a)  $f(x, y) = \frac{1}{x^2 + 4y^2}$
- b)  $f(x, y) = \frac{1}{x^2 - y^2}$
- c)  $f(x, y) = \ln(x + y)$
- d)  $f(x, y) = \tan(x - y)$
- e)  $f(x, y, z) = \frac{1}{x^2 + y^2 + z^2 - 25}$

## 2.2 Kontinuerlige funksjoner

I seksjon 5.1 i *Kalkulus* studerte vi  $\epsilon$ - $\delta$ -definisjonen av kontinuitet. Denne definisjonen er sannsynligvis ikke like populær blant alle, men den har mange fordeler, blant annet at den lett kan generaliseres til nye situasjoner. I denne seksjonen skal vi se hvordan den generaliseres til funksjoner av flere variable. Før vi begynner, trenger vi å vite litt om avstander og kuler i  $\mathbb{R}^n$ .

Akkurat som i planet og rommet lar vi *avstanden* mellom to punkter  $\mathbf{a}$  og  $\mathbf{x}$  i  $\mathbb{R}^n$  være lik lengden til vektoren  $\mathbf{x} - \mathbf{a}$  som forbinder dem, det vil si at avstanden er

$$|\mathbf{x} - \mathbf{a}| = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2 + \cdots + (x_n - a_n)^2}$$

Mengden

$$B(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < r\}$$

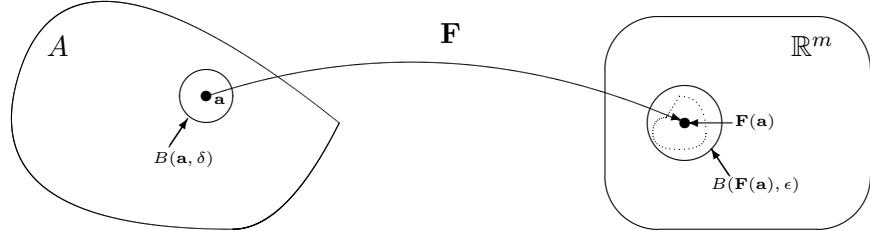
består av de punktene i  $\mathbb{R}^n$  som har avstand mindre enn  $r$  til punktet  $\mathbf{a}$ . Vi kaller  $B(\mathbf{a}, r)$  *kulen om  $\mathbf{a}$  med radius  $r$* . Legg merke til at i  $\mathbb{R}^3$  er dette virkelig en (åpen) kule i tradisjonell forstand, mens det i  $\mathbb{R}^2$  er en (åpen) sirkelskive og i  $\mathbb{R}$  et åpent intervall. Vi velger å bruke “kule” som et fellesord i alle dimensjoner, selv om det til å begynne med kan virke litt uvant når vi arbeider i planet eller på tallinjen. De fleste illustrasjonene våre vil være i planet, og der vil kuler fremstå som sirkler.

Vi kan nå definere kontinuitet for funksjoner av flere variable:

**Definisjon 2.2.1** Anta at  $A \subset \mathbb{R}^n$ , og at  $\mathbf{a} \in A$ . En funksjon  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er kontinuerlig i  $\mathbf{a}$  dersom det til enhver  $\epsilon > 0$  finnes en  $\delta > 0$  slik at

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon \text{ for alle } \mathbf{x} \in A \text{ slik at } |\mathbf{x} - \mathbf{a}| < \delta$$

Figuren nedenfor illustrerer definisjonen: Gitt en kule  $B(\mathbf{F}(\mathbf{a}), \epsilon)$  om punktet  $\mathbf{F}(\mathbf{a})$ , kan vi finne en kule  $B(\mathbf{a}, \delta)$  om punktet  $\mathbf{a}$  slik at bildet av  $B(\mathbf{a}, \delta)$  (markert med den stiplete kurven på figuren) ligger helt inni  $B(\mathbf{F}(\mathbf{a}), \epsilon)$ .



Figur 1: Kontinuitet i punktet  $\mathbf{a}$

Siden kontinuitet er definert på akkurat samme måte som for funksjoner av én variabel, har vi de samme reglene med (nesten) de samme bevisene.

**Setning 2.2.2** Anta at  $A \subset \mathbb{R}^n$ , og at funksjonene  $\mathbf{F}, \mathbf{G} : A \rightarrow \mathbb{R}^m$  er kontinuerlige i  $\mathbf{a} \in A$ . Da er  $\mathbf{F} + \mathbf{G}$ ,  $\mathbf{F} - \mathbf{G}$  og  $\mathbf{F} \cdot \mathbf{G}$  kontinuerlige i  $\mathbf{a}$ . Det er også  $\frac{\mathbf{F}}{\mathbf{G}}$  forutsatt at  $\mathbf{F}$  og  $\mathbf{G}$  tar verdier i  $\mathbb{R}$  (slik at divisjon gir mening) og  $\mathbf{G}(\mathbf{a}) \neq 0$ .

*Bevis:* Bevisene er akkurat som for funksjoner av én variabel (se setning 5.1.5 i *Kalkulus*), den eneste ekstra komplikasjonen er at vi noen steder må appellere til trekantulikheten (setning 1.2.4) og Schwarz' ulikhet (setning 1.2.3) for å få argumentene til å gå opp. Vi tar  $\mathbf{F} + \mathbf{G}$  og  $\mathbf{F} \cdot \mathbf{G}$  som eksempler — det første av disse bevisene er ganske enkelt, det andre er mer komplisert:

For å vise at  $\mathbf{F} + \mathbf{G}$  er kontinuerlig i  $\mathbf{a}$ , må vi vise at for hver  $\epsilon > 0$  finnes det en  $\delta > 0$  slik at hvis  $\mathbf{x} \in A$  og  $|\mathbf{x} - \mathbf{a}| < \delta$ , så er

$$|(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) - (\mathbf{F}(\mathbf{a}) + \mathbf{G}(\mathbf{a}))| < \epsilon$$

Vi stokker litt om på leddene slik at vi får  $\mathbf{F}$  og  $\mathbf{G}$  hver for seg, og bruker deretter trekantulikheten:

$$\begin{aligned} |(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) - (\mathbf{F}(\mathbf{a}) + \mathbf{G}(\mathbf{a}))| &= |(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})) + (\mathbf{G}(\mathbf{x}) + \mathbf{G}(\mathbf{a}))| \leq \\ &\leq |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| + |\mathbf{G}(\mathbf{x}) + \mathbf{G}(\mathbf{a})| \end{aligned}$$

Siden  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ , finnes det en  $\delta_1 > 0$  slik at  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \frac{\epsilon}{2}$  når  $\mathbf{x} \in A$  og  $|\mathbf{x} - \mathbf{a}| < \delta_1$ . Siden  $\mathbf{G}$  er kontinuerlig i  $\mathbf{a}$ , finnes det tilsvarende

en  $\delta_2 > 0$  slik at  $|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| < \frac{\epsilon}{2}$  når  $\mathbf{x} \in A$  og  $|\mathbf{x} - \mathbf{a}| < \delta_2$ . Lar vi  $\delta$  være det minste av de to tallene  $\delta_1$  og  $\delta_2$ , ser vi at når  $\mathbf{x} \in A$  og  $|\mathbf{x} - \mathbf{a}| < \delta$ , så er

$$\begin{aligned} |(\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) - (\mathbf{F}(\mathbf{a}) + \mathbf{G}(\mathbf{a}))| &\leq \\ |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| + |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Dermed har vi vist at  $\mathbf{F} + \mathbf{G}$  er kontinuerlig i  $\mathbf{a}$ .

La oss så vise kontinuitet av  $\mathbf{F} \cdot \mathbf{G}$ . Gitt en  $\epsilon > 0$ , må vi finne en  $\delta > 0$  slik at hvis  $\mathbf{x} \in A$  og  $|\mathbf{x} - \mathbf{a}| < \delta$ , så er  $|\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| < \epsilon$ . Vi legger til og trekker fra  $\mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x})$ , og bruker deretter trekantulikheten og Schwarz' ulikhet:

$$\begin{aligned} |\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| &= \\ |\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x}) + \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| &\leq \\ \leq |\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x})| + |\mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| &\leq \\ \leq |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x})| + |\mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| & \end{aligned}$$

Vi vil være i mål dersom vi kan vise at vi kan få begge uttrykkene  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x})|$  og  $|\mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})|$  mindre enn  $\frac{\epsilon}{2}$  ved å velge  $\mathbf{x}$  tilstrekkelig nær  $\mathbf{a}$ . Det siste ledet er det enkleste, så vi begynner med det. Hvis  $\mathbf{F}(\mathbf{a}) = \mathbf{0}$ , er ledet lik 0 og derfor mindre enn  $\frac{\epsilon}{2}$ . Vi kan derfor anta at  $\mathbf{F}(\mathbf{a}) \neq \mathbf{0}$ . Siden  $\mathbf{G}$  er kontinuerlig i  $\mathbf{a}$ , finnes det en  $\delta_1 > 0$  slik at hvis  $|\mathbf{x} - \mathbf{a}| < \delta_1$ , så er  $|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| < \frac{\epsilon}{2|\mathbf{F}(\mathbf{a})|}$ . Dermed er

$$|\mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| \leq |\mathbf{F}(\mathbf{a})| \frac{\epsilon}{2|\mathbf{F}(\mathbf{a})|} = \frac{\epsilon}{2}$$

Vi tar så for oss ledet  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x})|$ . Observer først at siden  $\mathbf{G}$  er kontinuerlig i  $\mathbf{a}$ , så finnes det en  $\delta_2 > 0$  slik at hvis  $|\mathbf{x} - \mathbf{a}| < \delta_2$ , så er  $|\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| < 1$ . Dette betyr at dersom  $|\mathbf{x} - \mathbf{a}| < \delta_2$ , så er (vi bruker trekantulikheten igjen)

$$|\mathbf{G}(\mathbf{x})| = |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a}) + \mathbf{G}(\mathbf{a})| \leq |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| + |\mathbf{G}(\mathbf{a})| \leq 1 + |\mathbf{G}(\mathbf{a})|$$

Siden  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ , finnes det en  $\delta_3 > 0$  slik at hvis  $|\mathbf{x} - \mathbf{a}| < \delta_3$ , så er  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| \leq \frac{\epsilon}{2(1+|\mathbf{G}(\mathbf{a})|)}$ . Er  $|\mathbf{x} - \mathbf{a}|$  mindre enn både  $\delta_1$  og  $\delta_2$ , er dermed

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x})| \leq \frac{\epsilon}{2(1+|\mathbf{G}(\mathbf{a})|)} (1 + |\mathbf{G}(\mathbf{a})|) = \frac{\epsilon}{2}$$

Velger vi  $\delta$  til å være det minste av tallene  $\delta_1$ ,  $\delta_2$  og  $\delta_3$ , ser vi dermed at når  $\mathbf{x} \in A$  og  $|\mathbf{x} - \mathbf{a}| < \delta$ , så er

$$|\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x}) - \mathbf{F}(\mathbf{a}) \cdot \mathbf{G}(\mathbf{a})| \leq |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x})| + |\mathbf{F}(\mathbf{a})| |\mathbf{G}(\mathbf{x}) - \mathbf{G}(\mathbf{a})| <$$

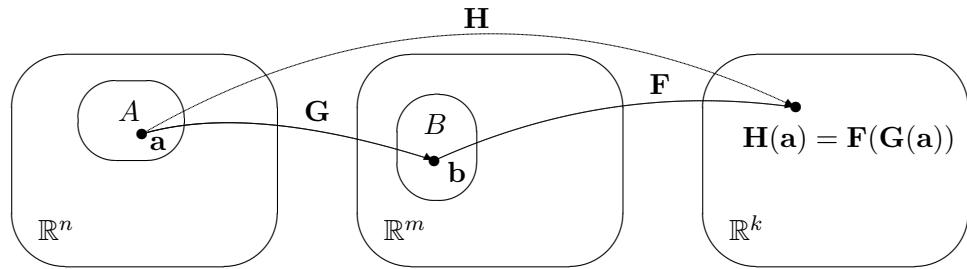
$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

og beviset er fullført.  $\square$

Det neste resultatet kjenner du også igjen fra teorien for funksjoner av én variabel:

**Setning 2.2.3** *Anta at vi har to mengder  $A \subset \mathbb{R}^n$ ,  $B \subset \mathbb{R}^m$ , og to funksjoner  $\mathbf{G} : A \rightarrow B$ ,  $\mathbf{F} : B \rightarrow \mathbb{R}^k$  (se figuren nedenfor). Dersom  $\mathbf{G}$  er kontinuerlig i punktet  $\mathbf{a}$ , og  $\mathbf{F}$  er kontinuerlig i punktet  $\mathbf{b} = \mathbf{G}(\mathbf{a})$ , så er den sammensatte funksjonen  $\mathbf{H}(x) = \mathbf{F}(\mathbf{G}(x))$  kontinuerlig i  $\mathbf{a}$ .*  $\square$

Beviset er akkurat som for funksjoner av én variabel (se setning 5.1.7 i *Kalkulus*) for hjelp). Figur 2 viser hvordan  $\mathbf{G}$ ,  $\mathbf{F}$  og  $\mathbf{H}$  virker.



Figur 2: Sammensetning av funksjoner

Når vi skal bruke reglene ovenfor til å vise at en funksjon  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  er kontinuerlig, lønner det seg ofte å skrive den på komponentform:

$$\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_m(\mathbf{x}))$$

Det neste resultatet forteller oss nemlig at  $\mathbf{F}$  er kontinuerlig hvis og bare hvis hver komponent  $F_i$  er kontinuerlig.

**Setning 2.2.4** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av  $n$  variable med komponenter*

$$\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_m(\mathbf{x}))$$

*Da er  $\mathbf{F}$  kontinuerlig i et punkt  $\mathbf{a} \in A$  hvis og bare hvis hver komponent  $F_i$  er kontinuerlig i  $\mathbf{a}$ .*

*Bevis:* La oss først anta at alle komponentene  $F_1, F_2, \dots, F_m$  er kontinuerlige i  $\mathbf{a}$ . Vi må vise at for enhver  $\epsilon > 0$ , finnes det en  $\delta > 0$  slik at hvis  $|\mathbf{x} - \mathbf{a}| < \delta$ , så er  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$ . Siden hver komponent  $F_i$  er kontinuerlig i  $\mathbf{a}$ , finnes det en  $\delta_i > 0$  slik at  $|F_i(\mathbf{x}) - F_i(\mathbf{a})| < \frac{\epsilon}{\sqrt{m}}$  når  $|\mathbf{x} - \mathbf{a}| < \delta_i$ . Vi velger  $\delta$  til å være det minste av tallene  $\delta_1, \delta_2, \dots, \delta_m$ . Hvis  $|\mathbf{x} - \mathbf{a}| < \delta$ , har vi da

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| = \sqrt{(F_1(\mathbf{x}) - F_1(\mathbf{a}))^2 + \dots + (F_m(\mathbf{x}) - F_m(\mathbf{a}))^2} \leq$$

$$\leq \sqrt{\left(\frac{\epsilon}{\sqrt{m}}\right)^2 + \cdots + \left(\frac{\epsilon}{\sqrt{m}}\right)^2} = \sqrt{m \left(\frac{\epsilon^2}{m}\right)} = \epsilon$$

Dette viser at  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ .

Anta nå omvendt at  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ , og at vi skal vise at komponenten  $F_i$  er kontinuerlig i  $\mathbf{a}$ . Gitt  $\epsilon > 0$ , må vi finne en  $\delta > 0$  slik at  $|F_i(\mathbf{x}) - F_i(\mathbf{a})| < \epsilon$  når  $|\mathbf{x} - \mathbf{a}| < \delta$ . Siden  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ , finnes det en  $\delta > 0$  slik at  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon$  når  $|\mathbf{x} - \mathbf{a}| < \delta$ . Men dermed er

$$\begin{aligned} |F_i(\mathbf{x}) - F_i(\mathbf{a})| &\leq \sqrt{(F_1(\mathbf{x}) - F_1(\mathbf{a}))^2 + \cdots + (F_m(\mathbf{x}) - F_m(\mathbf{a}))^2} = \\ &= |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| < \epsilon \end{aligned}$$

når  $|\mathbf{x} - \mathbf{a}| < \delta$ . Dette viser at  $F_i$  er kontinuerlig i  $\mathbf{a}$ .  $\square$

De enkleste funksjonene av flere variable (bortsett fra konstantfunksjonene) er de som bare gir oss en av variablene som funksjonsverdi, f.eks.

$$f(x, y, z, u) = y \quad \text{eller} \quad g(x, y, z) = z$$

Generelt kaller vi

$$k_i(x_1, x_2, \dots, x_n) = x_i$$

den  $i$ -te *koordinatfunksjonen*. Det er lett å se at disse koordinatfunksjonene er kontinuerlige, og det neste eksemplet forteller oss hvordan vi kan bruke dem og resultatene ovenfor til å vise at mer kompliserte funksjoner er kontinuerlige.

**Eksempel 1:** Vis at funksjonen  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  definert ved

$$\mathbf{F}(x, y, z) = \left( x^2 \sin(yz), \frac{e^{x^2+z}}{y^2+1} \right)$$

er kontinuerlig i punktet  $\mathbf{a} = (1, 0, -1)$ . Ifølge setning 2.2.4 er det nok å vise at begge funksjonene

$$F_1(x, y, z) = x^2 \sin(yz) \quad \text{og} \quad F_2(x, y, z) = \frac{e^{x^2+z}}{y^2+1}$$

er kontinuerlig i  $\mathbf{a}$ . Vi begynner med  $F_1(x, y, z)$ . Siden koordinatfunksjonene  $k_2(x, y, z) = y$  og  $k_3(x, y, z) = z$  er kontinuerlige, er produktet av dem  $f(x, y, z) = yz$  også kontinuerlig ifølge setning 2.2.2. Siden sinus er en kontinuerlig funksjon, må den sammensatte funksjonen  $\sin(yz)$  også være kontinuerlig ifølge setning 2.2.3. På tilsvarende måte ser vi at funksjonen  $g(x, y, z) = x^2$  er kontinuerlig (fordi den er lik  $k_1(x, y, z) \cdot k_1(x, y, z)$ )

som er et produkt av to kontinuerlige funksjoner), og dermed er produktet  $F_1(x, y, z) = x^2 \sin(yz)$  kontinuerlig.

Vi behandler  $F_2$  på samme måte: Funksjonen  $h(x, y, z) = x^2 + z$  er kontinuerlig siden den er bygget opp fra kontinuerlige koordinatfunksjoner ved hjelp av multiplikasjon og addisjon. Siden eksponentialfunksjonen er kontinuerlig, er da  $p(x, y, z) = e^{x^2+z}$  kontinuerlig. Ved et tilsvarende resonnement ser vi at nevneren  $y^2 + 1$  er kontinuerlig, og siden den er forskjellig fra 0, må brøken

$$F_2(x, y, z) = \frac{e^{x^2+z}}{y^2 + 1}$$

være kontinuerlig. ♣

Ved hjelp av teknikken i dette eksemplet er det lett å vise at funksjoner bygget opp ved hjelp av potensfunksjoner, eksponentialfunksjoner, logaritmer, trigonometriske funksjoner og arcusfunksjoner er kontinuerlige (men husk å sjekke at de er definert i punktet du er interessert i!)

Hittil har teorien for kontinuerlige funksjoner av flere variable vært forbløffende lik teorien for kontinuerlige funksjoner av én variabel. På grunn av den rikere geometrien finnes det imidlertid fenomener for flervariabel funksjoner som ikke har noen motsvarighet for funksjoner av én variabel. Her er et eksempel.

**Eksempel 2:** Funksjonen  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  er gitt ved

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{for } (x, y) \neq (0, 0) \\ 0 & \text{for } (x, y) = (0, 0) \end{cases}$$

Vi skal undersøke om  $f$  er kontinuerlig i  $(0, 0)$ . Uformelt sier en gjerne at  $f$  er kontinuerlig  $(0, 0)$  dersom  $f(x, y)$  nærmer seg  $f(0, 0)$  når  $(x, y)$  nærmer seg  $(0, 0)$ . La oss derfor se hva som skjer når  $(x, y)$  nærmer seg  $(0, 0)$  på forskjellige måter.

La oss først se på hva som skjer når  $(x, y)$  nærmer seg origo langs en skrålinje  $y = cx$  der  $c \neq 0$ ; dvs. vi har punkter  $(x, cx)$  der  $x$  går mot null. Setter vi inn i funksjonsuttrykket, får vi

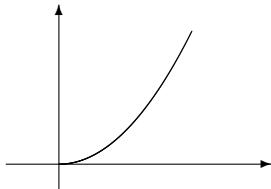
$$f(x, y) = f(x, cx) = \frac{x^2(cx)}{x^4 + cx^2} = \frac{cx}{x^2 + c} \xrightarrow{x \rightarrow 0} \frac{0}{c} = 0$$

Tilsvarende ser vi at når  $(x, y)$  nærmer seg  $(0, 0)$  langs  $x$ -aksen, så er punktene på formen  $(x, 0)$  der  $x \rightarrow 0$ , og vi får

$$f(x, y) = f(x, 0) = \frac{x^2 \cdot 0}{x^4 + 0^2} = 0$$

Akkurat det samme skjer om vi lar  $(x, y)$  nærme seg  $(0, 0)$  langs  $y$ -aksen.

Alt dette tyder på at  $f$  er kontinuerlig i  $(0, 0)$ , men det finnes jo andre måter å nærme seg et punkt på enn å følge en rett linje. La oss prøve å nærme oss  $(0, 0)$  langs parabelen  $y = x^2$  (se figuren). Vi ser altså på punkter  $(x, x^2)$  der  $x$  går mot 0.



Figur 3. Punkter som nærmer seg  $(0, 0)$  langs parabelen  $y = x^2$

Setter vi inn i funksjonsuttrykket, får vi

$$f(x, y) = f(x, x^2) = \frac{x^2 \cdot x^2}{x^4 + (x^2)^2} = \frac{1}{2}$$

Dette viser at  $f$  ikke er kontinuerlig i  $(0, 0)$  siden vi har punkter med verdien  $\frac{1}{2}$  så nær  $(0, 0)$  vi måtte ønske. Funksjonen virker altså å være kontinuerlig så lenge vi beveger oss langs rette linjer, men er det likevel ikke! ♣

Eksemplet ovenfor viser at det kan være vanskelig å få god oversikt over hvordan en funksjon av flere variable oppfører seg, og at det kanskje finnes flere geometriske muligheter enn det vi kan forestille oss. Det er i slike sammenhenger vi virkelig får nytte av abstrakte definisjoner av  $\epsilon$ - $\delta$ -typen; de gir oss muligheten til å føre vanntette bevis selv i tilfeller der vi ikke er sikre på om vi har fått full oversikt over alle geometriske snurrepiper!

Hittil har vi bare snakket om kontinuitet i et punkt. Vi avslutter denne seksjonen med definisjonen av en kontinuerlig funksjon — den er helt tilsvarende definisjonen for funksjoner av én variabel.

**Definisjon 2.2.5** En funksjon  $\mathbf{F}$  kalles kontinuerlig dersom den er kontinuerlig i alle punkter i sitt definisjonsområde.

### Oppgaver til seksjon 2.2

1. Vis at funksjonen  $f$  er kontinuerlig

- |                                     |                                   |
|-------------------------------------|-----------------------------------|
| a) $f(x, y) = x + y$                | d) $f(x, y) = e^{-x} \sin(x + y)$ |
| b) $f(x, y) = x^2y + y$             | e) $f(x, y, z) = x^2 + y^2 + z^2$ |
| c) $f(x, y) = \frac{xy}{1+x^2+y^2}$ |                                   |

2. Vis at funksjonene er kontinuerlige:

- a)  $\mathbf{F}(x, y, z) = (x^2z + y, x^2 \sin xyz, x^3)$
- b)  $\mathbf{G}(x, y, z, u) = (e^{xu+z^2}, z \cos xy^2u)$
- c)  $\mathbf{H}(x, y, z, u) = (x^y e^{xz^2}, z + u^2, x^2 + 3yzu)$

d)  $\mathbf{K}(x, y, z, u, v) = (\sin(xy + z^2v), 2uv)$

**3.** Vis at koordinatfunksjonene  $k_i(x_1, x_2, \dots, x_n) = x_i$  er kontinuerlige ved å bruke definisjonen av kontinuitet.

**4.** a) Anta det finnes en konstant  $M \in \mathbb{R}$  slik at  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \leq M|\mathbf{x} - \mathbf{y}|$  for alle  $\mathbf{x}, \mathbf{y}$  i definisjonsområdet  $D_{\mathbf{F}}$ . Vis at  $\mathbf{F}$  er kontinuerlig.

b) Anta at  $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$  for en matrise  $A$ . Vis at  $\mathbf{F}$  er kontinuerlig. (*Hint:* Husk setning 1.6.3.)

**5.** I denne oppgaven har du bruk for trekantulikheten som sier at hvis  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , så er  $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$ .

a) Vis at  $||\mathbf{x}| - |\mathbf{y}|| \leq |\mathbf{x} - \mathbf{y}|$  for alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ .

b) La  $\mathbf{a} \in \mathbb{R}^n$ . Vis at funksjonen  $f(\mathbf{x}) = |\mathbf{x} - \mathbf{a}|$  er kontinuerlig.

c) Vis at funksjon  $g(\mathbf{x}) = \frac{1}{|\mathbf{x} - \mathbf{a}|}$  er kontinuerlig der den er definert.

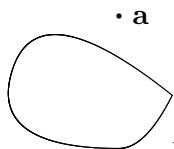
**6.** Bevis setning 2.2.2.

**7.** Bevis setning 2.2.3.

### 2.3 Grenseverdier

I teorien for funksjoner av én variabel opererer vi både med ensidige og tosidige grenser. Ensidige grenser er blant annet nyttig når vi skal avgjøre om en funksjon er kontinuerlig i enden av sitt definisjonsområde — vi bruker grenseverdien  $\lim_{x \rightarrow a^+} f(x)$  for å undersøke om  $f$  er kontinuerlig i det venstre endepunktet  $a$  av intervallet  $[a, b]$ , og grenseverdien  $\lim_{x \rightarrow b^-} f(x)$  for å undersøke om  $f$  er kontinuerlig i det høyre endepunktet  $b$ . Med flere variable er det så mange måter å nærme seg et randpunkt på at vi ikke kan ha ett grensebegrep for hver måte — vi må finne frem til et felles begrep som dekker alle tilfeller.

La oss først se litt på hvilke punkter det er naturlig å regne ut grenseverdier i. Vi ser på en funksjon  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  definert på en delmengde  $A$  av  $\mathbb{R}^n$ . Figur 1 viser én mulighet; her består  $A$  av en sammenhengende mengde pluss et isolert punkt  $\mathbf{a}$ .



Figur 1

Siden vi ikke kan nærme oss  $\mathbf{a}$  innenfor definisjonsmengden til  $\mathbf{F}$ , er det

ikke rimelig å definere grenseverdien til  $\mathbf{F}$  i punktet  $\mathbf{a}$ . Figur 2 viser en annen situasjon.



Figur 2

På denne figuren er omkretsen (“randen”) til  $A$  stiplet for å markere at punktene der ikke hører med til mengden. Dermed er  $\mathbf{a}$  ikke med i definisjonsmengden  $A$ , men det går fint an å nærme seg  $\mathbf{a}$  fra  $A$ . I dette tilfellet er det rimelig å definere grenseverdien til  $\mathbf{F}$  i punktet  $\mathbf{a}$ . Konklusjonen på disse observasjonene må bli at det er rimelig å definere grenseverdien til  $\mathbf{F}$  i punktet  $\mathbf{a}$  dersom det går an å nærme seg  $\mathbf{a}$  med punkter ( $\neq \mathbf{a}$ ) fra definisjonsmengden til  $\mathbf{F}$ , uansett om punktet  $\mathbf{a}$  selv ligger i definisjonsmengden eller ikke. Slike punkter kaller vi *oppnopningspunkter* for  $A$ . Her er den presise definisjonen:

**Definisjon 2.3.1** La  $A$  være en delmengde av  $\mathbb{R}^n$ . Et punkt  $\mathbf{a} \in \mathbb{R}^n$  kalles et oppnopningspunkt for  $A$  dersom enhver kule  $B(\mathbf{a}, r)$  om  $\mathbf{a}$  inneholder uendelig mange punkter fra  $A$ .

Vi er nå klare til å definere grenseverdien i et oppnopningspunkt:

**Definisjon 2.3.2** La  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  være en funksjon av  $n$  variable og anta at  $\mathbf{a} \in \mathbb{R}^n$  er et oppnopningspunkt for  $A$ . Vi sier at  $\mathbf{b} \in \mathbb{R}^m$  er grenseverdien for  $\mathbf{F}$  i punktet  $\mathbf{a}$  dersom det for hver  $\epsilon > 0$  finnes en  $\delta > 0$  slik at

$$|\mathbf{F}(\mathbf{x}) - \mathbf{b}| < \epsilon \text{ for alle } \mathbf{x} \in A \text{ slik at } 0 < |\mathbf{x} - \mathbf{a}| < \delta$$

Vi skriver  $\lim_{x \rightarrow \mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{b}$ . Dersom  $\mathbf{a}$  ikke er et oppnopningspunkt for  $A$ , er grenseverdien ikke definert.

Denne definisjonen ligner på den du finner for funksjoner av én variabel i definisjon 5.4.1 i *Kalkulus*, men skiller seg på et viktig punkt: Vi insisterer ikke lenger på at  $\mathbf{F}$  skal være definert i alle punkter i nærheten av  $\mathbf{a}$ , men kompenserer for dette ved bare å kreve at  $|\mathbf{F}(\mathbf{x}) - \mathbf{b}| < \epsilon$  skal holde for punkter  $\mathbf{x}$  som er med i definisjonsområdet  $A$  til  $\mathbf{F}$ .

Akkurat som kontinuitet kan grenseverdier studeres komponentvis:

**Setning 2.3.3** La  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  være en funksjon av  $n$  variable, og anta at  $\mathbf{a} \in \mathbb{R}^n$  er et oppnopningspunkt for  $A$ . Anta at komponentene til  $\mathbf{F}$  er gitt ved

$$\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_m(\mathbf{x}))$$

og la  $\mathbf{b}$  være en vektor med komponenter  $\mathbf{b} = (b_1, b_2, \dots, b_m)$ . Da er

$$\lim_{x \rightarrow \mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{b}$$

hvis og bare hvis

$$\lim_{x \rightarrow \mathbf{a}} F_i(\mathbf{x}) = b_i \text{ for alle } i$$

□

Beviset overlates til leserne. Ideen er den samme som i beviset for setning 2.2.4.

Vi har også de vanlige regnereglene for grenseverdier til summer, differenser, produkter og brøker:

**Setning 2.3.4 (Regneregler for grenseverdier)** *Anta at  $\mathbf{F}, \mathbf{G} : A \rightarrow \mathbb{R}^m$  er to funksjoner av  $n$  variable og at  $\mathbf{a} \in A$  er et opphopningspunkt for  $A$ . Dersom  $\lim_{x \rightarrow \mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{A}$  og  $\lim_{x \rightarrow \mathbf{a}} \mathbf{G}(\mathbf{x}) = \mathbf{B}$ , så er:*

- a)  $\lim_{x \rightarrow \mathbf{a}} (\mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})) = \mathbf{A} + \mathbf{B}$ .
- b)  $\lim_{x \rightarrow \mathbf{a}} (\mathbf{F}(\mathbf{x}) - \mathbf{G}(\mathbf{x})) = \mathbf{A} - \mathbf{B}$ .
- c)  $\lim_{x \rightarrow \mathbf{a}} (\mathbf{F}(\mathbf{x}) \cdot \mathbf{G}(\mathbf{x})) = \mathbf{A} \cdot \mathbf{B}$ .
- d)  $\lim_{x \rightarrow \mathbf{a}} \frac{\mathbf{F}(\mathbf{x})}{\mathbf{G}(\mathbf{x})} = \frac{\mathbf{A}}{\mathbf{B}}$  forutsatt at  $\mathbf{F}$  og  $\mathbf{G}$  tar verdier i  $\mathbb{R}$  og  $\mathbf{B} \neq 0$ .

□

Også disse bevisene overlates til leserne.

Det neste resultatet tar seg av sammenhengen mellom grenseverdier og kontinuitet. Legg merke til at det bare gjelder for kontinuitet i opphopningspunkter (men i et isolerte punkt er en funksjon alltid kontinuerlig, så der er det ikke så mye å vise!)

**Setning 2.3.5** *La  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  være en funksjon av  $n$  variable, og anta at  $\mathbf{a} \in A$  er et opphopningspunkt for  $A$ . Da er  $\mathbf{F}$  kontinuerlig i  $\mathbf{a}$  hvis og bare hvis  $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a})$*

*Bevis:* Sammenlign definisjonen av kontinuitet med definisjonen av grenseverdi. □

Resultatet ovenfor er nyttig når vi skal regne ut enkle grenseverdier:

**Eksempel 1:** Finn grenseverdien til

$$F(x, y) = (x^2 y, e^{-xy} \sin(\pi x))$$

når  $(x, y) \rightarrow (1, -2)$ .

Vi ser at funksjonen er kontinuerlig i  $(1, -2)$ , så

$$\lim_{(x,y) \rightarrow (1,-2)} F(x, y) = F(1, -2) = (1^2 \cdot (-2), e^{-1 \cdot (-2)} \sin(\pi \cdot 1)) = (-2, 0)$$

♣

I noen eksempler må vi forenkle uttrykket før vi går til grensen:

**Eksempel 2:** Finn grenseverdien til

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

når  $(x, y) \rightarrow (0, 0)$ .

Siden  $x^2 - y^2 = (x + y)(x - y)$  har vi

$$\begin{aligned} \lim_{(x,y) \rightarrow (0,0)} f(x, y) &= \lim_{(x,y) \rightarrow (0,0)} \frac{x^2 - y^2}{x - y} = \\ &= \lim_{(x,y) \rightarrow (0,0)} \frac{(x + y)(x - y)}{x - y} = \lim_{(x,y) \rightarrow (0,0)} (x + y) = 0 \end{aligned}$$



### Oppgaver til seksjon 2.3

1. Finn grenseverdiene

- a)  $\lim_{(x,y) \rightarrow (2,3)} (x^3 + 2xy)$
- b)  $\lim_{(x,y) \rightarrow (1, \frac{\pi}{2})} x^2 \sin(xy)$
- c)  $\lim_{(x,y) \rightarrow (1,0)} \frac{e^{x+y}}{x^2 + 3y}$
- d)  $\lim_{(x,y) \rightarrow (0,0)} \frac{\sin(xy)}{xy} \cdot \cos(x + y)$

2. Anta at  $A \subset \mathbb{R}^n$  og at  $\mathbf{a} \in \mathbb{R}^n$ . Anta at enhver kule  $B(\mathbf{a}, \epsilon)$  om  $\mathbf{a}$  inneholder minst ett element fra  $A$  forskjellig fra  $\mathbf{a}$ . Vis at  $\mathbf{a}$  er et opphopningspunkt for  $A$ .

3. Bevis setning 2.3.3.

4. Bevis setning 2.3.4

5. Bevis setning 2.3.5

## 2.4 Derivasjon av skalarfelt

Vi skal nå begynne å se på derivasjon av funksjoner av flere variable. For at det ikke skal være altfor mange komponenter å holde styr på, skal vi først derivere skalarfelt, dvs. funksjoner  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  der verdiene  $f(\mathbf{x})$  er tall. Før vi setter igang for alvor, trenger vi en definisjon: Hvis  $A$  er en delmengde av  $\mathbb{R}^n$ , kalles  $\mathbf{a} \in A$  et *indre punkt* i  $A$  dersom det finnes en  $\epsilon > 0$  slik at  $B(\mathbf{a}, \epsilon) \subset A$ . Dette betyr at dersom vi starter i  $\mathbf{a}$ , kan vi gå et lite stykke i en hvilken som helst retning uten å forlate  $A$ .

For en funksjon  $y = f(x)$  av én variabel forteller den deriverte  $f'(x)$  oss hvor fort funksjonen vokser i punktet  $x$  — går vi et lite skritt med lengde  $h$  langs  $x$ -aksen, vil funksjonsverdien øke med (omtrent)  $f'(x)h$ . For funksjoner

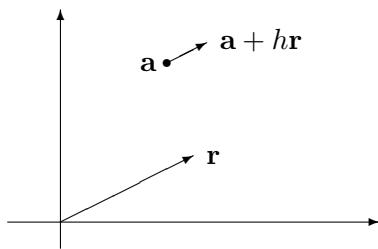
av flere variable er situasjonen mer komplisert; vi har flere akser å bevege oss langs, og vi kan ikke regne med at funksjonen stiger like mye uansett hvilken retning vi går i. Før vi regner ut stigningstallet til funksjonen, må vi derfor spesifisere hvilken retning vi er interessert i. Dette er idéen bak begrepet retningsderivert:

**Definisjon 2.4.1** *Anta at funksjonen  $f : A \rightarrow \mathbb{R}$  er definert på en delmengde  $A$  av  $\mathbb{R}^n$  og at  $\mathbf{a}$  er et indre punkt i  $A$ . Tenk på  $\mathbf{r} \in \mathbb{R}^n$  som en vektor. Den retningsderiverte til  $f$  i punktet  $\mathbf{a}$  og retningen  $\mathbf{r}$  er gitt ved*

$$f'(\mathbf{a}; \mathbf{r}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})}{h}$$

forutsatt at denne grensen eksisterer.

Figur 1 viser ideen bak definisjonen.



Figur 1

Punktene  $\mathbf{a} + h\mathbf{r}$  er de punktene vi kommer til hvis vi starter i  $\mathbf{a}$  og går i retning  $\mathbf{r}$ . Differansen  $f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})$  forteller oss hvor mye funksjonen øker når vi beveger oss i denne retning, og brøken

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

er økningen per lengdeenhet når vi bruker  $|\mathbf{r}|$  som måleenhet.

Legg merke til at vi bare har definert retningsderiverte i indre punkter. Det garanterer at uttrykket  $f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})$  alltid gir mening bare vi velger  $h$  liten nok. I punkter som ikke er indre, kan man ofte definere retningsderiverte i noen retninger, men ikke i andre. Vi skal ikke komme inn på dette her, men konsentrere oss om indre punkter.

**Eksempel 1:** La  $f(x, y) = x^2 + xy$ . Vi skal beregne den retningsderiverte  $f'(\mathbf{a}; \mathbf{r})$  når  $\mathbf{a} = (1, 0)$ ,  $\mathbf{r} = (2, 1)$ . Først observerer vi at

$$\mathbf{a} + h\mathbf{r} = (1, 0) + h(2, 1) = (1 + 2h, h),$$

som gir  $f(\mathbf{a} + h\mathbf{r}) = (1 + 2h)^2 + (1 + 2h)h = 1 + 5h + 6h^2$ . Tilsvarende er

$f(\mathbf{a}) = f(1, 0) = 1^2 + 1 \cdot 0 = 1$ . Vi får

$$\begin{aligned} f'(\mathbf{a}; \mathbf{r}) &= \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})}{h} = \lim_{h \rightarrow 0} \frac{(1 + 5h + 6h^2) - 1}{h} \\ &= \lim_{h \rightarrow 0} \frac{5h + 6h^2}{h} = \lim_{h \rightarrow 0} (5 + 6h) = 5. \end{aligned}$$

Hva betyr dette resultatet? Legg merke til at lengden til vektoren  $\mathbf{r}$  er  $|\mathbf{r}| = \sqrt{2^2 + 1^2} = \sqrt{5}$ . Dersom vi går et lite stykke  $h\sqrt{5}$  i retningen til vektoren  $\mathbf{r} = (2, 1)$ , vil funksjonsverdien stige med (omtrent)  $f'(\mathbf{a}; \mathbf{r}) \cdot h = 5 \cdot h$ . ♣

Det er lettest å forstå hva den retningsderiverte er dersom vektoren  $\mathbf{r}$  har lengde 1 – da er  $f'(\mathbf{a}; \mathbf{r})$  rett og slett stigningstallet til funksjonen i retning  $\mathbf{r}$  når vi måler med vanlige enheter. Ut ifra dette kan det være fristende å forutsette at  $|\mathbf{r}| = 1$  når vi regner med retningsderiverte, men dette viser seg å være upraktisk, blant annet fordi enhetsvektorer ofte inneholder stygge kvadratrøtter.

Så langt kan det se ut som om vi må bygge opp en ny derivasjonsteori helt fra bunnen av for å kunne beregne retningsderiverte til funksjoner av flere variable. Det er heldigvis ikke nødvendig; ved hjelp av såkalte partiellderiverte kan vi føre mye av teorien tilbake til vanlig derivasjon av funksjoner av én variabel. Før vi definerer partiellderiverte, er det lurt å bli enig om litt notasjon.

Den  $i$ -te enhetsvektoren  $\mathbf{e}_i$  i  $\mathbb{R}^n$  er vektoren

$$\mathbf{e}_i = (0, 0, \dots, 0, \underset{\substack{\uparrow \\ i\text{-te plass}}}{1}, 0, \dots, 0)$$

langs den  $i$ -te koordinataksen.

**Definisjon 2.4.2** La  $f: A \rightarrow \mathbb{R}$  være en funksjon av  $n$  variable, og la  $\mathbf{a}$  være et indre punkt i  $A$ . Den  $i$ -te partiellderiverte  $\frac{\partial f}{\partial x_i}(\mathbf{a})$  er den retningsderiverte av  $f$  i retning av den  $i$ -te enhetsvektoren  $\mathbf{e}_i$ ; det vil si

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = f'(\mathbf{a}; \mathbf{e}_i)$$

Andre notasjoner for  $\frac{\partial f}{\partial x_i}(\mathbf{a})$  er  $D_i f(\mathbf{a})$  og  $f_{x_i}(\mathbf{a})$ . De partiellderiverte er altså stigningstallene til funksjonen parallelt med koordinataksene. Skriver vi ut definisjonen i detalj, ser vi at

$$\begin{aligned} \frac{\partial f}{\partial x_i}(\mathbf{a}) &= f'(\mathbf{a}; \mathbf{e}_i) = \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h} = \\ &= \lim_{h \rightarrow 0} \frac{f(a_1, a_2, \dots, a_i + h, \dots, a_n) - f(a_1, a_2, \dots, a_i, \dots, a_n)}{h} \end{aligned}$$

Det siste uttrykket har en slående likhet med definisjonen av vanlig derivert. Underslår vi de variablene hvor det ikke skjer noen endring, ser vi at

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \rightarrow 0} \frac{f(a_i + h) - f(a_i)}{h}$$

Dette betyr at vi kan finne den partiellderiverte  $\frac{\partial f}{\partial x_i}$  ved å derivere uttrykket  $y = f(x_1, \dots, x_i, \dots, x_n)$  med hensyn på  $x_i$  mens vi later som om alle de andre variablene er konstanter.

**Eksempel 2:** Finn de partiellderiverte  $\frac{\partial f}{\partial x}$  og  $\frac{\partial f}{\partial y}$  til funksjonen

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

For å finne  $\frac{\partial f}{\partial x}$  deriverer vi uttrykket med hensyn på  $x$  mens vi later som om  $y$  er en konstant:

$$\frac{\partial f}{\partial x}(x, y) = 2x + y^3 + \cos(xy) \cdot y$$

For å finne  $\frac{\partial f}{\partial y}$  deriverer vi med hensyn på  $y$  mens vi holder  $x$  konstant:

$$\frac{\partial f}{\partial y}(x, y) = 0 + 3xy^2 + \cos(xy)x = 3xy^2 + x \cos(xy)$$



En funksjon  $f(x_1, x_2, \dots, x_n)$  av  $n$  variable har  $n$  partiellderiverte  $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}$ . Vi kan sette sammen disse til en vektor

$$\left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)$$

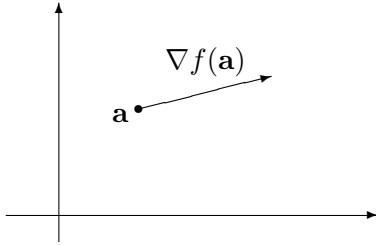
Denne vektoren er så viktig at den har fått sitt eget navn og sitt eget symbol.

**Definisjon 2.4.3** Anta at de partiellderiverte til  $f$  eksisterer i punktet  $\mathbf{a} \in \mathbb{R}^n$ . Da kallas

$$\nabla f(\mathbf{a}) = \left( \frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right)$$

gradienten til  $f$  i punktet  $\mathbf{a}$ .

Det er ofte lurt å tenke på gradienten  $\nabla f(\mathbf{a})$  som en vektor som starter i punktet  $\mathbf{a}$  slik som vist på figuren nedenfor. Som vi snart skal se (setning 2.4.7), får gradienten da en geometrisk betydning — den peker i den retningen hvor funksjonen vokser raskest, og lengden  $|\nabla f(\mathbf{a})|$  er lik stigningstallet i denne retningen.



Figur 2

**Eksempel 3:** Finn gradienten til

$$f(x, y, z) = x^2ye^{xz}$$

i punktet  $\mathbf{a} = (1, -2, 0)$ .

Vi må først finne de partiellderiverte. Deriverer vi mhp.  $x$  som om  $y$  og  $z$  er konstanter, får vi

$$\frac{\partial f}{\partial x} = 2xye^{xz} + x^2ye^{xz}z = xye^{xz}(2 + xz)$$

Deriverer vi mhp.  $y$  som om  $x$  og  $z$  er konstanter, får vi tilsvarende

$$\frac{\partial f}{\partial y} = x^2e^{xz}$$

Til slutt deriverer vi mhp.  $z$  som om  $x$  og  $y$  er konstanter:

$$\frac{\partial f}{\partial z} = x^2ye^{xz}x = x^3ye^{xz}$$

Gradienten i et generelt punkt er dermed

$$\nabla f(x, y, z) = (xye^{xz}(2 + xz), x^2e^{xz}, x^3ye^{xz})$$

I vårt punkt  $\mathbf{a} = (1, -2, 0)$  får vi

$$\begin{aligned} \nabla f(1, -2, 0) &= (1 \cdot (-2) \cdot e^{1 \cdot 0} \cdot (2 + 1 \cdot 0), 1^2 \cdot e^{1 \cdot 0}, 1^3 \cdot (-2) \cdot e^{1 \cdot 0}) = \\ &= (-4, 1, -2) \end{aligned}$$



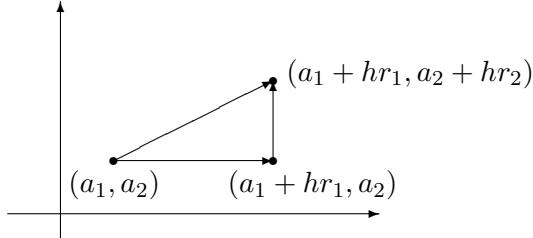
Vi har nå sett hvordan vi kan bruke våre vanlige derivasjonsregler til å regne ut partiellderiverte — vi bare deriverer som om de andre variablene var konstanter. Neste post på programmet er å vise hvordan vi kan regne ut retningsderiverte ved hjelp av partiellderiverte.

La oss begynne med å se på hva som skjer i to dimensjoner. Anta at vi ønsker å deritere  $f$  i punktet  $\mathbf{a} = (a_1, a_2)$  i retningen  $\mathbf{r} = (r_1, r_2)$ . Dersom  $h$  er en liten størrelse, vet vi at

$$f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2) = f'(\mathbf{a}; \mathbf{r}) \cdot h \quad (2.4.1)$$

når vi ser bort fra en størrelse som er liten sammenlignet med  $h$ .

Vi kan også beregne differansen  $f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2)$  på en annen måte. Istedentfor å gå direkte langs vektoren  $\mathbf{r}$  velger vi å gå parallelt med koordinataksene som vist på figur 2.



Figur 3

Vi ser at

$$\begin{aligned} f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2) &= \\ &= \overbrace{f(a_1 + hr_1, a_2 + hr_2) - f(a_1 + hr_1, a_2)}^{\text{økning parallelt med } y\text{-aksen}} + \\ &\quad + \overbrace{f(a_1 + hr_1, a_2) - f(a_1, a_2)}^{\text{økning parallelt med } x\text{-aksen}} \end{aligned}$$

Ser vi bort fra feil som er små sammenlignet med  $h$ , får vi videre

$$\begin{aligned} f(a_1 + hr_1, a_2) - f(a_1, a_2) &= \frac{\partial f}{\partial x}(\mathbf{a}) \cdot hr_1 \\ f(a_1 + hr_1, a_2 + hr_2) - f(a_1 + hr_1, a_2) &= \frac{\partial f}{\partial y}(\mathbf{a}) \cdot hr_2 \end{aligned}$$

Dette betyr at

$$f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2) = \frac{\partial f}{\partial x}(\mathbf{a})hr_1 + \frac{\partial f}{\partial y}(\mathbf{a})hr_2 \quad (2.4.2)$$

(fortsatt med en feil som er liten sammenlignet med  $h$ ). Sammenligner vi de to uttrykkene (2.4.1) og (2.4.2) vi nå har for  $f(a_1 + hr_1, a_2 + hr_2) - f(a_1, a_2)$ , ser vi at

$$f'(\mathbf{a}; \mathbf{r}) = \frac{\partial f}{\partial x}(\mathbf{a})r_1 + \frac{\partial f}{\partial y}(\mathbf{a})r_2$$

Gjennomfører vi et tilsvarende argument i  $n$  variable, får vi formelen

$$f'(\mathbf{a}; \mathbf{r}) = \frac{\partial f}{\partial x_1}(\mathbf{a})r_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})r_2 + \dots + \frac{\partial f}{\partial x_n}(\mathbf{a})r_n$$

der  $\mathbf{r} = (r_1, r_2, \dots, r_n)$ . Husker vi at gradienten til  $f$  er gitt ved

$$\nabla f(\mathbf{a}) = \left( \frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right),$$

kan vi skrive formelen ovenfor som et skalarprodukt

$$f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r} \quad (2.4.3)$$

Denne formelen gir oss en effektiv måte å regne ut retningsderiverte på; vi finner alle de partiellderiverte, setter dem sammen til en gradient, og tar skalarproduktet mellom gradienten og vektoren  $\mathbf{r}$ . Beregningen som ledet oss frem til (2.4.3), er imidlertid ingen streng utledning i matematisk forstand (vi har skrevet eksakt likhet = en rekke steder der vi bare hadde omtrentlig likhet), og det viser seg at det finnes funksjoner  $f$  som ikke oppfyller (2.4.3) til tross for at både den retningsderiverte og de partiellderiverte eksisterer (et eksempel på dette er funksjonen i eksempel 2 i seksjon 2.2, se oppgave 7). Disse funksjonene oppfører seg imidlertid så merkelig at vi ønsker å utelukke dem fra teorien vår. Vi skal derfor innføre et begrep *deriverbar funksjon* som fanger opp funksjoner med den oppførselen vi ønsker oss. Utgangspunktet for definisjonen er at vi ønsker at  $\nabla f(\mathbf{a}) \cdot \mathbf{r}$  skal være en god tilnærming til funksjonsdifferansen  $f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a})$  når  $\mathbf{r}$  er liten. Mer presist ønsker vi at “feilreddet”

$$\sigma(\mathbf{r}) = f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{r}$$

skal bli mindre og mindre sammenlignet med størrelsen til  $\mathbf{r}$ , dvs. at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{\sigma(\mathbf{r})}{|\mathbf{r}|} = 0$$

Vi får altså denne definisjonen:

**Definisjon 2.4.4** *Anta at  $f: A \rightarrow \mathbb{R}$  er definert på en delmengde  $A$  av  $\mathbb{R}^n$  og at  $\mathbf{a}$  er et indre punkt i  $A$ . Anta videre at alle de partiellderiverte til  $f$  eksisterer i punktet  $\mathbf{a}$ . Vi sier at  $f$  er deriverbar i  $\mathbf{a}$  dersom funksjonen*

$$\sigma(\mathbf{r}) = f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{r}$$

*går mot 0 hurtigere enn  $|\mathbf{r}|$ , dvs.*

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{\sigma(\mathbf{r})}{|\mathbf{r}|} = 0$$

**Kommentar:** Man kan lure på hvorfor vi ikke rett å slett sier at  $f$  er deriverbar i  $\mathbf{a}$  dersom  $f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}$  holder for alle  $\mathbf{r}$ . Det viser seg imidlertid at definisjonen ovenfor gir en glattere teori hvor delene passer bedre sammen. Det neste resultatet sier dessuten at vi får likheten  $f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}$  uansett.

**Setning 2.4.5** *Anta at  $f: A \rightarrow \mathbb{R}$  er deriverbar i  $\mathbf{a}$ . Da er  $f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}$  for alle  $\mathbf{r}$ .*

*Bevis:* Vi har

$$\begin{aligned} f(\mathbf{a}; \mathbf{r}) &= \lim_{h \rightarrow 0} \frac{f(\mathbf{a} + h\mathbf{r}) - f(\mathbf{a})}{h} = \lim_{h \rightarrow 0} \frac{\nabla f(\mathbf{a}) \cdot (h\mathbf{r}) + \sigma(h\mathbf{r})}{h} = \\ &= \lim_{h \rightarrow 0} \left( \nabla f(\mathbf{a}) \cdot \mathbf{r} + |\mathbf{r}| \frac{\sigma(h\mathbf{r})}{h|\mathbf{r}|} \right) = \nabla f(\mathbf{a}) \cdot \mathbf{r} \end{aligned}$$

siden  $\lim_{h \rightarrow 0} \frac{\sigma(h\mathbf{r})}{h|\mathbf{r}|} = 0$ .  $\square$

Vi tar med et eksempel på hvordan formel (2.4.3) kan brukes til å beregne retningsderiverte.

**Eksempel 4:** La oss anta at funksjonen

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

er deriverbar. Vi skal finne den retningsderiverte  $f'(\mathbf{a}; \mathbf{r})$  der  $\mathbf{a} = (1, 1, 1)$  og  $\mathbf{r} = (1, -1, 1)$ . La oss først finne gradienten til  $f$ :

$$\frac{\partial f}{\partial x} = 2xy; \quad \frac{\partial f}{\partial y} = x^2 - ze^{-yz}; \quad \frac{\partial f}{\partial z} = -ye^{-yz}$$

Dette gir  $\nabla f = (2xy, x^2 - ze^{-yz}, -ye^{-yz})$  og  $\nabla f(\mathbf{a}) = \nabla f(1, 1, 1) = (2, 1 - e^{-1}, -e^{-1})$ . Følgelig er

$$f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r} = (2, 1 - e^{-1}, -e^{-1}) \cdot (1, -1, 1) = 1$$



For å kunne bruke formel (2.4.3) trenger vi å vite at funksjonene våre er deriverbare. Det neste resultatet gir oss den informasjonen vi vanligvis trenger. Beiset er ganske langt og komplisert, og egner seg nok best for de ivrigste og flittigste. Et lite ord om terminologi: Vi sier at en funksjon er definert i en *omogn* om  $\mathbf{a}$  dersom det finnes en kule  $B(\mathbf{a}, \epsilon)$  om  $\mathbf{a}$  der funksjonen er definert (den kan godt være definert på et større område — poenget er at vi i hvert fall vil sikre oss at den er definert for alle punkter tilstrekkelig nær  $\mathbf{a}$ ).

**Teorem 2.4.6** *La  $f : A \rightarrow \mathbb{R}^m$  være en funksjon av  $n$  variable. Anta at alle de partiellederiverte  $\frac{\partial f}{\partial x_i}$  er definert i en omogn om  $\mathbf{a} \in A$ , og at de er kontinuerlige i  $\mathbf{a}$ . Da er  $f$  deriverbar i  $\mathbf{a}$ .*

*Bevis:* For at ikke notasjonen skal bli for overveldende, skal vi nøye oss med å bevise setningen for en funksjon  $f(x_1, x_2)$  av to variable. Beiset er en litt oppstrammet variant av det argumentet som ledet oss til formel (2.4.3).

Vi velger  $\mathbf{r}$  så liten at  $\mathbf{a} + \mathbf{r}$  ligger innenfor denkulen der vi vet at de partielle deriverte eksisterer. Vi lar  $\mathbf{a} = (a_1, a_2)$ ,  $\mathbf{r} = (r_1, r_2)$  og observerer at

$$\begin{aligned} f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a}) &= f(a_1 + r_1, a_2 + r_2) - f(a_1, a_2) \\ &= f(a_1 + r_1, a_2 + r_2) - f(a_1 + r_1, a_2) + \\ &\quad + f(a_1 + r_1, a_2) - f(a_1, a_2) \end{aligned}$$

Dette er samme type omskrivning som vi foretok da vi regnet oss frem til (2.4.3). Hvis vi tenker på  $x \rightarrow f(x, a_2)$  som en funksjon av én variabel, forteller middelverdisetningen (se *Kalkulus*, seksjon 6.2) oss at det finnes et punkt  $c$  mellom  $a_1$  og  $a_1 + r_1$  slik at

$$f(a_1 + r_1, a_2) - f(a_1, a_2) = \frac{\partial f}{\partial x_1}(c, a_2) r_1$$

Helt tilsvarende kan vi finne et punkt  $d$  mellom  $a_2$  og  $a_2 + r_2$  slik at

$$f(a_1 + r_1, a_2 + r_2) - f(a_1 + r_1, a_2) = \frac{\partial f}{\partial x_2}(a_1 + r_1, d) r_2$$

Kombinerer vi de resultatene vi nå har, ser vi at

$$f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a}) = \frac{\partial f}{\partial x_1}(c, a_2) r_1 + \frac{\partial f}{\partial x_2}(a_1 + r_1, d) r_2$$

Trekker vi fra  $\nabla f(\mathbf{a}) \cdot \mathbf{r} = \frac{\partial f}{\partial x_1}(a_1, a_2) r_1 + \frac{\partial f}{\partial x_2}(a_1, a_2) r_2$  på begge sider, får vi

$$\begin{aligned} f(\mathbf{a} + \mathbf{r}) - f(\mathbf{a}) - \nabla f(\mathbf{a}) \cdot \mathbf{r} &= \\ &= \frac{\partial f}{\partial x_1}(c, a_2) r_1 + \frac{\partial f}{\partial x_2}(a_1 + r_1, d) r_2 - \frac{\partial f}{\partial x_1}(a_1, a_2) r_1 - \frac{\partial f}{\partial x_2}(a_1, a_2) r_2 = \\ &= \left( \frac{\partial f}{\partial x_1}(c, a_2) - \frac{\partial f}{\partial x_1}(a_1, a_2) \right) r_1 + \left( \frac{\partial f}{\partial x_2}(a_1 + r_1, d) - \frac{\partial f}{\partial x_2}(a_1, a_2) \right) r_2 \end{aligned}$$

Sammenligner vi dette med definisjon 2.4.4, ser vi at

$$\sigma(\mathbf{r}) = \left( \frac{\partial f}{\partial x_1}(c, a_2) - \frac{\partial f}{\partial x_1}(a_1, a_2) \right) r_1 + \left( \frac{\partial f}{\partial x_2}(a_1 + r_1, d) - \frac{\partial f}{\partial x_2}(a_1, a_2) \right) r_2$$

Vår oppgave er å vise at  $\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \sigma(\mathbf{r}) = \mathbf{0}$ . Siden  $|r_1|, |r_2| \leq |\mathbf{r}|$ , får vi

$$|\sigma(\mathbf{r})| \leq \left( \left| \frac{\partial f}{\partial x_1}(c, a_2) - \frac{\partial f}{\partial x_1}(a_1, a_2) \right| + \left| \frac{\partial f}{\partial x_2}(a_1 + r_1, d) - \frac{\partial f}{\partial x_2}(a_1, a_2) \right| \right) |\mathbf{r}|$$

Deler vi på på  $|\mathbf{r}|$ , ser vi at

$$\frac{|\sigma(\mathbf{r})|}{|\mathbf{r}|} \leq \left( \left| \frac{\partial f}{\partial x_1}(c, a_2) - \frac{\partial f}{\partial x_1}(a_1, a_2) \right| + \left| \frac{\partial f}{\partial x_2}(a_1 + r_1, d) - \frac{\partial f}{\partial x_2}(a_1, a_2) \right| \right)$$

Dette uttrykket går mot 0 fordi de partiellderiverte er kontinuerlige i  $\mathbf{a} = (a_1, a_2)$ , og  $(c, a_2)$  og  $(a + r_1, d)$  nærmer seg  $(a_1, a_2)$  når  $\mathbf{r} \rightarrow \mathbf{0}$ . Dermed er teoremet bevist.  $\square$

Sammen med setning 2.4.5 forteller setningen ovenfor oss at så lenge de partiellderiverte er kontinuerlige, kan vi trygt bruke formelen

$$f'(\mathbf{a}; \mathbf{r}) = \nabla f(\mathbf{a}) \cdot \mathbf{r}.$$

Vi har allerede nevnt den geometriske tolkningen av gradienten — at gradienten i punktet  $\mathbf{a}$  peker i den retningen hvor funksjonen vokser hurtigst, og at stigningstallet i denne retningen er lik lengden til gradienten. Vi har nå de redskapene som trengs til å bevise dette.

**Setning 2.4.7** *Anta at  $f$  er deriverbar i  $\mathbf{a}$ . Da peker gradienten  $\nabla f(\mathbf{a})$  i den retningen hvor  $f$  vokser hurtigst i punktet  $\mathbf{a}$ , og stigningstallet til  $f$  i denne retningen er  $|\nabla f(\mathbf{a})|$ .*

*Bevis:* Hvis  $\mathbf{u}$  er en enhetsvektor (dvs.  $|\mathbf{u}| = 1$ ), så forteller den retningsderiverte  $f'(\mathbf{a}; \mathbf{u}) = \nabla f(\mathbf{a}) \cdot \mathbf{u}$  oss hvor fort funksjonen vokser i den retningen  $\mathbf{u}$  peker. Funksjonen  $f$  vokser derfor hurtigst i den retningen  $\mathbf{u}$  hvor  $\nabla f(\mathbf{a}) \cdot \mathbf{u}$  er størst. Siden  $\mathbf{u}$  er en enhetsvektor, forteller Schwarz' ulikhet (setning 1.2.3) oss at

$$|\nabla f(\mathbf{a}) \cdot \mathbf{u}| \leq |\nabla f(\mathbf{a})| |\mathbf{u}| = |\nabla f(\mathbf{a})|$$

med likhet bare hvis  $\mathbf{u}$  og  $\nabla f(\mathbf{a})$  er parallelle. Dette betyr at  $|\nabla f(\mathbf{a}) \cdot \mathbf{u}|$  er størst når  $\mathbf{u}$  og  $\nabla f(\mathbf{a})$  er parallelle. Nå er det to enhetsvektorer som er parallelle med  $\nabla f(\mathbf{a})$  — en som peker samme vei som  $\nabla f(\mathbf{a})$ , og en som peker motsatt vei. Det er lett å se at  $\nabla f(\mathbf{a}) \cdot \mathbf{u} = |\nabla f(\mathbf{a})|$  når  $\mathbf{u}$  og  $\nabla f(\mathbf{a})$  peker samme vei, og at  $\nabla f(\mathbf{a}) \cdot \mathbf{u} = -|\nabla f(\mathbf{a})|$  når  $\mathbf{u}$  og  $\nabla f(\mathbf{a})$  peker motsatt vei. Altså har  $f$  sitt største stigningstall  $|\nabla f(\mathbf{a})|$  i den retningen som  $\nabla f(\mathbf{a})$  peker.  $\square$

**Eksempel 5:** I hvilken retning vokser funksjonen  $f(x, y) = x^3y \sin(\pi xy)$  hurtigst når vi står i punktet  $(1, \frac{1}{2})$ ?

Vi må først finne gradienten. De partiellderiverte er

$$\frac{\partial f}{\partial x} = 3x^2y \sin(\pi xy) + x^3y \cos(\pi xy)(\pi y) = 3x^2y \sin(\pi xy) + \pi x^3y^2 \cos(\pi xy)$$

$$\frac{\partial f}{\partial y} = x^3 \sin(\pi xy) + x^3y \cos(\pi xy)(\pi x) = x^3 \sin(\pi xy) + \pi x^4y \cos(\pi xy)$$

Setter vi inn  $(x, y) = (1, \frac{1}{2})$ , får vi

$$\frac{\partial f}{\partial x} \left(1, \frac{1}{2}\right) = 3 \cdot 1^2 \cdot \frac{1}{2} \cdot \sin(\pi \cdot 1 \cdot \frac{1}{2}) + \pi \cdot 1^3 \cdot \left(\frac{1}{2}\right)^2 \cdot \cos(\pi \cdot 1 \cdot \frac{1}{2}) = \frac{3}{2}$$

og

$$\frac{\partial f}{\partial y} \left(1, \frac{1}{2}\right) = 1^3 \cdot \sin(\pi \cdot 1 \cdot \frac{1}{2}) + \pi \cdot 1^4 \cdot \frac{1}{2} \cdot \cos(\pi \cdot 1 \cdot \frac{1}{2}) = 1$$

Dermed er

$$\nabla f \left(1, \frac{1}{2}\right) = \left(\frac{3}{2}, 1\right)$$

Funksjonen vokser altså raskest i retningen  $(\frac{3}{2}, 1)$ , og stigningstallet i denne retningen er

$$|\nabla f \left(1, \frac{1}{2}\right)| = \sqrt{\left(\frac{3}{2}\right)^2 + 1^2} = \sqrt{\frac{9}{4} + 1} = \frac{1}{2}\sqrt{13}$$

♣

Vi tar med et litt mer teoretisk resultat som vi får bruk for siden. For funksjoner av én variabel vet vi at dersom den deriverter eksisterer i et punkt, så er funksjonen kontinuerlig i punktet. For funksjoner av flere variable er sammenhengen litt mer subtil — det kan faktisk hende at alle de retningsderiverte eksisterer i et punkt, men at funksjonen likevel ikke er kontinuerlig i punktet (se oppgave 7). Et funksjonen deriverbar i betydningen vi innførte i definisjon 2.4.4, er vi imidlertid på den sikre siden.

**Setning 2.4.8** *Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable. Dersom  $f$  er deriverbar i et punkt  $\mathbf{a} \in A$ , så er  $f$  kontinuerlig i  $\mathbf{a}$ .*

*Bevis:* Ifølge setning 2.3.5 er det nok å vise at  $\lim_{\mathbf{x} \rightarrow \mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$ . Setter vi  $\mathbf{r} = \mathbf{x} - \mathbf{a}$ , er dette det samme som å vise at  $\lim_{\mathbf{r} \rightarrow \mathbf{0}} f(\mathbf{a} + \mathbf{r}) = f(\mathbf{a})$ . Siden  $f$  er deriverbar i  $\mathbf{a}$ , vet vi at

$$f(\mathbf{a} + \mathbf{r}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{r} + \sigma(\mathbf{r})$$

der  $\frac{\sigma(\mathbf{r})}{|\mathbf{r}|} \rightarrow 0$  når  $\mathbf{r} \rightarrow \mathbf{0}$ . Dette betyr spesielt at  $\sigma(\mathbf{r}) \rightarrow 0$ , og dermed er

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} f(\mathbf{a} + \mathbf{r}) = \lim_{\mathbf{r} \rightarrow \mathbf{0}} (f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{r} + \sigma(\mathbf{r})) = f(\mathbf{a}) + 0 + 0 = f(\mathbf{a})$$

□

Helt til slutt skal vi se på eksempel som viser en typisk anvendelse av grader og partiellderiverte.

**Eksempel 6:** Anta at en tynn gass oppbevares i en beholder der vi kan justere volumet  $V$  og temperaturen  $T$ . Det viser seg da at trykket  $P$  er proporsjonalt med temperaturen  $T$  og omvendt proporsjonalt med volumet  $V$ , dvs.  $P = k \frac{T}{V}$  der  $k$  er en konstant som blant annet avhenger av hvor mye

gass det er i beholderen (dette forutsetter egentlig at temperaturen måles i grader Kelvin og ikke Celsius, men det behøver vi ikke bry oss om her). Vi kan tenke på trykket  $P$  som en funksjon av  $T$  og  $V$ :

$$P(T, V) = k \frac{T}{V}$$

Partielllderiverer vi dette uttrykket, får vi

$$\frac{\partial P}{\partial T} = k \frac{1}{V} \quad \text{og} \quad \frac{\partial P}{\partial V} = -k \frac{T}{V^2}$$

Gradienten er dermed

$$\nabla P(T, V) = \left( k \frac{1}{V}, -k \frac{T}{V^2} \right)$$

Det er ofte naturlig å spørre hvor mye trykket endrer seg dersom vi gir temperaturen et lite tillegg  $\Delta T$  og volumet et lite tillegg  $\Delta V$ . Det eksakte uttrykket for dette tillegget er selvfølgelig

$$\Delta P = P(T + \Delta T, V + \Delta V) - P(T, V),$$

men denne differansen er tung å arbeide med. Vi vet imidlertid at den har en god tilnærming i skalarproduktet av gradienten  $\nabla P$  med tilvekstvektoren  $(\Delta T, \Delta V)$ :

$$\begin{aligned} \Delta P &= P(T + \Delta T, V + \Delta V) - P(T, V) \approx \nabla P(T, V) \cdot (\Delta T, \Delta V) = \\ &= \frac{\partial P}{\partial T} \Delta T + \frac{\partial P}{\partial V} \Delta V = k \frac{1}{V} \Delta T - k \frac{T}{V^2} \Delta V \end{aligned}$$

Ved hjelp av denne formelen er det lett å anslå hvor mye trykket endrer seg når vi regulerer volumet og temperaturen.

La oss nå anta at vi regulerer temperaturen og volumet kontinuerlig, og at endringene  $\Delta P$ ,  $\Delta T$  og  $\Delta V$  har foregått i løpet av et lite tidsintervall  $\Delta t$ , Endringen per tidsenhet er da

$$\frac{\Delta P}{\Delta t} \approx k \frac{1}{V} \frac{\Delta T}{\Delta t} - k \frac{T}{V^2} \frac{\Delta V}{\Delta t}$$

med bedre tilnærming dess mindre  $\Delta t$  er. Lar vi  $\Delta t \rightarrow 0$ , får vi

$$P'(t) = k \frac{1}{V} T'(t) - k \frac{T}{V^2} V'(t)$$

Ved hjelp av denne formelen kan vi regne ut hvor fort trykket endrer seg dersom vi kjenner endringshastighetene til temperaturen og volumet. Formelen er en forsmak på *kjerneregelen for funksjoner av flere variable*. Vi skal se nærmere på denne regelen i seksjon 2.7. ♣

### Oppgaver til seksjon 2.4

**1.** Finn de partielle deriverte til  $f$ .

a) $f(x, y) = x^3y + 3xy^4$	e) $f(x, y, z) = (x + y)e^{-z}$
b) $f(x, y) = \frac{x^2+x^3}{y}$	f) $f(x, y, z) = \frac{z^2 \tan x}{1+y^2}$
c) $f(x, y) = \cos(x + y^2)$	g) $f(x, y, z) = z \arctan(x + y)$
d) $f(x, y) = x^2 \ln(xy^2)$	h) $f(x, y, z, u) = (z^2 + u)e^{-x+3y}$

**2.** Finn gradienten til funksjonen:

- a)  $f(x, y) = x^2y$
- b)  $f(x, y, z) = x \cos(xy^2z)$
- c)  $f(u, v, w) = we^{u \cos v}$
- d)  $f(z_1, z_2, z_3) = z_3 \arctan(z_1 z_2) + e^{z_3}$

**3.** Finn den retningsderiverte  $f'(\mathbf{a}; \mathbf{r})$ :

a) $f(x, y) = 3xy + y^2$ ; $\mathbf{a} = (1, 2)$ ; $\mathbf{r} = (3, -1)$
b) $f(x, y) = \ln(x + y^2)$ ; $\mathbf{a} = (1, 0)$ ; $\mathbf{r} = (-1, 1)$
c) $f(x, y, z) = x^2y + z^2$ ; $\mathbf{a} = (1, 0, 1)$ ; $\mathbf{r} = (1, 1, -1)$
d) $f(x, y, z) = z \sin(xy)$ ; $\mathbf{a} = (\frac{\pi}{2}, 1, 0)$ ; $\mathbf{r} = (2, 0, -1)$

**4.** I hvilken retning vokser funksjonen hurtigst i det angitte punktet

- a)  $f(x, y) = -x^2y + 7y^3$ ;  $\mathbf{a} = (4, -3)$
- b)  $f(x, y, z) = (x^2 - y^2)e^z$ ;  $\mathbf{a} = (1, -1, 3)$
- c)  $f(x, y, z, u) = xuz^2 - y^2zu$ ;  $\mathbf{a} = (1, 0, -2, 3)$

**5.** Volumet til en sylinder med radius  $r$  og høyde  $h$  er  $V = \pi r^2 h$ . Når høyden og radien varierer, kan vi tenke på dette som en funksjon i to variable  $V(r, h) = \pi r^2 h$ . Forklar at når radien endrer seg fra  $r$  til  $r + \Delta r$  og høyden endrer seg fra  $h$  til  $h + \Delta h$ , så er endringen i  $V$  tilnærmet gitt ved

$$\Delta V \approx \frac{\partial V}{\partial r} \cdot \Delta r + \frac{\partial V}{\partial h} \cdot \Delta h = 2\pi rh \Delta r + \pi r^2 \Delta h.$$

Anta at du har en sylinder hvor du vet at radien ligger mellom 2 m og 2.05 m og hvor høyden ligger mellom 5 m og 5.05 m. Bruk formelen ovenfor til å anslå usikkerheten i volumet.

**6.** BMI (body mass index) er en indikator for undervekt og overvekt. For å finne din BMI tar du vekten din (målt i kilo) og deler på kvadratet av høyden din (målt i meter). Du kan tenke på BMI som en funksjon av to variable

$$f(v, h) = \frac{v}{h^2}$$

a) Vis at dersom  $\Delta v$  og  $\Delta h$  er små endringer i vekt og høyde, så er endringen i BMI gitt ved

$$\Delta f(v, h) \approx \frac{\Delta v}{h^2} - 2 \frac{v}{h^3} \Delta h$$

b) En tommelfingerregel sier at for hver ekstra centimeter du har i høyde, kan du “tåle” en ekstra kilo uten at BMI-en din endrer seg særlig. Bruk formelen i a) til å undersøke hvor godt dette passer for personer med forskjellig vekt og høyde.

**7.** Vi skal se mer på funksjonen  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  fra eksempel 2 i seksjon 2.2. Husk at denne funksjonen er gitt ved

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{for } (x, y) \neq (0, 0) \\ 0 & \text{for } (x, y) = (0, 0) \end{cases}$$

- a) Vis at  $\frac{\partial f}{\partial x}(0, 0) = 0$  og  $\frac{\partial f}{\partial y}(0, 0) = 0$ . Hva er  $\nabla f(0, 0)$ ?
- b) Bruk definisjonen av retningsderivert til å vise at  $f(\mathbf{0}; \mathbf{r}) = \frac{r_1^2}{r_2}$  der  $\mathbf{r} = (r_1, r_2)$ ,  $r_2 \neq 0$ .
- c) Vis at for denne funksjonen gjelder ikke likheten  $f(\mathbf{0}; \mathbf{r}) = \nabla f(\mathbf{0}) \cdot \mathbf{r}$ .
- d) Vis at alle de retningsderiverte til  $f$  eksisterer i  $\mathbf{0}$ , men at funksjonen hverken er kontinuerlig eller deriverbar i punktet.

## 2.5 Partiellderiverte av høyere orden

Fra teorien for funksjoner av én variabel vet vi at det ofte er nyttig eller nødvendig å derivere mer enn én gang. Også i flervariabel teori er det ofte nyttig å arbeide med annenderiverte, tredjederiverte osv. Den store forskjellen er at vi har så mange måter å derivere på.

**Eksempel 1:** La  $f(x, y) = x^2 y^3 + y^2$ . Vi har to partiellderiverte av første orden

$$\frac{\partial f}{\partial x} = 2xy^3 \quad \text{og} \quad \frac{\partial f}{\partial y} = 3x^2y^2 + 2y$$

Når vi skal regne ut annenderiverte, har vi mange valg. Vi kan for eksempel derivere  $\frac{\partial f}{\partial x}$  med hensyn på  $x$  en gang til:

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right) = \frac{\partial}{\partial x} (2xy^3) = 2y^3.$$

Vi kan også derivere  $\frac{\partial f}{\partial x}$  med hensyn på  $y$ :

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right) = \frac{\partial}{\partial y} (2xy^3) = 6xy^2.$$

I tillegg kan vi derivere  $\frac{\partial f}{\partial y}$  med hensyn på både  $x$  og  $y$ :

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial y} \right) = \frac{\partial}{\partial x} (3x^2y^2 + 2y) = 6xy^2$$

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial y} \right) = \frac{\partial}{\partial y} (3x^2y^2 + 2y) = 6x^2y + 2$$

Vi har altså fire annenordens partiellderiverte for denne funksjonen:  $\frac{\partial^2 f}{\partial x^2}$ ,  $\frac{\partial^2 f}{\partial y \partial x}$ ,  $\frac{\partial^2 f}{\partial x \partial y}$  og  $\frac{\partial^2 f}{\partial y^2}$ . ♣

Den generelle notasjonen skulle fremgå av eksemplet ovenfor —

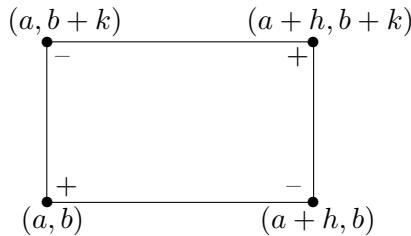
$$\frac{\partial^n f}{\partial x_{i_n} \dots \partial x_{i_2} \partial x_{i_1}}$$

er den funksjonen vi får ved å derivere funksjonen  $f$   $n$  ganger, først med hensyn på variabelen  $x_{i_1}$ , så med hensyn på variabelen  $x_{i_2}$  osv.

I eksemplet ovenfor så vi at de to “blandede” partiellderiverte  $\frac{\partial^2 f}{\partial x \partial y}$  og  $\frac{\partial^2 f}{\partial y \partial x}$  var like. Dette er ikke en universell regel; det finnes funksjoner  $f$  slik at  $\frac{\partial^2 f}{\partial x \partial y}$  og  $\frac{\partial^2 f}{\partial y \partial x}$  er forskjellige, men for de fleste vi støter på i praksis, vil de blandede partiellderiverte være like. Den neste setningen viser at dette gjelder dersom de annenordens partiellderiverte eksisterer i en omegn rundt punktet  $\mathbf{a}$  og er kontinuerlige i  $\mathbf{a}$ . Beviset er krevende og minner om beviset for teorem 2.4.6. I oppgave 4 finner du et eksempel på en funksjon der de blandede partiellderiverte ikke er like.

**Setning 2.5.1** La  $f(x_1, \dots, x_n)$  være en funksjon av  $n$  variable. Anta at  $\frac{\partial^2 f}{\partial x_i \partial x_j}$  og  $\frac{\partial^2 f}{\partial x_j \partial x_i}$  eksisterer i en omegn om punktet  $\mathbf{a}$  og er kontinuerlige i  $\mathbf{a}$ . Da er  $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$ .

*Bevis:* For å forenkle notasjonen antar vi at  $f(x, y)$  er en funksjon av to variable, og at  $\frac{\partial^2 f}{\partial x \partial y}$  og  $\frac{\partial^2 f}{\partial y \partial x}$  eksisterer i en omegn om punktet  $(a, b)$  og er kontinuerlige i  $(a, b)$ . Anta at tallene  $h, k$  er så små at hele rektangelet i figur 1 ligger i det området der de blandede partiellderiverte eksisterer.



Figur 1

La

$$\Delta(h, k) = f(a + h, b + k) - f(a, b + k) - f(a + h, b) + f(a, b)$$

der vi har kombinert funksjonsverdiene i hjørnene på rektanglet vårt ved å bruke fortregnene vist på figuren. Vi skal vise at grenseverdien

$$\lim_{(h,k) \rightarrow (0,0)} \frac{\Delta(h, k)}{h \cdot k}$$

er lik både  $\frac{\partial^2 f}{\partial x \partial y}(a, b)$  og  $\frac{\partial^2 f}{\partial y \partial x}(a, b)$ .

Først skritt er å bruke middelverdisetningen på funksjonen

$$g(x) = f(x, b + k) - f(x, b).$$

Vi får

$$g(a + h) - g(a) = g'(c) \cdot h$$

for en  $c$  mellom  $a$  og  $a + h$ . Setter vi inn den opprinnelige funksjonen, ser vi at

$$\begin{aligned} f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b) &= \\ &= \left[ \frac{\partial f}{\partial x}(c, b + k) - \frac{\partial f}{\partial x}(c, b) \right] h \end{aligned}$$

Dette kan også skrives

$$\Delta(h, k) = \left[ \frac{\partial f}{\partial x}(c, b + k) - \frac{\partial f}{\partial x}(c, b) \right] h$$

Neste steg er å bruke middelverdisetningen på funksjonen

$$G(y) = \frac{\partial f}{\partial x}(c, y).$$

Vi får

$$G(b + k) - G(b) = G'(d) \cdot k$$

for en  $d$  mellom  $b$  og  $b + k$ . Dette kan også skrives

$$\frac{\partial f}{\partial x}(c, b + k) - \frac{\partial f}{\partial x}(c, b) = \frac{\partial^2 f}{\partial y \partial x}(c, d) \cdot k$$

Kombinerer vi formlene våre, ser vi at

$$\Delta(h, k) = \frac{\partial^2 f}{\partial y \partial x}(c, d) \cdot hk$$

Siden  $\frac{\partial^2 f}{\partial y \partial x}$  er kontinuerlig i  $(a, b)$ , vil  $\frac{\partial^2 f}{\partial y \partial x}(c, d) \rightarrow \frac{\partial^2 f}{\partial y \partial x}(a, b)$  når  $(h, k) \rightarrow 0$ .  
Følgelig er

$$\lim_{(h,k) \rightarrow (0,0)} \frac{\Delta(h, k)}{hk} = \frac{\partial^2 f}{\partial y \partial x}(a, b)$$

For å vise at også  $\lim_{(h,k) \rightarrow (0,0)} \frac{\Delta(h, k)}{hk} = \frac{\partial^2 f}{\partial x \partial y}(a, b)$ , bytter vi om på rollene til variablene  $x$  og  $y$  i argumentet ovenfor. Vi starter med å bruke middelverdisetningen på funksjonen

$$\gamma(y) = f(a + h, y) - f(a, y)$$

og fortsetter på akkurat samme måte som ovenfor. Detaljene overlates til leserne.  $\square$

At blandede partiellderiverte av annen orden er like, medfører også at blandede partiellderiverte av høyere orden er like dersom de inneholder like mange derivasjoner med hensyn på hver variabel. Dersom  $f$  har kontinuerlige fjerdederiverte, kan vi for eksempel vise at

$$\frac{\partial^4 f}{\partial x \partial y \partial z \partial x} = \frac{\partial^4 f}{\partial z \partial y \partial x \partial x}$$

på følgende måte:

$$\frac{\partial^4 f}{\partial x \partial y \partial z \partial x} = \frac{\partial^4 f}{\partial x \partial z \partial y \partial x} = \frac{\partial^4 f}{\partial z \partial x \partial y \partial x} = \frac{\partial^4 f}{\partial z \partial y \partial x \partial x}$$

Overbevis deg selv om at du kan begrunne disse overgangene.

## Oppgaver til seksjon 2.5

**1.** Regn ut de annenordens partiellderiverte til funksjonene:

- a)  $f(x, y) = 3x^2y + 2y^2x$
- b)  $f(x, y) = x \sin y$
- c)  $f(x, y) = x^2e^{x-y}$
- d)  $f(x, y, z) = x^2z - y^2z^2$

**2.** Regn ut de partiellderiverte:

- a)  $\frac{\partial^3 f}{\partial x \partial z \partial x}$  når  $f(x, y, z) = x^2ye^{xz}$ .
- b)  $\frac{\partial^4 f}{\partial y \partial z \partial x \partial z}$  når  $f(x, y, z) = x^2y^3 \cos xyz$ .

**3.** Gjennomfør den siste delen av beviset for setning 2.5.1 (dvs. at  $\lim_{(h,k) \rightarrow (0,0)} \frac{\Delta(h,k)}{hk} = \frac{\partial^2 f}{\partial x \partial y}(a, b)$ ).

**4.** I denne oppgaven skal vi se på en funksjon  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  slik at  $\frac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \frac{\partial^2 f}{\partial y \partial x}(0, 0)$ . Funksjonen er gitt ved

$$f(x, y) = \begin{cases} \frac{x^3y - xy^3}{x^2 + y^2} & \text{når } (x, y) \neq (0, 0) \\ 0 & \text{når } (x, y) = (0, 0) \end{cases}$$

a) Vis at  $f(x, 0) = 0$  for alle  $x$  og at  $f(0, y) = 0$  for alle  $y$ . Bruk dette til å vise at  $\frac{\partial f}{\partial x}(0, 0) = 0$  og  $\frac{\partial f}{\partial y}(0, 0) = 0$ .

b) Vis at for  $(x, y) \neq (0, 0)$  er

$$\frac{\partial f}{\partial x}(x, y) = \frac{y(x^4 + 4x^2y^2 - y^4)}{(x^2 + y^2)^2}$$

$$\frac{\partial f}{\partial y}(x, y) = -\frac{x(y^4 + 4x^2y^2 - x^4)}{(x^2 + y^2)^2}$$

c) Vis at  $\frac{\partial^2 f}{\partial y \partial x}(0, 0) = -1$  ved å bruke

$$\frac{\partial^2 f}{\partial y \partial x}(0, 0) = \lim_{h \rightarrow 0} \frac{\frac{\partial f}{\partial x}(0, h) - \frac{\partial f}{\partial x}(0, 0)}{h}$$

Vis på tilsvarende måte at  $\frac{\partial^2 f}{\partial x \partial y}(0, 0) = 1$ .

## 2.6 Derivasjon av vektorvaluerte funksjoner

Hittil har vi sett på derivasjon av funksjoner som tar verdier i  $\mathbb{R}$ . Vi skal nå utvide teorien vår til også å omfatte funksjoner som tar verdier i  $\mathbb{R}^m$  der  $m$  er større enn 1, altså til *vektorvaluerte* funksjoner. Husk at en funksjon  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  kan skrives på komponentform

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_m(\mathbf{x}) \end{pmatrix}$$

der  $F_1, F_2, \dots, F_m$  er funksjoner med verdier i  $\mathbb{R}$  (i denne seksjonen er det lurt å tenke på de fleste vektorer som søylevektorer siden vi skal multiplisere dem med matriser). Dersom alle disse funksjonen lar seg derivere, kan vi samle alle de partiellderiverte i en stor matrise, den såkalte *Jacobi-matrisen* til  $\mathbf{F}$ :

$$\mathbf{F}'(\mathbf{a}) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{a}) & \frac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

Legg merke til hvordan denne matrisen er bygget opp: I første linje har vi alle partiellderiverte av førstekomponent  $F_1$ , i annen linje alle partiellderiverte av annenkomponent  $F_2$  osv. Sagt på en annen måte: Første linje i Jacobi-matrisen er gradienten til  $F_1$ , andre linje er gradienten til  $F_2$  osv. Dersom  $\mathbf{F}$  er et skalarfelt (og altså tar verdier i  $\mathbb{R}$ ), har Jacobi-matrisen bare én linje og er identisk med gradienten  $\nabla \mathbf{F}$ .

Jacobi-matrisen spiller på mange måter den samme rollen for funksjoner av flere variable som den deriverte gjør for funksjoner av én variabel, og vi har derfor valgt en notasjon  $\mathbf{F}'(\mathbf{a})$  som ligner på den notasjonen vi er vant til.

Hvis man først har lært seg å partiellderivere, er det ingen kunst (men en del arbeid!) å finne en Jacobi-matrise.

**Eksempel 1:** Finn Jacobi-matrisen til funksjonen

$$\mathbf{F}(x, y) = \begin{pmatrix} xy^3 \\ e^{x+y^2} \\ 3x^2y \end{pmatrix}$$

I dette tilfellet er

$$F_1(x, y) = xy^3, \quad F_2(x, y) = e^{x+y^2} \quad \text{og} \quad F_3(x, y) = 3x^2y$$

Vi partiellderiverer og får

$$\begin{aligned} \frac{\partial F_1}{\partial x} &= y^3 & \text{og} & \quad \frac{\partial F_1}{\partial y} = 3xy^2 \\ \frac{\partial F_2}{\partial x} &= e^{x+y^2} & \text{og} & \quad \frac{\partial F_2}{\partial y} = 2ye^{x+y^2} \\ \frac{\partial F_3}{\partial x} &= 6xy & \text{og} & \quad \frac{\partial F_3}{\partial y} = 3x^2 \end{aligned}$$

Jacobi-matrisen blir dermed

$$\mathbf{F}'(x, y) = \begin{pmatrix} y^3 & 3xy^2 \\ e^{x+y^2} & 2ye^{x+y^2} \\ 6xy & 3x^2 \end{pmatrix}$$



Vi skal nå forsøke å finne ut hvordan vi bør definere deriverbarhet for vektorvaluerte funksjoner  $\mathbf{F}$ . La oss gå tilbake til den generelle Jacobi-matrisen

$$\mathbf{F}'(\mathbf{a}) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{a}) & \frac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

Ganger vi denne med en søylevektor

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

får vi

$$\mathbf{F}'(\mathbf{a})\mathbf{r} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_n}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_n}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1}(\mathbf{a}) & \frac{\partial F_m}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_m}{\partial x_n}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix} = \begin{pmatrix} \nabla F_1(\mathbf{a}) \cdot \mathbf{r} \\ \nabla F_2(\mathbf{a}) \cdot \mathbf{r} \\ \vdots \\ \nabla F_m(\mathbf{a}) \cdot \mathbf{r} \end{pmatrix}$$

der vi har brukt at første rad i  $\mathbf{F}'(\mathbf{a})$  er gradienten til  $F_1$ , andre rad er gradienten til  $F_2$  osv. Dersom  $F_1, F_2, \dots, F_n$  er deriverbare funksjoner, er

$$\nabla F_1(\mathbf{a}) \cdot \mathbf{r} \approx F_1(\mathbf{a} + \mathbf{r}) - F_1(\mathbf{a})$$

$$\nabla F_2(\mathbf{a}) \cdot \mathbf{r} \approx F_2(\mathbf{a} + \mathbf{r}) - F_2(\mathbf{a})$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$\nabla F_m(\mathbf{a}) \cdot \mathbf{r} \approx F_m(\mathbf{a} + \mathbf{r}) - F_m(\mathbf{a})$$

med bedre tilnærming dess mindre  $\mathbf{r}$  er. Altså er

$$\mathbf{F}'(\mathbf{a})\mathbf{r} \approx \begin{pmatrix} F_1(\mathbf{a} + \mathbf{r}) - F_1(\mathbf{a}) \\ F_2(\mathbf{a} + \mathbf{r}) - F_2(\mathbf{a}) \\ \vdots \\ F_m(\mathbf{a} + \mathbf{r}) - F_m(\mathbf{a}) \end{pmatrix} = \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a})$$

Dette gir oss en indikasjon på hvordan vi skal definere deriverbarhet for vektorvaluerte funksjoner — vi ønsker at avviket i tilnærmingen

$$\mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) \approx \mathbf{F}'(\mathbf{a})\mathbf{r}$$

skal være lite sammenlignet med størrelsen til  $\mathbf{r}$ :

**Definisjon 2.6.1** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av  $n$  variable og at  $\mathbf{a}$  er et indre punkt i  $A$ . Vi sier at  $\mathbf{F}$  er deriverbar i  $\mathbf{a}$  dersom funksjonen*

$$\boldsymbol{\sigma}(\mathbf{r}) = \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{r}$$

*går mot null fortære enn  $|\mathbf{r}|$ , dvs. at*

$$\lim_{\mathbf{r} \rightarrow 0} \frac{1}{|\mathbf{r}|} \boldsymbol{\sigma}(\mathbf{r}) = \mathbf{0}$$

*(legg merke til at  $\boldsymbol{\sigma}$  nå er en vektorvaluert funksjon med verdier i  $\mathbb{R}^m$ ).*

Mange vil kanskje frykte at vi nå må begynne helt forfra med å utvikle en teori for deriverbarhet akkurat som vi gjorde for skalarfelt i seksjon 2.4, men takket være den neste setningen er det unødvendig.

**Setning 2.6.2** En funksjon  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er deriverbar i et indre punkt  $\mathbf{a} \in A$  hvis og bare hvis hver komponent  $F_i$  er deriverbar i  $\mathbf{a}$ .

*Bevis:* Vi har

$$\begin{aligned}\boldsymbol{\sigma}(\mathbf{r}) &= \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{r} = \\ &= \begin{pmatrix} F_1(\mathbf{a} + \mathbf{r}) - F_1(\mathbf{a}) - \nabla F_1(\mathbf{a}) \cdot \mathbf{r} \\ F_2(\mathbf{a} + \mathbf{r}) - F_2(\mathbf{a}) - \nabla F_2(\mathbf{a}) \cdot \mathbf{r} \\ \vdots \\ F_m(\mathbf{a} + \mathbf{r}) - F_m(\mathbf{a}) - \nabla F_m(\mathbf{a}) \cdot \mathbf{r} \end{pmatrix} = \begin{pmatrix} \sigma_1(\mathbf{r}) \\ \sigma_2(\mathbf{r}) \\ \vdots \\ \sigma_m(\mathbf{r}) \end{pmatrix}\end{aligned}$$

Dermed er

$$\frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r}) = \begin{pmatrix} \frac{\sigma_1(\mathbf{r})}{|\mathbf{r}|} \\ \frac{\sigma_2(\mathbf{r})}{|\mathbf{r}|} \\ \vdots \\ \frac{\sigma_m(\mathbf{r})}{|\mathbf{r}|} \end{pmatrix}$$

og vi vet fra setning 2.3.3 at  $\frac{1}{|\mathbf{r}|}\boldsymbol{\sigma}(\mathbf{r}) \rightarrow \mathbf{0}$  hvis og bare hvis  $\frac{\sigma_i(\mathbf{r})}{|\mathbf{r}|} \rightarrow 0$  for alle  $i$ . Dette betyr at  $\mathbf{F}$  er deriverbar hvis og bare hvis hver  $F_i$  er deriverbar.  $\square$

Kombinerer vi dette resultatet med setning 2.4.6, får vi:

**Korollar 2.6.3** Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av  $n$  variable og at  $\mathbf{a}$  er et indre punkt i  $A$ . Dersom alle komponentene  $\frac{\partial F_i}{\partial x_j}$  i Jacobi-matrisen er definert i en omegn rundt  $\mathbf{a}$  og er kontinuerlige i  $\mathbf{a}$ , så er  $\mathbf{F}$  deriverbar i  $\mathbf{a}$ .

*Bevis:* Ifølge setning 2.4.6 er hver av komponentene  $F_i$  til  $\mathbf{F}$  deriverbar i  $\mathbf{a}$ , og ifølge setningen ovenfor er da  $\mathbf{F}$  deriverbar i  $\mathbf{a}$ .  $\square$

**Eksempel 2:** Vis at funksjonen

$$\mathbf{F}(x, y) = \begin{pmatrix} xy^3 \\ e^{x+y^2} \\ 3x^2y \end{pmatrix}$$

fra eksempel 1 er deriverbar.

Vi har allerede regnet ut Jacobi-matrisen til  $\mathbf{F}$ , og komponentene er åpenbart kontinuerlige overalt. Altså er  $\mathbf{F}$  deriverbar ifølge korollaret.  $\square$

Den neste setningen kan se litt underlig ut, men vi kommer til å ha stor glede av den. Setningen sier at ingen annen matrise  $B$  kan “stjele jobben” til Jacobi-matrisen  $\mathbf{F}'(\mathbf{a})$  — dersom  $B$  tilfredsstiller en betingelse av samme type som  $\mathbf{F}'(\mathbf{a})$  oppfyller i definisjon 2.6.1, så må  $B$  være lik  $\mathbf{F}'(\mathbf{a})$ . Det finnes også andre tolkninger av setningen som vi skal komme tilbake til senere.

**Setning 2.6.4** La  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  være en funksjon av  $n$  variable og la  $\mathbf{a}$  være et indre punkt i  $A$ . Anta at det finnes en  $m \times n$ -matrise  $B$  slik at funksjonen

$$\hat{\sigma}(\mathbf{r}) = \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - B\mathbf{r}$$

tilfredsstiller

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \hat{\sigma}(\mathbf{r}) = \mathbf{0}.$$

Da er  $\mathbf{F}$  deriverbar i  $\mathbf{a}$  og

$$\mathbf{F}'(\mathbf{a}) = B$$

*Bevis:* Vi velger  $\mathbf{r} = r\mathbf{e}_i$  der

$$\mathbf{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

er den  $i$ -te enhetsvektoren og der  $r \in \mathbb{R}$ . Da er

$$\mathbf{F}(\mathbf{a} + r\mathbf{e}_i) - \mathbf{F}(\mathbf{a}) = B(r\mathbf{e}_i) + \hat{\sigma}(r\mathbf{e}_i)$$

Deler vi på  $r$  og bruker at  $|r\mathbf{e}_i| = r$ , får vi

$$\frac{\mathbf{F}(\mathbf{a} + r\mathbf{e}_i) - \mathbf{F}(\mathbf{a})}{r} = B\mathbf{e}_i + \frac{\hat{\sigma}(r\mathbf{e}_i)}{|r\mathbf{e}_i|}$$

Når  $r \rightarrow 0$ , vil det siste leddet på høyre side gå mot  $\mathbf{0}$  ifølge antagelsen, så høyresiden går mot

$$B\mathbf{e}_i = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{mi} \end{pmatrix}$$

Da må venstresiden konvergere mot det samme, og vi får

$$\lim_{r \rightarrow 0} \frac{\mathbf{F}(\mathbf{a} + r\mathbf{e}_i) - \mathbf{F}(\mathbf{a})}{r} = \begin{pmatrix} \lim_{r \rightarrow 0} \frac{F_1(\mathbf{a} + r\mathbf{e}_i) - F_1(\mathbf{a})}{r} \\ \lim_{r \rightarrow 0} \frac{F_2(\mathbf{a} + r\mathbf{e}_i) - F_2(\mathbf{a})}{r} \\ \vdots \\ \lim_{r \rightarrow 0} \frac{F_m(\mathbf{a} + r\mathbf{e}_i) - F_m(\mathbf{a})}{r} \end{pmatrix} = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{mi} \end{pmatrix}$$

Ved definisjonen av partiellderiverte betyr dette at  $\frac{\partial F_1}{\partial x_i} = b_{1i}$ ,  $\frac{\partial F_2}{\partial x_i} = b_{2i}$  og så videre. Altså eksisterer alle de partiellderiverte og  $\mathbf{F}'(\mathbf{a}) = B$ . Deriverbarhet følger nå av at  $\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \hat{\sigma}(\mathbf{r}) = \mathbf{0}$ .  $\square$

### Oppgaver til seksjon 2.6

1. Finn Jacobi-matrisen til funksjonen.

- a)  $\mathbf{F}(x, y) = (x^2y, x + y^2)$ .
- b)  $\mathbf{F}(x, y, z) = (e^{x^2y+z}, xyz^2)$ .
- c)  $\mathbf{F}(x, y) = (x \arctan(xy), x \ln y, xy \cos y^2)$ .
- d)  $\mathbf{F}(x, y, z, u) = (xy \sin(xu^2), z^2u)$ .

## 2.7 Kjerneregelen

For funksjoner av én variabel sier kjerneregelen at den deriverte til den sammensatte funksjonen  $h(x) = f(g(x))$  er gitt ved  $h'(x) = f'(g(x))g'(x)$ . Kjerneregelen for funksjoner av flere variable har akkurat samme form; der som  $\mathbf{H}$  er den sammensatte funksjonen  $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$ , så er

$$\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x}) \quad (2.7.1)$$

Selv om formen er den samme, er innholdet mer komplisert — uttrykkene  $\mathbf{H}'(\mathbf{x})$ ,  $\mathbf{F}'(\mathbf{G}(\mathbf{x}))$  og  $\mathbf{G}'(\mathbf{x})$  er nå matriser, og formelen sier at Jacobi-matrisen til  $\mathbf{H}$  i punktet  $\mathbf{x}$  er lik matriseproduktet av Jacobi-matrisen til  $\mathbf{F}$  i punktet  $\mathbf{G}(\mathbf{x})$  og Jacobi-matrisen til  $\mathbf{G}$  i punktet  $\mathbf{x}$ .

Formel (2.7.1) uttrykker kjerneregelen i en kortfattet form som er lett å huske, men som ikke er så praktisk å bruke i utregninger. Vi skal derfor nærme oss kjerneregelen fra en litt annen synsvinkel og heller komme til matriseformuleringen i (2.7.1) etter hvert. Vi tenker oss at vi har en funksjon (et skalarfelt)  $f(u_1, u_2, \dots, u_m)$  av  $m$  variable. For hver variabel  $u_i$

substituerer vi en funksjon  $g_i(x_1, x_2, \dots, x_n)$  slik at vi får en sammensatt funksjon

$$h(x_1, x_2, \dots, x_n) = f(g_1(x_1, x_2, \dots, x_n), g_2(x_1, x_2, \dots, x_n), \dots, g_m(x_1, x_2, \dots, x_n))$$

av  $n$  variable. Spørsmålet er: dersom vi kjenner alle de partiellderiverte til de opprinnelige funksjonene  $f, g_1, g_2, \dots, g_m$ , kan vi da finne de partiellderiverte til den sammensatte funksjonen  $h$ ? Svaret er ja, men før vi skriver opp formelen som viser oss hvordan vi kan regne ut de partiellderiverte til  $h$ , er det lurt å komprimere notasjonen litt slik at vi slipper å arbeide med altfor lange uttrykk. Dersom vi bruker vektornotasjon og skriver

$$\mathbf{u} = (u_1, u_2, \dots, u_m)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_n)$$

og

$$\mathbf{g}(\mathbf{x}) = ((g_1(x_1, x_2, \dots, x_n), g_2(x_1, x_2, \dots, x_n), \dots, g_m(x_1, x_2, \dots, x_n))),$$

kan definisjonen av  $h$  skrives

$$h(\mathbf{x}) = f(\mathbf{g}(\mathbf{x}))$$

Dersom  $\mathbf{g}$  er deriverbar i punktet  $\mathbf{x}$  og  $f$  er deriverbar i punktet  $\mathbf{u} = \mathbf{g}(\mathbf{x})$ , viser det seg at de partiellderiverte til  $h$  er gitt ved

$$\frac{\partial h}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u}) \frac{\partial g_1}{\partial x_i}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u}) \frac{\partial g_2}{\partial x_i}(\mathbf{x}) + \dots + \frac{\partial f}{\partial u_m}(\mathbf{u}) \frac{\partial g_m}{\partial x_i}(\mathbf{x}) \quad (2.7.2)$$

Denne formelen kaller vi *kjerneregelen på komponentform*, mens (2.7.1) kalles *kjerneregelen på matriseform* — vi skal senere se at selv om formen er forskjellig, så uttrykker de to formlene det samme.

Legg merke til at vi i (2.7.2) evaluerer de partiellderiverte til  $f$  i punktet  $\mathbf{u} = \mathbf{g}(\mathbf{x})$ ; det betyr at etter at vi har regnet ut de partiellderiverte til  $f$ , må vi erstatte  $u_1$  med  $g_1(x_1, \dots, x_n)$ ,  $u_2$  med  $g_2(x_1, \dots, x_n)$  osv. Legg også merke til at det er et system i hvilke derivasjoner vi gjør i formel (2.7.2); vi deriverer med hensyn på *alle* “mellomvariablene”  $u_1, u_2, \dots, u_m$ , men bare med hensyn på én av “grunnvariablene”  $x_1, x_2, \dots, x_n$ , nemlig den  $x_i$ -en som inngår i den opprinnelige derivasjonen av  $h$ .

Ting blir klarere hvis vi ser på et eksempel:

**Eksempel 1:** Vi skal se på tilfellet der

$$f(u_1, u_2) = 2u_1 u_2^2$$

og

$$g_1(x_1, x_2, x_3) = x_1 x_2 \sin x_3$$

$$g_2(x_1, x_2, x_3) = 3x_1^2 x_2 x_3$$

Vi finner den sammensatte funksjonen  $h = f(\mathbf{g}(\mathbf{x}))$  ved å substituere  $u_1 = g_1(x_1, x_2, x_3) = x_1 x_2 \sin x_3$  og  $u_2 = g_2(x_1, x_2, x_3) = 3x_1^2 x_2 x_3$  inn i uttrykket for  $f$ :

$$\begin{aligned} h(x_1, x_2, x_3) &= f(u_1, u_2) = 2u_1 u_2^2 = \\ &= 2(x_1 x_2 \sin x_3)(3x_1^2 x_2 x_3)^2 = 18x_1^5 x_2^3 x_3^2 \sin x_3 \end{aligned}$$

Vi kan selvfølgelig finne de partiellderiverte til  $h$  ved å derivere dette uttrykket på vanlig måte, men la oss se hvordan vi kan bruke kjerneregelen isteden. Ifølge (2.7.2) er

$$\frac{\partial h}{\partial x_1}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u}) \frac{\partial g_1}{\partial x_1}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u}) \frac{\partial g_2}{\partial x_1}(\mathbf{x})$$

Regner vi ut de partiellderiverte til  $f$ ,  $g_1$  og  $g_2$  og setter inn i dette uttrykket, får vi

$$\frac{\partial h}{\partial x_1}(\mathbf{x}) = (2u_2^2)(x_2 \sin x_3) + (4u_1 u_2)(6x_1 x_2 x_3)$$

Helt til slutt setter vi inn uttrykkene  $u_1 = g_1(x_1, x_2, x_3) = x_1 x_2 \sin x_3$  og  $u_2 = g_2(x_1, x_2, x_3) = 3x_1^2 x_2 x_3$ :

$$\begin{aligned} \frac{\partial h}{\partial x_1}(\mathbf{x}) &= (2(3x_1^2 x_2 x_3)^2)(x_2 \sin x_3) + (4(x_1 x_2 \sin x_3)(3x_1^2 x_2 x_3))(6x_1 x_2 x_3) = \\ &= 18x_1^4 x_2^3 x_3^2 \sin x_3 + 72x_1^4 x_2^3 x_3^2 \sin x_3 = 90x_1^4 x_2^3 x_3^2 \sin x_3 \end{aligned}$$

Det er god trening i å finne  $\frac{\partial h}{\partial x_2}(\mathbf{x})$  og  $\frac{\partial h}{\partial x_3}(\mathbf{x})$  på tilsvarende måte. ♣

**Bemerkning:** Hvis du ser nøyere på eksemplet ovenfor, vil du oppdage at vi har regnet ut  $\frac{\partial h}{\partial x_1}(\mathbf{x})$  på en svært tungvinn måte — det går mye raske re å bare derivere uttrykket  $h(x_1, x_2, x_3) = 18x_1^5 x_2^3 x_3^2 \sin x_3$  på vanlig måte! Dette er ganske typisk; kjerneregelen i flere variable er faktisk ikke en særlig effektiv metode for å regne ut partiellderiverte til funksjoner gitt ved formler — regelens styrke ligger isteden i at den er et utmerket redskap til å utlede *generelle* sammenhenger mellom varierende størrelser. Før vi ser nærmere på dette, kan det være lurt å ta med enda et konkret eksempel slik at vi blir litt bedre kjent med kjerneregelen.

**Eksempel 2:** I eksemplet ovenfor het de variable  $u_1, u_2$  og  $x_1, x_2, x_3$  akkurat som i vår generelle formel (2.7.2). Når man bruker kjerneregelen i andre fag, har de variable vanligvis helt andre navn, og det er viktig at vi da greier å kjenne igjen de forskjellige delene av kjerneregelen. La oss nå tenke oss at vi har en funksjon

$$f(s, t) = e^s t^2$$

og at  $s$  og  $t$  er funksjoner av underliggende variable  $u$  og  $v$

$$s = S(u, v) = 3uv^3 \quad \text{og} \quad t = T(u, v) = \cos(uv)$$

Vi vil bruke kjerneregelen til å finne de partiellderiverte til den sammensatte funksjonen

$$h(u, v) = f(S(u, v), T(u, v))$$

I dette tilfellet sier formel (2.7.2)

$$\frac{\partial h}{\partial u}(u, v) = \frac{\partial f}{\partial s}(s, t) \frac{\partial S}{\partial u}(u, v) + \frac{\partial f}{\partial t}(s, t) \frac{\partial T}{\partial u}(u, v)$$

og

$$\frac{\partial h}{\partial v}(u, v) = \frac{\partial f}{\partial s}(s, t) \frac{\partial S}{\partial v}(u, v) + \frac{\partial f}{\partial t}(s, t) \frac{\partial T}{\partial v}(u, v)$$

(husk at vi skal derivere med hensyn på begge "mellomvariablene"  $s$  og  $t$ , men bare med hensyn på den "opprikkelige" grunnvariablen'  $u$  eller  $v$ ). Regner vi ut de partiellderiverte og setter inn  $s = S(u, v) = 3uv^3$  og  $t = T(u, v) = \cos(uv)$ , får vi

$$\begin{aligned} \frac{\partial h}{\partial u}(u, v) &= e^s t^2 3v^3 + 2e^s t(-\sin(uv))v = \\ &= e^{3uv^3} (\cos(uv))^2 3v^3 - 2e^{3uv^3} \cos(uv) \sin(uv)v = \\ &= v e^{3uv^3} \cos(uv) \left( 3v^2 \cos(uv) - 2 \sin(uv) \right) \end{aligned}$$

Den andre partiellderiverte  $\frac{\partial h}{\partial v}(u, v)$  finner vi på tilsvarende måte.  $\square$

Vi skal ta med et par eksempler på hvordan kjerneregelen brukes i andre fag. Før vi gjør det, kan det være greit med noen ord om notasjon. Når vi bruker kjerneregelen i praktiske situasjoner, blir uttrykkene ofte lange og uoversiktlige. For å lette lesbarheten er det vanlig å utelate punktene som funksjonene evalueres i — man skriver altså

$$\frac{\partial h}{\partial x_i} = \frac{\partial f}{\partial u_1} \frac{\partial g_1}{\partial x_i} + \frac{\partial f}{\partial u_2} \frac{\partial g_2}{\partial x_i} + \cdots + \frac{\partial f}{\partial u_m} \frac{\partial g_m}{\partial x_i}$$

istedenfor

$$\frac{\partial h}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u}) \frac{\partial g_1}{\partial x_i}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u}) \frac{\partial g_2}{\partial x_i}(\mathbf{x}) + \cdots + \frac{\partial f}{\partial u_m}(\mathbf{u}) \frac{\partial g_m}{\partial x_i}(\mathbf{x})$$

og setter først inn verdiene til  $\mathbf{x}$  og  $\mathbf{u}$  når man har behov for dem.

**Eksempel 3:** I økonomi er utbyttet til et firma avhengig av forskjellige faktorer. For en bonde kan vi tenke oss at utbyttet er avhengig av tre faktorer: arbeidsinnsatsen  $a$ , den løpende kapitalinvesteringen  $k$ , og arealet av dyrkbar mark  $m$ . Utbyttet  $U$  er dermed en funksjon

$$U = F(a, k, m)$$

Vi tenker oss at de tre faktorene varierer med tiden:  $a = A(t)$ ,  $k = K(t)$  og  $m = M(t)$ . Kjerneregelen forteller oss nå hvordan utbyttet  $U$  varierer med tiden:

$$\frac{\partial U}{\partial t} = \frac{\partial F}{\partial a} \frac{\partial A}{\partial t} + \frac{\partial F}{\partial k} \frac{\partial K}{\partial t} + \frac{\partial F}{\partial m} \frac{\partial M}{\partial t}$$

der vi må huske at vi skal evaluere de partiellderiverte  $\frac{\partial F}{\partial a}$ ,  $\frac{\partial F}{\partial k}$ ,  $\frac{\partial F}{\partial m}$  i punktet  $(A(t), K(t), M(t))$ . Legg merke til at i denne formelen er  $U$ ,  $A$ ,  $K$  og  $M$  bare avhengige av én variabel  $t$ , og vi kan derfor skrive vanlige deriverte istedenfor partiellderiverte:

$$U'(t) = \frac{\partial F}{\partial a} A'(t) + \frac{\partial F}{\partial k} K'(t) + \frac{\partial F}{\partial m} M'(t)$$

I økonomiske modeller antar man ofte at utbyttefunksjonen  $F$  er en *Cobb-Douglasfunksjon*, dvs. at den har formen

$$F(a, k, m) = C a^\alpha k^\beta m^\gamma$$

der  $C$ ,  $\alpha$ ,  $\beta$  og  $\gamma$  er positive konstanter. Deriverer vi dette uttrykket, får vi

$$\frac{\partial F}{\partial a} = C \alpha a^{\alpha-1} k^\beta m^\gamma, \quad \frac{\partial F}{\partial k} = C \beta a^\alpha k^{\beta-1} m^\gamma, \quad \frac{\partial F}{\partial m} = C \gamma a^\alpha k^\beta m^{\gamma-1}$$

som innsatt i uttrykket ovenfor gir (etter litt opprydning):

$$U'(t) = C A(t)^\alpha K(t)^\beta M(t)^\gamma \left( \alpha \frac{A'(t)}{A(t)} + \beta \frac{K'(t)}{K(t)} + \gamma \frac{M'(t)}{M(t)} \right)$$

I en økonomisk situasjon med begrensede ressurser vil man ofte overveie å styrke noen innsatsområder på bekostning av andre. Bonden vår kan f.eks. overveie å utvide det dyrkede arealet  $m$  ved å kjøpe tilleggsjord, men er da tvunget til å redusere kapitalinvesteringen  $k$ . Om slike omprioriteringer er lønnsomme eller ikke, viser seg i fortegnet til den deriverte  $U'(t)$  av utbyttefunksjonen. Ifølge formelen ovenfor er dette fortegnet bestemt av fortegnet til uttrykket

$$\alpha \frac{A'(t)}{A(t)} + \beta \frac{K'(t)}{K(t)} + \gamma \frac{M'(t)}{M(t)}$$

( $C A(t)^\alpha K(t)^\beta M(t)^\gamma$  er alltid positiv). Fortegnet til den deriverte av utbyttefunksjonen er altså bestemt av en kombinasjon av de *relative veksthastigheten*  $\frac{A'(t)}{A(t)}$ ,  $\frac{K'(t)}{K(t)}$  og  $\frac{M'(t)}{M(t)}$  og Cobb-Douglasponentene  $\alpha$ ,  $\beta$  og  $\gamma$ . For en

økonom med trening i å tolke Cobb-Douglaseksponenter gir et slikt uttrykk økonomisk mening.  $\square$

**Eksempel 4:** Når en gass oppbevares i en beholder, er trykket  $P$  en funksjon av temperaturen  $T$  og volumet  $V$ ; vi har altså

$$P = f(T, V)$$

Hvilken funksjon  $f$  det er naturlig å bruke, avhenger både av hvilken gass vi ser på, og av hvor stor nøyaktighet vi ønsker. Foreløpig lar vi  $f$  være uspesifisert slik at regningene våre gjelder for alle modeller. Dersom volumet og temperaturen varierer med tiden  $t$ , vil også trykket variere med tiden

$$P(t) = f(T(t), V(t))$$

Vi kan finne den deriverte av denne funksjonen ved å bruke kjerneregelen (husk at  $P(t)$ ,  $T(t)$  og  $V(t)$  bare avhenger av én variabel  $t$  slik at vi kan bruke vanlige deriverte  $P'(t)$ ,  $T'(t)$  og  $V'(t)$  istedenfor partiellderiverte  $\frac{\partial P}{\partial t}$ ,  $\frac{\partial T}{\partial t}$  og  $\frac{\partial V}{\partial t}$ ):

$$P'(t) = \frac{\partial f}{\partial T} T'(t) + \frac{\partial f}{\partial V} V'(t)$$

Dersom gassen er tynn og kretene mellom gasspartiklene er svake (fysikerne kaller dette en *ideell gass*), kan man anta at trykket  $P$  er proporsjonal med temperaturen  $T$  og omvendt proporsjonal med volumet  $V$ , dvs.  $P = k \frac{T}{V}$  der  $k$  er en konstant som blant annet avhenger av hvor mye gass det er i beholderen. Vi har altså

$$P = f(T, V) = k \frac{T}{V}$$

I dette tilfellet er

$$\frac{\partial f}{\partial T} = k \frac{1}{V} \quad \text{og} \quad \frac{\partial f}{\partial V} = -k \frac{T}{V^2}$$

Setter vi dette inn i formelen for  $P'(t)$ , får vi

$$P'(t) = k \frac{1}{V} T'(t) - k \frac{T}{V^2} V'(t)$$

Vi har tidligere regnet oss frem til denne formelen på en litt annen måte (se eksempel 6 i seksjon 2.4).

Regning med partiellderiverte og kjerneregelen står sentralt i termodynamikken, og utregningene ovenfor er bare et enkelt eksempel.  $\clubsuit$

Vi begynte denne seksjonen med å skrive opp kjerneregelen på matriseform:

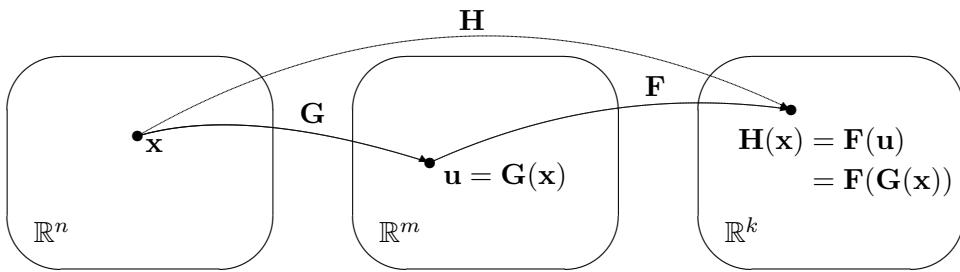
$$\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x})) \mathbf{G}'(\mathbf{x}),$$

men siden har vi bare arbeidet med regelen på komponentform:

$$\frac{\partial h}{\partial x_i}(\mathbf{x}) = \frac{\partial f}{\partial u_1}(\mathbf{u}) \frac{\partial g_1}{\partial x_i}(\mathbf{x}) + \frac{\partial f}{\partial u_2}(\mathbf{u}) \frac{\partial g_2}{\partial x_i}(\mathbf{x}) + \dots + \frac{\partial f}{\partial u_m}(\mathbf{u}) \frac{\partial g_m}{\partial x_i}(\mathbf{x})$$

La oss nå se litt på sammenhengen mellom de to formene.

Når vi arbeider med kjerneregelen på matriseform, tenker vi oss at vi starter med to funksjoner  $\mathbf{G} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^k$  og at vi danner den sammensatte funksjonen  $\mathbf{H} : \mathbb{R}^n \rightarrow \mathbb{R}^k$  ved  $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$  som vist på figuren.



Skriver vi ut komponentene til  $\mathbf{H}'(\mathbf{x})$ ,  $\mathbf{F}'(\mathbf{G}(\mathbf{x}))$ , og  $\mathbf{G}'(\mathbf{x})$ , får kjerne-regelen  $\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x})$  denne formen:

$$\begin{aligned} & \left( \begin{array}{cccc} \frac{\partial H_1}{\partial x_1}(\mathbf{x}) & \frac{\partial H_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial H_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial H_2}{\partial x_1}(\mathbf{x}) & \frac{\partial H_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial H_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial H_k}{\partial x_1}(\mathbf{x}) & \frac{\partial H_k}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial H_k}{\partial x_n}(\mathbf{x}) \end{array} \right) = \\ & = \left( \begin{array}{cccc} \frac{\partial F_1}{\partial u_1}(\mathbf{G}(\mathbf{x})) & \frac{\partial F_1}{\partial u_2}(\mathbf{G}(\mathbf{x})) & \cdots & \frac{\partial F_1}{\partial u_m}(\mathbf{G}(\mathbf{x})) \\ \frac{\partial F_2}{\partial u_1}(\mathbf{G}(\mathbf{x})) & \frac{\partial F_2}{\partial u_2}(\mathbf{G}(\mathbf{x})) & \cdots & \frac{\partial F_2}{\partial u_m}(\mathbf{G}(\mathbf{x})) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_k}{\partial u_1}(\mathbf{G}(\mathbf{x})) & \frac{\partial F_k}{\partial u_2}(\mathbf{G}(\mathbf{x})) & \cdots & \frac{\partial F_k}{\partial u_m}(\mathbf{G}(\mathbf{x})) \end{array} \right) \left( \begin{array}{cccc} \frac{\partial G_1}{\partial x_1}(\mathbf{x}) & \frac{\partial G_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial G_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial G_2}{\partial x_1}(\mathbf{x}) & \frac{\partial G_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial G_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial G_m}{\partial x_1}(\mathbf{x}) & \frac{\partial G_m}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial G_m}{\partial x_n}(\mathbf{x}) \end{array} \right) \end{aligned}$$

Den  $ij$ -te komponenten i den første matrisen er  $\frac{\partial H_i}{\partial x_j}(\mathbf{x})$ . Denne komponenten må være skalarproduktet av den  $i$ -te linjen i den andre matrisen og den  $j$ -te søylen i den tredje matrisen, altså:

$$\frac{\partial H_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_i}{\partial u_1}(\mathbf{G}(\mathbf{x})) \frac{\partial G_1}{\partial x_j}(\mathbf{x}) + \frac{\partial F_i}{\partial u_2}(\mathbf{G}(\mathbf{x})) \frac{\partial G_2}{\partial x_j}(\mathbf{x}) + \cdots + \frac{\partial F_i}{\partial u_m}(\mathbf{G}(\mathbf{x})) \frac{\partial G_m}{\partial x_j}(\mathbf{x})$$

Dette er ikke noe annet enn kjerneregelen på komponentform anvendt på funksjonen  $H_i(\mathbf{x}) = F_i(\mathbf{G}(\mathbf{x}))$ . Vi har dermed vist at kjerneregelen på komponentform rett og slett er det vi får når vi ganger ut matriseproduktet i formelen  $\mathbf{H}'(\mathbf{x}) = \mathbf{F}'(\mathbf{G}(\mathbf{x}))\mathbf{G}'(\mathbf{x})$  og sjekker hva som skjer med hver enkelt komponent. Når vi nå skal bevise kjerneregelen, holder det derfor å bevise den på matriseform.

**Teorem 2.7.1 (Kjerneregelen på matriseform)** *Anta at vi har to mengder  $A \subset \mathbb{R}^n$ ,  $B \subset \mathbb{R}^m$  og to funksjoner  $\mathbf{G} : A \rightarrow B$ ,  $\mathbf{F} : B \rightarrow \mathbb{R}^k$ . Dersom  $\mathbf{G}$  er deriverbar i punktet  $\mathbf{a} \in A$ , og  $\mathbf{F}$  er deriverbar i punktet  $\mathbf{b} = \mathbf{G}(\mathbf{a})$ , så er den sammensatte funksjonen  $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$  deriverbar i  $\mathbf{a}$ , og Jacobimatrisen til  $\mathbf{H}$  er gitt ved*

$$\mathbf{H}'(\mathbf{a}) = \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a})$$

*Bevis:* Dette beviset er ikke så vanskelig når man har forstått ideen, men det er langt og litt kronglete. Ideen er å bruke setning 2.6.4 med  $\mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a})$  i rollen som matrisen  $B$ . Alt vi da behøver å vise, er at funksjonen

$$\hat{\boldsymbol{\sigma}}(\mathbf{r}) = \mathbf{H}(\mathbf{a} + \mathbf{r}) - \mathbf{H}(\mathbf{a}) - (\mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a}))\mathbf{r} \quad (2.7.3)$$

tilfredsstiller kravet

$$\lim_{\mathbf{r} \rightarrow 0} \frac{1}{|\mathbf{r}|} \hat{\boldsymbol{\sigma}}(\mathbf{r}) = \mathbf{0} \quad (2.7.4)$$

For å vise dette, må vi bruke at  $\mathbf{G}$  og  $\mathbf{F}$  er deriverbare. Siden  $\mathbf{G}$  er deriverbar i  $\mathbf{a}$ , vet vi at funksjonen

$$\boldsymbol{\sigma}_1(\mathbf{r}) = \mathbf{G}(\mathbf{a} + \mathbf{r}) - \mathbf{G}(\mathbf{a}) - \mathbf{G}'(\mathbf{a})\mathbf{r} \quad (2.7.5)$$

tilfredsstiller kravet

$$\lim_{\mathbf{r} \rightarrow 0} \frac{1}{|\mathbf{r}|} \boldsymbol{\sigma}_1(\mathbf{r}) = \mathbf{0} \quad (2.7.6)$$

Siden  $\mathbf{F}$  er deriverbar i  $\mathbf{b} = \mathbf{G}(\mathbf{a})$ , vet vi tilsvarende at funksjonen

$$\begin{aligned} \boldsymbol{\sigma}_2(\mathbf{s}) &= \mathbf{F}(\mathbf{b} + \mathbf{s}) - \mathbf{F}(\mathbf{b}) - \mathbf{F}'(\mathbf{b})\mathbf{s} = \\ &= \mathbf{F}(\mathbf{G}(\mathbf{a}) + \mathbf{s}) - \mathbf{F}(\mathbf{G}(\mathbf{a})) - \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{s} \end{aligned} \quad (2.7.7)$$

tilfredsstiller kravet

$$\lim_{\mathbf{s} \rightarrow 0} \frac{1}{|\mathbf{s}|} \boldsymbol{\sigma}_2(\mathbf{s}) = \mathbf{0} \quad (2.7.8)$$

Vi setter  $\mathbf{s} = \mathbf{G}(\mathbf{a} + \mathbf{r}) - \mathbf{G}(\mathbf{a})$ . Da er  $\mathbf{G}(\mathbf{a} + \mathbf{r}) = \mathbf{G}(\mathbf{a}) + \mathbf{s}$ , og fra (2.7.5) ser vi dessuten at  $\mathbf{s} = \mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})$ . Dermed er

$$\mathbf{H}(\mathbf{a} + \mathbf{r}) = \mathbf{F}(\mathbf{G}(\mathbf{a} + \mathbf{r})) = \mathbf{F}(\mathbf{G}(\mathbf{a}) + \mathbf{s}) = \mathbf{F}(\mathbf{G}(\mathbf{a})) + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{s} + \boldsymbol{\sigma}_2(\mathbf{s})$$

der vi i siste trinn har brukt (2.7.7). Siden  $\mathbf{s} = \mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})$ , gir dette videre

$$\begin{aligned}\mathbf{H}(\mathbf{a} + \mathbf{r}) &= \mathbf{F}(\mathbf{G}(\mathbf{a})) + \mathbf{F}'(\mathbf{G}(\mathbf{a})) \left( \mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r}) \right) + \boldsymbol{\sigma}_2(\mathbf{s}) = \\ &= \mathbf{H}(\mathbf{a}) + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\mathbf{G}'(\mathbf{a})\mathbf{r} + \mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s})\end{aligned}$$

Sammenligner vi dette med (2.7.3), ser vi at

$$\hat{\boldsymbol{\sigma}}(\mathbf{r}) = \mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s})$$

Ifølge (2.7.4) er vår oppgave dermed å vise at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \hat{\boldsymbol{\sigma}}(\mathbf{r}) = \lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) + \boldsymbol{\sigma}_2(\mathbf{s}) \right) = \mathbf{0} \quad (2.7.9)$$

Vi tar ett ledd av gangen. Fra setning 1.6.3 vet vi at

$$|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r})| \leq \|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\| |\boldsymbol{\sigma}_1(\mathbf{r})|$$

der  $\|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\|$  bare er et fast tall (normen til matrisen  $\mathbf{F}'(\mathbf{G}(\mathbf{a}))$ ). Dermed har vi

$$\left| \frac{1}{|\mathbf{r}|} \left( \mathbf{F}'(\mathbf{G}(\mathbf{a}))\boldsymbol{\sigma}_1(\mathbf{r}) \right) \right| \leq \|\mathbf{F}'(\mathbf{G}(\mathbf{a}))\| \frac{|\boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|} \rightarrow 0$$

siden  $\frac{\boldsymbol{\sigma}_1(\mathbf{r})}{|\mathbf{r}|} \rightarrow \mathbf{0}$  når  $\mathbf{r} \rightarrow \mathbf{0}$ . Dette viser at den første delen av uttrykket i (2.7.9) går mot  $\mathbf{0}$ . Den andre delen

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \boldsymbol{\sigma}_2(\mathbf{s}) = \mathbf{0}$$

er litt verre. Vi legger først merke til at siden  $\boldsymbol{\sigma}_2(\mathbf{0}) = \mathbf{0}$ , har vi ingen problemer med tilfellet  $\mathbf{s} = \mathbf{0}$ . Vi kan derfor anta at  $\mathbf{s} \neq \mathbf{0}$ , og multiplisere uttrykket  $\frac{1}{|\mathbf{r}|} \boldsymbol{\sigma}_2(\mathbf{s})$  med  $|\mathbf{s}|$  i teller og nevner:

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \boldsymbol{\sigma}_2(\mathbf{s}) = \lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{|\mathbf{s}|}{|\mathbf{r}|} \cdot \frac{\boldsymbol{\sigma}_2(\mathbf{s})}{|\mathbf{s}|}$$

Når  $\mathbf{r} \rightarrow \mathbf{0}$ , vil  $\mathbf{s} = \mathbf{G}(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r}) \rightarrow \mathbf{0}$ , og dette medfører at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{\boldsymbol{\sigma}_2(\mathbf{s})}{|\mathbf{s}|} = \lim_{\mathbf{s} \rightarrow \mathbf{0}} \frac{\boldsymbol{\sigma}_2(\mathbf{s})}{|\mathbf{s}|} = \mathbf{0}$$

ifølge (2.7.8). Det gjenstår derfor å vise at faktoren  $\frac{|\mathbf{s}|}{|\mathbf{r}|}$  er begrenset og ikke kan gå mot uendelig. Bruker vi at  $\mathbf{s} = \mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})$ , får vi

$$\frac{|\mathbf{s}|}{|\mathbf{r}|} = \frac{|\mathbf{G}'(\mathbf{a})\mathbf{r} + \boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|} \leq \frac{|\mathbf{G}'(\mathbf{a})\mathbf{r}|}{|\mathbf{r}|} + \frac{|\boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|} \leq \|\mathbf{G}'(\mathbf{a})\| + \frac{|\boldsymbol{\sigma}_1(\mathbf{r})|}{|\mathbf{r}|}$$

som er begrenset siden  $\frac{\sigma_1(\mathbf{r})}{|\mathbf{r}|} \rightarrow \mathbf{0}$ . Dermed har vi vist at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \hat{\sigma}(\mathbf{r}) = \lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}'(\mathbf{G}(\mathbf{a})) \sigma_1(\mathbf{r}) + \sigma_2(\mathbf{s}) \right) = \mathbf{0}$$

og beviset er fullført.  $\square$

**Bemerkning:** Legg merke til at kjerneregelen garanterer at  $\mathbf{H}$  er deriverbar i  $\mathbf{a}$  så lenge  $\mathbf{G}$  er deriverbar i  $\mathbf{a}$  og  $\mathbf{F}$  er deriverbar i  $\mathbf{b} = \mathbf{G}(\mathbf{a})$ . Siden deriverbarhet kan være vanskelig å sjekke, er denne garantien ofte til stor hjelp i teoretisk arbeid.

Vi skriver også opp den presise formuleringen av kjerneregelen på komponentform:

**Teorem 2.7.2 (Kjerneregelen på komponentform)** *Anta at vi har to mengder  $A \subset \mathbb{R}^n$ ,  $B \subset \mathbb{R}^m$  og to funksjoner  $\mathbf{G} : A \rightarrow B$ ,  $\mathbf{F} : B \rightarrow \mathbb{R}^k$ . Dersom  $\mathbf{G}$  er deriverbar i punktet  $\mathbf{a} \in A$  og  $\mathbf{F}$  er deriverbar i punktet  $\mathbf{b} = \mathbf{G}(\mathbf{a})$ , så er den sammensatte funksjonen  $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$  deriverbar i  $\mathbf{a}$ , og de partiellderiverte til  $\mathbf{H}$  er gitt ved*

$$\begin{aligned} \frac{\partial H_i}{\partial x_j}(\mathbf{a}) &= \sum_{p=1}^m \frac{\partial F_i}{\partial u_p}(\mathbf{G}(\mathbf{a})) \frac{\partial G_p}{\partial x_j}(\mathbf{a}) = \\ &= \frac{\partial F_i}{\partial u_1}(\mathbf{G}(\mathbf{a})) \frac{\partial G_1}{\partial x_j}(\mathbf{a}) + \frac{\partial F_i}{\partial u_2}(\mathbf{G}(\mathbf{a})) \frac{\partial G_2}{\partial x_j}(\mathbf{a}) + \cdots + \frac{\partial F_i}{\partial u_m}(\mathbf{G}(\mathbf{a})) \frac{\partial G_m}{\partial x_j}(\mathbf{a}) \end{aligned}$$

*Bevis:* Som vi allerede har sett, følger denne formuleringen direkte fra matriseformen når man utfører matrisemultiplikasjonen og sammenligner koefisientene på begge sider av likhetstegnet.  $\square$

### Oppgaver til seksjon 2.7

1. La  $f(u, v) = u^2 + v$ ,  $g(x, y) = 2xy$ ,  $h(x, y) = x + y^2$ . Bruk kjerneregelen til å finne de partiellderiverte av  $k(x, y) = f(g(x, y), h(x, y))$ .
2. La  $f(u, v) = ue^{-v}$ ,  $g(x, y, z) = 2xy + z$ ,  $h(x, y, z) = 2y(z + x)$ . Bruk kjerneregelen til å finne de partiellderiverte av  $k(x, y, z) = f(g(x, y, z), h(x, y, z))$ .
3. Bruk kjerneregelen til å regne ut  $\frac{\partial h}{\partial x_2}$  og  $\frac{\partial h}{\partial x_3}$  i eksempel 1.
4. Bruk kjerneregelen til å regne ut  $\frac{\partial h}{\partial v}$  i eksempel 2.
5. Vi har to funksjoner  $\mathbf{G} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  og  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ . Anta at  $\mathbf{G}(1, -2) = (1, 2, 3)$  og at

$$\mathbf{G}'(1, -2) = \begin{pmatrix} 1 & -2 \\ 3 & 1 \\ 2 & -1 \end{pmatrix}, \quad \mathbf{F}'(1, 2, 3) = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 2 \end{pmatrix}$$

Finn Jacobi-matrisen til den sammensatte funksjonen  $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$  i punktet  $(1, -2)$ .

- 6.** Vi har to funksjoner  $\mathbf{G} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  og  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . Anta at  $\mathbf{G}(-1, -2, 1) = (2, 4)$  og at

$$\mathbf{G}'(-1, -2, 1) = \begin{pmatrix} 4 & -2 & 0 \\ 1 & 3 & -1 \end{pmatrix}, \quad \mathbf{F}'(2, 4) = \begin{pmatrix} 2 & -3 \\ 0 & 2 \end{pmatrix}$$

Finn Jacobi-matrisen til den sammensatte funksjonen  $\mathbf{H}(\mathbf{x}) = \mathbf{F}(\mathbf{G}(\mathbf{x}))$  i punktet  $(-1, -2, 1)$ .

- 7.** To vareslag konkurrerer om det samme markedet. Etterspørselen  $E_1$  etter det første vareslaget varierer med prisene  $p_1$  og  $p_2$  på begge vareslagene.

Vi har altså en funksjon  $E_1 = E_1(p_1, p_2)$ . Anta at vi vet hvordan prisene  $p_1 = p_1(t)$  og  $p_2 = p_2(t)$  varierer med tiden. Vis at etterspørselens variasjon med tiden kan uttrykkes ved

$$\frac{dE_1}{dt} = \frac{\partial E_1}{\partial p_1} p'_1(t) + \frac{\partial E_1}{\partial p_2} p'_2(t)$$

- 8.** Temperaturen  $T$  i et område avhenger av posisjonen; vi kan tenke oss at den er gitt som en funksjon  $T = f(x, y)$  av to variable der  $x$  og  $y$  er vanlige koordinater. Vi innfører nå polarkoordinater  $r$  og  $\theta$  på vanlig måte slik at  $x = r \cos \theta$ ,  $y = r \sin \theta$ . Vi får da temperaturen som en funksjon  $T = f(r \cos \theta, r \sin \theta)$  av  $r$  og  $\theta$ .

- a) Vis at

$$\begin{aligned} \frac{\partial T}{\partial r} &= \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta \\ \frac{\partial T}{\partial \theta} &= -\frac{\partial f}{\partial x} r \sin \theta + \frac{\partial f}{\partial y} r \cos \theta \end{aligned}$$

- b) En radiomerket fugl beveger seg i området. Radiosignalene viser hvordan avstanden  $r$  og vinkelen  $\theta$  varierer med tiden; vi har  $r = g(t)$  og  $\theta = h(t)$ . Vis at temperaturendringene fuglen opplever er gitt ved

$$T'(t) = \left( \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta \right) g'(t) + \left( -\frac{\partial f}{\partial x} r \sin \theta + \frac{\partial f}{\partial y} r \cos \theta \right) h'(t)$$

der vi må sette inn  $r = g(t)$ ,  $\theta = h(t)$ ,  $x = g(t) \cos h(t)$ ,  $y = g(t) \sin h(t)$ .

- 9.** La  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  være en deriverbar funksjon av  $n + 1$  variable og anta at det finnes en deriverbar funksjon  $g : \mathbb{R}^n \rightarrow \mathbb{R}$  slik at

$$f(x_1, x_2, \dots, x_n, g(x_1, x_2, \dots, x_n)) = 0$$

for alle  $x_1, x_2, \dots, x_n$ . (Tenk deg at  $y = g(x_1, x_2, \dots, x_n)$  er det uttrykket du får dersom du løser ligningen  $f(x_1, x_2, \dots, x_n, y) = 0$  med hensyn på  $y$ ).

- a) Vis at

$$\frac{\partial g}{\partial x_i}(x_1, x_2, \dots, x_n) = -\frac{\frac{\partial f}{\partial x_i}(x_1, x_2, \dots, x_n, g(x_1, x_2, \dots, x_n))}{\frac{\partial f}{\partial x_{n+1}}(x_1, x_2, \dots, x_n, g(x_1, x_2, \dots, x_n))}$$

- b) La  $f(x, y) = x^2 + y^2 - R^2$  der  $R$  er en positiv konstant, og anta at  $y = g(x)$  er en deriverbar funksjon slik at  $f(x, g(x)) = 0$ . Vis at

$$g'(x) = -\frac{x}{g(x)}$$

Gi en geometrisk tolkning av resultatet.

- c) La  $f(x, y, z) = x^2 + y^2 + z^2 - R^2$  der  $R$  er en positiv konstant og anta at  $z = g(x, y)$  er en deriverbar funksjon slik at  $f(x, y, g(x, y)) = 0$ . Vis at

$$\frac{\partial g}{\partial x}(x, y) = -\frac{x}{g(x, y)}$$

og

$$\frac{\partial g}{\partial y}(x, y) = -\frac{y}{g(x, y)}$$

Gi en geometrisk tolkning av resultatet.

## 2.8 Lineæravbildninger

I de to siste seksjonene i dette kapitlet skal vi se på to spesielle og nært beslektede typer av funksjoner fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$  — *lineæravbildninger* og *affinavbildninger*. Disse funksjonstypene knytter teorien i dette kapitlet nærmere til teorien for matriser og vektorer i forrige kapittel.

Vi begynner med definisjonen av lineæravbildninger.

**Definisjon 2.8.1** En funksjon  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  kalles en *lineæravbildning* dersom vi for alle  $c \in \mathbb{R}$  og alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  har:

- (i)  $\mathbf{T}(c\mathbf{x}) = c\mathbf{T}(\mathbf{x})$
- (ii)  $\mathbf{T}(\mathbf{x} + \mathbf{y}) = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$

De aller fleste funksjoner fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$  er ikke lineæravbildninger, men disse funksjonene er allikevel så viktige at det er en hel gren av matematikken som hovedsakelig handler om dem — denne grenen kalles *lineær algebra*.

La oss begynne med en enkel og nyttig generalisering av definisjonen.

**Setning 2.8.2** Anta at  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  er en lineæravbildning. Da er

$$\mathbf{T}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k) = c_1\mathbf{T}(\mathbf{x}_1) + c_2\mathbf{T}(\mathbf{x}_2) + \cdots + c_k\mathbf{T}(\mathbf{x}_k)$$

for alle tall  $c_1, c_2, \dots, c_k \in \mathbb{R}$  og alle vektorer  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbb{R}^n$ .

*Bevis:* Vi spalter av ett og ett ledd. Siden vi kan oppfatte  $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k$  som en sum av to ledd  $c_1\mathbf{x}_1$  og  $c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k$ , har vi

$$\begin{aligned} \mathbf{T}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k) &= \mathbf{T}(c_1\mathbf{x}_1) + \mathbf{T}(c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k) = \\ &= c_1\mathbf{T}(\mathbf{x}_1) + \mathbf{T}(c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k) \end{aligned}$$

der vi har brukt de to punktene i definisjonen av lineæravbildning. Vi kan nå spalte av leddet  $c_2\mathbf{x}_2$  på akkurat samme måte, og fortsetter vi slik, står vi til slutt igjen med

$$\mathbf{T}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k) = c_1\mathbf{T}(\mathbf{x}_1) + c_2\mathbf{T}(\mathbf{x}_2) + \cdots + c_k\mathbf{T}(\mathbf{x}_k)$$

□

Det neste resultatet viser oss at lineæravbildninger finnes:

**Setning 2.8.3** *Anta at  $A$  er en  $m \times n$ -matrise. Da er funksjonen  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  definert ved*

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

*en lineæravbildning.*

*Bevis:* Etter regnereglene for matrisemultiplikasjon er  $\mathbf{T}(c\mathbf{x}) = A(c\mathbf{x}) = c(A\mathbf{x}) = c\mathbf{T}(\mathbf{x})$  og  $\mathbf{T}(\mathbf{x} + \mathbf{y}) = A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$ . □

**Eksempel 1:** Hvis  $A$  er  $2 \times 3$ -matrisen

$$A = \begin{pmatrix} 2 & -1 & 3 \\ 1 & -4 & 2 \end{pmatrix},$$

får vi en lineæravbildning  $\mathbf{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  ved

$$\mathbf{T}(x, y, z) = \begin{pmatrix} 2 & -1 & 3 \\ 1 & -4 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x - y + 3z \\ x - 4y + 2z \end{pmatrix}$$

♣

Den neste setningen er nok mer overraskende — den sier at det ikke finnes andre lineæravbildninger enn de som er gitt av matriser!

**Setning 2.8.4** *Anta at  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  er en lineæravbildning. Da finnes det en  $m \times n$ -matrise  $A$  slik at*

$$\mathbf{T}(\mathbf{x}) = A\mathbf{x} \quad \text{for alle } \mathbf{x} \in \mathbb{R}^n$$

*Matrisen  $A$  er gitt ved at den  $j$ -te søylen er lik  $\mathbf{T}(\mathbf{e}_j)$  der  $\mathbf{e}_j$  er den  $j$ -te enhetsvektoren*

$$\mathbf{e}_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \quad \leftarrow j\text{-te komponent}$$

*Bevis:* Vi begynner med å sette navn på komponentene til  $\mathbf{T}(\mathbf{e}_j)$  (som du vil se, er navnvalget inspirert av resultatet vi skal frem til!):

$$\mathbf{T}(\mathbf{e}_j) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{mj} \end{pmatrix}$$

Siden enhver vektor  $\mathbf{x} \in \mathbb{R}^n$  kan uttrykkes ved hjelp av enhetsvektorene på denne måten

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \\ &= x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n, \end{aligned}$$

har vi ifølge setningen ovenfor

$$\begin{aligned} \mathbf{T}(\mathbf{x}) &= \mathbf{T}(x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + \cdots + x_n \mathbf{e}_n) = x_1 \mathbf{T}(\mathbf{e}_1) + x_2 \mathbf{T}(\mathbf{e}_2) + \cdots + x_n \mathbf{T}(\mathbf{e}_n) = \\ &= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = A\mathbf{x} \end{aligned}$$

der

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

Dermed er setningen bevist.  $\square$

Vi skal nå se hvordan vi kan bruke resultatet ovenfor på refleksjoner og rotasjoner i planet. Dette er viktige eksempler i geometri.

**Eksempel 2:** Vi skal nå se på avbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  som avbilder enhver vektor på sitt speilbilde om  $x$ -aksen. Det er lett å overbevise seg om

at  $T$  er en lineæravbildning, og vi skal bruke setningen ovenfor til å finne den tilhørende matrisen. Siden speilingen er om  $x$ -aksen, har vi

$$\mathbf{T}(\mathbf{e}_1) = \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{og} \quad \mathbf{T}(\mathbf{e}_2) = -\mathbf{e}_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Ifølge setningen ovenfor får vi matrisen til  $T$  ved å bruke disse vektorene som søylevektorer:

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

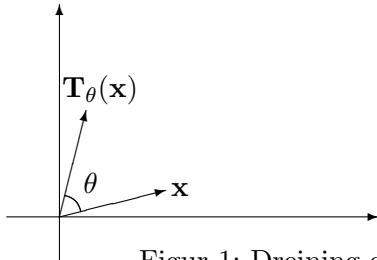
Hvis  $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$ , er dermed

$$\mathbf{T}(\mathbf{x}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}$$

♣

I eksemplet ovenfor er avbildningen så enkel at det ikke ville by på noe problem å skrive opp formelen for  $\mathbf{T}(\mathbf{x})$  direkte uten å gå veien om  $\mathbf{T}(\mathbf{e}_1)$  og  $\mathbf{T}(\mathbf{e}_2)$ . I litt mer kompliserte eksempler blir imidlertid regningene atskillig mer oversiktlige om vi går veien om  $\mathbf{T}(\mathbf{e}_1)$  og  $\mathbf{T}(\mathbf{e}_2)$ .

**Eksempel 3:** Avbildningen  $\mathbf{T}_\theta : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er gitt ved at den dreier enhver vektor  $\mathbf{x}$  en vinkel  $\theta$  i positiv omløpsretning (se figur 1).



Figur 1: Dreining en vinkel  $\theta$

Det er ikke vanskelig å overbevise seg om at  $\mathbf{T}_\theta$  er en lineæravbildning (f.eks. sier regelen  $\mathbf{T}_\theta(\mathbf{x} + \mathbf{y}) = \mathbf{T}_\theta(\mathbf{x}) + \mathbf{T}_\theta(\mathbf{y})$  i dette tilfellet at dersom vi først legger sammen to vektorer  $\mathbf{x}$  og  $\mathbf{y}$ , og så dreier resultatet en vinkel  $\theta$ , så får vi det samme som om vi først dreier begge vektorene en vinkel  $\theta$ , og så legger sammen de nye vektorene). For å finne matrisen til  $\mathbf{T}_\theta$  må vi beregne  $\mathbf{T}_\theta(\mathbf{e}_1)$  og  $\mathbf{T}_\theta(\mathbf{e}_2)$ . Dreier vi enhetsvektoren  $\mathbf{e}_1$  en vinkel  $\theta$ , ender den i punktet  $\begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$  (dette er bare definisjonen av cosinus og sinus til generelle vinkler). Dette betyr altså at

$$\mathbf{T}_\theta(\mathbf{e}_1) = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$$

Siden  $\mathbf{e}_2$  ligger en vinkel  $\pi/2$  foran  $\mathbf{e}_1$  når vi dreier i positiv retning, får vi

$$\mathbf{T}_\theta(\mathbf{e}_2) = \begin{pmatrix} \cos(\theta + \frac{\pi}{2}) \\ \sin(\theta + \frac{\pi}{2}) \end{pmatrix} = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix}$$

(for å få til den siste overgangen kan du enten resonnere geometrisk eller bruke formlene for sinus og cosinus til en sum). Matrisen til lineæravbildningen  $\mathbf{T}_\theta$  er altså

$$A_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

Dersom vi dreier en vektor  $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$  en vinkel  $\theta$  i positiv retning, får vi dermed en ny vektor

$$\mathbf{x}' = A_\theta \mathbf{x} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \cos \theta - y \sin \theta \\ x \sin \theta + y \cos \theta \end{pmatrix}$$

Vi har nå løst den opprinnelige oppgaven vår som var å finne matrisen til lineæravbildningen  $\mathbf{T}_\theta$ . La oss gå litt videre for å se hva som skjer når vi kombinerer to rotasjoner. Dersom vi dreier en annen vinkel  $\phi$ , får vi selvfølgelig matrisen

$$A_\phi = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}$$

La oss nå anta at vi først dreier en vinkel  $\theta$  og deretter en vinkel  $\phi$ . I alt har vi da dreiet en vinkel  $\phi + \theta$  tilsvarende matrisen

$$A_{\phi+\theta} = \begin{pmatrix} \cos(\phi + \theta) & -\sin(\phi + \theta) \\ \sin(\phi + \theta) & \cos(\phi + \theta) \end{pmatrix}$$

På den annen side vet vi at når vi gjør to transformasjoner etter hverandre, svarer dette til å multiplisere de tilhørende matrisene med hverandre. Med andre ord må

$$A_{\phi+\theta} = A_\phi A_\theta$$

Skriver vi ut komponentene, ser vi at

$$A_{\phi+\theta} = \begin{pmatrix} \cos(\phi + \theta) & -\sin(\phi + \theta) \\ \sin(\phi + \theta) & \cos(\phi + \theta) \end{pmatrix}$$

og

$$\begin{aligned} A_\phi A_\theta &= \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \\ &= \begin{pmatrix} \cos \phi \cos \theta - \sin \phi \sin \theta & -\sin \phi \cos \theta - \cos \phi \sin \theta \\ \sin \phi \cos \theta + \cos \phi \sin \theta & \cos \phi \cos \theta - \sin \phi \sin \theta \end{pmatrix} \end{aligned}$$

Sammenligner du komponentene i de to uttrykkene, vil du gjenkjenne formlene for sinus og cosinus til en sum. Vi har altså brukt matriser til å gi et nytt bevis for disse formlene.



### Eigenverdier

Det er ofte et omstendelig arbeid å regne ut  $\mathbf{T}(\mathbf{x})$  for en lineæravbildning  $\mathbf{T}$  og en vektor  $\mathbf{x}$ . For noen vektorer går det imidlertid raskt — alt vi behøver å gjøre, er å gange vektoren med et tall  $\lambda$ . Slike vektorer kalles *eigenvektorer*. Eigenvektorer kan bare finnes når  $\mathbf{T}$  avbilder et rom inn i seg selv, altså når den går fra et rom  $\mathbb{R}^n$  til det *samme* rommet  $\mathbb{R}^n$ . Her er den formelle definisjonen:

**Definisjon 2.8.5** *Anta at  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  er en lineæravbildning. Vi kaller  $\mathbf{x} \neq \mathbf{0}$  en eigenvektor for  $\mathbf{T}$  dersom det finnes et tall  $\lambda$  slik at*

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

Tallet  $\lambda$  kaller vi eigenverdien til  $\mathbf{x}$ .

**Bemerkning:** I definisjonen ovenfor har vi knyttet eigenvektorer til lineæravbildninger, men vi kunne like godt ha knyttet dem til matriser. Vektoren  $\mathbf{x} \neq \mathbf{0}$  er en eigenvektor for matrisen  $A$  dersom det finnes et tall  $\lambda$  slik at  $A\mathbf{x} = \lambda\mathbf{x}$ .

**Eksempel 4:** La oss undersøke om  $\mathbf{a} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$  er en eigenvektor for lineæravbildningen  $\mathbf{T}(\mathbf{x}) = A\mathbf{x}$  der  $A = \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix}$ . Vi har

$$\mathbf{T}(\mathbf{a}) = \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 10 \\ -5 \end{pmatrix} = 5\mathbf{a}$$

så  $\mathbf{a}$  er en eigenvektor med eigenverdi 5.

Gjør vi tilsvarende beregninger for  $\mathbf{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ , får vi

$$\mathbf{T}(\mathbf{b}) = \begin{pmatrix} 1 & -8 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -6 \\ -3 \end{pmatrix} = (-3)\mathbf{b}$$

Altså er  $\mathbf{b}$  en eigenvektor med eigenverdi  $-3$ .

Anta nå at en tredje vektor  $\mathbf{c}$  kan skrives som en lineærkombinasjon av eigenvektorene  $\mathbf{a}$  og  $\mathbf{b}$ , dvs. at det finnes tall  $x$  og  $y$  slik at  $\mathbf{c} = x\mathbf{a} + y\mathbf{b}$  (det viser seg faktisk at alle vektorer i  $\mathbb{R}^2$  kan skrives på denne måten). Da er

$$\mathbf{T}(\mathbf{c}) = \mathbf{T}(x\mathbf{a} + y\mathbf{b}) = x\mathbf{T}(\mathbf{a}) + y\mathbf{T}(\mathbf{b}) = 5x\mathbf{a} - 3y\mathbf{b}$$

Det viser seg altså at ikke bare eigenvektorene selv, men også deres lineærkombinasjoner kan behandles på en enkel måte. 

Observasjonen i slutten av eksemplet ovenfor er viktig. Det viser seg at for “de fleste” lineæravbildninger  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  finnes det egenvektorer  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  slik at *enhver* vektor  $\mathbf{v}$  kan skrives som en lineærkombinasjon

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

Hvis egenverdiene er henholdsvis  $\lambda_1, \lambda_2, \dots, \lambda_n$ , får vi da

$$\begin{aligned}\mathbf{T}(\mathbf{v}) &= \mathbf{T}(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n) = \\ &= x_1\mathbf{T}(\mathbf{v}_1) + x_2\mathbf{T}(\mathbf{v}_2) + \cdots + x_n\mathbf{T}(\mathbf{v}_n) = \\ &= x_1\lambda_1\mathbf{v}_1 + x_2\lambda_2\mathbf{v}_2 + \cdots + x_n\lambda_n\mathbf{v}_n\end{aligned}$$

Bruker vi  $\mathbf{T}$  på begge side av dette uttrykket, får vi på tilsvarende måte

$$\mathbf{T}^2(\mathbf{v}) = \mathbf{T}(\mathbf{T}(\mathbf{v})) = x_1\lambda_1^2\mathbf{v}_1 + x_2\lambda_2^2\mathbf{v}_2 + \cdots + x_n\lambda_n^2\mathbf{v}_n$$

Fortsetter vi på denne måten, får vi generelt

$$\mathbf{T}^k(\mathbf{v}) = x_1\lambda_1^k\mathbf{v}_1 + x_2\lambda_2^k\mathbf{v}_2 + \cdots + x_n\lambda_n^k\mathbf{v}_n$$

Denne formelen forteller oss at dersom vi kjenner egenvektorene og egenverdiene til en lineæravbildning  $\mathbf{T}$ , så har vi god oversikt både over  $\mathbf{T}$  selv og over alle dens potenser. I mange anvendelser forteller  $\mathbf{T}^k(\mathbf{x})$  hvordan et system utvikler seg når vi starter i en tilstand  $\mathbf{x}$  og lar tiden gå ( $\mathbf{T}^1(\mathbf{x})$  er tilstanden etter ett tidsintervall,  $\mathbf{T}^2(\mathbf{x})$  tilstanden etter to tidsintervaller osv.) Formelen ovenfor forteller oss da at veksten til systemet hovedsakelig er bestemt av den største egenverdien (eller, for å være helt presis, den egenverdien som har størst tallverdi).

Vi skal vente til kapittel 4 med å forklare hvordan man kan finne egenverdier i praksis. Det er likevel viktig å vite litt om begrepet og dets anvendelser allerede nå siden det vil gjøre det lettere å forstå hensikten med mye av teorien som kommer senere.

**Bemerkning:** Ser du nøyere på definisjon 2.8.5, vil du se at den er litt upresis når det gjelder hva slags tall egenverdien  $\lambda$  skal være og hva slags vektor egenvektoren  $\mathbf{x}$  skal være. Det viser seg at det finnes reelle matriser som har kompleks egenverdier og egenvektorer. Om man vil “regne med” disse avhenger av problemstillingen man ser på — i noen tilfeller er det nyttig å ha dem med, i andre tilfeller må man ekskludere dem. Situasjonen minner om den vi har for ligninger av  $n$ -te grad; noen ganger er det nyttig å ha med de kompleks løsningene og andre ganger ikke.

**Eksempel 5:** Dersom  $\theta$  ikke er multiplum av  $\pi$ , viser det seg at rotasjonsmatrisene

$$A_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

ikke har reelle egenvektorer og egenverdier (kan du forklare dette geometrisk?) De har imidlertid de komplekse egenvektorene

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -i \end{pmatrix} \quad \text{og} \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ i \end{pmatrix}$$

med tilhørende egenverdier  $\lambda_1 = \cos \theta + i \sin \theta$  og  $\lambda_2 = \cos \theta - i \sin \theta$ . La oss sjekke dette for  $\lambda_1$  og  $\mathbf{v}_1$ :

$$\begin{aligned} A_\theta \mathbf{v}_1 &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} \cos \theta + i \sin \theta \\ \sin \theta - i \cos \theta \end{pmatrix} = \\ &= (\cos \theta + i \sin \theta) \begin{pmatrix} 1 \\ -i \end{pmatrix} = \lambda_1 \mathbf{v}_1 \end{aligned}$$

Du kan på tilsvarende måte sjekke at  $A_\theta \mathbf{v}_2 = \lambda_2 \mathbf{v}_2$ .



### Oppgaver til seksjon 2.8

- 1.** Finn matrisen til lineæravbildningen  $\mathbf{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  gitt ved

$$\mathbf{T}(x, y, z) = \begin{pmatrix} 2x - y + z \\ -x + y - 3z \end{pmatrix}$$

- 2.** En lineæravbildning  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^4$  tilfredsstiller

$$\mathbf{T}(\mathbf{e}_1) = \begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \end{pmatrix} \quad T(\mathbf{e}_2) = \begin{pmatrix} 0 \\ -2 \\ 4 \\ 7 \end{pmatrix}$$

Finn matrisen til  $\mathbf{T}$ .

- 3.** La  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ . Lineæravbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  tilfredsstiller  $\mathbf{T}(\mathbf{a}) = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ ,  $\mathbf{T}(\mathbf{b}) = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$ . Finn  $\mathbf{T}(3\mathbf{a} - 2\mathbf{b})$ .

- 4.** Lineæravbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  avbilder ethvert punkt på sitt speilbilde om  $y$ -aksen. Finn matrisen til  $\mathbf{T}$ .

- 5.** Lineæravbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  fordobler alle annenkomponenter, men endrer ikke førstekomponenter. Finn matrisen til  $\mathbf{T}$ .

- 6.** Lineæravbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  gjør alle vektorer dobbelt så lange og dreier dem en vinkel  $\theta$  i positiv retning. Finn matrisen til  $\mathbf{T}$ .

- 7.** Lineæravbildningen  $\mathbf{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  avbilder alle vektorer på sin projeksjon ned i  $xy$ -planet. Finn matrisen til  $\mathbf{T}$ .

**8.** Lineæravbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  speiler alle vektorer om  $x$ -aksen og dreier dem deretter en vinkel  $\theta$  i positiv retning. Finn matrisen til  $\mathbf{T}$  (det kan være lurt å tenke på  $\mathbf{T}$  som sammensetningen av to enklere avbildninger).

**9.** La  $A_\theta$  være rotasjonsmatrisen i eksempel 3. Forklar at  $A_{-\theta}$  er den inverse matrisen til  $A_\theta$  uten å regne. Kontroller ved å regne ut  $A_\theta A_{-\theta}$ .

**10.**  $\mathcal{L}$  er linjen gjennom origo som fremkommer når vi dreier  $x$ -aksen en vinkel  $\theta$  i positiv retning. Lineæravbildningen  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  speiler alle punkter om linjen  $\mathcal{L}$ . La  $A_\phi$  være matrisen til avbildningen som dreier alle vektorer en vinkel  $\phi$  i positiv retning, og la  $B$  være matrisen til avbildningen som speiler alle punkter om  $x$ -aksen. Forklar at matrisen  $C$  til  $\mathbf{T}$  er gitt ved

$$C = A_\theta B A_{-\theta}$$

og bruk denne formelen til å finne  $C$ .

**11.** La  $\mathbf{a} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$  og  $\mathbf{b} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ .

a) Finn tall  $x, y, z, u$  slik at  $\mathbf{e}_1 = x\mathbf{a} + y\mathbf{b}$  og  $\mathbf{e}_2 = z\mathbf{a} + u\mathbf{b}$ .

b) Lineæravbildningen  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  tilfredsstiller  $T(\mathbf{a}) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $T(\mathbf{b}) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ . Finn  $T(\mathbf{e}_1)$  og  $T(\mathbf{e}_2)$ .

c) Finn matrisen til  $\mathbf{T}$ .

**12.** a) Finn tall  $x, y$  slik at  $\mathbf{e}_1 = x\mathbf{a} + y\mathbf{b}$  der  $\mathbf{a}, \mathbf{b}$  er som i eksempel 4. Finn  $\mathbf{T}^4(\mathbf{e}_1)$  der  $\mathbf{T}$  er lineæravbildningen i eksemplet.

b) Finn tall  $u, v$  slik at  $\mathbf{e}_2 = u\mathbf{a} + v\mathbf{b}$ . Finn  $\mathbf{T}^4(\mathbf{e}_2)$ . Hva er matrisen til  $\mathbf{T}^4$ ?

**13.** Fullfør eksempel 5 ved å sjekke at  $\mathbf{v}_2$  er en egenvektor med egenverdi  $\lambda_2$ .

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

a) Vis at  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  er en egenvektor for  $A$  med egenverdi  $\lambda_1 = 3$ .

b) Vis at  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$  er en egenvektor for  $A$  med egenverdi  $\lambda_1 = -1$ .

c) La  $\mathbf{a} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ . Finn tall  $x, y$  slik at  $\mathbf{a} = x\mathbf{v}_1 + y\mathbf{v}_2$ . Regn ut  $A^{10}\mathbf{a}$ .

**15.** Anta at  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  tilfredsstiller

$$\mathbf{F}(c\mathbf{x} + d\mathbf{y}) = c\mathbf{F}(\mathbf{x}) + d\mathbf{F}(\mathbf{y})$$

for alle  $c, d \in \mathbb{R}$  og alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ . Vis at  $\mathbf{F}$  er en lineæravbildning.

**16.** Anta at  $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^2$  ( $\mathbf{a}_1, \mathbf{a}_2 \neq \mathbf{0}$ ) ikke er parallele, og la  $\mathbf{b}_1, \mathbf{b}_2$  være to vektorer i  $\mathbb{R}^2$ . Vis at de finnes nøyaktig én lineæravbildning  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  slik at  $\mathbf{T}(\mathbf{a}_1) = \mathbf{b}_1$

og  $\mathbf{T}(\mathbf{a}_2) = \mathbf{b}_2$ .

## 2.9 Affinavbildninger og lineariseringer

Som vi så i eksemplene i forrige seksjon, er lineæravbildninger ofte nyttige når vi skal beskrive geometriske transformasjoner som refleksjoner og rotasjoner. De har imidlertid en stor svakhet; siden en lineæravbildning alltid tilfredsstiller  $\mathbf{T}(\mathbf{0}) = \mathbf{0}$ , kan ikke lineæravbildninger brukes til å forskyve figurer i planet. Vi skal nå utvide klassen av avbildninger slik at vi også kan behandle forskyvninger (eller *translasjoner* som matematikere liker å kalle dem).

**Definisjon 2.9.1** En funksjon  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  kalles en *affinavbildning* dersom det finnes en  $m \times n$ -matrise  $A$  og en vektor  $\mathbf{c} \in \mathbb{R}^m$  slik at

$$\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c} \quad \text{for alle } \mathbf{x} \in \mathbb{R}^n$$

Vi kaller  $A$  matrisen til  $\mathbf{F}$  og  $\mathbf{c}$  konstantleddet til  $\mathbf{F}$ .

Vi ser at lineæravbildninger rett og slett er affinavbildninger med konstantledd  $\mathbf{c}$  lik  $\mathbf{0}$ . Vi ser også at translasjonen  $\mathbf{F}(\mathbf{x}) = \mathbf{x} + \mathbf{c}$  som forskyver alle vektorer en distanse  $\mathbf{c}$ , er en affinavbildning siden den kan skrives

$$\mathbf{F}(\mathbf{x}) = I_n \mathbf{x} + \mathbf{c}$$

(husk at  $I_n$  er  $n \times n$ -identitetsmatrisen og at  $I_n \mathbf{x} = \mathbf{x}$  for alle  $\mathbf{x} \in \mathbb{R}^n$ ).

En viktig egenskap ved affinavbildninger er at de avbilder rette linjer på rette linjer. La oss være helt sikre på at vi skjønner hva dette betyr. Anta at  $\mathcal{L}$  er den rette linjen i  $\mathbb{R}^n$  som går gjennom punktet  $\mathbf{a}$  og har retningsvektor  $\mathbf{b}$  (det betyr at  $\mathcal{L}$  er samlingen av alle punkter på formen  $\mathbf{r}(t) = \mathbf{a} + t\mathbf{b}$  slik vi så i seksjon 1.2). Dersom  $\mathbf{F}$  er en kontinuerlig funksjon fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ , kan vi bruke  $\mathbf{F}$  på alle punktene som ligger på linjen  $\mathcal{L}$ . Vi får da en samling av punkter i  $\mathbb{R}^m$  som vi kaller *bildet av  $\mathcal{L}$  under  $\mathbf{F}$* . Vanligvis vil dette bildet være en kurve i  $\mathbb{R}^m$ . Når  $\mathbf{F}$  er en affinavbildning, er denne kurven en rett linje.

**Setning 2.9.2** Anta at  $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$  er en affinavbildning fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ , og la  $\mathbf{r}(t) = \mathbf{a} + t\mathbf{b}$  være parameterfremstillingen til en linje  $\mathcal{L}$  i  $\mathbb{R}^n$ . Dersom  $A\mathbf{b} \neq \mathbf{0}$ , vil bildet av  $\mathcal{L}$  under  $\mathbf{F}$  være linjen i  $\mathbb{R}^m$  som går gjennom punktet  $\mathbf{F}(\mathbf{a})$  og har retningsvektor  $A\mathbf{b}$ .

*Bevis:* Vi ser at

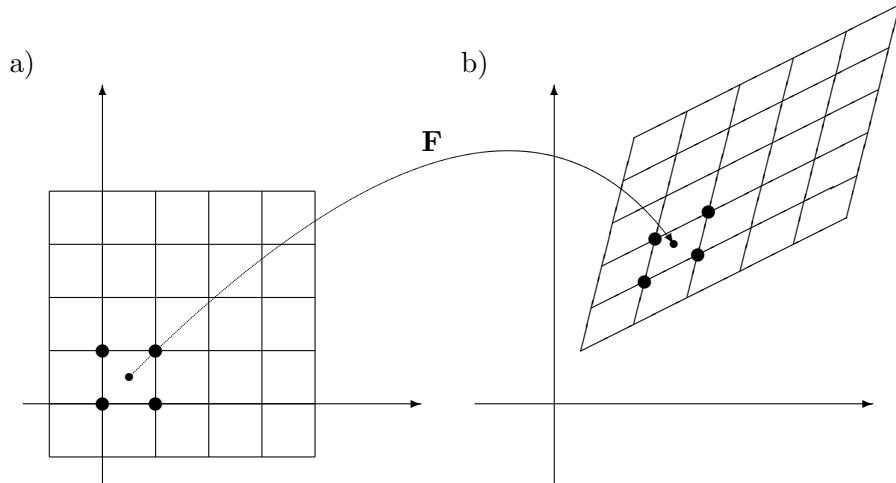
$$\mathbf{F}(\mathbf{r}(t)) = A(\mathbf{a} + t\mathbf{b}) + \mathbf{c} = A\mathbf{a} + \mathbf{c} + t(A\mathbf{b}) = \mathbf{F}(\mathbf{a}) + t(A\mathbf{b})$$

som er parameterfremstillingen til en rett linje om går gjennom punktet  $F(\mathbf{a})$  og har retningsvektor  $A\mathbf{b}$ .  $\square$

**Bemerkning.** Dersom  $A\mathbf{b} = \mathbf{0}$ , ser vi at bildet av  $\mathcal{L}$  degenererer til ett eneste punkt. Legg også merke til at dersom vi bruker  $\mathbf{F}$  på to parallelle linjer (dvs. to linjer med samme retningsvektor  $\mathbf{b}$ ), så blir også de resulterende linjene parallelle (fordi de får samme retningsvektor  $A\mathbf{b}$ ).

### Determinanten som forstørrelsesfaktor

Som vi akkurat har sett, er bildet av en rett linje under en affinavbildning selv en rett linje. Vi vet også at affinavbildninger avbilder parallele linjer på parallele linjer. Figur 1 illustrerer dette for en affinavbildning  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ; rutenettet i punkt a) avbildes på det forskjøvede og fordreide rutenettet i punkt b). Kvadratet med markerte hjørner i a) avbildes på parallelogrammet med markerte hjørner i b).



Figur 1: En affinavbildning  $\mathbf{F}$  anvendt på et rutenett

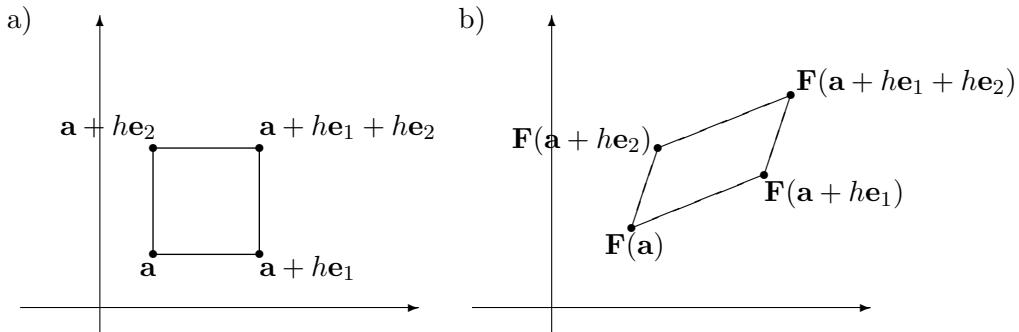
Et spørsmål som ofte dukker opp, er hvor mye avbildningen  $F$  forstørrer eller forminsker arealet: Hvor stort er arealet til parallelogrammet vi ender opp med, sammenlignet med arealet til kvadratet vi startet med? Figur 2 viser et kvadrat før og etter vi har brukt  $\mathbf{F}$  på det. Arealet til kvadratet i punkt a) er  $h^2$ . Parallellogrammet i b) er utspent av vektorene  $\mathbf{F}(\mathbf{a} + h\mathbf{e}_1) - \mathbf{F}(\mathbf{a})$  og  $\mathbf{F}(\mathbf{a} + h\mathbf{e}_2) - \mathbf{F}(\mathbf{a})$ . La oss se nærmere på disse størrelsene.

Siden  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er en affinavbildning, er den på formen

$$\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$$

der  $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$  er en  $2 \times 2$ -matrise og  $\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$  er en vektor i

$\mathbb{R}^2$  (siden det er matrisemultiplikasjon involvert, må vi skrive vektorene som søylevektorer). Vi ser nå at



Figur 2: Bildet av et kvadrat under  $\mathbf{F}$

$$\mathbf{F}(\mathbf{a} + h\mathbf{e}_1) - \mathbf{F}(\mathbf{a}) = (A(\mathbf{a} + h\mathbf{e}_1) + \mathbf{c}) - (A\mathbf{a} + \mathbf{c}) = hA\mathbf{e}_1 = h \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$$

$$\mathbf{F}(\mathbf{a} + h\mathbf{e}_2) - \mathbf{F}(\mathbf{a}) = (A(\mathbf{a} + h\mathbf{e}_2) + \mathbf{c}) - (A\mathbf{a} + \mathbf{c}) = hA\mathbf{e}_2 = h \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

Parallellogrammet er derfor utspent av vektorene  $(ha_{11}, ha_{21})$  og  $(ha_{12}, ha_{22})$ , og har — ifølge setning 1.8.1 — areal

$$\left| \begin{pmatrix} ha_{11} & ha_{21} \\ ha_{12} & ha_{22} \end{pmatrix} \right| = |h^2 a_{11}a_{22} - h^2 a_{12}a_{21}| = h^2 |\det(A)|$$

Vi ser altså at arealet har endret seg med en faktor  $|\det(A)|$ ; tallverdien til determinanten er *forstørrelsesfaktoren* til affinavbildningen  $F$ .

Vi kan gjennomføre akkurat det samme resonnementet i det tredimensjonale tilfellet ved å dele rommet opp i små terninger med sider parallelle med aksene. En affinavbildning  $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$  vil avbilde disse terningene på parallelepiper, og volmet til parallelepipedene vil være  $|\det(A)|$  ganger volumet til terningene. La oss oppsummere resultatene.

**Setning 2.9.3** *Dersom  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er en affinavbildning med matrise  $A$ , så forstørrer  $\mathbf{F}$  arealer med en faktor  $|\det(A)|$ . Dersom  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  er en affinavbildning med matrise  $A$ , så forstørrer  $\mathbf{F}$  volumer med en faktor  $|\det(A)|$ .*  $\square$

**Bemerkning:** Det kan se ut som vi tar i litt vel kraftig i setningen ovenfor — strengt tatt har vi vel bare vist at  $|\det(A)|$  er forstørrelsesfaktoren for arealet til kvadrater, og ikke for mer generelle arealer? Det viser seg imidlertid at alle andre mengder vi kan definere arealet til, kan tilnærmes med kvadrater, og resultatet gjelder derfor generelt. Tilsvarende gjelder for terninger i det tredimensjonale tilfellet. Vi skal komme grundigere tilbake til disse spørsmålene i kapittel 6.

Det er naturlig å spørre om fortegnet til determinanten har en geometrisk tolkning også for avbildninger. Det har det: Fortegnet er positivt dersom  $\mathbf{F}$  bevarer orienteringen til enhetsvektorene  $\mathbf{i}$  og  $\mathbf{j}$  ( $\mathbf{i}, \mathbf{j}$  og  $\mathbf{k}$  i det tredimensjonale tilfellet) og negativt om  $\mathbf{F}$  reverserer orienteringen.

### Linearisering

Til slutt i denne seksjonen skal vi se hvordan vi kan bruke affinavbildninger til å generalisere tangentbegrepet fra funksjoner  $f : \mathbb{R} \rightarrow \mathbb{R}$  til funksjoner  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Vi skal først se på Jacobi-matrisen til en affinavbildning  $\mathbf{F}$ . Anta at  $\mathbf{F}$  er gitt ved  $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$  der

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

og

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}$$

Skriver vi ut definisjonen  $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$  på komponentform, får vi

$$\begin{aligned} \mathbf{F}(\mathbf{x}) &= \begin{pmatrix} F_1(\mathbf{x}) \\ F_2(\mathbf{x}) \\ \vdots \\ F_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix} = \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n + c_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n + c_m \end{pmatrix} \end{aligned}$$

Deriverer vi den  $i$ -te komponenten  $F_i$  med hensyn på den  $j$ -te variabelen  $x_j$ , får vi

$$\frac{\partial F_i}{\partial x_j} = \frac{\partial}{\partial x_j}(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n + c_i) = a_{ij}$$

Dette betyr at Jacobi-matrisen til en affinavbildning  $\mathbf{F}$  rett og slett er matrisen  $A$  til  $\mathbf{F}$ .

**Setning 2.9.4** *Anta at affinavbildningen  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  er gitt ved  $\mathbf{F}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$ . Da er Jacobi-matrisen til  $\mathbf{F}$  lik matrisen  $A$  til  $\mathbf{F}$ .*  $\square$

Vi er nå klare til å studere hvordan affinavbildninger kan brukes til å tilnærme mer generelle funksjoner. Husk at hvis  $f : \mathbb{R} \rightarrow \mathbb{R}$  er en funksjon av én variabel, er tangenten  $T_a f$  i punktet  $a$  gitt ved

$$T_a f(x) = f(a) + f'(a)(x - a)$$

Starter vi isteden med en funksjon  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , kan vi på tilsvarende måte definere en funksjon  $T_{\mathbf{a}} \mathbf{F}$  ved å etterligne uttrykket ovenfor:

$$T_{\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

Ganger vi ut parentesen, får vi

$$T_{\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a} + \mathbf{F}'(\mathbf{a})\mathbf{x}$$

som viser at  $T_{\mathbf{a}} \mathbf{F}$  er en affinavbildung med matrise  $\mathbf{F}'(\mathbf{a})$  og konstantledd  $\mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a}$ .

**Definisjon 2.9.5** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av  $n$  variable som er deriverbar i punktet  $\mathbf{a}$ . Affinavbildunga  $T_{\mathbf{a}} \mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  gitt ved*

$$T_{\mathbf{a}} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

kalles lineariseringen til  $\mathbf{F}$  i punktet  $\mathbf{a}$ .

Legg merke til at  $T_{\mathbf{a}} \mathbf{F}(\mathbf{a}) = \mathbf{F}(\mathbf{a})$ . Siden vi allerede har observert at matrisen til affinavbildunga  $T_{\mathbf{a}} \mathbf{F}$  er  $\mathbf{F}'(\mathbf{a})$ , vet vi fra setning 2.9.4 at Jacobi-matrisen til  $T_{\mathbf{a}} \mathbf{F}$  er  $\mathbf{F}'(\mathbf{a})$ . Dette betyr at lineariseringa  $T_{\mathbf{a}} \mathbf{F}$  har samme verdi og samme deriverte i punktet  $\mathbf{a}$  som den opprinnelige funksjonen  $\mathbf{F}$ . Dette er akkurat samme egenskap som tangenten har for funksjoner av én variabel. En annen viktig (og nært beslektet) egenskap ved tangenten er at den er linjen som smyger seg tøttest inntil funksjonsgrafen til  $f$  i nærheten av punktet  $a$ . Vi skal smart vise en tilsvarende egenskap for lineariseringen — at den er affinavbildunga som ligger tøttest opptil  $\mathbf{F}$  i nærheten av  $\mathbf{a}$ . Men før vi gjør dette, kan det være lurt å kikke på et eksempel.

**Eksempel 1:** Vi skal finne lineariseringen til funksjonen  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  gitt ved

$$\mathbf{F}(x, y, z) = \begin{pmatrix} x^2 y \\ 2xyz \end{pmatrix}$$

i punktet  $\mathbf{a} = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}$ . La oss først regne ut Jacobi-matrisen til  $\mathbf{F}$ :

$$\mathbf{F}'(x, y, z) = \begin{pmatrix} \frac{\partial F_1}{\partial x} & \frac{\partial F_1}{\partial y} & \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} & \frac{\partial F_2}{\partial y} & \frac{\partial F_2}{\partial z} \end{pmatrix} = \begin{pmatrix} 2xy & x^2 & 0 \\ 2yz & 2xz & 2xy \end{pmatrix}$$

I punktet  $\mathbf{a}$  har vi dermed

$$\begin{aligned}\mathbf{F}'(1, -2, -1) &= \begin{pmatrix} 2 \cdot 1 \cdot (-2) & 1^2 & 0 \\ 2 \cdot (-2) \cdot (-1) & 2 \cdot 1 \cdot (-1) & 2 \cdot 1 \cdot (-2) \end{pmatrix} = \\ &= \begin{pmatrix} -4 & 1 & 0 \\ 4 & -2 & -4 \end{pmatrix}\end{aligned}$$

Vi har også

$$\mathbf{F}(1, -2, -1) = \begin{pmatrix} 1^2 \cdot (-2) \\ 2 \cdot 1 \cdot (-2) \cdot (-1) \end{pmatrix} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$$

Vi kan nå regne ut lineariseringen. Med  $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$  får vi:

$$\begin{aligned}T_{\mathbf{a}}\mathbf{F}(\mathbf{x}) &= \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a}) = \\ &= \begin{pmatrix} -2 \\ 4 \end{pmatrix} + \begin{pmatrix} -4 & 1 & 0 \\ 4 & -2 & -4 \end{pmatrix} \left( \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix} \right) = \\ &= \begin{pmatrix} -2 \\ 4 \end{pmatrix} + \begin{pmatrix} -4x + y \\ 4x - 2y - 4z \end{pmatrix} - \begin{pmatrix} -6 \\ 12 \end{pmatrix} = \begin{pmatrix} -4x + y + 4 \\ 4x - 2y - 4z - 8 \end{pmatrix}\end{aligned}$$

Altså er

$$T_{\mathbf{a}}\mathbf{F}(x, y, z) = \begin{pmatrix} -4x + y + 4 \\ 4x - 2y - 4z - 8 \end{pmatrix}$$



La oss vende tilbake til resultatet vi annonserte ovenfor, det som sier at lineariseringen  $T_{\mathbf{a}}\mathbf{F}$  er den affinavbildningen som ligger tettest opp til  $\mathbf{F}$  i nærheten av  $\mathbf{a}$ . Et punkt  $\mathbf{x}$  i nærheten av  $\mathbf{a}$  kan vi skrive  $\mathbf{x} = \mathbf{a} + \mathbf{r}$  der  $\mathbf{r}$  er liten. Vi ønsker å vise at differensen  $\mathbf{F}(\mathbf{a} + \mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r})$  er liten sammenlignet med  $|\mathbf{r}|$  for små  $\mathbf{r}$ , og at tilnærmingen blir bedre og bedre når  $\mathbf{r}$  går mot  $\mathbf{0}$ . En naturlig måte å formulere dette på er å kreve at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r}) \right) = \mathbf{0}$$

Bruker vi at

$$T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})((\mathbf{a} + \mathbf{r}) - \mathbf{a}) = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})\mathbf{r}$$

ser vi at

$$\mathbf{F}(\mathbf{a} + \mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r}) = \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{r} = \boldsymbol{\sigma}(\mathbf{r})$$

der  $\boldsymbol{\sigma}$  er som i definisjon 2.6.1. Nå er det ikke så vanskelig å formulere og bevise resultatet.

**Teorem 2.9.6** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av  $n$  variable som er deriverbar i punktet  $\mathbf{a}$ , og la  $T_{\mathbf{a}}\mathbf{F}$  være lineariseringen til  $\mathbf{F}$  i  $\mathbf{a}$ . Da er*

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r}) \right) = \mathbf{0}$$

*Det finnes ingen annen affinavbildning  $\mathbf{G} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  slik at*

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{G}(\mathbf{a} + \mathbf{r}) \right) = \mathbf{0}$$

*Bevis:* Det følger av definisjonen av deriverbarhet (definisjon 2.6.1) og regningene ovenfor at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - T_{\mathbf{a}}\mathbf{F}(\mathbf{a} + \mathbf{r}) \right) = \lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \boldsymbol{\sigma}(\mathbf{r}) = \mathbf{0}$$

Dermed er første del av teoremet bevist. For å bevise den andre delen, antar vi at  $\mathbf{G}(\mathbf{x})$  er en affinavbildning slik at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{G}(\mathbf{a} + \mathbf{r}) \right) = \mathbf{0}$$

Vi må vise at  $\mathbf{G} = T_{\mathbf{a}}\mathbf{F}$ . Siden  $\mathbf{G}$  er en affinavbildning, vet vi at  $\mathbf{G}(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$  for en matrise  $A$  og en vektor  $\mathbf{c}$ . Setter vi dette inn i uttrykket ovenfor, får vi

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{G}(\mathbf{a} + \mathbf{r}) \right) = \lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - A\mathbf{a} - Ar - \mathbf{c} \right) = \mathbf{0}$$

Skal dette være mulig, må  $\mathbf{F}(\mathbf{a} + \mathbf{r}) - A\mathbf{a} - Ar - \mathbf{c} \rightarrow \mathbf{0}$  når  $\mathbf{r} \rightarrow \mathbf{0}$ , og det er bare tilfellet hvis  $\mathbf{F}(\mathbf{a}) = A\mathbf{a} + \mathbf{c}$ . Setter vi  $A\mathbf{a} + \mathbf{c} = \mathbf{F}(\mathbf{a})$  inn i det siste uttrykket ovenfor, ser vi at

$$\lim_{\mathbf{r} \rightarrow \mathbf{0}} \frac{1}{|\mathbf{r}|} \left( \mathbf{F}(\mathbf{a} + \mathbf{r}) - \mathbf{F}(\mathbf{a}) - Ar \right) = \mathbf{0}$$

og ifølge setning 2.6.4 er det bare mulig dersom  $A = \mathbf{F}'(\mathbf{a})$ . Siden vi allerede vet at  $\mathbf{F}(\mathbf{a}) = A\mathbf{a} + \mathbf{c}$ , følger det at  $\mathbf{c} = \mathbf{F}(\mathbf{a}) - A\mathbf{a} = \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a}$ . Dermed er

$$\mathbf{G}(\mathbf{x}) = A\mathbf{x} + \mathbf{c} = \mathbf{F}'(\mathbf{a})\mathbf{x} + \mathbf{F}(\mathbf{a}) - \mathbf{F}'(\mathbf{a})\mathbf{a} = \mathbf{F}(\mathbf{a}) + \mathbf{F}'(\mathbf{a})(\mathbf{x} - \mathbf{a}) = T_{\mathbf{a}}\mathbf{F}(\mathbf{x})$$

□

Det som er viktigst å ta med seg fra teoremet ovenfor, er at lineariseringen  $T_{\mathbf{a}}\mathbf{F}$  er en svært god tilnærming til  $\mathbf{F}$  i området rundt  $\mathbf{a}$ . Dette skal vi få bruk for både når vi studerer Newtons metode i flere variable i kapittel

5, og når vi studerer skifte av variable i multiple integraler (dvs. integraler av funksjoner av flere variable) i kapittel 6.

**Bemerkning:** Du har kanskje lurt på hvorfor  $T_{\mathbf{a}}\mathbf{F}$  kalles en *linearisering* og ikke en *affinisering* —  $T_{\mathbf{a}}\mathbf{F}$  er tross alt en affinavbildning og ikke (bortsett fra i helt spesielle tilfeller) en lineæravbildning. Dessverre er bruken av ordparet lineær/affin svært usystematisk og forvirrende; ordene bruktes om hverandre på en måte som ikke er lett å få oversikt over. Et annet eksempel på denne vaklingen får du om du betrakter en funksjon  $f : \mathbb{R} \rightarrow \mathbb{R}$  på formen  $f(x) = ax + c$ . Når man studerer funksjoner av én variabel, er alle enige om å kalte dette en lineær funksjon, men ifølge terminologien vi nå har innført, er den bare en lineæravbildning hvis  $c = 0$ ; i alle andre tilfeller må vi nøye oss med å kalte den en affinavbildning! Heldigvis er det sjeldent at den vakkende terminologibruken skaper misforståelser, men det er greit å vite om den slik at man kan være litt på vakt.

### Oppgaver til seksjon 2.9

1. Finn matrisen og konstantleddet til affinavbildningen

$$\mathbf{F}(x, y, z) = \begin{pmatrix} 2x - 3y + z - 7 \\ -x + z - 2 \end{pmatrix}$$

2.  $\mathcal{L}$  er linjen i  $\mathbb{R}^3$  med parametrisering

$$\mathbf{r}(t) = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} + t \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

og  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  er affinavbildningen

$$\mathbf{F}(x, y, z) = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 3 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

Finn en parametrisering av bildet av  $\mathcal{L}$  under  $\mathbf{F}$ .

3. Finn lineariseringen til funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} x^2 y \\ x y + x \end{pmatrix}$$

i punktet  $\mathbf{a} = (-2, 1)$ .

4. Finn lineariseringen til funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} x \sin(xy) \\ x e^y \\ 2x^3 + y \end{pmatrix}$$

i punktet  $\mathbf{a} = (2, 0)$ .

**5.** En affinavbildning  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  tilfredsstiller  $\mathbf{F}(0, 0) = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ,  $\mathbf{F}(1, 0) = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$  og  $\mathbf{F}(0, 1) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ . Finn matrisen og konstantleddet til  $\mathbf{F}$ .

**6.** Affinavbildningen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  dreier enhver vektor en vinkel  $\frac{\pi}{4}$  i positiv retning, og flytter den deretter en distanse  $(3, -1)$ . Finn matrisen og konstantleddet til  $\mathbf{F}$ .

**7.** a) Affinavbildningen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  avbilder ethvert punkt på sitt speilbilde om den vertikale linjen  $x = 3$ . Finn matrisen og konstantleddet til  $\mathbf{F}$ .

b) Affinavbildningen  $\mathbf{G} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  avbilder ethvert punkt på sitt speilbilde om den horisontale linjen  $y = -2$ . Finn matrisen og konstantleddet til  $\mathbf{G}$ .

**8.** Affinavbildningen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  avbilder ethvert punkt på sitt speilbilde om linjen  $y = x + 1$ . Finn matrisen og konstantleddet til  $\mathbf{F}$ .

**9.** Anta at punktene  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^2$  ikke ligger på samme rette linje, og la  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  være tre punkter i  $\mathbb{R}^2$ . Vis at det finnes nøyaktig én affinavbildning  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  slik at  $\mathbf{F}(\mathbf{a}_1) = \mathbf{b}_1$ ,  $\mathbf{F}(\mathbf{a}_2) = \mathbf{b}_2$  og  $\mathbf{F}(\mathbf{a}_3) = \mathbf{b}_3$ .



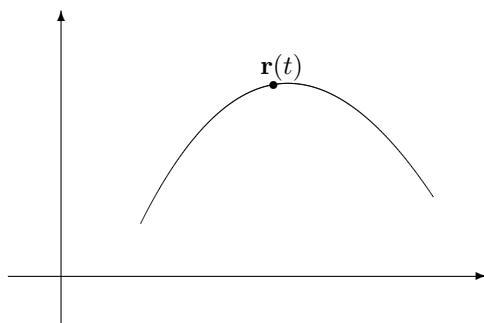
# Kapittel 3

## Kurver og flater

I dette kapitlet skal vi benytte teorien vi utviklet i kapittel 1 og 2 til å studere kurver og flater. Kapitlet inneholder dessuten en del anvendelser i fysikk. Disse kan nok virke litt skremmende hvis du ikke har tatt fysikk i videregående skole, men det er absolutt verdt å bruke litt tid på dem. Teorien vi skal se på, ble i stor grad utviklet for å beskrive fysiske fenomener, og det er ingen andre anvendelser som i samme grad illustrerer slagkraften i resultatene.

### 3.1 Parametriserte kurver

I videregående skole har du lært om parametriserte kurver i planet og rommet. Nå skal vi gå et skritt videre og studere parametriserte kurver i  $\mathbb{R}^n$ . Vi skal også se på noen viktige anvendelser som du sannsynligvis ikke har vært borti tidligere.



Figur 1: Et punkt  $\mathbf{r}(t)$  beveger seg langs en kurve

La oss begynne med litt repetisjon. En parametrisert kurve i planet er rett og slett en funksjon  $\mathbf{r}(t) = (x(t), y(t))$ . Ofte (men ikke bestandig) er det lurt å tenke på  $t$  som tiden, og  $\mathbf{r}(t)$  som posisjonen ved tiden  $t$  til et punkt som beveger seg langs en kurve (se figur 1). På tilsvarende måte er en

parametrisert kurve i  $\mathbb{R}^3$  en funksjon  $\mathbf{r}(t) = (x(t), y(t), z(t))$  som beskriver hvordan et punkt beveger seg langs en kurve i rommet.

**Eksempel 1:** Parametriseringen

$$\mathbf{r}(t) = (\cos t, \sin t) \quad t \in [0, \pi]$$

fremstiller en halvsirkel om origo med radius 1. Startpunktet er  $(1, 0)$  og slutt punktet  $(-1, 0)$ . En annen parametrisering av samme kurve får vi ved å sette

$$\mathbf{r}(t) = (-t, \sqrt{1 - t^2}) \quad t \in [-1, 1] \quad \clubsuit$$

La oss også se på en parametrisert kurve i rommet.

**Eksempel 2:** Vi setter

$$\mathbf{r}(t) = (\cos t, \sin t, t) \quad t \in [0, \infty)$$

Denne kurven ser ut som en spiralfjær. Hadde vi bare hatt de to første koordinatene, ville kurven ha beskrevet en sirkulær bevegelse i planet akkurat som i eksemplet ovenfor, men den tredje koordinaten trekker kurven oppover i en spiralbevegelse. Prøv å finne ut hvordan kurven ser ut og tegn en skisse av den. ♣

Vi er nå klar for den formelle definisjonen av en parametrisert kurve i  $\mathbb{R}^n$ .

**Definisjon 3.1.1** En parametrisert kurve i  $\mathbb{R}^n$  er en kontinuerlig funksjon  $\mathbf{r} : I \rightarrow \mathbb{R}^n$  der  $I \subset \mathbb{R}$  er et intervall. Vi skriver ofte funksjonen på komponentform

$$\mathbf{r}(t) = (x_1(t), x_2(t), \dots, x_n(t))$$

Parametriserte kurver kalles også vektorvaluerte funksjoner.

Husk fra kapittel 2 (setning 2.2.4) at funksjonen  $\mathbf{r}$  er kontinuerlig hvis og bare hvis hver komponent  $x_1, x_2, \dots, x_n$  er kontinuerlig.

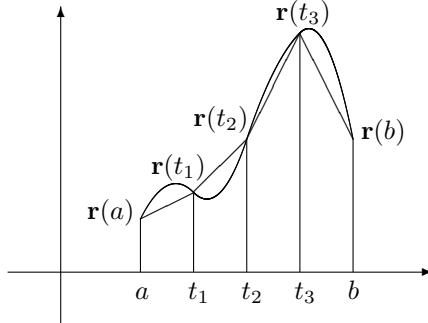
Selv om “parametrisert kurve” i utgangspunktet er et geometrisk begrep, er det mange anvendelser som ikke har så mye med geometri å gjøre.

**Eksempel 3:** Et grossistfirma fører  $n$  vareslag. Hvis  $x_i(t)$  er lagerbeholdningen av vareslag  $i$  ved tiden  $t$ , kan hele varebeholdningen beskrives ved

$$\mathbf{r}(t) = (x_1(t), x_2(t), \dots, x_n(t))$$

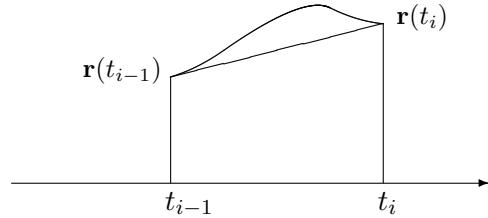
I dette tilfellet faller det mer naturlig å kalle  $\mathbf{r}$  en “vektorvaluert funksjon” enn en “parametrisert kurve”. ♣

Et naturlig spørsmål er hvordan man finner lengden til en parametrisert kurve. Dersom vi beveger oss langs den parametriserte kurven fra  $t = a$  til  $t = b$ , hvor langt har vi da gått?



Figur 2: Tilnærming til buelengden

Figur 2 viser en naturlig måte å nærme seg problemet på — vi deler intervallet  $[a, b]$  med delepunkter  $a = t_0 < t_1 < t_2 \dots < t_N = b$  og regner ut lengden til den brudne kurven fra  $\mathbf{r}(a)$  til  $\mathbf{r}(t_1)$  videre til  $\mathbf{r}(t_2)$  osv. inntil vi når  $\mathbf{r}(b)$ . Lar vi oppdelingen bli finere og finere, er det naturlig å tenke seg at lengde til den brudne kurven nærmer seg lengden til den opprinnelige kurven. Vår oppgave blir derfor å finne et matematisk uttrykk for denne grenseverdien.



Figur 3: Tilnærming til buelengden

Figur 3 viser et nærbilde av den  $i$ -te delen av en slik kurve fra  $\mathbf{r}(t_{i-1})$  til  $\mathbf{r}(t_i)$ . Lengden av denne delen av den brudne kurven er gitt ved

$$|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})| = \sqrt{(x_1(t_i) - x_1(t_{i-1}))^2 + \dots + (x_n(t_i) - x_n(t_{i-1}))^2}$$

Hvis vi ganger og deler med  $(t_i - t_{i-1})$ , ser vi at dette uttrykket er lik

$$\begin{aligned} & \sqrt{\left(\frac{x_1(t_i) - x_1(t_{i-1})}{t_i - t_{i-1}}\right)^2 + \dots + \left(\frac{x_n(t_i) - x_n(t_{i-1})}{t_i - t_{i-1}}\right)^2} (t_i - t_{i-1}) \approx \\ & \approx \sqrt{x'_1(t_i)^2 + \dots + x'_n(t_i)^2} (t_i - t_{i-1}) \end{aligned}$$

Den totale lengden til den brudne veien er derfor

$$\begin{aligned} \sum_{i=1}^N \sqrt{\left(\frac{x_1(t_i) - x_1(t_{i-1})}{t_i - t_{i-1}}\right)^2 + \cdots + \left(\frac{x_n(t_i) - x_n(t_{i-1})}{t_i - t_{i-1}}\right)^2} (t_i - t_{i-1}) &\approx \\ \approx \sum_{i=1}^N \sqrt{x'_1(t_i)^2 + \cdots + x'_n(t_i)^2} (t_i - t_{i-1}) \end{aligned}$$

(forutsatt at  $x_1, x_2, \dots, x_n$  er deriverbare). Dette er en Riemann-sum for funksjonen  $\sqrt{x'_1(t)^2 + \cdots + x'_n(t)^2}$ , og gjør vi oppdelingen av intervallet finere og finere, nærmer dette uttrykket seg derfor

$$\int_a^b \sqrt{x'_1(t)^2 + \cdots + x'_n(t)^2} dt$$

Følgende definisjon er derfor fornuftig:

**Definisjon 3.1.2** *Anta at funksjonene  $x_1, x_2, \dots, x_n$  er deriverbare med kontinuerlige deriverte. Da er buelengden til den parametriserte kurven  $\mathbf{r}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  fra  $a$  til  $b$*

$$L(a, b) = \int_a^b \sqrt{x'_1(t)^2 + x'_2(t)^2 + \cdots + x'_n(t)^2} dt$$

**Bemerkning:** Definisjonen ovenfor forutsetter at komponentfunksjonene  $x_1, x_2, \dots, x_n$  har kontinuerlige deriverte. Fra et teoretisk synspunkt er dette en svakhet — vi burde også være i stand til å måle lengden til en kurve som ikke har kontinuerlige deriverte. Oppgave 21 skisserer en mer generell måte å angripe problemet på.

**Eksempel 4:** La oss bruke formelen ovenfor til å regne ut omkretsen til en sirkel. Bruker vi parametriseringen

$$\mathbf{r}(t) = (x(t), y(t)) = (\cos t, \sin t) \quad t \in [0, 2\pi]$$

får vi  $x'(t) = -\sin t$  og  $y'(t) = \cos t$ . Dette gir

$$\begin{aligned} L(0, 2\pi) &= \int_0^{2\pi} \sqrt{x'(t)^2 + y'(t)^2} dt \\ &= \int_0^{2\pi} \sqrt{(-\sin t)^2 + (\cos t)^2} dt = \int_0^{2\pi} 1 dt = 2\pi \end{aligned}$$



På tilsvarende måte kan vi regne ut buelengden til en omdreining av spiralen i eksempel 2:

**Eksempel 5:** Siden

$$\mathbf{r}(t) = (x(t), y(t), z(t)) = (\cos t, \sin t, t),$$

har vi

$$x'(t) = -\sin t \quad y'(t) = \cos t \quad z'(t) = 1$$

Dette gir

$$\begin{aligned} L(0, 2\pi) &= \int_0^{2\pi} \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} dt \\ &= \int_0^{2\pi} \sqrt{(-\sin t)^2 + (\cos t)^2 + 1^2} dt = \int_0^{2\pi} \sqrt{2} dt = 2\sqrt{2}\pi \end{aligned}$$



**Bemerkning:** Den samme kurven kan parametriseres på mange måter (i eksempel 1 så vi to måter å parametrisere en halvsirkel på), og geometriske egenskaper slik som buelengde burde være uavhengig av hvilken parametrisering vi velger. I teorien for kurver finnes det derfor en rekke resultater som sier at forskjellige geometriske egenskaper er uavhengige av parametriseringen. Vi skal ikke komme nærmere inn på dette her, men har tatt med noen eksempler i oppgavene.

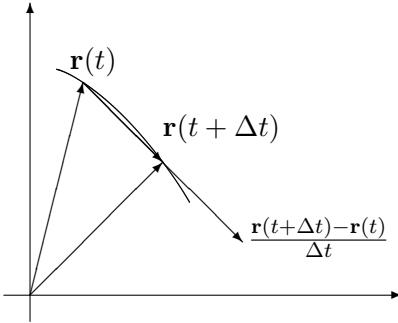
Vi skal nå se hvordan vi kan finne farten til gjenstanden vår. Fra tiden  $a$  til tiden  $t$  har den tilbakelagt en strekning  $s(t)$  gitt ved  $s(t) = L(a, t) = \int_a^t \sqrt{x'_1(r)^2 + x'_2(r)^2 + \dots + x'_n(r)^2} dr$ . Deriverer vi dette uttrykket med hensyn på  $t$ , får vi (ifølge analysens fundamentalteorem)

$$s'(t) = \sqrt{x'_1(t)^2 + x'_2(t)^2 + \dots + x'_n(t)^2}$$

Siden farten er den deriverte av strekningen med hensyn på tiden, forteller dette oss at farten til gjenstanden vår ved tiden  $t$  er

$$v(t) = s'(t) = \sqrt{x'_1(t)^2 + x'_2(t)^2 + \dots + x'_n(t)^2}$$

Formelen ovenfor gir oss farten ved tidspunktet  $t$  som en skalar størrelse  $v(t)$ . I mange sammenhenger er vi ikke bare interessert i hvor fort en gjenstand beveger seg, men også hvilken retning den beveger seg i. Vi er altså interessert i å oppfatte hastigheten som en vektor som har både størrelse og retning. La oss tenke gjennom problemet fra begynnelsen av.



Figur 4: Tilnærming til tangenten

I løpet av et lite tidsintervall  $[t, t + \Delta t]$  vil partikkelen flytte seg fra  $\mathbf{r}(t)$  til  $\mathbf{r}(t + \Delta t)$ . Forflytningen er altså  $\mathbf{r}(t + \Delta t) - \mathbf{r}(t)$  og den gjennomsnittlige forflytningen per tidsenhet er  $(\mathbf{r}(t + \Delta t) - \mathbf{r}(t))/\Delta t$ . Dersom  $\Delta t$  er liten, vil vektoren  $\mathbf{r}(t + \Delta t) - \mathbf{r}(t)$  (og dermed den parallele vektoren  $(\mathbf{r}(t + \Delta t) - \mathbf{r}(t))/\Delta t$ ) nesten være en tangent til kurven, og denne tilnærmingen blir bedre og bedre dess mindre  $\Delta t$  er (se figur 4).

Den deriverte vektoren

$$\begin{aligned} \mathbf{r}'(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \left( \frac{x_1(t + \Delta t) - x_1(t)}{\Delta t}, \frac{x_2(t + \Delta t) - x_2(t)}{\Delta t}, \dots, \frac{x_n(t + \Delta t) - x_n(t)}{\Delta t} \right) = \\ &= (x'_1(t), x'_2(t), \dots, x'_n(t)) \end{aligned}$$

blir derfor både en tangent til kurven og en beskrivelse av forflytning per tidsenhet ved tidspunktet  $t$ . Dette er motivasjonen for følgende definisjon.

**Definisjon 3.1.3** *Anta at funksjonene  $x_1, x_2, \dots, x_n$  er deriverbare i punktet  $t$ . Da sier vi at den parametriserte kurven  $\mathbf{r}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  er deriverbar i  $t$ , og at den deriverte er*

$$\mathbf{v}(t) = \mathbf{r}'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))$$

*I situasjoner der  $\mathbf{r}(t)$  representerer posisjonen til en gjenstand ved tiden  $t$ , kaller vi  $\mathbf{v}(t)$  for hastigheten til gjenstanden.*

**Bemerkning:** Legg merke til terminologien — vektorstørrelsen  $\mathbf{v}(t)$  kaller vi *hastigheten*, mens tallet (skalaren)  $v(t)$  kaller vi *farten*. Dette pleide å være standard terminologi i matematikk- og fysikkbøker, men i senere år har det blitt mer og mer vanlig å bruke betegnelsen *fart* om både vektorstørrelsen  $\mathbf{r}(t)$  og skalartørrelsen  $v(t)$ . Jeg synes den gamle terminologien er oversiktlig og grei og holder derfor fast på den i dette kompendiet (det er en tilsvarende distinksjon på engelsk; *farten* kalles “speed” og *hastigheten*

kalles “velocity”). Legg forøvrig merke til at  $|\mathbf{v}(t)| = v(t)$  slik at farten er lengden til hastighetsvektoren.

Det er på tide med et eksempel:

**Eksempel 6:** Finn hastigheten og farten til den parametriserte kurven  $\mathbf{r}(t) = (t \cos t, t \sin t)$  (prøv å finne ut hvordan denne kurven ser ut!). Vi får

$$\mathbf{v}(t) = \mathbf{r}'(t) = ((t \sin t)', (t \cos t)') = (\sin t + t \cos t, \cos t - t \sin t)$$

Farten blir

$$\begin{aligned} v(t) &= |\mathbf{v}(t)| = \sqrt{(\sin t + t \cos t)^2 + (\cos t - t \sin t)^2} \\ &= \sqrt{\sin^2 t + 2t \sin t \cos t + t^2 \cos^2 t + \cos^2 t - 2t \sin t \cos t + t^2 \sin^2 t} \\ &= \sqrt{1 + t^2} \end{aligned}$$



De vanlige derivasjonsreglene gjelder også for derivasjon av vektorvaluerte funksjoner.

**Setning 3.1.4** Dersom  $\mathbf{r}_1(t)$  og  $\mathbf{r}_2(t)$  er to deriverbare parametriserte kurver, gjelder:

- (i)  $(\mathbf{r}_1(t) + \mathbf{r}_2(t))' = \mathbf{r}'_1(t) + \mathbf{r}'_2(t)$
- (ii)  $(\mathbf{r}_1(t) - \mathbf{r}_2(t))' = \mathbf{r}'_1(t) - \mathbf{r}'_2(t)$
- (iii)  $(\mathbf{r}_1(t) \cdot \mathbf{r}_2(t))' = \mathbf{r}'_1(t) \cdot \mathbf{r}_2(t) + \mathbf{r}_1(t) \cdot \mathbf{r}'_2(t)$
- (iv)  $(\mathbf{r}_1(t) \times \mathbf{r}_2(t))' = \mathbf{r}'_1(t) \times \mathbf{r}_2(t) + \mathbf{r}_1(t) \times \mathbf{r}'_2(t)$
- (v) Dersom  $\mathbf{r}(t)$  er en deriverbar parametrisert kurve og  $u(t)$  er en deriverbar funksjon, er  $(u(t)\mathbf{r}(t))' = u'(t)\mathbf{r}(t) + u(t)\mathbf{r}'(t)$

*Bevis:* Vi tar (iii) som et eksempel (de andre bevises på lignende måte). Dersom

$$\mathbf{r}_1(t) = (x_1(t), x_2(t), \dots, x_n(t))$$

og

$$\mathbf{r}_2(t) = (y_1(t), y_2(t), \dots, y_n(t))$$

så er  $\mathbf{r}_1(t) \cdot \mathbf{r}_2(t) = x_1(t)y_1(t) + x_2(t)y_2(t) + \dots + x_n(t)y_n(t)$ . Bruker vi den vanlige regelen for den deriverte til et produkt (husk at  $x_1, x_2, \dots, x_n$  og  $y_1, y_2, \dots, y_n$  er vanlige funksjoner fra  $\mathbb{R}$  til  $\mathbb{R}$ ), får vi:

$$\begin{aligned} (\mathbf{r}_1(t) \cdot \mathbf{r}_2(t))' &= (x_1(t)y_1(t) + x_2(t)y_2(t) + \dots + x_n(t)y_n(t))' \\ &= x'_1(t)y_1(t) + x_1(t)y'_1(t) + x'_2(t)y_2(t) + x_2(t)y'_2(t) + \dots + x'_n(t)y_n(t) + x_n(t)y'_n(t) \end{aligned}$$

$$\begin{aligned}
&= (x'_1(t)y_1(t) + x'_2(t)y_2(t) + \cdots + x'_n(t)y_n(t)) + \\
&+ (x_1(t)y'_1(t) + x_2(t)y'_2(t) + \cdots + x_n(t)y'_n(t)) = \\
&= \mathbf{r}'_1(t) \cdot \mathbf{r}_2(t) + \mathbf{r}_1(t) \cdot \mathbf{r}'_2(t)
\end{aligned}$$

□

Som et eksempel på hvordan regnereglene kan brukes i mer teoretisk arbeid, tar vi med en konsekvens av (iii) som vi skal få bruk for senere:

**Korollar 3.1.5** *Dersom  $|\mathbf{r}(t)|$  er konstant, så er  $\mathbf{r}(t)$  og  $\mathbf{r}'(t)$  ortogonale.*

*Bevis:* Vi vet  $\mathbf{r}(t) \cdot \mathbf{r}(t) = |\mathbf{r}(t)|^2 = C$  er konstant. Bruker vi punkt (iii) i setningen ovenfor til å derivere  $\mathbf{r}(t) \cdot \mathbf{r}(t)$ , får vi dermed:

$$2\mathbf{r}(t) \cdot \mathbf{r}'(t) = 0$$

Følgelig er  $\mathbf{r}(t)$  og  $\mathbf{r}'(t)$  ortogonale. □

**Bemerkning:** Legg merk til at korollaret ovenfor har en geometrisk tolkning: Beveger vi oss på et kuleskall, vil alltid hastigheten (som er tangent tilkulen) stå normalt på posisjonsvektoren (som er en radius i kulen).

Den dobbeltderiverte til  $\mathbf{r}(t)$  er

$$\mathbf{a}(t) = \mathbf{v}'(t) = (x''_1(t), x''_2(t), \dots, x''_n(t))$$

Dersom  $\mathbf{r}(t)$  representerer posisjonen til en gjenstand ved tiden  $t$ , kalles  $\mathbf{a}(t)$  for *akselerasjonen*. Dette er en vektor som forteller oss hvordan hastigheten endrer seg, både i størrelse og retning. I dagliglivet er det vanligere å snakke om akselerasjon i en litt annen betydning, nemlig som fartsendring per tidsenhett, dvs. som  $a(t) = v'(t)$  (legg merke til at vi her deriverer skalarstørrelsen  $v(t)$  og ikke vektorstørrelsen  $\mathbf{v}(t)$ ). Vi skal kalle  $a(t)$  for *baneakselerasjonen* ved tiden  $t$ . Det er naturlig å spørre om sammenhengen mellom vektoren  $\mathbf{a}(t)$  og skalaren  $a(t)$ . Mange vil kanskje tippe at  $|\mathbf{a}(t)| = a(t)$ , men følgende eksempel viser at det ikke er tilfellet.

**Eksempel 7:** La

$$\mathbf{r}(t) = (\cos t, \sin t)$$

Da er

$$\mathbf{v}(t) = (-\sin t, \cos t)$$

som gir  $v(t) = |\mathbf{v}(t)| = 1$ . Følgelig er  $a(t) = v'(t) = 0$ , mens

$$\mathbf{a}(t) = (-\cos t, -\sin t)$$

Dette gir  $|\mathbf{a}(t)| = 1$ . Vi ser altså at  $0 = a(t) \neq |\mathbf{a}(t)| = 1$ . Det er ikke så vanskelig å forstå hva som foregår. Gjenstanden vår går i en sirkelbane med konstant fart. Siden farten er konstant, er baneakselerasjonen 0. Hastigheten skifter imidlertid retning hele tiden, og det medfører at  $\mathbf{a}(t)$  er forskjellig fra **0**.



Forutsatt at  $\mathbf{v}(t) \neq \mathbf{0}$ , kan vi definere *enhetstangentvektoren*  $\mathbf{T}(t)$  ved

$$\mathbf{T}(t) = \frac{\mathbf{v}(t)}{v(t)}$$

Som navnet sier, har denne vektoren lengde 1 for alle  $t$ . Snur vi på ligningen ovenfor, får vi  $\mathbf{v}(t) = v(t)\mathbf{T}(t)$ , og deriverer vi dette uttrykket, ser vi at (husk setning 3.1.4(v)):

$$\mathbf{a}(t) = \mathbf{v}'(t) = v'(t)\mathbf{T}(t) + v(t)\mathbf{T}'(t) = a(t)\mathbf{T}(t) + v(t)\mathbf{T}'(t)$$

Ifølge korollar 3.1.5 står  $\mathbf{T}'(t)$  normalt på  $\mathbf{T}(t)$  (og dermed på  $\mathbf{v}(t)$ ). Vi har dermed vist:

**Setning 3.1.6** *Dersom  $\mathbf{v}(t) \neq \mathbf{0}$ , kan akselerasjonen  $\mathbf{a}(t)$  dekomponeres i to ortogonale vektorer*

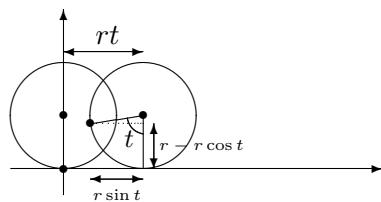
$$\mathbf{a}(t) = a(t)\mathbf{T}(t) + v(t)\mathbf{T}'(t)$$

der  $a(t)\mathbf{T}(t)$  er parallel med tangenten og  $v(t)\mathbf{T}'(t)$  står normalt på tangenten.

Vi ser hva som skjer; baneakselerasjonen  $a(t)$  måler hvor mye farten  $v(t)$  endrer seg, mens  $v(t)\mathbf{T}'(t)$  måler hvor mye retningen endrer seg. Det er mulig å finne et mer informativt uttrykk for leddet  $v(t)\mathbf{T}'(t)$ , men vi skal ikke komme nærmere inn på dette her.

La oss avslutte dette kapittelet med to eksempler som viser hvordan vi kan bruke parametriserte kurver til å beskrive fenomener i virkeligheten.

**Eksempel 8:** Hvilken kurve beskriver et punkt på et hjul når hjulet ruller bortover? Tenk deg av du har festet en refleksbrikke til et sykkeldekk og vil finne kurven som det lysende punktet beskriver i nattemørket.



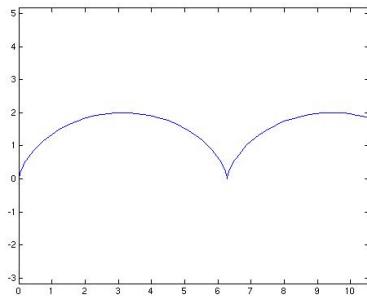
Figur 5: Et punkt på et rullende hjul.

Vi tenker oss at  $x$ -aksen er bakken som hjulet ruller på, og at punktet vårt starter i origo. Etter at hjulet har dreiet seg en vinkel  $t$ , har hjulet flyttet seg en distanse  $rt$  mot høyre, der  $r$  er radien i hjulet. Koordinatene til det lysende punktet blir dermed  $(rt - r \sin t, r - r \cos t)$  (se figur 5).

Vi har dermed parameterfremstillingen

$$\mathbf{r}(t) = (rt - r \sin t, r - r \cos t)$$

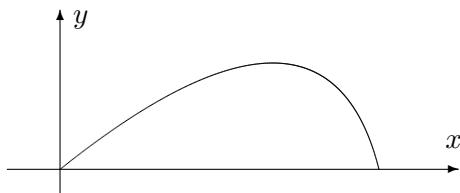
Denne kurven kalles en *sykloide* (se figur 6). ♣



Figur 6: En sykloide

Det neste eksemplet forutsetter litt kunnskaper om fysikk og differensiellligninger.

**Eksempel 9:** I dette eksemplet skal vi studere en stein som kastes (eller et prosjektil som skytes ut) under luftmotstand. Vi skal beskrive bevegelsen i et koordinatsystem som vist på figur 7 der  $x$ -aksen ligger vannrett og  $y$ -aksen loddrett.



Figur 7: Kast med luftmotstand.

Vi trenger litt kunnskaper om fysikk. Dersom den totale kraften som virker på prosjektilet er  $\mathbf{F}$  og akselerasjonen er  $\mathbf{a}$ , så forteller Newtons annen lov oss at  $\mathbf{F} = m\mathbf{a}$  der  $m$  er massen til steinen. Det er to krefter som virker på steinen i luften. Den ene er tyngdekraften som er gitt ved  $-mg\mathbf{j}$  der  $g$  er tyngdens akselerasjon ( $g \approx 9.8m/s^2$ ), og der  $\mathbf{j} = (0, 1)$ . Den andre kraften er luftmotstanden som vi skal anta er lik  $-k\mathbf{v}$  der  $k$  er en konstant og  $\mathbf{v}$  er hastigheten til steinen (i virkeligheten er luftmotstand en komplisert affære og vår formel er bare én av flere mulige tilnærminger).

Vi lar  $\mathbf{r}(t) = (x(t), y(t))$  være posisjonen til steinen ved tiden  $t$ . Da er  $\mathbf{v}(t) = (x'(t), y'(t))$  og  $\mathbf{a}(t) = (x''(t), y''(t))$ . Newtons lov  $\mathbf{F} = m\mathbf{a}$  kan nå skrives

$$m\mathbf{a}(t) = -mg\mathbf{j} - k\mathbf{v}$$

Deler vi på  $m$  og ser på første- og annenkomponenten hver for seg, får vi:

$$x''(t) = -\frac{k}{m}x'(t) \quad \text{og} \quad y''(t) = -g - \frac{k}{m}y'(t)$$

For å finne kurven  $\mathbf{r}(t)$  må vi løse disse differensialligningene med passende begynnelsesbetingelser. Vi antar at bevegelsen starter i origo med hastighet  $\mathbf{v}(0) = (u_1, u_2)$ . Da blir begynnelsesbetingelsene  $x(0) = 0$ ,  $x'(0) = u_1$ ,  $y(0) = 0$ ,  $y'(0) = u_2$ . Løser vi differensialligningene med disse begynnelsesbetingelsene (gjør det!), får vi

$$\begin{aligned} x(t) &= \frac{mu_1}{k} \left( 1 - e^{-\frac{kt}{m}} \right) \\ y(t) &= -\frac{mg}{k}t + \left( \frac{mu_2}{k} + \frac{m^2g}{k^2} \right) \left( 1 - e^{-\frac{kt}{m}} \right) \end{aligned}$$

Vi har dermed parametriseringen

$$\mathbf{r}(t) = \frac{mu_1}{k} \left( 1 - e^{-\frac{kt}{m}} \right) \mathbf{i} + \left( -\frac{mg}{k}t + \left( \frac{mu_2}{k} + \frac{m^2g}{k^2} \right) \left( 1 - e^{-\frac{kt}{m}} \right) \right) \mathbf{j}$$

Legger du denne formelen inn i MATLAB (se nedenfor) eller en lommeregner og velger verdier for  $m$  og  $k$ , kan du f.eks. eksperimentere med hvilken utkastvinkel som gir den største kastlengden (sett  $u_1 = v_0 \cos(\alpha)$ ,  $u_2 = v_0 \sin(\alpha)$  der  $v_0$  er passende utgangsfart, og undersøk hvordan kastlengden varierer med vinkelen  $\alpha$ ). ♣

I fysikk og beslektede fag brukes parametriserte kurver til å beskrive hvordan objekter beveger seg, for eksempel hvordan planeter, stjerner og meteorer beveger seg i verdensrommet. Eksempelet ovenfor gir en liten følelse av hvordan man kommer frem til slike parameterfremstillinger: Ved å bruke Newtons lov  $\mathbf{F} = m\mathbf{a}$  får vi satt opp differensialligninger som forbinder akcelerasjonen  $\mathbf{x}''(t) = \mathbf{a}(t)$ , hastigheten  $\mathbf{x}'(t) = \mathbf{v}(t)$  og posisjonen  $\mathbf{x}(t)$ . Løser vi disse differensialligningene med passende begynnelsesbetingelser, finner vi parametriseringen.

**MATLAB-kommentar:** Du kan få MATLAB til å tegne parametriserte kurver i planet ved å bruke kommandoen `plot` på vanlig måte. Vil du tegne sirkelen

$$\mathbf{r}(t) = (\cos t, \sin t) \quad t \in [0, 2\pi]$$

kan du for eksempel taste

```
t=linspace(0,2*pi,100); % dette gir deg 100 jevnt
                           % fordelte punkter fra 0 til 2π
x=cos(t);
y=sin(t);
plot(x,y)
axis('equal')
```

Kommandoen `axis('equal')` sørger for at MATLAB bruker samme lengdeenhett langs begge aksene — utslerter du den, blir sirkelen seende ut som en ellipse. Kommandoen virker tilbake på det siste plottet du har utført, og bør derfor brukes etter kommandoen `plot(x,y)`.

For å lage kurver i rommet, kan du bruke kommandoen `plot3` på tilsvarende måte. Vil du plotte kurven

$$\mathbf{r}(t) = (\cos t, \sin t, t^2) \quad t \in [0, 2\pi],$$

kan du f.eks. taste

```
t=linspace(0,2*pi,100);
x=cos(t);
y=sin(t);
z=t.^2;
plot3(x,y,z)
axis('equal')
```

### Oppgaver til seksjon 3.1

1. En parametrisert kurve er gitt ved  $\mathbf{r}(t) = (t^3, t^2)$ . Finn  $\mathbf{v}(t)$ ,  $v(t)$ ,  $\mathbf{a}(t)$  og  $a(t)$ .
2. En kurve er parametrisert ved  $\mathbf{r}(t) = (\cos t, t \sin t)$ . Finn  $\mathbf{v}(t)$ ,  $v(t)$ ,  $\mathbf{a}(t)$  og  $a(t)$ .
3. Finn hastigheten og akselerasjonen når  $\mathbf{r}(t) = (t, e^{-t}, \sin t)$ .
4. Finn hastigheten og akselerasjonen når  $\mathbf{r}(t) = (\ln t, t^2, \cos t)$ .
5. Bruk MATLAB til å tegne kurvene:
  - a)  $\mathbf{r}(t) = t \cos t \mathbf{i} + t \sin t \mathbf{j}$ ,  $t \in [0, 6\pi]$
  - b)  $\mathbf{r}(t) = 5 \cos t \mathbf{i} + 3 \sin t \mathbf{j}$ ,  $t \in [0, 2\pi]$
  - c)  $\mathbf{r}(t) = \sin(2t) \cos t \mathbf{i} + \sin(2t) \sin t \mathbf{j}$ ,  $t \in [0, 2\pi]$
  - d)  $\mathbf{r}(t) = t \cos t \mathbf{i} + t \sin t \mathbf{j} + t \mathbf{k}$ ,  $t \in [0, 6\pi]$
  - e)  $\mathbf{r}(t) = t \mathbf{i} + \sin t \mathbf{j} + \cos t \mathbf{k}$ ,  $t \in [-20, 20]$
6. Bruk MATLAB til å tegne kurven  $\mathbf{r}(t)$  i eksempel 9. Velg  $m = 1$ ,  $g = 9.8$  og eksperimenter med forskjellige verdier for  $k$ ,  $u_1$  og  $u_2$ .
7. En kurve er parametrisert ved

$$\mathbf{r}(t) = (a \cos t, b \sin t) \quad t \in [0, 2\pi]$$

- a) Vis at denne kurven er ellipsen med ligning  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$
- b) Finn hastigheten, farten og akselerasjonen.
- c) Vis at omkretsen til ellipsen er  $\int_0^{2\pi} \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} dt$ . Sett  $a = 5$ ,  $b = 3$  og finn omkretsen ved å bruke numerisk integrasjon på en lommeregner eller en datamaskin.

**8.** Finn buelengden til kurven

$$\mathbf{r}(t) = (t^2, t^3) \quad t \in [0, 10]$$

**9.** En kurve er parametrisert ved  $\mathbf{r}(t) = (t, \ln(\cos t))$  for  $t \in [0, \pi/4]$ .

- a) Finn hastigheten  $\mathbf{v}(t)$  og farten  $v(t)$ .
- b) Finn buelengden (Hint: For å integrere  $\frac{1}{\cos x}$  kan det være nyttig å bruke at  $\frac{1}{\cos x} = \frac{\cos x}{1 - \sin^2(x)}$ ).

**10.** Vi har  $\mathbf{r}(t) = (2 \cos t, \sqrt{2} \sin t, \sqrt{2} \sin t)$ .

- a) Finn hastigheten, farten og akselerasjonen.
- b) Finn buelengden fra  $t = 0$  til  $t = 2\pi$ .
- c) Vis at kurven ligger på en kuleflate med sentrum i origo.
- d) Vis at kurven ligger i planet  $y - z = 0$ .
- e) Hva slags kurve fremstiller  $\mathbf{r}$ ? Bruk MATLAB til å tegne kurven.

**11.** Vi har  $\mathbf{r}(t) = (t \cos t, t \sin t, t)$

- a) Finn hastigheten, farten og akselerasjonen.
- b) Vis at buelengden fra  $t = 0$  til  $t = 2\pi$  er  $\int_0^{2\pi} \sqrt{2 + t^2} dt$ . Bruk numerisk integrasjon til å beregne dette integralet.
- c) Løs integralet i b) ved regning. Bruk substitusjonen  $t = \frac{e^u - e^{-u}}{\sqrt{2}}$ .

**12.** La  $\mathbf{r}(t)$  være sykloiden i eksempel 8.

- a) Finn hastigheten og akselerasjonen.
- b) Vis at lengden punktet gjennomløper mens hjulet dreier en gang rundt, er  $r\sqrt{2} \int_0^{2\pi} \sqrt{1 - \cos t} dt$
- c) Forklar hvorfor  $\sqrt{1 - \cos t} = \frac{|\sin t|}{\sqrt{1 + \cos t}}$ .
- d) Regn ut integralet i b).

**13.** Bevis (i), (ii), (iv) og (v) i setning 3.1.4.

**14.** En partikkel går i en sirkelbane med radius  $r$  om origo. Farten er konstant lik  $v$ . Partikkelen starter i punktet  $(0, 1)$  ved tiden  $t = 0$  og beveger seg mot urviserne.

- a) Vis at posisjonen ved tiden  $t$  er  $\mathbf{r}(t) = (r \cos(\frac{vt}{r}), r \sin(\frac{vt}{r}))$ .

b) Vis at  $\mathbf{a}(t) = -(\frac{v}{r})^2 \mathbf{r}(t)$ .

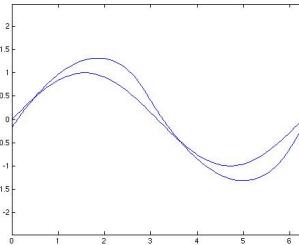
**15.** En kanonkule skytes ut med en fart  $v_0$ . I utskytingsøyeblikket danner kulens bane en vinkel  $\alpha$  med horisontalplanet. Kulens posisjon etter  $t$  sekunder kaller vi  $\mathbf{r}(t) = (x(t), y(t))$ . Dersom vi kan se bort fra luftmotstanden, vil  $x(t) = v_0 t \cos \alpha$  og  $y(t) = v_0 t \sin \alpha - \frac{1}{2} g t^2$  der  $g$  er tyngdens akselerasjon.

- a) Finn  $\mathbf{v}$  og  $\mathbf{a}$ .
- b) Hvor høyt over bakken erkulen på det høyeste?
- c) Hvor langt kan kanonen skyte (vi antar at bakken er horisontal)?

**16.** Når er steinen i eksempel 9 i det høyeste punktet på banen? Hvor høyt er dette punktet?

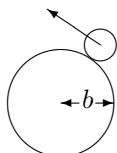
**17.** Avstanden mellom det stedet der bakhjulet til en sykkel berører bakken, og det stedet der forhjulet berører bakken, er 1 meter. Når vi sykler, etterlater både forhjulet og bakhjulet et spor i bakken.

- a) Anta at sporet bakhjulet etterlater seg, er gitt ved  $\mathbf{r}_1(t)$ . Vis at sporet forhjulet etterlater seg, har parametrisering  $\mathbf{r}_2(t) = \mathbf{r}_1(t) + \mathbf{T}_1(t)$ , der  $\mathbf{T}_1(t)$  er enhetstangentvektoren til  $\mathbf{r}_1(t)$ .
- b) Anta at bakhjulet følger kurven  $\mathbf{r}_1(t) = (t, \sin t)$ . Finn parametriseringen  $\mathbf{r}_2(t)$  til kurven som forhjulet følger.
- c) Bruk MATLAB til å tegne kurvene  $\mathbf{r}_1$  og  $\mathbf{r}_2$  i samme koordinatsystem.



- d) Figuren ovenfor viser sporene etter en sykkel som har vinglet forbi. Kjørte sykkelen fra venstre mot høyre eller i motsatt retning?

**18.** Et hjul med radius  $a$  ruller på utsiden av en sirkel med radius  $b$  (se figuren). Finn en parameterfremstilling for den kurven et punkt på hjulet følger. Du kan selv velge hvordan du vil legge koordinatsystemet og hvor startpunktet er.



**19.** (Eksamensoppgave i MAT 100A/C, 8/12-2000) I denne oppgaven er  $\mathbf{r}(t) = (t, t^2)$  en parametrisert kurve og  $\mathbf{b} = (0, y)$ ,  $y > 0$ , er et punkt på den positive  $y$ -aksen.

- a) Skissér kurven og finn  $\mathbf{r}'(t)$ . Vis at den deriverte til funksjonen

$$f(t) = |\mathbf{r}(t) - \mathbf{b}|^2$$

kan skrives  $f'(t) = 2\mathbf{r}'(t) \cdot (\mathbf{r}(t) - \mathbf{b})$ .

- b) Vi er interessert i å finne de punktene på kurven som ligger nærmest  $\mathbf{b}$ . Vis at dersom  $\mathbf{r}(t_0)$  er et slik punkt, så er

$$\mathbf{r}'(t_0) \cdot (\mathbf{r}(t_0) - \mathbf{b}) = 0$$

Forklar denne likningen geometrisk.

- c) Finn de punktene på kurven som ligger nærmest  $\mathbf{b}$ .

**20.** En parametrisert kurve er gitt ved  $\mathbf{r}(t) = (x(t), y(t))$ ,  $a \leq t \leq b$ , der  $x$  og  $y$  har kontinuerlige deriverte  $x'$ ,  $y'$ . Anta at  $g : [c, d] \rightarrow [a, b]$  er en voksende funksjon med kontinuerlig derivert og at  $g(c) = a$ ,  $g(d) = b$ .

- a) Forklar at  $\mathbf{s}(t) = \mathbf{r}(g(t))$ ,  $c \leq t \leq d$ , er en annen parametrisering av den samme kurven.

I resten av denne oppgaven skal vi vise at et par grunnleggende geometriske egenskaper til kurven er de samme uansett hvilken av de to parametriseringene vi velger.

- b) La  $\mathbf{a} = \mathbf{s}(t_0) = \mathbf{r}(g(t_0))$  være et punkt på kurven. Vi kan regne ut to tangentvektorer i punktet  $\mathbf{a}$ , nemlig  $\mathbf{s}'(t_0)$  og  $\mathbf{r}'(g(t_0))$ . Vis at disse vektorene er parallelle (vi godtar at den ene eller begge er lik  $\mathbf{0}$ ).
- c) Vis at buelengden til kurven blir den samme uansett hvilken av de to parametriseringene vi velger.

**21.** En partikkel beveger seg i et kraftfelt der kraften hele tiden er rettet mot eller fra origo (dette gjelder blant annet partikler i et gravitasjonsfelt eller et elektrisk felt der massen eller ladningen er koncentrert i origo). Ifølge Newtons annen lov er  $\mathbf{F} = m\mathbf{a}$ , så akselerasjonen er også rettet mot eller fra origo. Det betyr at akselerasjonen ved tiden  $t$  er gitt ved  $\mathbf{a}(t) = k(t)\mathbf{r}(t)$  der  $k(t)$  er en skalar størrelse.

- a) Vis at  $\frac{d}{dt}[\mathbf{r}(t) \times \mathbf{v}(t)] = \mathbf{0}$ .
- b) Forklar hvorfor  $\mathbf{r}(t) \times \mathbf{v}(t) = \mathbf{c}$  der  $\mathbf{c}$  er en konstant vektor (uavhengig av  $t$ ).
- c) Vis at partikkelen hele tiden beveger seg i planet gjennom punktene  $\mathbf{0}$ ,  $\mathbf{r}(0)$  og  $\mathbf{v}(0)$ .

**22.** Anta at  $\mathbf{v} : [a, b] \rightarrow \mathbb{R}^n$  er en parametrisert kurve.

- a) La  $a = t_0 < t_1 < \dots < t_N = b$  være en partisjon av  $[a, b]$ . Vis at

$$\left| \sum_{i=1}^N \mathbf{v}(t_i)(t_i - t_{i-1}) \right| \leq \sum_{i=1}^N |\mathbf{v}(t_i)| (t_i - t_{i-1})$$

- b) Vi definerer

$$\int_a^b \mathbf{v}(t) dt = \left( \int_a^b v_1(t) dt, \int_a^b v_2(t) dt, \dots, \int_a^b v_n(t) dt \right)$$

dvs. at vi integrerer  $\mathbf{v}$  komponentvis. Vis at

$$\left| \int_a^b \mathbf{v}(t) dt \right| \leq \int_a^b |\mathbf{v}(t)| dt$$

- c) Anta at  $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$  er en parametrisert kurve med kontinuerlig derivert  $\mathbf{v}(t) = \mathbf{r}'(t)$ . Vis at

$$|\mathbf{r}(b) - \mathbf{r}(a)| \leq \int_a^b v(t) dt$$

der som vanlig  $v(t) = |\mathbf{v}(t)|$ .

**23.** Definisjonen vi har gitt av buelengde er grei for regnetekniske formål, men den har teoretiske svakheter. I denne oppgaven skal vi se på en mer tilfredsstillende måte å definere buelengde på. Deler av oppgaven bygger på oppgave 20 ovenfor, og det kan derfor være lurt å regne den først.

Vi starter med en parametrisert kurve  $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$ . For hver partisjon

$$\Pi = \{a = t_0 < t_1 < t_2 < \dots < t_N = b\}$$

av  $[a, b]$ , lar vi

$$l(\Pi) = \sum_{i=1}^N |\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})|$$

være lengden til den tilhørende brudne stien. Vi definerer nå

$$\Lambda(a, b) = \sup\{l(\Pi) \mid \Pi \text{ er en partisjon av } [a, b]\}$$

der vi setter  $\Lambda(a, b) = \infty$  hvis mengden er ubegrenset. Dersom  $\Lambda(a, b)$  er endelig, sier vi at den parametriserte kurven er *rektifiserbar* og definerer lengden til  $\mathbf{r}$  til å være  $\Lambda(a, b)$ .

- a) Anta at  $\mathbf{r}$  er rektifiserbar. Vis at dersom  $a < c < b$  så er  $\Lambda(a, b) = \Lambda(a, c) + \Lambda(c, b)$
- b) I resten av oppgaven antar vi at  $\mathbf{r}$  er deriverbar og at den deriverte  $\mathbf{r}'(t) = \mathbf{v}(t)$  er kontinuerlig. Vis at for alle partisjoner  $\Pi$  er  $l(\Pi) \leq \int_a^b v(t) dt$ , og at  $\mathbf{r}$  derfor er rektifiserbar (her kan du få bruk for oppgave 20).
- c) Definer  $s : [a, b] \rightarrow \mathbb{R}$  ved  $s(t) = \Lambda(a, t)$ . I neste punkt skal vi vise at  $s'(t) = v(t)$ . Forklar at dette vil medføre at  $s(t) = \int_a^t v(r) dr$ .
- d) Forklar ulikhettene

$$\left| \frac{\mathbf{r}(t+h) - \mathbf{r}(t)}{h} \right| \leq \frac{s(t+h) - s(t)}{h} \leq \frac{1}{h} \int_t^{t+h} v(s) ds$$

(du vil sannsynligvis få bruk for resultatene i oppgave 20) og vis at de medfører at  $s'(t) = v(t)$ .

### 3.2 Kjerneregelen for parametriserte kurver

I seksjon 2.7 studerte vi kjerneregelen både på matrise- og koordinatform. Vi skal nå kikke litt på hvordan regelen blir seende ut når kjernefunksjonen er en parametrisert kurve. Det er ingen ny matematikk i dette, bare litt notasjon som gjør kjerneregelen enklere å bruke i praksis.

**Setning 3.2.1** *Hvis den parametriserte kurven  $\mathbf{r}(t) = (x_1(t), \dots, x_n(t))$  er deriverbar i punktet  $t \in I$ , og skalarfeltet  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  er deriverbart i punktet  $\mathbf{r}(t)$ , så er funksjonen  $u(t) = f(\mathbf{r}(t))$  deriverbar i  $t$ , og*

$$u'(t) = \frac{\partial f}{\partial x_1}(\mathbf{r}(t))x'_1(t) + \frac{\partial f}{\partial x_2}(\mathbf{r}(t))x'_2(t) + \dots + \frac{\partial f}{\partial x_n}(\mathbf{r}(t))x'_n(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

*Beweis:* Bruker vi kjerneregelen på komponentform på uttrykket

$$u(t) = f(\mathbf{r}(t)) = f(x_1(t), x_2(t), \dots, x_n(t))$$

får vi

$$\begin{aligned} u'(t) &= \frac{\partial f}{\partial x_1}(\mathbf{r}(t)) \frac{\partial x_1(t)}{\partial t} + \frac{\partial f}{\partial x_2}(\mathbf{r}(t)) \frac{\partial x_2(t)}{\partial t} + \dots + \frac{\partial f}{\partial x_n}(\mathbf{r}(t)) \frac{\partial x_n(t)}{\partial t} = \\ &= \frac{\partial f}{\partial x_1}(\mathbf{r}(t))x'_1(t) + \frac{\partial f}{\partial x_2}(\mathbf{r}(t))x'_2(t) + \dots + \frac{\partial f}{\partial x_n}(\mathbf{r}(t))x'_n(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) \end{aligned}$$

der vi i siste skritt bruker at  $\nabla f(\mathbf{r}(t)) = (\frac{\partial f}{\partial x_1}(\mathbf{r}(t)), \frac{\partial f}{\partial x_2}(\mathbf{r}(t)), \dots, \frac{\partial f}{\partial x_n}(\mathbf{r}(t)))$  og  $\mathbf{r}'(t) = (x'_1(t), x'_2(t), \dots, x'_n(t))$   $\square$

Som det fremgår av beviset, er setningen ovenfor bare et spesialtilfelle av den generelle kjerneregelen. Siden gradienten har en viktig geometrisk betydning, er det imidlertid lurt å være klar over skrivemåten  $u'(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$ .

**Eksempel 1.** Et primitivt havdyr svømmer slik at det alltid befinner seg et sted der havtemperaturen er  $20^\circ$ . Anta at  $T(x, y, z)$  er havtemperaturen i punktet  $(x, y, z)$ , og at  $\mathbf{r}(t)$  er posisjonen til dyret ved tiden  $t$ . Vi skal vise at hastigheten  $\mathbf{r}'(t)$  alltid står normalt på gradienten  $\nabla T(\mathbf{r}(t))$  til temperaturfunksjonen.

Siden gradienten peker i den retningen hvor temperaturen vokser raskest, virker det ikke unaturlig at dyret må bevege seg “på tvers” av gradienten for å holde temperaturen konstant. For å vise dette observerer vi at hvis  $u(t)$  er havtemperaturen i dyrets posisjon ved tiden  $t$ , så er på den ene side  $u(t) = 20$ , og på den annen side

$$u(t) = T(\mathbf{r}(t))$$

Deriverer vi disse uttrykkene, får vi hhv.  $u'(t) = 0$  og

$$u'(t) = \nabla T(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

Dermed er  $\nabla T(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = 0$  og følgelig står  $\nabla T(\mathbf{r}(t))$  og  $\mathbf{r}'(t)$  normalt på hverandre. ♣

Ofte trenger vi resultatet i setning 3.2.1 på en litt utvidet form fordi funksjonen  $f$  ikke bare avhenger av posisjonen  $\mathbf{r}(t)$ , men også av tiden  $t$ . Vi ønsker altså å derivere et uttrykk av typen

$$u(t) = f(\mathbf{r}(t), t)$$

Ved å bruke kjerneregelen på akkurat samme måte som ovenfor, ser vi at

$$u'(t) = \frac{\partial f}{\partial x_1}(\mathbf{r}(t), t)x'_1(t) + \cdots + \frac{\partial f}{\partial x_n}(\mathbf{r}(t), t)x'_n(t) + \frac{\partial f}{\partial t}(\mathbf{r}(t), t) \quad (3.2.1)$$

(siden den deriverte av  $t$  mhp.  $t$  er 1, får ikke det siste ledet en faktor av typen  $x'_1(t), \dots, x'_n(t)$ ).

Vi tar også med en variant av kjerneregelen for vektorvaluerte funksjoner. Dersom

$$\mathbf{F}(\mathbf{x}, t) = \begin{pmatrix} F_1(\mathbf{x}, t) \\ F_2(\mathbf{x}, t) \\ \vdots \\ F_m(\mathbf{x}, t) \end{pmatrix}$$

er en funksjon av  $n + 1$  variable med verdier i  $\mathbb{R}^m$ , skal vi bruke notasjonen

$$\frac{\partial \mathbf{F}}{\partial x_i}(\mathbf{x}, t) = \begin{pmatrix} \frac{\partial F_1}{\partial x_i}(\mathbf{x}, t) \\ \frac{\partial F_2}{\partial x_i}(\mathbf{x}, t) \\ \vdots \\ \frac{\partial F_m}{\partial x_i}(\mathbf{x}, t) \end{pmatrix} \quad \text{og} \quad \frac{\partial \mathbf{F}}{\partial t}(\mathbf{x}, t) = \begin{pmatrix} \frac{\partial F_1}{\partial t}(\mathbf{x}, t) \\ \frac{\partial F_2}{\partial t}(\mathbf{x}, t) \\ \vdots \\ \frac{\partial F_m}{\partial t}(\mathbf{x}, t) \end{pmatrix}$$

**Setning 3.2.2** La  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  være en funksjon av  $n + 1$  variable. Dersom den parametriserte kurven  $\mathbf{r}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  er deriverbar i punktet  $t$ , og  $\mathbf{F}$  er deriverbar i punktet  $(\mathbf{r}(t), t)$ , så er den sammesatte funksjonen  $\mathbf{U}(t) = \mathbf{F}(\mathbf{r}(t), t)$  deriverbar i  $t$  og

$$\mathbf{U}'(t) = \frac{\partial \mathbf{F}}{\partial x_1}(\mathbf{r}(t), t)x'_1(t) + \cdots + \frac{\partial \mathbf{F}}{\partial x_n}(\mathbf{r}(t), t)x'_n(t) + \frac{\partial \mathbf{F}}{\partial t}(\mathbf{r}(t), t)$$

*Bevis:* Den  $i$ -te komponenten til  $\mathbf{U}$  er

$$U_i(t) = F_i(\mathbf{r}(t), t)$$

Bruker vi formel (3.2.1) på dette uttrykket, får vi

$$U'_i(t) = \frac{\partial F_i}{\partial x_1}(\mathbf{r}(t), t)x'_1(t) + \cdots + \frac{\partial F_i}{\partial x_n}(\mathbf{r}(t), t)x'_n(t) + \frac{\partial F_i}{\partial t}(\mathbf{r}(t), t)$$

Kombinerer vi alle komponentene, får vi

$$\mathbf{U}'(t) = \frac{\partial \mathbf{F}}{\partial x_1}(\mathbf{r}(t), t)x'_1(t) + \cdots + \frac{\partial \mathbf{F}}{\partial x_n}(\mathbf{r}(t), t)x'_n(t) + \frac{\partial \mathbf{F}}{\partial t}(\mathbf{r}(t), t) \quad \square$$

Denne varianten av kjerneregelen brukes mye i fysikk og hydrodynamikk. Her er et typisk eksempel:

**Eksempel 2:** I en strømmende væske er hastigheten i punktet  $(x, y, z)$  ved tiden  $t$  gitt ved

$$\mathbf{F}(x, y, z, t) = \begin{pmatrix} F_1(x, y, z, t) \\ F_2(x, y, z, t) \\ F_3(x, y, z, t) \end{pmatrix}$$

Vi skal finne akselerasjonen til en partikkel i væsken.

Dersom  $\mathbf{r}(t) = (x(t), y(t), z(t))$  er posisjonen til partikkelen ved tiden  $t$ , vet vi at hastigheten er gitt ved

$$\mathbf{v}(t) = \mathbf{F}(\mathbf{r}(t), t)$$

Akselerasjonen  $\mathbf{a}(t)$  er den tidsderiverte til hastigheten, så ifølge setningen ovenfor er

$$\mathbf{a}(t) = \mathbf{v}'(t) = \frac{\partial \mathbf{F}}{\partial x}(\mathbf{r}(t), t)x'(t) + \frac{\partial \mathbf{F}}{\partial y}(\mathbf{r}(t), t)y'(t) + \frac{\partial \mathbf{F}}{\partial z}(\mathbf{r}(t), t)z'(t) + \frac{\partial \mathbf{F}}{\partial t}(\mathbf{r}(t), t)$$

Siden  $\mathbf{v}(t) = \mathbf{F}(\mathbf{r}(t), t)$ , har vi

$$x'(t) = F_1(\mathbf{r}(t), t), \quad y'(t) = F_2(\mathbf{r}(t), t) \quad \text{og} \quad z'(t) = F_3(\mathbf{r}(t), t)$$

Setter vi dette inn i uttrykket ovenfor, får vi

$$\begin{aligned} \mathbf{a}(t) &= \frac{\partial \mathbf{F}}{\partial x}(\mathbf{r}(t), t)F_1(\mathbf{r}(t), t) + \frac{\partial \mathbf{F}}{\partial y}(\mathbf{r}(t), t)F_2(\mathbf{r}(t), t) + \\ &+ \frac{\partial \mathbf{F}}{\partial z}(\mathbf{r}(t), t)F_3(\mathbf{r}(t), t) + \frac{\partial \mathbf{F}}{\partial t}(\mathbf{r}(t), t) \end{aligned}$$

Dette uttrykket for akselerasjonen til en partikkel i et hastighetsfelt kan kobles sammen med Newtons annen lov, og blir da et av utgangspunktene for de grunnleggende ligningene i hydrodynamikken. ♣

### Middelverdisetningen for funksjoner av flere variable

Til slutt i denne seksjonen tar vi med en utvidelse av middelverdisetningen (setning 6.2.3 i *Kalkulus*) til funksjoner av flere variable. Vi skal få bruk for dette resultatet flere steder i kapittel 5.

#### **Setning 3.2.3 (Middelverdisetning for funksjoner av flere variable)**

Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $m$  variable, og at  $f$  er deriverbar i et område som inneholder linjestykket mellom punktene  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ . Da finnes det et punkt  $\mathbf{c}$  på linjestykket fra  $\mathbf{a}$  til  $\mathbf{b}$  slik at

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a})$$

*Bevis:* Definer en funksjon  $g : [0, 1] \rightarrow \mathbb{R}$  av én variabel ved

$$g(t) = f(\mathbf{r}(t))$$

der  $\mathbf{r}(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ ,  $t \in [0, 1]$ , er en parametrisering av linjestykket fra  $\mathbf{a}$  til  $\mathbf{b}$ . Ved kjerneregelen er

$$g'(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = \nabla f(\mathbf{r}(t)) \cdot (\mathbf{b} - \mathbf{a})$$

Etter den vanlige middelverdisetningen (*Kalkulus*, setning 6.2.3) finnes det et tall  $c$  mellom 0 og 1 slik at

$$g(1) - g(0) = g'(c) = \nabla f(\mathbf{r}(c)) \cdot (\mathbf{b} - \mathbf{a})$$

Setter vi  $\mathbf{c} = \mathbf{r}(c)$ , har vi dermed

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a})$$

og setningen er bevist. □

### Oppgaver til seksjon 3.2

1. Anta at  $f(x, y) = x^2y^3$  og  $\mathbf{r}(t) = t^2\mathbf{i} + 3t\mathbf{j}$ . Regn ut  $g'(t)$  når  $g(t) = f(\mathbf{r}(t))$ .
2. Anta at  $f(x, y) = x^2e^{xy^2}$  og  $\mathbf{r}(t) = \sin t\mathbf{i} + \cos t\mathbf{j}$ . Regn ut  $g'(t)$  når  $g(t) = f(\mathbf{r}(t))$ .
3. Anta at  $f(x, y, z) = x^2z - y \sin(yz)$  og  $\mathbf{r}(t) = e^t\mathbf{i} + t\mathbf{j} + \cos t^2\mathbf{k}$ . Regn ut  $g'(t)$  når  $g(t) = f(\mathbf{r}(t))$ .
4. Anta at  $f(x, y, t) = ty^2 \ln(x^2 + 1)$  og  $\mathbf{r}(t) = t^3\mathbf{i} + (3t + 1)\mathbf{j}$ . Regn ut  $g'(t)$  når  $g(t) = f(\mathbf{r}(t), t)$ .
5. Anta at  $\mathbf{F}(x, y) = \begin{pmatrix} x^2y \\ xy + x \end{pmatrix}$  og at  $\mathbf{r}(t) = \sin t\mathbf{i} + \cos t\mathbf{j}$ . Regn ut  $\mathbf{G}'(t)$  når  $\mathbf{G}(t) = \mathbf{F}(\mathbf{r}(t))$ .

- 6.** Anta at  $\mathbf{F}(x, y, t) = \begin{pmatrix} x^2t + y \\ \cos(xy) \\ t^2x \end{pmatrix}$  og at  $\mathbf{r}(t) = (1 - t^2)\mathbf{i} + \sin t\mathbf{j}$ . Regn ut  $\mathbf{G}'(t)$  når  $\mathbf{G}(t) = \mathbf{F}(\mathbf{r}(t), t)$ .

- 7.** Temperaturen i et punkt  $(x, y)$  ved tiden  $t$  er

$$f(x, y, t) = 20 + 2t - x^2 + y^2$$

En person befinner seg ved tiden  $t$  i punktet

$$\mathbf{r}(t) = (3t - \frac{t^2}{4})\mathbf{i} + (2t + \frac{t^2}{8})\mathbf{j}$$

Er temperaturen som personen opplever økende eller avtagende ved tiden  $t = 1$ ?

- 8.** Anta at  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  har kontinuerlige annenordens partiellderiverte, og at  $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j}$  der  $x$  og  $y$  er to ganger deriverbare. La  $g(t) = f(\mathbf{r}(t))$ . Vis at

$$\begin{aligned} g''(t) &= \frac{\partial^2 f}{\partial x^2}(\mathbf{r}(t)) x'(t)^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(\mathbf{r}(t)) x'(t)y'(t) + \frac{\partial^2 f}{\partial y^2}(\mathbf{r}(t)) y'(t)^2 + \\ &\quad + \frac{\partial f}{\partial x}(\mathbf{r}(t)) x''(t) + \frac{\partial f}{\partial y}(\mathbf{r}(t)) y''(t) \end{aligned}$$

### 3.3 Linjeintegraler for skalarfelt

I denne og den neste seksjonen skal vi se på to forskjellige metoder for å integrere en funksjon langs en kurve  $\mathcal{C}$ . De to integraltypene betegnes med henholdsvis

$$\int_{\mathcal{C}} f \, ds \quad \text{og} \quad \int_{\mathcal{C}} \mathbf{F} \, d\mathbf{r}$$

og de har litt forskjellige bruksområder.

Vi starter med integraltypen  $\int_{\mathcal{C}} f \, ds$ , og tenker oss at  $\mathcal{C}$  er en kurve parametrisert ved

$$\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$$

La oss videre tenke oss at  $\mathcal{C}$  er en tråd med varierende tykkelse, og at *tetheten* til tråden i et punkt  $\mathbf{x}$  er  $f(\mathbf{x})$ . Det betyr at hvis vi kutter ut en liten bit av tråden med lengde  $s$  rundt punktet  $\mathbf{x}$ , så er massen til den delen tilnærmet lik  $f(\mathbf{x})s$ . Hvis vi har lyst til å finne et tilnærmet uttrykk for hele massen til tråden, deler vi den opp i flere smådeler med lengde  $s_1, s_2, \dots, s_N$  og velger punkter  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  på de forskjellige delene. Massen til tråden vil da være tilnærmet lik

$$f(\mathbf{x}_1)s_1 + f(\mathbf{x}_2)s_2 + \cdots + f(\mathbf{x}_N)s_N = \sum_{i=1}^N f(\mathbf{x}_i)s_i$$

med bedre tilnærming dess finere oppdelingen er. Dette minner sterkt om Riemann-summen til et integral, og gir oss et hint om at massen kan uttrykkes som et integral. La oss ta et skritt tilbake og begynne med parametriseringen

$$\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$$

Vi deler opp intervallet  $[a, b]$  på den vanlige måten

$$a = t_0 < t_1 < \dots < t_N = b$$

og velger et punkt  $c_i \in [t_{i-1}, t_i]$  i det  $i$ -te delintervallet. Den delen av snoren som er parametrisert av dette delintervallet, vil ha en lengde som er omtrent  $v(c_i)(t_i - t_{i-1})$  (strekning er fart ganger tid) og en tetthet som er omtrent  $f(\mathbf{r}(c_i))$ . Massen til denne delen av tråden er altså tilnærmet lik  $f(\mathbf{r}(c_i))v(c_i)(t_i - t_{i-1})$ , og massen til hele tråden blir tilnærmet lik

$$\sum_{i=1}^N f(\mathbf{r}(c_i))v(c_i)(t_i - t_{i-1})$$

Dette er en Riemann-sum for funksjonen  $f(\mathbf{r}(t))v(t)$  og konvergerer mot integralet

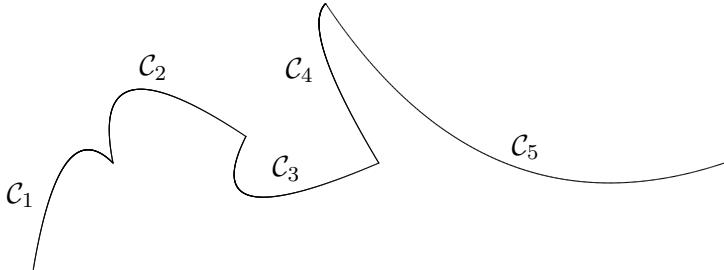
$$\int_a^b f(\mathbf{r}(t))v(t) dt$$

når oppdelingen blir finere og finere. Massen til tråden er derfor gitt ved

$$M = \int_a^b f(\mathbf{r}(t))v(t) dt$$

Argumentet ovenfor er ikke et matematisk bevis (vi har ikke gitt en matematisk definisjon av begrepene som inngår, og vi har vært temmelig uforsiktige med begrepet “tilnærmet lik”), men det gir oss en sterk indikasjon på at integraler av typen  $\int_a^b f(\mathbf{r}(t))v(t) dt$  er nyttige når vi skal regne ut totaleffekten av noe som varierer langs en kurve. Vi skal ta dette som utgangspunkt for vår definisjon av integraler av typen  $\int_C f ds$ , men før vi ser på den formelle definisjonen, må vi ble enige om hva slags kurver vi skal arbeide med.

Vi skal si at en parametrisering  $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$  er *glatt* dersom funksjonen  $\mathbf{r}$  er kontinuerlig på  $[a, b]$  og den deriverte  $\mathbf{r}'(t)$  er kontinuerlig på det indre intervallet  $(a, b)$ . Hvis kurven er glatt, varierer altså hastighetsvektoren på en kontinuerlig måte når vi flytter oss langs kurven. En kontinuerlig parametrisering  $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$  er *stykkevis glatt* dersom vi kan dele  $[a, b]$  i et endelig antall delintervaller slik at  $\mathbf{r}$  er glatt på hvert av disse delintervallene. På en stykkevis glatt kurve kan det finnes unntakspunkter der hastighetsvektoren ikke eksisterer eller varierer på en diskontinuerlig måte. Figuren nedenfor viser et typisk eksempel på en stykkevis glatt kurve; kurven  $\mathcal{C}$  er satt sammen av de glatte delene  $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$  og  $\mathcal{C}_5$ .

Figur 1: En stykkevis glatt kurve  $\mathcal{C}$ 

**Definisjon 3.3.1** Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable, og at  $\mathbf{r} : [a, b] \rightarrow A$  er en stykkevis glatt parametrisering av en kurve  $\mathcal{C}$ . Linjeintegralet  $\int_{\mathcal{C}} f ds$  er definert ved

$$\int_{\mathcal{C}} f ds = \int_a^b f(\mathbf{r}(t)) v(t) dt$$

forutsatt at dette integralet eksisterer som et vanlig eller uegentlig integral.

Legg merke til at  $v(t)$  kan gå mot uendelig når vi nærmer oss et av punktene der  $\mathbf{r}$  ikke er glatt, og at integralet derfor kan divergere. I de eksemplene vi vanligvis skal arbeide med, vil  $v(t)$  være begrenset, og dette problemet vil ikke oppstå.

La oss se på et eksempel på hvordan vi regner ut et linjeintegral.

**Eksempel 1:** La oss regne ut  $\int_{\mathcal{C}} f ds$  når  $\mathcal{C}$  er enhetssirkelen parametrisert ved

$$\mathbf{r}(t) = \cos t \mathbf{i} + \sin t \mathbf{j} \quad t \in [0, 2\pi]$$

og  $f(x, y) = xy^2$ .

Vi regner først ut  $v(t) = \sqrt{x'(t)^2 + y'(t)^2} = \sqrt{(-\sin t)^2 + (\cos t)^2} = 1$ . Siden  $f(\mathbf{r}(t)) = f(\cos t, \sin t) = \cos t \sin^2 t$ , har vi

$$\int_{\mathcal{C}} f ds = \int_0^{2\pi} \cos t \sin^2 t \cdot 1 dt = \int_0^{2\pi} \cos t \sin^2 t dt$$

Vi løser dette integralet ved å sette  $u = \sin t$ . Da er  $du = \cos t dt$ ,  $u(0) = 0$ ,  $u(2\pi) = 0$ , så vi får

$$\int_{\mathcal{C}} f ds = \int_0^{2\pi} \cos t \sin^2 t dt = \int_0^0 u^2 du = 0$$

(du kan også se dette direkte fra integralet  $\int_0^{2\pi} \cos t \sin^2 t dt$  ved symmetri betraktnng). ♣

Kurveintegraller tilfredsstiller de samme reglene som vanlige integraler:

**Setning 3.3.2** Anta at  $\mathbf{r}$  er en stykkevis glatt parametriseing og at  $f, g$  er to kontinuerlige funksjoner slik at integralene  $\int_C f ds$  og  $\int_C g ds$  eksisterer. Da er

- (i)  $\int_C (f + g) ds = \int_C f ds + \int_C g ds$
- (ii)  $\int_C (f - g) ds = \int_C f ds - \int_C g ds$
- (iii)  $\int_C af ds = a \int_C f ds$  for alle  $a \in \mathbb{R}$ .

Disse resultatene følger direkte fra tilsvarende resultater for vanlige integraler, og vi tar ikke med bevisene her. Det neste resultatet forteller oss at vi kan regne ut et linjeintegral ved å dele opp kurven i mindre deler og regne ut integralet over hver del. Også i dette tilfellet følger resultatet direkte fra tilsvarende resultat for vanlige integraler, og vi sløyfer beviset.

**Setning 3.3.3** Anta at  $\mathbf{r}$  er en stykkevis glatt parametriseing av kurven  $C$  og at  $f$  er en kontinuerlig funksjon slik at integralet  $\int_C f ds$  eksisterer. Dersom

$$a = t_0 < t_1 < \dots < t_m = b$$

er en partisjon av  $[a, b]$ , og  $C_i$  er kurven parametrisert ved  $\mathbf{r} : [t_{i-1}, t_i] \rightarrow \mathbb{R}^n$  (vi deler altså kurven opp i  $m$  biter og lar  $C_i$  være den  $i$ -te biten), så er

$$\int_C f ds = \int_{C_1} f ds + \int_{C_2} f ds + \dots + \int_{C_m} f ds$$

Den samme kurven kan parametriseres på forskjellige måter, og det er naturlig å spørre om integralet  $\int_C f ds$  får samme verdi når vi bruker to forskjellige parametriseringer. Som vi nå skal se, er svaret “ja” forutsatt at vi har en fornuftig definisjon av hva det vil si at to parametriseringer fremstiller den samme kurven. La oss begynne med én parametrisering  $\mathbf{r}_1 : [a, b] \rightarrow \mathbb{R}^n$ , og se hva som skal til for å lage enn annen parametrisering  $\mathbf{r}_2 : [c, d] \rightarrow \mathbb{R}^n$  av den samme kurven. Hvis vi tenker på  $t$  som tiden, vil den første parametriseringen være i punktet  $\mathbf{r}_1(t)$  ved tiden  $t$ . Den andre parametriseringen må være i dette punktet ved et tidspunkt  $\phi(t)$  i intervallet  $[c, d]$ . Vi får altså sammenhengen

$$\mathbf{r}_2(\phi(t)) = \mathbf{r}_1(t)$$

der  $\phi$  er en funksjon med definisjonsmengde  $[a, b]$  og verdimengde  $[c, d]$ . Funksjonen  $\phi$  må være injektiv (hvis ikke tillater vi den ene parametriseringen å gå frem og tilbake langs kurven) og siden parametriseringene skal være (stykkevis) deriverbare med kontinuerlige deriverte, er det naturlig å kreve at  $\phi$  også er deriverbar med kontinuerlig derivert. Av litt mer tekniske grunner skal vi anta at  $\phi'(t) \neq 0$  for alle  $t \in [a, b]$ . I utgangspunktet kan det virke naturlig å anta at  $\phi$  er voksende slik at de to parametriseringene gjennomløper kurven i samme retning, men det viser seg å være lurt også å tillate avtagende  $\phi$  (i så fall vil  $\mathbf{r}_2$  gjennomløpe kurven i motsatt retning av  $\mathbf{r}_1$ ). La oss oppsummere denne diskusjonen i en definisjon.

**Definisjon 3.3.4** Anta at  $\mathbf{r}_1 : [a, b] \rightarrow \mathbb{R}^n$  og  $\mathbf{r}_2 : [c, d] \rightarrow \mathbb{R}^n$  er to stykkevis glatte parametriseringer. Vi sier at  $\mathbf{r}_1$  og  $\mathbf{r}_2$  er ekvivalente dersom det finnes en funksjon  $\phi : [a, b] \rightarrow [c, d]$  slik at:

- (i)  $\mathbf{r}_2(\phi(t)) = \mathbf{r}_1(t)$  for alle  $t \in [a, b]$ .
- (ii)  $\phi$  er kontinuerlig med verdimengde  $[c, d]$ .
- (iii)  $\phi'$  er kontinuerlig og forskjellig fra 0 på intervallet  $(a, b)$ .

Dersom  $\phi$  er strengt voksende, sier vi at  $\mathbf{r}_1$  og  $\mathbf{r}_2$  har samme orientering; dersom  $\phi$  er strengt avtagende, sier vi at de har motsatt orientering.

Legg merke til at siden funksjonen  $\phi'$  er kontinuerlig og forskjellig fra 0, må den (ifølge skjæringssetningen) enten være positiv overalt eller negativ overalt. Det betyr at  $\phi$  enten er strengt voksende eller strengt avtagende.

Vi kan nå bevise resultatet vi ønsker oss:

**Setning 3.3.5** Anta at  $\mathbf{r}_1 : [a, b] \rightarrow \mathbb{R}^n$  og  $\mathbf{r}_2 : [c, d] \rightarrow \mathbb{R}^n$  er to ekvivalente, stykkevis glatte parametriseringer av kurven  $\mathcal{C}$ . Da har integralet  $\int_{\mathcal{C}} f ds$  samme verdi uansett hvilken av parametriseringene vi bruker.

*Bevis:* Vi skal anta at parametriseringene er glatte; det generelle tilfellet får vi ved å skjøte sammen de glatte delene av stykkevis glatte parametriseringer. Vi skal også anta at funksjonen  $\phi$  er voksende; beviset for avtagende  $\phi$  går på samme måte, men er litt mer komplisert fordi vi må bytte fortegn to ganger underveis.

Bruker vi parametriseringen  $\mathbf{r}_1$ , har vi

$$I_1 = \int_{\mathcal{C}} f ds = \int_a^b f(\mathbf{r}_1(t)) v_1(t) dt$$

Siden  $\mathbf{r}_1(t) = \mathbf{r}_2(\phi(t))$ , gir kjerneregelen

$$\mathbf{v}_1(t) = \mathbf{r}'_1(t) = \mathbf{r}'_2(\phi(t))\phi'(t) = \mathbf{v}_2(\phi(t))\phi'(t)$$

Siden  $\phi'(t) > 0$ , gir dette  $v_1(t) = |\mathbf{v}_1(t)| = |\mathbf{v}_2(\phi(t))\phi'(t)| = v_2(\phi(t))\phi'(t)$ . Vi setter inn i uttrykket for  $I_1$  ovenfor, og får

$$I_1 = \int_{\mathcal{C}} f ds = \int_a^b f(\mathbf{r}_1(t)) v_1(t) dt = \int_a^b f(\mathbf{r}_2(\phi(t))) v_2(\phi(t)) \phi'(t) dt$$

Hvis vi nå innfører en ny variabel  $u = \phi(t)$ , får vi  $du = \phi'(t) dt$ . Siden  $\phi(a) = c$  og  $\phi(b) = d$ , har vi dermed

$$I_1 = \int_a^b f(\mathbf{r}_2(\phi(t))) v_2(\phi(t)) \phi'(t) dt = \int_c^d f(\mathbf{r}_2(u)) v_2(u) du$$

Det siste integralet er nøyaktig det vi får dersom vi prøver å regne ut linjeintegralet  $\int_{\mathcal{C}} f ds$  ved hjelp av parametriseringen  $\mathbf{r}_2$ , og beviset er dermed

fullført. □

Siden parametriseringen ikke spiller noen rolle, skal vi ofte referere til integralet  $\int_{\mathcal{C}} f \, ds$  uten å spesifisere hviken parametrisering vi tenker på. I det neste eksemplet får vi anledning til å bruke en del av de egenskapene vi nå har sett på.

**Eksempel 2:** Regn ut integralet  $\int_{\mathcal{C}} f \, ds$  når  $f(x, y) = x + y^2$  og  $\mathcal{C}$  er omkretsen til trekanten med hjørner  $(0, 0)$ ,  $(1, 0)$  og  $(0, 1)$ .

I dette eksemplet arbeider vi med en *lukket kurve*, dvs. en kurve som starter og ender i samme punkt. Det følger fra resultatene ovenfor (undersøk hvordan!) at det ikke spiller noen rolle hvilket punkt vi bruker som start- og endepunkt. Vi velger derfor å begynne i  $(0, 0)$ . Kurven  $\mathcal{C}$  består av tre naturlige deler; det horisontale linjestykket  $\mathcal{C}_1$  fra  $(0, 0)$  til  $(1, 0)$ , det skrå linjestykket  $\mathcal{C}_2$  fra  $(1, 0)$  til  $(0, 1)$  og det vertikale linjestykket  $\mathcal{C}_3$  fra  $(0, 1)$  til  $(0, 0)$ . Ifølge setning 3.3.3 er

$$\int_{\mathcal{C}} f \, ds = \int_{\mathcal{C}_1} f \, ds + \int_{\mathcal{C}_2} f \, ds + \int_{\mathcal{C}_3} f \, ds$$

og vi kan derfor ta for oss ett linjestykke av gangen.

Vi begynner med  $\mathcal{C}_1$ . Den enkleste parametriseringen av denne kurven er

$$\mathbf{r}_1(t) = t \mathbf{i} + 0 \mathbf{j} = t \mathbf{i} \quad t \in [0, 1]$$

Deriverer vi, får vi

$$\mathbf{v}_1(t) = 1 \mathbf{i} = \mathbf{i}$$

som gir  $v_1(t) = 1$ . Dermed er

$$\int_{\mathcal{C}_1} f \, ds = \int_0^1 f(\mathbf{r}_1(t)) v_1(t) \, dt = \int_0^1 (t + 0^2) \cdot 1 \, dt = \int_0^1 t \, dt = \frac{1}{2}$$

Neste skritt er å se på  $\mathcal{C}_2$ . En grei parametrisering av denne linjen er

$$\mathbf{r}_2(t) = (1 - t) \mathbf{i} + t \mathbf{j} \quad t \in [0, 1]$$

Vi får

$$\mathbf{v}_2(t) = -\mathbf{i} + \mathbf{j}$$

som gir  $v_2(t) = \sqrt{2}$ . Dermed er

$$\begin{aligned} \int_{\mathcal{C}_2} f \, ds &= \int_0^1 f(\mathbf{r}_2(t)) v_2(t) \, dt = \int_0^1 ((1 - t) + t^2) \cdot \sqrt{2} \, dt = \\ &= \sqrt{2} \int_0^1 (1 - t + t^2) \, dt = \sqrt{2} \left[ t - \frac{t^2}{2} + \frac{t^3}{3} \right]_0^1 = \frac{5}{6} \sqrt{2} \end{aligned}$$

Til slutt tar vi for oss  $\mathcal{C}_3$ . Vi bruker parametriseringen

$$\mathbf{r}_3(t) = 0\mathbf{i} + t\mathbf{j} \quad t \in [0, 1]$$

(den parametriserer riktignok kurven “gal vei” i forhold til den retningen vi hittil har gått i, men det spiller ingen rolle). Vi får

$$\mathbf{v}_3(t) = \mathbf{j}$$

og  $v_3(t) = 1$ . Dermed er

$$\int_{\mathcal{C}_3} f \, ds = \int_0^1 f(\mathbf{r}_3(t)) v_3(t) \, dt = \int_0^1 (0 + t^2) \cdot 1 \, dt = \int_0^1 t^2 \, dt = \frac{1}{3}$$

Kombinerer vi alle disse resultatene, får vi

$$\int_{\mathcal{C}} f \, ds = \frac{1}{2} + \frac{5}{6}\sqrt{2} + \frac{1}{3} = \frac{5}{6}(1 + \sqrt{2})$$

□

Vi avslutter med et eksempel på hvordan linjeintegraler kan brukes i praktiske situasjoner.

**Eksempel 3:** En bil befinner seg ved tiden  $t \in [0, 2]$  i punktet

$$\mathbf{r}(t) = 30t\mathbf{i} + 40t\mathbf{j} + 8t^{\frac{3}{2}}\mathbf{k}$$

Tiden er målt i timer og alle avstander i kilometer. Bensinforbruket i liter per kilometer er  $0.05 + 0.002v$  der  $v$  er farten. Vår oppgave er å finne det totale bensinforbruket.

Dersom bilen kjører med fart  $v$  i et tidsrom  $\Delta t$ , tilbakelegger den en strekning  $\Delta s = v\Delta t$ , og bensinforbruket i denne perioden er dermed

$$(0.05 + 0.002v)\Delta s = (0.05 + 0.002v)v\Delta t$$

Bruker vi Riemann-summer, ser vi at det totale forbruket i løpet av kjøreturen er

$$F = \int_{\mathcal{C}} (0.05 + 0.002v) \, ds = \int_0^2 (0.05 + 0.002v)v \, dt$$

For å finne  $v$ , regner vi først ut

$$\mathbf{v}(t) = 30\mathbf{i} + 40\mathbf{j} + 12t^{\frac{1}{2}}\mathbf{k}$$

som gir

$$v(t) = \sqrt{2500 + 144t}$$

Forbruket blir dermed

$$\begin{aligned} F &= \int_0^2 (0.05\sqrt{2500 + 144t} + 0.002(2500 + 144t)) dt = \\ &= \int_0^2 (0.05\sqrt{2500 + 144t} + 5 + 0.288t) dt = \\ &= \left[ \frac{0.05}{144} \cdot \frac{2}{3} (2500 + 144t)^{\frac{3}{2}} + 5t + 0.144t^2 \right]_0^2 \approx 15.7 \end{aligned}$$

Regningene i dette eksemplet er ikke direkte festlige, men slik er det ofte med linjeintegraler — utregningene blir stygge dersom de i det hele tatt lar seg utføre for hånd.



### Oppgaver til seksjon 3.3

**1.** Regn ut linjeintegralet  $\int_C f ds$  når  $f(x, y) = x$  og  $C$  er parametrisert ved

$$\mathbf{r}(t) = \sin t \mathbf{i} + \cos t \mathbf{j} \quad t \in [0, 2\pi]$$

**2.** Regn ut linjeintegralet  $\int_C f ds$  når  $f(x, y) = xy$  og  $C$  er parametrisert ved

$$\mathbf{r}(t) = 3t \mathbf{i} + 4t \mathbf{j} \quad t \in [0, 2]$$

**3.** Regn ut linjeintegralet  $\int_C f ds$  når  $f(x, y, z) = z \cos(xy)$  og  $C$  er parametrisert ved

$$\mathbf{r}(t) = 3t \mathbf{i} + 4t \mathbf{j} + 5t \mathbf{k} \quad t \in [0, \sqrt{\pi}]$$

**4.** Regn ut linjeintegralet  $\int_C f ds$  når  $f(x, y, z) = xz$  og  $C$  er parametrisert ved

$$\mathbf{r}(t) = 2t^3 \mathbf{i} + 3\sqrt{2}t^2 \mathbf{j} + 6t \mathbf{k} \quad t \in [0, 1]$$

**5.** Regn ut linjeintegralet  $\int_C f ds$  når  $f(x, y, z) = z$  og  $C$  er parametrisert ved

$$\mathbf{r}(t) = t \sin t \mathbf{i} + t \cos t \mathbf{j} + t \mathbf{k} \quad t \in [0, 2\pi]$$

**6.** Regn ut linjeintegralet  $\int_C f ds$  når  $f(x, y, z) = xyz$  og  $C$  er parametrisert ved

$$\mathbf{r}(t) = e^t \mathbf{i} - e^{-t} \mathbf{j} + \sqrt{2}t \mathbf{k} \quad t \in [0, 1]$$

**7.** Bevis setning 3.3.2.

**8.** Bevis setning 3.3.3.

**9.** Gjennomfør beviset for setning 3.3.5 når de to parametriseringene har motsatt orientering.

**10.** En prosjektert vei har form som kurven

$$\mathbf{r}(t) = (2t - t^2) \mathbf{i} + \frac{8}{3}t^{\frac{3}{2}} \mathbf{j} \quad \text{for } t \in [0, 1]$$

når den sees ovenfra (alle avstander er målt i kilometer). Utbygningskostnadene varierer langs veien på grunn av ulikheter i terrenget, og man regner med at prisen per kilometer er gitt ved en funksjonen av typen  $p(x, y) = K(10 + y)$  der  $K$  er en konstant. Finn de totale utbygningskostnadene.

**11.** (Eksamens i MA 105, 24/5-1991) En vei er parametrisert ved

$$\mathbf{r}(t) = \frac{t^2}{2} \mathbf{i} + \frac{2\sqrt{2}}{9} t^{3/2} \mathbf{j} + \frac{t}{9} \mathbf{k} \quad 1 \leq t \leq 7$$

(alle avstander er målt i kilometer). En bil som kjører langs veien har et bensinforbruk som avhenger av hvor bratt veien er — den bruker  $\frac{1}{15} + \frac{1}{2} \frac{dz}{ds}$  liter per kilometer ( $s$  er buelengden). Finn det totale bensinforbruket.

**12.** En kurve i polarkordinater er gitt ved en funksjon  $r = f(\theta)$  der  $\theta \in [a, b]$  (kurven består altså av alle de punktene som har polarkoordinater  $r$  og  $\theta$  der  $r = f(\theta)$  og  $\theta \in [a, b]$ ).

a) Vis at kurven har parametriseringen

$$\mathbf{r}(\theta) = f(\theta) \cos \theta \mathbf{i} + f(\theta) \sin(\theta) \mathbf{j} \quad \text{der } \theta \in [a, b]$$

b) Vis at farten er gitt ved

$$v(\theta) = \sqrt{f(\theta)^2 + f'(\theta)^2}$$

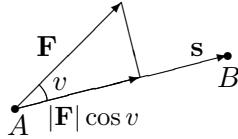
c) En kurve er gitt i polarkoordinater ved  $f(\theta) = \sin \theta$  der  $\theta \in [0, \pi]$ . Skisser kurven og regn ut buelengden.

d) Regn ut linjeintegralet  $\int_{\mathcal{C}} g \, ds$  der  $\mathcal{C}$  er kurven i punkt c) og  $g(x, y) = xy$ .

## 3.4 Linjeintegraler for vektorfelt

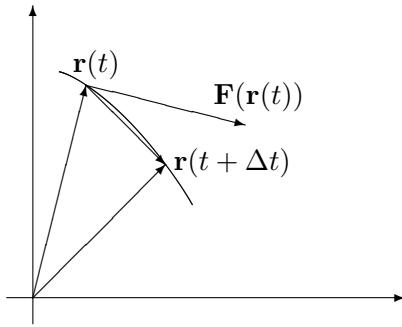
Vi skal nå studere den andre typen linjeintegral  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$ . Disse integralene har sine viktigste anvendelser i fysikk, og for at alle skal forstå de grunnleggende problemstillingene, begynner vi med et lynkurs i kraft og arbeid.

På ungdomsskolen lærte du at arbeid er kraft ganger vei; dersom du bruker en kraft  $F$  til å flytte en gjenstand en strekning  $s$ , så utfører kraften et arbeid  $W = Fs$ . Dette er riktig så sant du trekker med konstant kraft i bevegelsens retning. Dersom kraften ikke virker i bevegelsens retning, er det bare den komponenten av kraften som peker i bevegelsens retning som bidrar til arbeidet. Dersom kraften er gitt ved vektoren  $\mathbf{F}$ , og strekningen er gitt ved vektoren  $\mathbf{s}$ , er arbeidet  $W$  derfor lik  $|\mathbf{F}| |\mathbf{s}| \cos(v)$  der  $v$  er vinkelen mellom  $\mathbf{F}$  og  $\mathbf{s}$  (se figur 1). Siden  $|\mathbf{F}| |\mathbf{s}| \cos(v) = \mathbf{F} \cdot \mathbf{s}$ , er arbeidet  $W$  følgelig gitt av skalarproduktet  $W = \mathbf{F} \cdot \mathbf{s}$ .



Figur 1: Kraften  $\mathbf{F}$  utfører et arbeid  $W = |\mathbf{F}||\mathbf{s}| \cos(v) = \mathbf{F} \cdot \mathbf{s}$  fra  $A$  til  $B$

Spørsmålet vi skal se på, er hva som skjer når kraften varierer og bevegelsen skjer langs en kurve. Vi tenker oss at gjenstanden som beveger seg, er i punktet  $\mathbf{r}(t)$  ved tiden  $t$ , og at kraften som da virker på den, er gitt ved  $\mathbf{F}(\mathbf{r}(t))$ . I løpet av et kort tidsrom  $\Delta t$  vil gjenstanden forflytte seg til punktet  $\mathbf{r}(t + \Delta t)$  (se figur 2).



Figur 2: Arbeidet i tidsrommet fra  $t$  til  $t + \Delta t$

Arbeidet i dette tidsrommet må med god tilnærming være lik

$$\mathbf{F}(\mathbf{r}(t)) \cdot (\mathbf{r}(t + \Delta t) - \mathbf{r}(t))$$

Ganger og deler vi med  $\Delta t$ , ser vi at

$$\mathbf{F}(\mathbf{r}(t)) \cdot (\mathbf{r}(t + \Delta t) - \mathbf{r}(t)) = \mathbf{F}(\mathbf{r}(t)) \cdot \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t} \Delta t \approx \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) \Delta t$$

Vi tenker oss nå at hele arbeidet skjer i løpet av tidsintervallet  $[a, b]$ , og at vi har delt opp dette intervallet i små delintervaller på vanlig måte

$$a = t_0 < t_1 < t_2 < \dots < t_N = b$$

I det  $i$ -te delintervallet  $[t_{i-1}, t_i]$  utfører kraften da et arbeid som ifølge beregningene ovenfor er tilnærmet lik

$$\mathbf{F}(\mathbf{r}(t_i)) \cdot \mathbf{r}'(t_i)(t_i - t_{i-1})$$

Det totale arbeidet er dermed tilnærmet lik

$$\sum_i^N \mathbf{F}(\mathbf{r}(t_i)) \cdot \mathbf{r}'(t_i)(t_i - t_{i-1})$$

med bedre og bedre tilnærming dess finere oppdelingen blir. Vi ser at dette uttrykket er en Riemann-sum for funksjonen  $\mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$ , og det konvergerer derfor mot

$$\int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt$$

når oppdelingene blir finere og finere. Ut i fra slike betrakninger *definerer* fysikerne det arbeidet som en kraft  $\mathbf{F}$  utfører når en gjenstand flyttes langs en kurve  $\mathbf{r}$ , til å være

$$W = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt$$

Matematikerne bruker formelen til å definere en ny type linjeintegral (som blant annet kan brukes til å regne ut arbeid):

**Definisjon 3.4.1** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^n$  er en kontinuerlig funksjon av  $n$  variable, og at  $\mathbf{r} : [a, b] \rightarrow A$  er en stykkevis glatt parametrisering av en (orientert) kurve  $\mathcal{C}$ . Da er linjeintegralet  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  definert ved*

$$\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt$$

*forutsatt at integralet til høyre eksisterer som et vanlig eller uegentlig integral.*

**Bemerkning:** Legg merke til at selv om både  $\mathbf{F}(\mathbf{r}(t))$  og  $\mathbf{r}'(t)$  er vektorer, så er  $\mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$  en skalar, og

$$\int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt$$

er derfor et helt vanlig integral. Det viser seg (som vi snart skal se) at så lenge vi gjennomløper kurven i samme retning, spiller det ingen rolle hvilken parametrisering vi bruker til å regne ut kurveintegralet  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$ , men bytter vi gjennomløpsretning, skifter integralet fortegn. I definisjonen har vi derfor kalt  $\mathcal{C}$  en *orientert* kurve for å understreke at gjennomløpsretningen må være bestemt før det gir mening å regne ut integralet  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$ .

I definisjonen ovenfor er  $\mathbf{F}$  en funksjon fra (en delmengde av)  $\mathbb{R}^n$  til  $\mathbb{R}^n$ , dvs. at definisjonsmengden og verdimengden har samme dimensjon  $n$ . En slik funksjon kalles ofte et *vektorfelt*, spesielt når man tenker på anvendelser i geometri og fysikk.

La oss regne ut et linjeintegral:

**Eksempel 1:** Vi skal regne ut integralet  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  når

$$\mathbf{F}(x, y, z) = -x \mathbf{i} + yz \mathbf{j} + z \mathbf{k}$$

og

$$\mathbf{r}(t) = \cos t \mathbf{i} + \sin t \mathbf{j} + t \mathbf{k}, \quad t \in [0, 2\pi]$$

Vi regner først ut

$$\mathbf{r}'(t) = -\sin t \mathbf{i} + \cos t \mathbf{j} + \mathbf{k}$$

og observerer at

$$\mathbf{F}(\mathbf{r}(t)) = -\cos t \mathbf{i} + t \sin t \mathbf{j} + t \mathbf{k}$$

Dermed er

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{r} &= \int_0^{2\pi} \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt = \\ &= \int_0^{2\pi} (-\cos t \mathbf{i} + t \sin t \mathbf{j} + t \mathbf{k}) \cdot (-\sin t \mathbf{i} + \cos t \mathbf{j} + \mathbf{k}) dt = \\ &= \int_0^{2\pi} (\sin t \cos t + t \sin t \cos t + t) dt \end{aligned}$$

Nå gjenstår det bare vanlig integrasjon. Det første integralet  $\int \sin t \cos t dt$  kan løses ved substitusjon, men vi velger å bruke at  $\sin 2t = 2 \sin t \cos t$  isteden. Det gir

$$\int \sin t \cos t dt = \frac{1}{2} \int \sin 2t dt = -\frac{1}{4} \cos 2t + C$$

Bruker vi den samme omskrivningen i det andre integralet, ser vi at

$$\int t \sin t \cos t dt = \frac{1}{2} \int t \sin 2t dt$$

Her kan vi bruke delvis integrasjon med  $u = t$  og  $v' = \frac{1}{2} \sin 2t$ . Det gir  $u' = 1$  og  $v = -\frac{1}{4} \cos 2t$ , og dermed

$$\begin{aligned} \int t \sin t \cos t dt &= \frac{1}{2} \int t \sin 2t dt = -\frac{t}{4} \cos 2t + \frac{1}{4} \int \cos 2t dt = \\ &= -\frac{t}{4} \cos 2t + \frac{1}{8} \sin 2t + C \end{aligned}$$

Det tredje integralet er enkelt:

$$\int t dt = \frac{t^2}{2} + C$$

Kombinerer vi alle resultatene våre, har vi nå

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{r} &= \int_0^{2\pi} (\sin t \cos t + t \sin t \cos t + t) dt = \\ &= \left[ -\frac{1}{4} \cos 2t - \frac{t}{4} \cos 2t + \frac{1}{8} \sin 2t + \frac{t^2}{2} \right]_0^{2\pi} = -\frac{\pi}{2} + 2\pi^2 \end{aligned}$$
♣

Eksemplet ovenfor er ganske typisk; den verste jobben når man skal regne ut et linjeintegral, er ofte å løse de vanlige integralene man sitter igjen med til slutt!

De vanlige regnereglene for integraler gjelder også for linjeintegraler av vektorfelt.

**Setning 3.4.2** *Anta at  $\mathbf{r}$  er en stykkevis glatt parametrisering av en kurve  $C$  og at  $\mathbf{F}, \mathbf{G}$  er to kontinuerlige vektorfelt slik at integralene  $\int_C \mathbf{F} \cdot d\mathbf{r}$  og  $\int_C \mathbf{G} \cdot d\mathbf{r}$  eksisterer. Da er*

- (i)  $\int_C (\mathbf{F} + \mathbf{G}) \cdot d\mathbf{r} = \int_C \mathbf{F} \cdot d\mathbf{r} + \int_C \mathbf{G} \cdot d\mathbf{r}$
- (ii)  $\int_C (\mathbf{F} - \mathbf{G}) \cdot d\mathbf{r} = \int_C \mathbf{F} \cdot d\mathbf{r} - \int_C \mathbf{G} \cdot d\mathbf{r}$
- (iii)  $\int_C a\mathbf{F} \cdot d\mathbf{r} = a \int_C \mathbf{F} \cdot d\mathbf{r}$  for alle  $a \in \mathbb{R}$ .

Disse resultatene følger direkte fra tilsvarende resultater for vanlige integraler, og vi tar ikke med bevisene her. Det neste resultatet forteller oss at vi kan regne ut et linjeintegral ved å dele opp kurven i mindre deler og regne ut integralet over hver del. Også i dette tilfellet følger resultatet direkte fra tilsvarende resultat for vanlige integraler, og vi sløyfer beviset.

**Setning 3.4.3** *Anta at  $\mathbf{r}$  er en stykkevis glatt parametrisering av kurven  $C$  og at  $\mathbf{F}$  er et kontinuerlig vektorfelt slik at integralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  eksisterer. Dersom*

$$a = t_0 < t_1 < \dots < t_m = b$$

*er en partisjon av  $[a, b]$ , og  $C_i$  er kurven parametrisert ved  $\mathbf{r} : [t_{i-1}, t_i] \rightarrow \mathbb{R}^n$  (vi deler altså kurven opp i m biter og lar  $C_i$  være den i-te biten), så er*

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_{C_1} \mathbf{F} \cdot d\mathbf{r} + \int_{C_2} \mathbf{F} \cdot d\mathbf{r} + \dots + \int_{C_m} \mathbf{F} \cdot d\mathbf{r}$$

Som allerede nevnt, er kurveintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  uavhengig av parametriseringen så lenge vi gjennomløper kurven i samme retning. Her er det presise resultatet:

**Setning 3.4.4** *Anta at  $\mathbf{r}_1 : [a, b] \rightarrow \mathbb{R}^n$  og  $\mathbf{r}_2 : [c, d] \rightarrow \mathbb{R}^n$  er to ekvivalente, stykkevis glatte parametriseringer av kurven  $C$ . Dersom de to parametriseringene har samme orientering, får integralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  samme verdi uansett hvilken av dem vi bruker. Dersom parametriseringene har motsatt orientering, får integralene samme tallverdi, men motsatt fortegn.*

*Bevis:* Vi skal anta at parametriseringene er glatte; det generelle tilfellet får vi ved å skjøte sammen de glatte delene av stykkevis glatte parametriseringer. For å se hvor minusen kommer fra, tar vi for oss tilfellet der parametriseringene har motsatt orientering. Vi har altså to parametriseringer

$$\mathbf{r}_1 : [a, b] \rightarrow \mathbb{R}^n$$

$$\mathbf{r}_2 : [c, d] \rightarrow \mathbb{R}^n$$

forbundet med en funksjon  $\phi : [a, b] \rightarrow [c, d]$  slik at

$$\mathbf{r}_2(\phi(t)) = \mathbf{r}_1(t)$$

Siden parametriseringene har motsatt orientering, er  $\phi$  avtagende og  $\phi(a) = d$ ,  $\phi(b) = c$ .

Bruker vi parametriseringen  $\mathbf{r}_1$  til å regne ut integralet, har vi

$$I_1 = \int_C \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}_1(t)) \cdot \mathbf{r}'_1(t) dt$$

Siden  $\mathbf{r}_1(t) = \mathbf{r}_2(\phi(t))$ , gir kjerneregelen

$$\mathbf{r}'_1(t) = \mathbf{r}'_2(\phi(t))\phi'(t)$$

Vi setter inn i uttrykket for  $I_1$  ovenfor, og får

$$I_1 = \int_a^b \mathbf{F}(\mathbf{r}_1(t)) \cdot \mathbf{r}'_1(t) dt = \int_a^b \mathbf{F}(\mathbf{r}_2(\phi(t))) \cdot \mathbf{r}'_2(\phi(t)) \phi'(t) dt$$

Hvis vi nå innfører en ny variabel  $u = \phi(t)$ , får vi  $du = \phi'(t) dt$ . Siden  $\phi(a) = d$  og  $\phi(b) = c$ , har vi dermed

$$\begin{aligned} I_1 &= \int_a^b \mathbf{F}(\mathbf{r}_2(\phi(t))) \cdot \mathbf{r}'_2(\phi(t)) \phi'(t) dt = \int_d^c \mathbf{F}(\mathbf{r}_2(u)) \cdot \mathbf{r}'_2(u) du = \\ &= - \int_c^d \mathbf{F}(\mathbf{r}_2(u)) \cdot \mathbf{r}'_2(u) du \end{aligned}$$

Det siste integralet er nøyaktig det vi får dersom vi prøver å regne ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  ved hjelp av parametriseringen  $\mathbf{r}_2$ , og beviset er dermed fullført.  $\square$

Siden parametriseringen ikke spiller noen rolle, skal vi ofte referere til integralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  uten å spesifisere hviken parametrisering vi tenker på. Vi må imidlertid spesifisere retningen dersom det ikke fremgår fra sammenhengen.

Det neste eksemplet vi skal se på, er litt mer teoretisk, men det er slike "teoretiske" anvendelser som viser seg å ha de største praktiske konsekvensene. For virkelig å verdsette eksemplet er det sikkert en fordel å kunne litt fysikk, men her kommer et lynkurs i det aller viktigste. En gjenstand med masse  $m$  som beveger seg med en fart  $v$ , har en *kinetisk energi*  $\frac{1}{2}mv^2$ . Eksemplet viser at når en kraft virker på en gjenstand, går det tilførte arbeidet med til øke gjenstandens kinetiske energi.

**Eksempel 2:** Vi har en partikkel med masse  $m$ . I løpet av tidsrommet  $[a, b]$  forflytter partikkelen seg langs en bane  $\mathcal{C}$ , påvirket av en (total) kraft  $\mathbf{F}$ . Vi skal vise at

$$W = \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \frac{1}{2}mv(b)^2 - \frac{1}{2}mv(a)^2$$

dvs. at det tilførte arbeidet er lik endringen i kinetisk energi.

Ifølge Newtons annen lov er  $\mathbf{F} = m\mathbf{a}$ , der  $\mathbf{a} = \mathbf{v}' = \mathbf{r}''$  er akselerasjonen. Dermed er

$$W = \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = m \int_{\mathcal{C}} \mathbf{a} \cdot d\mathbf{r} = m \int_a^b \mathbf{a}(t) \cdot \mathbf{v}(t) dt$$

Vi observerer nå at siden  $v(t)^2 = \mathbf{v}(t) \cdot \mathbf{v}(t)$ , så sier regneregel 3.1.4(iii) at

$$(v(t)^2)' = (\mathbf{v}(t) \cdot \mathbf{v}(t))' = 2\mathbf{a}(t) \cdot \mathbf{v}(t)$$

Dermed er

$$W = m \int_a^b \mathbf{a}(t) \cdot \mathbf{v}(t) dt = \frac{1}{2}m \int_a^b (v(t)^2)' dt = \frac{1}{2}mv(b)^2 - \frac{1}{2}mv(a)^2$$

ifølge analysens fundamentalteorem. Vi skal arbeide videre med dette eksemplet i neste seksjon. ♣

Vi har nå sett på to typer linjeintegraler  $\int_{\mathcal{C}} f ds$  og  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  som virker nært beslektet til tross for at de har sine ulikheter. Det er faktisk mulig å oppfatte linjeintegraler av typen  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  som spesialtilfeller av den andre typen. Vi tar da utgangspunkt i definisjonen

$$\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt$$

og innfører enhetstangentvektoren

$$\mathbf{T}(t) = \frac{\mathbf{r}'(t)}{v(t)}$$

Skriver vi denne ligningen på formen  $\mathbf{r}'(t) = \mathbf{T}(t)v(t)$  og setter inn i formelen ovenfor, får vi

$$\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{T}(t)v(t) dt = \int_{\mathcal{C}} \mathbf{F} \cdot \mathbf{T} ds$$

Linjeintegralet av vektorfeltet  $\mathbf{F}$  er altså det samme som linjeintegralet av skalarfeltet  $\mathbf{F} \cdot \mathbf{T}$ . På grunn av denne sammenhengen bruker enkelte bøker notasjonen  $\int_{\mathcal{C}} \mathbf{F} \cdot \mathbf{T} ds$  for linjeintegraler av vektorfelt.

### Oppgaver til seksjon 3.4

**1.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = y\mathbf{i} + x\mathbf{j}$ , og kurven  $C$  er parametrisert ved

$$\mathbf{r}(t) = 2t\mathbf{i} - 3t\mathbf{j} \quad \text{der } t \in [1, 3]$$

**2.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = x^2\mathbf{i} + xy\mathbf{j}$ , og kurven  $C$  er parametrisert ved

$$\mathbf{r}(t) = \cos t\mathbf{i} - \sin t\mathbf{j} \quad \text{der } t \in [0, \frac{\pi}{2}]$$

**3.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y, z) = zy\mathbf{i} + x^2\mathbf{j} + xz\mathbf{k}$ , og kurven  $C$  er parametrisert ved

$$\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + t^3\mathbf{k} \quad \text{der } t \in [0, 2]$$

**4.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y, z) = \frac{z}{x}\mathbf{i} + y\mathbf{j} + x\mathbf{k}$ , og kurven  $C$  er parametrisert ved

$$\mathbf{r}(t) = e^t\mathbf{i} + \ln t\mathbf{j} + \sin t\mathbf{k} \quad \text{der } t \in [1, 2]$$

**5.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y, z) = yz\mathbf{i} + x\mathbf{j} + xy\mathbf{k}$ , og kurven  $C$  er parametrisert ved

$$\mathbf{r}(t) = t\mathbf{i} + \arctan t\mathbf{j} + t\mathbf{k} \quad \text{der } t \in [0, 1]$$

**6.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = x\mathbf{i} + y\mathbf{j}$ , og  $C$  er sirkelen med sentrum i origo og radius 5.  $C$  skal gjennomløpes i positiv retning (dvs. mot klokken).

**7.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = x^2y\mathbf{i} + xy\mathbf{j}$ , og  $C$  er den delen av parabelen  $y = x^2$  som tilsvarer  $x \in [-2, 2]$ . Kurven skal parametriseres fra venstre mot høyre.

**8.** La  $C$  være omkretsen til trekanten med hjørner i punktene  $(0, 0)$ ,  $(\pi, 0)$  og  $(\pi, \pi)$ . Regn ut  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = \cos x \sin y\mathbf{i} + x\mathbf{j}$  og  $C$  er positivt orientert.

**9.** Bevis setning 3.4.2.

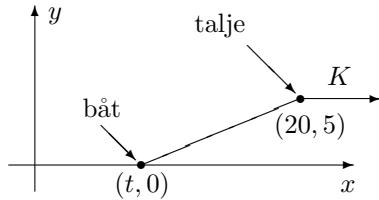
**10.** Bevis setning 3.4.3.

**11.** Gjenomfør beviset for setning 3.4.4 når de to parametriseringene har samme orientering.

**12.** Anta at kurven  $C$  er lukket. Vis at integralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  har samme verdi uansett hvilket punkt på kurven vi bruker som start-/stoppsted (forutsatt at orienteringen er den samme).

**13.** Helt til slutt i seksjonen påstod vi at linjeintegral av typen  $\int_C \mathbf{F} \cdot d\mathbf{r}$  kan oppfattes som integral av typen  $\int_C f ds$ ; alt vi behøver å gjøre er å velge  $f = \mathbf{F} \cdot \mathbf{T}$ . Hvordan kan dette være mulig når den ene typen integral er avhengig av orienteringen til kurven mens den andre typen ikke er det?

**14.** Figuren viser en båt som blir dratt bortover en flat strand med et tau. Tauet går gjennom en talje som er festet i punktet  $(20, 5)$ , og trekraften i tauet er konstant lik  $K$ . Posisjonen til båten ved tiden  $t \in [0, 20]$  er  $(t, 0)$ .

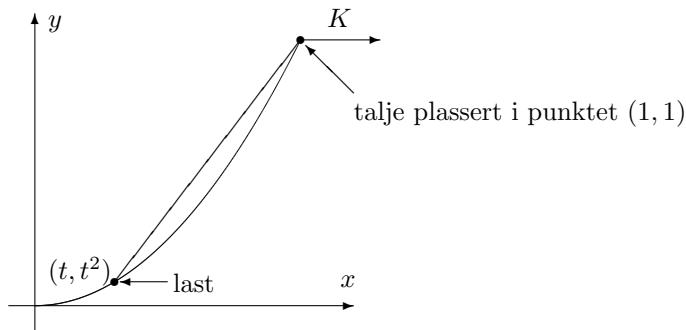


- a) Vis at arbeidet som kraften utfører er gitt ved

$$W = K \int_0^{20} \frac{20-t}{\sqrt{25+(20-t)^2}} dt$$

- b) Finn  $W$ .

**15.** Figuren viser en last som skal dras opp en islagt skråning ved hjelp av et tau. Med passende valgte enheter befinner lasten seg ved tiden  $t$  i punktet  $\mathbf{r}(t) = t \mathbf{i} + t^2 \mathbf{j}$  der  $t \in [0, 1]$ . Trekraften i tauet er konstant lik  $K$ .



- a) Vis at kraften som virker på lasten ved tiden  $t$  er

$$\mathbf{K}(t) = \frac{K}{\sqrt{2+2t+t^2}} \left( \mathbf{i} + (1+t) \mathbf{j} \right)$$

- b) Vis at arbeidet som kraften  $\mathbf{K}$  utfører på lasten fra bunn til topp er gitt ved:

$$W = K \int_0^1 \frac{1+2t+2t^2}{\sqrt{2+2t+t^2}} dt$$

- c) Deriver uttrykket  $(t-1)\sqrt{t^2+2t+2}$ .  
d) Bruk resultatet i punkt c) til å regne ut arbeidet  $W$ .

- e) (Forutsetter at du kjenner de hyperbolske funksjonene sinh og cosh; se kapittel 7.7 i *Kalkulus*). Løs integralet

$$W = \int_0^1 \frac{1+2t+2t^2}{\sqrt{2+2t+t^2}} dt$$

ved hjelp av substitusjonen  $\sinh u = t + 1$ .

### 3.5 Gradienter og konservative felt

Fra forrige seksjon husker vi at et *vektorfelt* er en funksjon  $\mathbf{F}$  fra (en delmengde av)  $\mathbb{R}^n$  inn i  $\mathbb{R}^n$ . Det er en type vektorfelt vi har arbeidet ganske mye med allerede, nemlig *gradienter*. Husk at dersom  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$  er et (deriverbart) skalarfelt, er gradienten til  $\phi$  gitt ved;

$$\nabla\phi(\mathbf{x}) = \left( \frac{\partial\phi}{\partial x_1}(\mathbf{x}), \frac{\partial\phi}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial\phi}{\partial x_n}(\mathbf{x}) \right)$$

Linjeintegraler av gradernter har en ganske spesiell egenskap.

**Setning 3.5.1** *Anta at  $\phi : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable med kontinuerlig gradient. Dersom  $\mathbf{r} : [a, b] \rightarrow A$  parametriserer en stykkevis glatt kurve  $C$  som begynner i punktet  $\mathbf{a}$  og ender i punktet  $\mathbf{b}$  (dvs.  $\mathbf{r}(a) = \mathbf{a}$  og  $\mathbf{r}(b) = \mathbf{b}$ ), så er*

$$\int_C \nabla\phi \cdot d\mathbf{r} = \phi(\mathbf{b}) - \phi(\mathbf{a})$$

*Bevis:* Vi skal først vise resultatet når  $C$  er glatt. For å unngå problemer i endepunktene ser vi først på en litt kortere kurve  $C_{c,d}$  gitt ved  $\mathbf{r} : [c, d] \rightarrow A$  der  $a < c < d < b$ . Ifølge kjerneregelen (se versjonen i setning 3.2.1) er

$$(\phi(\mathbf{r}(t)))' = \nabla\phi(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

for alle  $t \in [c, d]$ , og dermed er

$$\int_{C_{c,d}} \nabla\phi \cdot d\mathbf{r} = \int_c^d \nabla\phi(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt = \int_c^d (\phi(\mathbf{r}(t))') dt = \phi(\mathbf{r}(d)) - \phi(\mathbf{r}(c))$$

ifølge analysens fundamentalteorem. Tar vi grenseverdiene når  $c \rightarrow a$  og  $d \rightarrow b$ , får vi

$$\int_C \nabla\phi \cdot d\mathbf{r} = \phi(\mathbf{b}) - \phi(\mathbf{a})$$

Anta så at  $C$  er en stykkevis glatt kurve satt sammen av de glatte delene  $C_1, C_2, \dots, C_N$ , og la  $\mathbf{a} = \mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N = \mathbf{b}$  være start- og endepunktene til disse kurvene slik at  $C_i$  starter i  $\mathbf{a}_{i-1}$  og ender i  $\mathbf{a}_i$ . Siden vi allerede har vist resultatet for glatte kurver, vet vi at

$$\int_{C_i} \nabla\phi \cdot d\mathbf{r} = \phi(\mathbf{a}_i) - \phi(\mathbf{a}_{i-1})$$

Dermed er

$$\begin{aligned} \int_{\mathcal{C}} \nabla \phi \cdot d\mathbf{r} &= \int_{\mathcal{C}_1} \nabla \phi \cdot d\mathbf{r} + \int_{\mathcal{C}_2} \nabla \phi \cdot d\mathbf{r} + \cdots + \int_{\mathcal{C}_N} \nabla \phi \cdot d\mathbf{r} = \\ &= (\phi(\mathbf{a}_1) - \phi(\mathbf{a}_0)) + (\phi(\mathbf{a}_2) - \phi(\mathbf{a}_1)) + \cdots + (\phi(\mathbf{a}_N) - \phi(\mathbf{a}_{N-1})) = \\ &\quad \phi(\mathbf{a}_N) - \phi(\mathbf{a}_0) = \phi(\mathbf{b}) - \phi(\mathbf{a}) \end{aligned}$$

og setningen er bevist.  $\square$

Setningen ovenfor forteller oss at når vi integrerer en gradient fra et punkt  $\mathbf{a}$  til et punkt  $\mathbf{b}$ , så spiller det ingen rolle hvilken kurve  $\mathcal{C}$  vi følger; svaret blir alltid  $\phi(\mathbf{b}) - \phi(\mathbf{a})$ . Vi sier at linjeintegralet  $\int_{\mathcal{C}} \nabla \phi \cdot d\mathbf{r}$  er *uavhengig av veien*. Spesielt ser vi at integralet langs en lukket kurve (dvs. en kurve som begynner og ender i samme punkt) alltid er 0.

**Eksempel 1:** Dersom vi deriverer skalarfeltet  $\phi(x, y, z) = yz^2e^x$ , får vi

$$\mathbf{F}(x, y, z) = \nabla \phi(x, y, z) = yz^2e^x \mathbf{i} + z^2e^x \mathbf{j} + 2yze^x \mathbf{k}$$

Hvis  $\mathcal{C}$  er kurven parametrisert ved

$$\mathbf{r}(t) = t \mathbf{i} + 2e^{-t} \mathbf{j} + t^2 \mathbf{k}, \quad t \in [0, 2]$$

vet vi at

$$\begin{aligned} \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} &= \phi(\mathbf{r}(2)) - \phi(\mathbf{r}(0)) = \phi(2, 2e^{-2}, 4) - \phi(0, 2, 0) = \\ &= 2e^{-2} \cdot 4^2 \cdot e^2 - 2 \cdot 0^2 \cdot e^0 = 32 \end{aligned}$$

Vi kan altså regne ut linjeintegralet uten å integrere.  $\clubsuit$

La oss innføre litt terminologi.

**Definisjon 3.5.2** Dersom vektorfeltet  $\mathbf{F}$  er lik gradienten til et skalarfelt  $\phi$  i et område  $A$  (vi har altså  $\mathbf{F}(\mathbf{x}) = \nabla \phi(\mathbf{x})$  for alle  $\mathbf{x} \in A$ ), sier vi at  $\mathbf{F}$  er konservativt i  $A$ , og vi kaller  $\phi$  en potensialfunksjon for  $\mathbf{F}$  i  $A$ .

**Bemerkning:** Ordet “konservativt” innebærer ikke at feltene våre har en spesiell politisk overbevisning — navnet skyldes at konservative kraftfelt bevarer (*konserverer*) energi. Vi skal vise dette til slutt i seksjonen. Legg for øvrig merke til at potensialfunksjonene til et vektorfelt bare er definert opp til en konstant; er  $\phi$  en potensialfunksjon for  $\mathbf{F}$ , er  $\phi + C$  det også.

To av de viktigste kreftene i naturen, gravitasjonskraft og elektrostatisk kraft, har samme matematiske form. Dersom kilden til kraften er plassert i origo, er kraften i punktet  $\mathbf{x} \neq \mathbf{0}$  gitt ved

$$\mathbf{F}(\mathbf{x}) = \frac{k}{|\mathbf{x}|^3} \mathbf{x}$$

der  $k$  er en konstant. Det neste eksemplet viser at disse kreftene er konervative med potensialfunksjon

$$\phi(\mathbf{x}) = -\frac{k}{|\mathbf{x}|}$$

**Eksempel 2:** Dersom vi deriverer funksjonen

$$\phi(\mathbf{x}) = -\frac{k}{|\mathbf{x}|} = -k(x_1^2 + x_2^2 + \cdots + x_i^2 + \cdots + x_n^2)^{-\frac{1}{2}}$$

med hensyn på  $x_i$ , får vi

$$\frac{\partial \phi}{\partial x_i} = -k\left(-\frac{1}{2}\right)(x_1^2 + x_2^2 + \cdots + x_i^2 + \cdots + x_n^2)^{-\frac{3}{2}} 2x_i = \frac{k}{|\mathbf{x}|^3} x_i$$

Dermed er

$$\nabla \phi(\mathbf{x}) = \frac{k}{|\mathbf{x}|^3} (x_1, x_2, \dots, x_i, \dots, x_n) = \frac{k}{|\mathbf{x}|^3} \mathbf{x}$$

akkurat som vi skulle vise. ♣

At de grunnleggende fysiske kreftene er konservative, har store konsekvenser både teoretisk og beregningsmessig. På den beregningsmessige siden vet vi nå at dersom en partikkel beveger seg fra punkt  $\mathbf{a}$  til punkt  $\mathbf{b}$  påvirket av en kraft  $\mathbf{F}(\mathbf{x}) = \frac{k}{|\mathbf{x}|^3} \mathbf{x}$ , så er arbeidet rett og slett gitt ved  $\phi(\mathbf{b}) - \phi(\mathbf{a}) = \frac{k}{|\mathbf{a}|} - \frac{k}{|\mathbf{b}|}$ . Fra et teoretisk synspunkt er resultatet enda større og viktigere — vi skal se et eksempel på dette mot slutten av seksjonen.

Det er to viktige spørsmål vi nå må se på: Hvordan avgjør vi om et vektorfelt er konservativt, og hvordan finner vi i så fall en potensialfunksjon? Vi begynner med en nødvendig betingelse.

**Setning 3.5.3** *Anta at  $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), \dots, F_n(\mathbf{x}))$  er et vektorfelt med kontinuerlige partiellderiverte. Dersom  $\mathbf{F}$  er konservativt i området  $A$ , er*

$$\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$$

*for alle  $\mathbf{x} \in A$  og alle indekser  $i, j$ .*

*Bevis:* Siden  $\mathbf{F}$  er konservativt, finnes det en potensialfunksjon  $\phi$  slik at  $F_i = \frac{\partial \phi}{\partial x_i}$ . Dermed er

$$\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial^2 \phi}{\partial x_j \partial x_i}$$

Tilsvarende er

$$\frac{\partial F_j}{\partial x_i}(\mathbf{x}) = \frac{\partial^2 \phi}{\partial x_i \partial x_j}$$

Siden blandede partiellderiverte er like, følger setningen.  $\square$

Legg merke til at setningen ovenfor ikke garanterer at det finnes en potensialfunksjon når betingelsene  $\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$  er oppfylt — alt den sier er at når disse betingelsene *ikke* er oppfylt, så kan feltet *ikke* være konservativt. I det neste eksemplet er betingelsene oppfylt (i alle punkter unntatt origo), men feltet er allikevel ikke konservativt.

**Eksempel 3:** Vi ser på feltet

$$\mathbf{F}(x, y) = F_1(x, y) \mathbf{i} + F_2(x, y) \mathbf{j} = -\frac{y}{x^2 + y^2} \mathbf{i} + \frac{x}{x^2 + y^2} \mathbf{j}$$

for  $(x, y) \neq (0, 0)$ . Regner vi ut de partiellderiverte, får vi

$$\frac{\partial F_1}{\partial y} = \frac{y^2 - x^2}{(x^2 + y^2)^2}$$

og

$$\frac{\partial F_2}{\partial x} = \frac{y^2 - x^2}{(x^2 + y^2)^2}$$

Betingelsene for de partiellderiverte er altså oppfylt, men feltet er likevel ikke konservativt i  $\mathbb{R}^2 \setminus \{0, 0\}$ . For å vise dette, integrerer vi  $\mathbf{F}$  en gang rundt enhetssirkelen. Dersom feltet var konservativt, ville dette integralet blitt 0 ifølge setning 3.5.1, men bruker vi parametriseringen

$$\mathbf{r}(t) = \cos t \mathbf{i} + \sin t \mathbf{j} \quad t \in [0, 2\pi]$$

får vi:

$$\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_0^{2\pi} (-\sin t \mathbf{i} + \cos t \mathbf{j}) \cdot (-\sin t \mathbf{i} + \cos t \mathbf{j}) dt = \int_0^{2\pi} 1 dt = 2\pi$$

Dette viser at feltet ikke er konservativt i  $\mathbb{R}^2 \setminus \{0, 0\}$ .  $\clubsuit$

Eksemplet ovenfor er ikke fullt så problematisk som man kanskje skulle tro — vanskelighetene skyldes at vi har forsøkt å finne en potensialfunksjon på et område som ikke er *enkeltsammenhengende*. At et åpent område  $A$  er enkeltsammenhengende, betyr at det er sammenhengende, og at enhver

lukket, kontinuerlig kurve i  $A$  kan snurpes sammen til et punkt uten at den forlater  $A$ . Området  $\mathbb{R}^2 \setminus \{(0, 0)\}$ , som vår funksjon  $\mathbf{F}$  er definert på, er ikke enkeltsammenhengende siden en kurve som omslutter origo, ikke kan snurpes sammen til et punkt uten å forlate området.

Man kan vise at et vektorfelt  $\mathbf{F}$  med kontinuerlige partiellderiverte har en potensialfunksjon på et åpent, enkeltsammenhengende område  $A \subset \mathbb{R}^n$  hvis og bare hvis betingelsene  $\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$  er oppfylt for alle  $\mathbf{x} \in A$  (beviset er litt for vanskelig for oss, men vi skal vise at resultatet holder i stjerneformede områder i seksjon 5.3). Spesielt betyr dette at dersom  $\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$  er oppfylt for alle  $\mathbf{x} \in \mathbb{R}^n$ , så har  $\mathbf{F}$  en potensialfunksjon på hele  $\mathbb{R}^n$ .

Neste post på programmet er å studere hvordan man finner potensialfunksjoner når de eksisterer. Vi viser fremgangsmåten gjennom et eksempel:

**Eksempel 4:** Vi skal finne en potensialfunksjon til vektorfeltet

$$\mathbf{F}(x, y, z) = (2xyz + y)\mathbf{i} + (x^2z + x)\mathbf{j} + (x^2y + 1)\mathbf{k}$$

Det er lett å sjekke at  $\frac{\partial F_1}{\partial y} = \frac{\partial F_2}{\partial x}$ ,  $\frac{\partial F_1}{\partial z} = \frac{\partial F_3}{\partial x}$  og  $\frac{\partial F_2}{\partial z} = \frac{\partial F_3}{\partial y}$  overalt, så ifølge det vi nettopp har lært, har  $\mathbf{F}$  en potensialfunksjon  $\phi$ . Denne funksjonen må tilfredsstille

$$\begin{aligned}\frac{\partial \phi}{\partial x} &= F_1(x, y, z) = 2xyz + y \\ \frac{\partial \phi}{\partial y} &= F_2(x, y, z) = x^2z + x \\ \frac{\partial \phi}{\partial z} &= F_3(x, y, z) = x^2y + 1\end{aligned}$$

Integrerer vi den første av disse ligningene med hensyn på  $x$ , ser vi at  $\phi$  må være på formen

$$\phi(x, y, z) = x^2yz + xy + C_1(y, z)$$

der  $C_1(y, z)$  er en funksjon som bare avhenger av  $y$  og  $z$  (en slik funksjon blir borte når vi deriverer mhp.  $x$ ). Integrerer vi den andre av ligningene med henyn på  $y$ , ser vi at  $\phi$  må være på formen

$$\phi(x, y, z) = x^2yz + xy + C_2(x, z)$$

der  $C_2(x, z)$  er en funksjon som bare avhenger av  $x$  og  $z$ . Integrerer vi den tredje av ligningene med henyn på  $z$ , ser vi at  $\phi$  må være på formen

$$\phi(x, y, z) = x^2yz + z + C_3(x, y)$$

der  $C_3(x, y)$  er en funksjon som bare avhenger av  $x$  og  $y$ . Samler vi sammen ledd fra hver av disse ligningene, ser vi at

$$\phi(x, y, z) = x^2yz + xy + z$$

tilfredsstiller alle kravene (det tilsvarer å velge  $C_1(y, z) = z$ ,  $C_2(x, z) = z$  og  $C_3(x, y) = xy$ ). Det er lett å sjekke ved derivasjon at  $\phi$  er en potensialfunksjon til  $\mathbf{F}$ . ♣

Vi skal avslutte med et litt mer teoretisk eksempel. Dersom  $\mathbf{F}$  er et konservativt kraftfelt med potensialfunksjon  $\phi$ , så kalles funksjonen  $E_p(\mathbf{x}) = -\phi(\mathbf{x})$  den *potensielle energien* (det kan virke litt merkelig med minustegnet, men det er naturlig ut i fra fysiske betraktninger — man ønsker at energien skal være større dess mer “kraftfull” situasjonen er). Den totale energien til en funksjon er summen av potensiell og kinetisk energi; dersom en partikkel befinner seg i punktet  $\mathbf{x}$  med hastighet  $\mathbf{v}$ , er altså den totale energien  $E = -\phi(\mathbf{x}) + \frac{1}{2}mv^2$ . Vi skal nå vise at når en partikkel beveger seg i et konservativt kraftfelt, er den totale energien bevart. Dette er et av de viktigste prinsippene i klassisk mekanikk.

**Eksempel 5:** Vi ser på en partikkel som beveger seg langs en kurve  $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^n$  i et kraftfelt  $\mathbf{F}$ . I eksempel 2 i forrige seksjon så vi at det arbeidet kraften utfører på partikkelen, fører til en endring i den kinetiske energien

$$W = \int_C \mathbf{F} \cdot d\mathbf{r} = \frac{1}{2}mv(b)^2 - \frac{1}{2}mv(a)^2$$

Dersom kraften er konservativ ( $\mathbf{F} = \nabla\phi$ ), har vi i tillegg

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \phi(\mathbf{r}(b)) - \phi(\mathbf{r}(a))$$

Kombinerer vi disse ligningene, ser vi at

$$-\phi(\mathbf{r}(b)) + \frac{1}{2}mv(b)^2 = -\phi(\mathbf{r}(a)) + \frac{1}{2}mv(a)^2$$

Siden  $E_p(\mathbf{x}) = -\phi(\mathbf{x})$ , er dette det samme som

$$E_p(\mathbf{r}(b)) + \frac{1}{2}mv(b)^2 = E_p(\mathbf{r}(a)) + \frac{1}{2}mv(a)^2$$

som viser at energien er den samme før og etter bevegelsen. ♣

### Oppgaver til seksjon 3.5

I oppgave 1 til 6 skal du avgjøre om feltet er konservativt og i så fall finne en potensialfunksjon.

1.  $\mathbf{F}(x, y) = (2xy + 2x)\mathbf{i} + x^2\mathbf{j}$

2.  $\mathbf{F}(x, y) = (2xe^y)\mathbf{i} + (x^2e^y + x)\mathbf{j}$

**3.**  $\mathbf{F}(x, y) = (\sin xy + xy \cos xy + 3) \mathbf{i} + (x^2 \cos xy - 2) \mathbf{j}$

**4.**  $\mathbf{F}(x, y, z) = (y^2 z + z) \mathbf{i} + (2xyz - 2) \mathbf{j} + (xy^2 + x) \mathbf{k}$

**5.**  $\mathbf{F}(x, y, z) = (y \sin z + 2z) \mathbf{i} + (x \sin z + 3) \mathbf{j} + (xy \cos z + 2x) \mathbf{k}$

**6.**  $\mathbf{F}(x, y, z) = x^2 e^x y \mathbf{i} + x^3 e^x y \mathbf{j} + xyz e^x \mathbf{k}$

**7.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = 2xy \mathbf{i} + x^2 \mathbf{j}$  og  $C$  er kurven parametrisert ved

$$\mathbf{r}(t) = 2t \cos t \mathbf{i} + \sin t \mathbf{j} \quad t \in [0, \frac{\pi}{2}]$$

**8.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = (\cos(xy) - xy \sin(xy)) \mathbf{i} - x^2 \sin(xy) \mathbf{j}$  og  $C$  er kurven parametrisert ved

$$\mathbf{r}(t) = -t \cos t \mathbf{i} + (\sin t - \cos t) \mathbf{j} \quad t \in [0, \pi]$$

**9.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y) = y^2 e^{xy^2} \mathbf{i} + (2xye^{xy^2} + 1) \mathbf{j}$  og  $C$  sirkelen med sentrum i  $(1, -1)$  og radius 5 (vi gir  $C$  positiv orientering).

**10.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y, z) = (y^2 z + 2xy) \mathbf{i} + (2xyz + x^2) \mathbf{j} + (xy^2 + 1) \mathbf{k}$  og  $C$  er kurven parametrisert ved

$$\mathbf{r}(t) = t \mathbf{i} + t^2 \mathbf{j} + t \sin \frac{\pi t}{2} \mathbf{k} \quad t \in [0, 1]$$

**11.** Regn ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  når  $\mathbf{F}(x, y, z) = ze^{xz+y} \mathbf{i} + (e^{xz+y} + 2z) \mathbf{j} + (xe^{xz+y} + 2y) \mathbf{k}$  og  $C$  er skjæringskurven kurven mellom kulen  $x^2 + y^2 + z^2 = 25$  og planet  $x - 2y + 3z = 1$ . Kurven har positiv orientering sett ovenfra.

**12.** I denne oppgaven skal vi se nærmere på vektorfeltet

$$\mathbf{F}(x, y) = -\frac{y}{x^2 + y^2} \mathbf{i} + \frac{x}{x^2 + y^2} \mathbf{j}$$

i eksempel 3.

- a) La  $\phi_1(x, y) = \arctan \frac{y}{x} + C$  der  $C$  er en konstant. Vis at  $\nabla \phi_1(x, y) = \mathbf{F}(x, y)$  når  $x \neq 0$ .
- b) Regn ut  $\int_C \mathbf{F} \cdot d\mathbf{r}$  der  $C$  er en glatt kurve som ligger til høyre for  $y$ -aksen og som starter i punktet  $(1, -1)$  og ender i  $(3, 3)$ .
- c) La  $\phi_2(x, y) = -\arctan \frac{x}{y} + C$  der  $C$  er en konstant. Vis at  $\nabla \phi_2(x, y) = \mathbf{F}(x, y)$  når  $y \neq 0$ .
- d) Bruk MATLAB eller en lommeregner til å tegne grafene til  $\phi_1$  og  $\phi_2$  (husk at  $\arctan$  heter `atan` i MATLAB).
- e) Finn sammenhengen mellom  $\arctan \frac{y}{x}$  og  $-\arctan \frac{x}{y}$  (det kan lønne seg å se på hver kvadrant for seg).

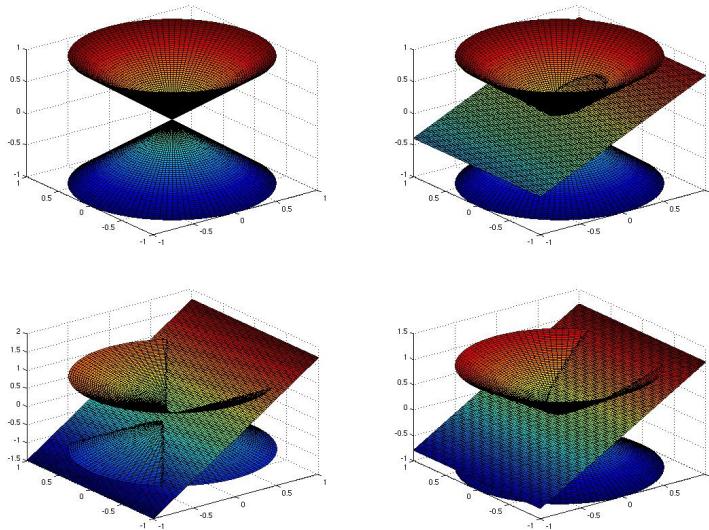
- f) Finn en potensialfunksjon  $\phi$  for  $\mathbf{F}$  i området

$$A = \{(x, y) \in \mathbb{R}^2 \mid y \text{ ligger ikke på den negative } y\text{-aksen}\}$$

Forklar hvorfor du ikke kan utvide denne funksjonen  $\phi$  til en kontinuerlig funksjon på hele  $\mathbb{R}^2$ .

## 3.6 Kjeglesnitt

I denne seksjonen skal vi se på en familie av kurver som dukker opp i mange sammenhenger — parabler, ellipser og hyperbler. Med et fellesnavn kalles disse kurvene *kjeglesnitt* fordi de fremkommer som snittkurver når man skjærer over en kjegle på skrå. Figur 1 viser hvordan dette foregår. Vi starter med en dobbelkjegle som vist øverst i venstre hjørne. Skjærer vi denne dobbelkjeglen med et plan som er *mindre bratt* en sidekanten i kjeglen, får vi figuren øverst til høyre. Snittkurven mellom kjeglen og planet er en *ellipse*. I figuren nedest til venstre har vi skåret dobbelkjeglen med et plan som er *brattere* enn kjeglekanten. Planet skjærer nå både den nedre og øvre delen av kjeglen, og vi får en skjæringskurve som består av to deler — en *hyperbel*. I den siste figuren har vi skåret over kjeglen med et plan som er *parallel*t med sidekanten. Skjæringskurven blir i dette tilfellet en *parabel*.

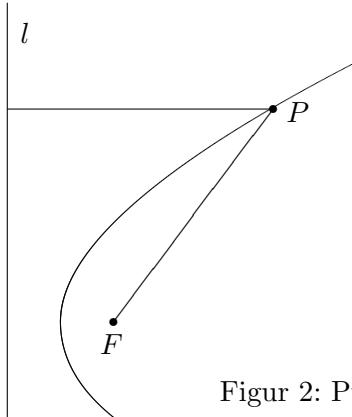


Figur 1: Romgeometrisk definisjon av kjeglesnitt

Selv om den romgeometriske definisjonen ovenfor forklarer ordet *kjeglesnitt*, er det i våre dager vanligere å behandle disse kurvene med plangeometriske metoder. Vi skal følge denne tradisjonen, men helt på slutten av seksjonen skal vi ta oss tid til å vise at det er en sammenheng mellom de romgeometriske definisjonene ovenfor og de plangeometriske definisjonen vi ellers benytter oss av.

## Parabler

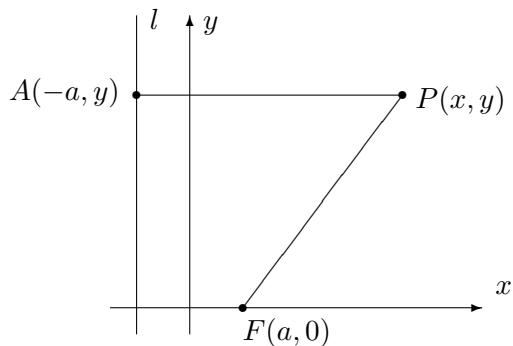
Vi starter med parablene som på mange måter er de enkleste kjeglesnittene. Tenk deg at du har en linje  $l$  og et punkt  $F$  i planet. Vi antar at  $F$  ikke ligger på  $l$ , og er på jakt etter de punktene  $P$  som ligger like langt fra  $l$  som fra  $F$ . Disse punktene danner en kurve, og denne kurven kaller vi *parabelen med brennpunkt  $F$  og styrelinje  $l$* .



Figur 2: Punkter som ligger like langt fra  $l$  som fra  $F$

For å finne en ligning som beskriver kurven, legger vi inn et koordinatsystem som vist på figur 3 med  $x$ -akse gjennom  $F$  loddrett på  $l$ , og med  $y$ -akse midtveis mellom  $l$  og  $F$ . Kaller vi avstanden fra  $F$  til  $l$  for  $2a$ , vil  $F$  ha koordinater  $(a, 0)$  og  $l$  ha ligning  $x = -a$ . På figuren er avstanden fra  $l$  til  $P$  lik avstanden fra  $A$  til  $P$ , dvs.  $x + a$ . Ifølge Pythagoras er avstanden fra  $F$  til  $P$  lik  $\sqrt{(x - a)^2 + y^2}$ . Disse avstandene skal være like, dvs.

$$x + a = \sqrt{(x - a)^2 + y^2}$$



Figur 3: Ligningen for en parabel

Kvadrerer vi på begge sider, får vi

$$(x + a)^2 = (x - a)^2 + y^2$$

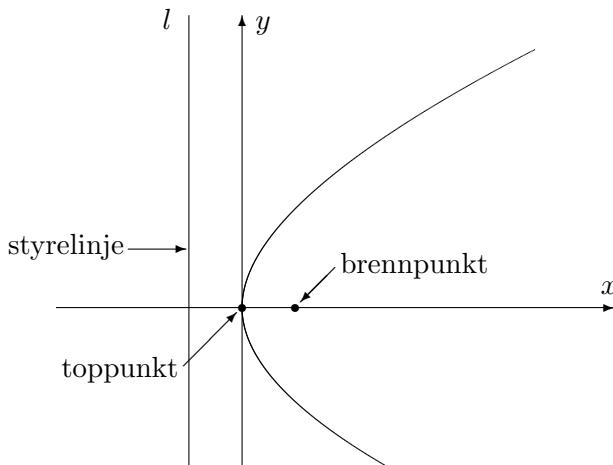
Ganger vi ut og forkorter, sitter vi igjen med

$$y^2 = 4ax$$

Alle punkter på parabelen må altså tilfredsstille denne ligningen, og det er lett å sjekke at alle punkter som tilfredsstiller ligningen, ligger på parabelen (det har altså ikke kommet med noen “falske løsninger” når vi kvadrerte ligningen). Vi har dermed vist:

**Setning 3.6.1** *Parabelen med brennpunkt  $F(a, 0)$  og styrelinje  $x = -a$  har ligning*

$$y^2 = 4ax$$



Figur 4: Parabel med brennpunkt og toppunkt

Punktet på parabelen som ligger nærmest styrelinjen, kaller vi for *toppunktet* (selv om det ikke ligger på toppen!). Avstanden  $a$  fra brennpunktet til toppunktet kaller vi *brennvidden* til parabelen (se figur 4).

I de parablene vi hittil har sett på, ligger toppunktet i origo, men det går selvfølgelig an å flytte parabelen slik at toppunktet får en annen posisjon. Flytter vi toppunktet til  $(m, n)$ , får parabelen ligning

$$(y - n)^2 = 4a(x - m)$$

**Eksempel 1:** Vis at ligningen

$$y^2 + 4y - 8x + 20 = 0$$

beskriver en parabel, og finn toppunkt og brennpunkt.

Vi fullfører først kvadratet:

$$y^2 + 4y - 8x + 20 = y^2 + 4y + 4 - 4 - 8x + 20 = (y + 2)^2 - 8x + 16$$

Ligningen kan dermed skrives

$$(y + 2)^2 - 8x + 16 = 0 \iff (y + 2)^2 = 8(x - 2)$$

som beskriver en parabel med toppunkt  $(2, -2)$  og brennvidde 2. Brennpunktet ligger i  $(4, -2)$ . 

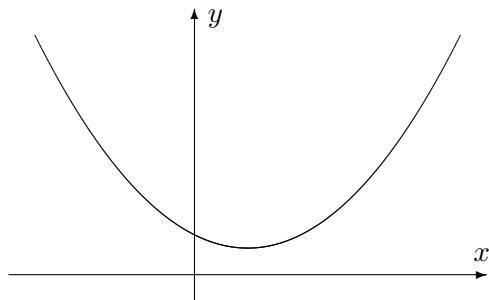
For å få en parabel som vender den andre veien (dvs. med åpningen mot venstre), trenger vi et minustegn:

$$(y - n)^2 = -4a(x - m)$$

Bytter vi om rollene til  $x$  og  $y$ , dvs. at vi lar

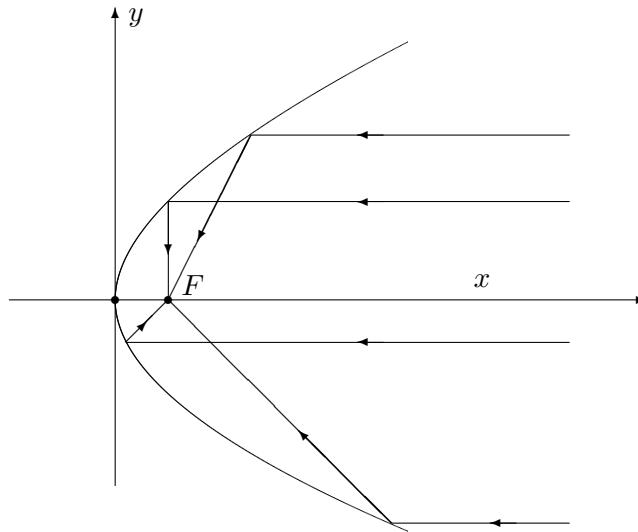
$$(x - m)^2 = 4a(y - n),$$

får vi en liggende parabel som i figur 5:



Figur 5: En parabelen på formen  $(x - m)^2 = 4a(y - n)$

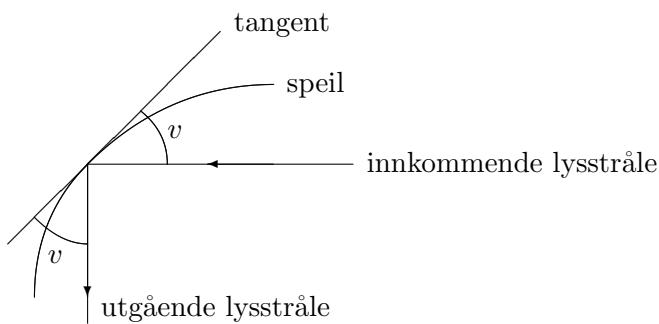
Putter vi på et minus slik at vi får  $(x - m)^2 = -4a(y - n)$ , vil parabelen snu slik at den har toppunktet øverst (sic!) og åpningen nedover.



Figur 6: Lysstråler reflekteres gjennom brennpunktet

Vi skal nå se på en viktig egenskap ved parabler. Vi tenker oss at parabelen i figur 6 er belagt med speil på innsiden. Lysstråler kommer inn parallelt med aksen og reflekteres når de treffer parabelen. Det viser seg at alle lysstrålene da blir reflektert gjennom brennpunktet  $F$  (det er nettopp derfor det heter *brennpunkt*).

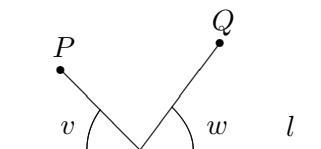
Denne *refleksjonsegenskapen* for parabler brukes i parabolantennener for å samle alle signalene som treffer antennen i ett område. Egenskapen brukes også i billykter og lommelykter, men da i motsatt retning — plasserer man lyspæren i brennpunktet, vil det parabolske speilet reflektere alt lyset i samme retning.



Figur 7: Vinkler ved refleksjon

For å bevise refleksjonsegenskapen trenger vi tre ingredienser.

1. Først litt fysikk: Når en lysstråle treffer et kurveformet speil, reflekteres den slik at den innkommende og utgående strålen danner samme vinkel  $v$  med tangenten, se figur 7. (Du husker kanskje fra naturfagtimene at “innfallsvinkel er lik utfallsvinkel”?)
2. Så litt matematikk som ligner forbløffende: Anta at du har en linje  $l$  og to punkter  $P$  og  $Q$  på samme side av  $l$ . Den korteste veien fra  $P$  til  $Q$  via  $l$  er den som gjør vinklene  $v$  og  $w$  på figur 8 like.



Figur 8: Korteste vei

Dette kan du vise både geometrisk og som et uoppstilt minimumsproblem (se *Kalkulus*, oppgave 7.1.18).

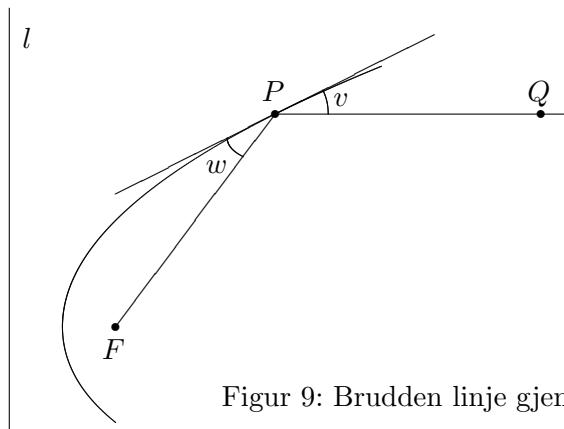
3. Vi vet at en parabel består av de punktene som har samme avstand til styrelinjen som til brennpunktet. Punktene som ligger utenfor pa-

rabelen har kortere vei til styrelinjen enn til brennpunktet, mens de som ligger inni parabelen har kortest vei til brennpunktet. Dette er intuitivt ganske opplagt, men kan du bevise det?

Vi er nå klare til å bevise refleksjonsegenskapen.

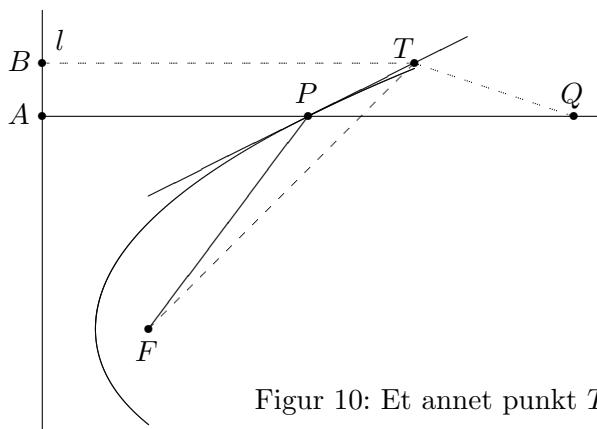
**Setning 3.6.2 (Refleksjonsegenskap for parabler)** *Enhver stråle som kommer inn parallelt med aksen til en parabel, reflekteres gjennom brennpunktet.*

*Bevis:*



Figur 9: Brudden linje gjennom  $Q$  og brennpunktet

På figur 9 har vi tegnet en brudden linje som begynner parallelt med aksen, går gjennom et punkt  $Q$ , treffer parabelen i  $P$  og fortsetter til brennpunktet  $F$ . For å vise at dette er veien en stråle vil følge, må vi ifølge punkt 1 ovenfor vise at vinklene  $v$  og  $w$  er like. Ifølge punkt 2 ovenfor er det tilfellet dersom  $P$  er det punktet på tangenten som gjør den samlede veien  $|QP| + |PF|$  minst mulig.



Figur 10: Et annet punkt  $T$  på tangenten

På figur 10 har vi tegnet inn et annet punkt  $T$  på tangenten. Vår oppgave er å vise at  $|QT| + |TF| > |QP| + |PF|$ . Som hjelp har vi tegnet inn punktet  $A$  der strålen hadde truffet styrelinjen om den ikke var blitt reflektert. Vi har også tegnet inn punktet  $B$  som ligger på styrelinjen i samme høyde som  $T$ . Per definisjon av parabel er  $|AP| = |PF|$ , og følgelig er  $|QP| + |PF| = |QA|$ . Siden  $T$  ligger på utsiden av parabelen, er  $|BT| < |TF|$  ifølge punkt 3 ovenfor. Det betyr at  $|QT| + |TF| > |QT| + |TB|$ . Men  $|QT| + |TB|$  er åpenbart større enn  $|QA|$  siden  $QA$  er den korteste veien fra  $Q$  til linjen  $l$ . Kombinerer vi disse resultatene, får vi

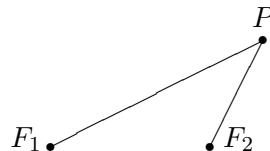
$$|QP| + |PF| = |QA| < |QT| + |TB| < |QT| + |TF|$$

som er det vi måtte vise.  $\square$

Vi skal senere i denne seksjonen se på lignende refleksjonsegenskaper for ellipser og hyperbler.

## Ellipser

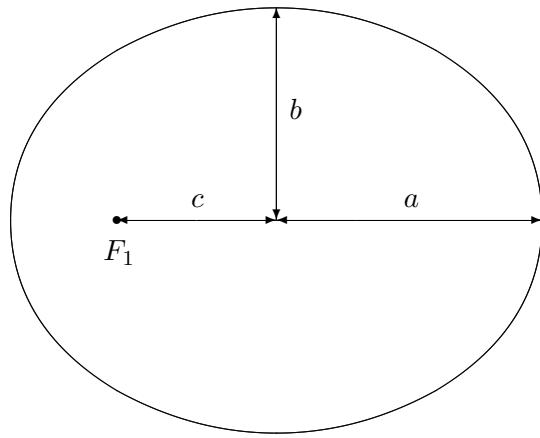
Anta at vi har to punkter  $F_1$  og  $F_2$  i planet. Vi ser på alle punkter  $P$  slik at summen av avstandene  $|PF_1|$  og  $|PF_2|$  er lik et fast tall  $2a$  (for at det skal finnes slike punkter må  $2a$  være større enn avstanden mellom  $F_1$  og  $F_2$ ).



Figur 11: Definisjon av ellipse

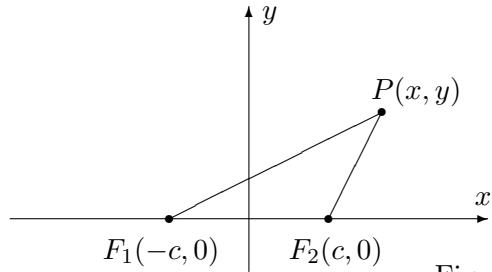
Samlingen av alle disse punktene kalles *ellipsen med brennpunkter  $F_1$  og  $F_2$  og store halvakse  $a$* . Du kan tegne ellipsen ved å ta en hyssing med lengde  $2a$ , feste endene med tegnestift i punktene  $F_1$  og  $F_2$ , dra hyssingen stram med spissen av en blyant og så føre blyanten rundt på papiret mens du hele tiden holder hyssingen stram.

Figur 12 viser en fullt tegnet ellipse. Avstanden fra sentrum i ellipsen til brennpunktene kalles for *brennvidden* og betegnes gjerne med  $c$ , mens  $b = \sqrt{a^2 - c^2}$  kalles *den lille halvaksen*. En ellipse ser ut som en flatklemt sirkel, og grunnen til at man kaller  $a$  og  $b$  for henholdsvis store og lille halvakse, ser du på figuren;  $a$  er den *lengste* avstanden fra sentrum til et punkt på ellipsen, og  $b$  er den *korteste* avstanden fra sentrum til et punkt på ellipsen.



Figur 12: Ellipse med halvakser og brennvidde

For å finne en formel for ellipsen vår legger vi inn et koordinatsystem med origo i sentrum av ellipsen,  $x$ -aksen langs store halvakse, og  $y$ -aksen langs lille halvakse (se figur 13).



Figur 13: Ligning for en ellipse

Vi ser at  $|PF_1| = \sqrt{(x+c)^2 + y^2}$  og  $|PF_2| = \sqrt{(x-c)^2 + y^2}$ . Punktet  $P$  ligger altså på ellipsen dersom

$$\sqrt{(x+c)^2 + y^2} + \sqrt{(x-c)^2 + y^2} = 2a$$

For å forenkle uttrykket fører vi den ene kvadratroten over på den andre siden og kvadrerer:

$$\left( \sqrt{(x+c)^2 + y^2} \right)^2 = \left( 2a - \sqrt{(x-c)^2 + y^2} \right)^2$$

Dette gir først

$$(x+c)^2 + y^2 = 4a^2 - 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2$$

men multpliserer vi ut parentesene og forkorter, sitter vi igjen med

$$a^2 - cx = a\sqrt{(x-c)^2 + y^2}$$

Vi kvadrerer en gang til:

$$a^4 - 2a^2cx + c^2x^2 = a^2(x - c)^2 + a^2y^2$$

Rydder vi opp og forkorter, sitter vi igjen med

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2)$$

Siden  $b^2 = a^2 - c^2$ , kan dette skrives

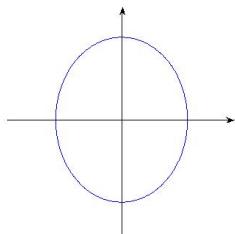
$$b^2x^2 + a^2y^2 = a^2b^2$$

Til slutt deler vi med  $a^2b^2$  og får

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Vi har dermed vist at alle punktene som ligger på ellipsen tilfredsstiller ligningen  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ , og det viser seg også at de punktene som tilfredsstiller denne ligningen, faktisk ligger på ellipsen (kvadreringene våre har altså ikke ført til falske løsninger).

I utregningene ovenfor er  $a > b$  (fordi  $b$  er definert som  $\sqrt{a^2 - c^2}$ ), men ligningen  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$  gir også mening når  $b$  er større enn  $a$ . Den fremstiller da en ellipse der  $x$ - og  $y$ -aksen har byttet roller — den store halvaksen har nå lengde  $b$  og ligger langs  $y$ -aksen, mens den lille halvaksen har lengde  $a$  og ligger langs  $x$ -aksen. Brennpunktene ligger på  $y$ -aksen i posisjonene  $(0, c)$  og  $(0, -c)$  der  $c = \sqrt{b^2 - a^2}$ . Figur 14 viser en ellipse av denne typen.



Figur 14: En ellipse  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$  der  $b > a$

La oss oppsummere resultatene våre:

**Setning 3.6.3 Ligningen**

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

fremstiller en ellipse med sentrum i origo og halvaksen  $a$  og  $b$ . Dersom  $a > b$ , er brennpunktene  $(c, 0)$  og  $(-c, 0)$  der  $c = \sqrt{a^2 - b^2}$ . Dersom  $a < b$ , er brennpunktene  $(0, c)$  og  $(0, -c)$  der  $c = \sqrt{b^2 - a^2}$ . Dersom  $a = b$  er ellipsen en sirkel med radius  $r = a = b$ . Brennpunktene faller da sammen og ligger i sentrum av sirkelen.

En ellipse kan selvfølgelig ha sentrum andre steder enn origo. Ligningen

$$\frac{(x-m)^2}{a^2} + \frac{(y-n)^2}{b^2} = 1$$

fremstiller en ellipse med sentrum i punktet  $(m, n)$  og halvakser  $a, b$ .

**Eksempel 2:** Vi skal vise at ligningen

$$9x^2 + 4y^2 - 36x + 24y + 36 = 0$$

fremstiller en ellipse, og finne sentrum, halvakser og brennpunkter.

Planen er å skrive ligningen på formen  $\frac{(x-m)^2}{a^2} + \frac{(y-n)^2}{b^2} = 1$  slik at vi kan lese av sentrum og halvakser. Vi begynner med å gjøre kvadratene i  $x$  og  $y$  fullstendige:

$$\begin{aligned} 9x^2 + 4y^2 - 36x + 24y + 36 &= 9x^2 - 36x + 4y^2 + 24y + 36 = \\ &= 9(x^2 - 4x) + 4(y^2 + 6y) + 36 = 9(x^2 - 4x + 4) - 36 + 4(y^2 + 6y + 9) - 36 + 36 = \\ &= 9(x - 2)^2 + 4(y + 3)^2 - 36 \end{aligned}$$

Ligningen kan dermed skrives

$$9(x - 2)^2 + 4(y + 3)^2 = 36$$

For å få det ønskede ett-tallet på høyresiden, deler vi med 36:

$$\frac{(x - 2)^2}{4} + \frac{(y + 3)^2}{9} = 1$$

Denne ligningen kan også skrives

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

og dermed ser vi at vi har en ellipse med sentrum i  $(2, -3)$  med halvakser  $a = 2$ ,  $b = 3$ . Siden  $b$  er den største halvaksen, blir brennvidden  $c = \sqrt{b^2 - a^2} = \sqrt{3^2 - 2^2} = \sqrt{5}$ . Brennpunktene ligger dermed i punktene  $F_1(2, -3 - \sqrt{5})$  og  $F_2(2, -3 + \sqrt{5})$  (husk at den store halvaksen er parallel med  $y$ -aksen). ♣

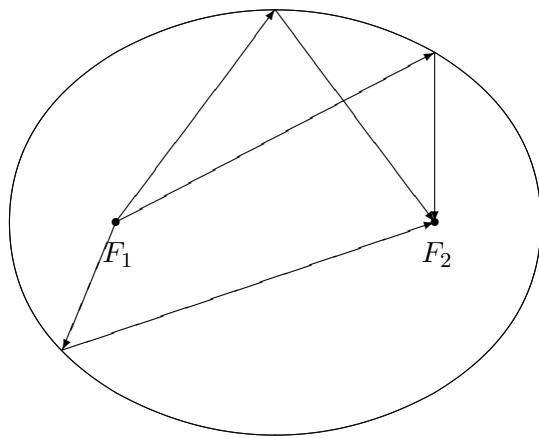
Det er ikke vanskelig å parametrisere en ellipse. Ved å sette inn i ligningen  $\frac{(x-m)^2}{a^2} + \frac{(y-n)^2}{b^2} = 1$  ser du at

$$\mathbf{r}(t) = (m + a \cos t) \mathbf{i} + (n + b \sin t) \mathbf{j} \quad t \in [0, 2\pi]$$

er en parametrisering av ellipsen med sentrum i  $(m, n)$  og halvakser  $a$  og  $b$ .

Figur 15 viser refleksjonsegenskapen til ellipser; en stråle som starter i det ene brennpunktet, reflekteres gjennom det andre brennpunktet. Du kan

opp leve dette prinsippet i praksis i enkelte ovale rom (f.eks. "Whispering Gallery" i St. Paul's Cathedral i London); en lavmælt samtale nær det ene brennpunktet, oppfattes tydelig i nærheten av det andre brennpunktet mange meter unna. Har du et fat eller en kjele som er (tilnærmet) ellipseformet, kan du eksperimentere med refleksjonsegenskapen. Hvis du fyller litt vann i bunnen og slipper en dråpe i det ene brennpunktet, vil du få en bølge som brer seg utover til den treffer kanten, og så samles i det andre brennpunktet.



Figur 15: Refleksjonsegenskapen for ellipser

Beviset for refleksjonsegenskapen for ellipser er enklere enn beviset for refleksjonsegenskapen for parabler, men bygger på den samme ideen. I tillegg til det vi allerede vet (det kan være lurt å repetere de tre punktene om refleksjon som vi skrev opp da vi gjennomgikk parabelen), trenger vi en observasjon til: Ellipsen med brennpunkter  $F_1$  og  $F_2$  og store halvakse  $a$ , består av de punktene  $P$  slik at den samlede avstandene fra  $P$  til brennpunktene er nøyaktig  $2a$ , dvs:

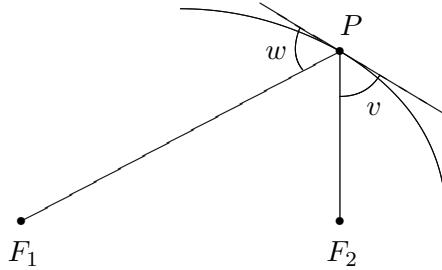
$$|PF_1| + |PF_2| = 2a$$

De punktene som ligger på utsiden av ellipsen, har en samlet avstand som er større en  $2a$ , mens de som ligger på innsiden av ellipsen har en samlet avstand som er mindre enn  $2a$  (dette virker intuitivt rimelig, men kan du bevise det?)

**Setning 3.6.4 (Refleksjonsegenskapen for ellipser)** *En stråle som går ut fra det ene brennpunktet til en ellipse, reflekteres gjennom det andre.*

*Bevis:* På figur 16 har vi tegnet den brudne linjen fra det ene brennpunktet  $F_1$ , via et punkt  $P$  på ellipsen til det andre brennpunktet  $F_2$ . For å vise at dette er veien en lysstråle vil følge, må vi vise at vinklene  $v$  og  $w$  er like store. Det er tilfellet dersom  $P$  er det punktet på tangenten som har minst

samlet avstand til  $F_1$  og  $F_2$ , men det er opplagt — alle andre punkter på tangenten ligger utenfor ellipsen, og har derfor en samlet avstand som er større enn  $2a$ .  $\square$



Figur 16: Bevis for refleksjonsegenkapen for ellipse

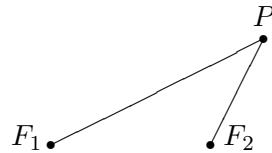
Ellipser har forskjellig form, noen er langstrakte, mens andre er nesten sirkelformede. *Eksentrisiteten* er et tradisjonelt mål på hvor langstrakt en ellipse er. Den er definert ved

$$e = \frac{\text{brennvidden}}{\text{store halvakse}} = \frac{c}{a}$$

En sirkel har eksentrisitet 0, mens en svært langstrakt ellipse har en eksentrisitet tett oppunder 1.

## Hyperbler

Vi skal nå se på den siste klassen av kjeglesnitt — hyperbler. Definisjonen ligner på definisjonen til ellipser, men vi bruker differenser istedenfor summer.



Figur 17: Definisjon av hyperbel

Vi starter med to punkter  $F_1$  og  $F_2$  i planet, og ser på alle punkter  $P$  slik at avstandene  $|PF_1|$  og  $|PF_2|$  har en fast differanse. Kaller vi denne differensen  $2a$ , er vi altså interessert i alle punkter  $P$  slik at

$$|PF_1| - |PF_2| = \pm 2a$$

(vi skriver  $\pm$  foran  $2a$  siden vi ikke er interessert i hvilken av avstandene som er størst, bare at differensen mellom den største og den minste er  $2a$ ).

Som figur 18 viser, er en hyperbel ikke en sammenhengende kurve, men består av to deler. Den høyre av disse *hyperbelgrenene* består av de punktene

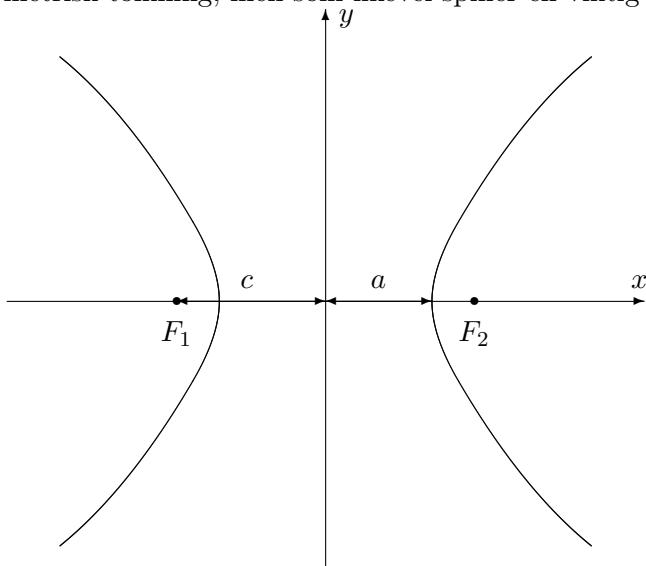
$P$  der

$$|PF_1| - |PF_2| = 2a$$

mens den venstre består av de punktene  $P$  der

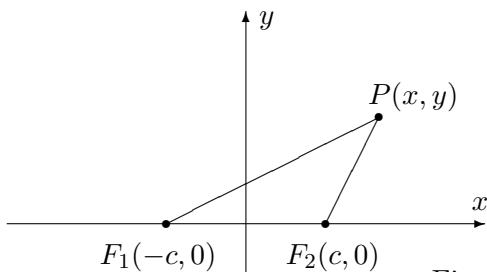
$$|PF_1| - |PF_2| = -2a$$

Vi kaller  $a$  for *halvaksen*. På figuren er  $a$  avstanden fra sentrum i hyperbelen til “snuten” av hver av hyperbelgrenene. Som for ellipsen er *brennvidden*  $c$  avstanden fra sentrum til brennpunktene. For hyperbler er  $c > a$ , og vi definerer hjelpestørrelsen  $b = \sqrt{c^2 - a^2}$  som ikke har noen umiddelbar geometrisk tolkning, men som likevel spiller en viktig rolle.



Figur 18: Hyperbel med halvakse og brennvidde

For å finne formelen til en hyperbel, legger vi inn et kordinatsystem som vist på figur 19.



Figur 19: Ligning for en hyperbel

Vi ser at  $|PF_1| = \sqrt{(x + c)^2 + y^2}$  og  $|PF_2| = \sqrt{(x - c)^2 + y^2}$ . Punktet  $P$  ligger altså på hyperbelen dersom

$$\sqrt{(x + c)^2 + y^2} - \sqrt{(x - c)^2 + y^2} = \pm 2a$$

Akkurat som for ellipsen kan vi forenkle dette uttrykket ved å føre den ene kvadratroten over på den andre siden og kvadrere:

$$\left(\sqrt{(x+c)^2 + y^2}\right)^2 = \left(\pm 2a + \sqrt{(x-c)^2 + y^2}\right)^2$$

Dette gir først

$$(x+c)^2 + y^2 = 4a^2 \pm 4a\sqrt{(x-c)^2 + y^2} + (x-c)^2 + y^2$$

men multipliserer vi ut parentesene og forkorter, sitter vi igjen med

$$a^2 - cx = \mp a\sqrt{(x-c)^2 + y^2}$$

Vi kvadrerer en gang til:

$$a^4 - 2a^2cx + c^2x^2 = a^2(x-c)^2 + a^2y^2$$

Rydder vi opp og forkorter, sitter vi igjen med

$$(a^2 - c^2)x^2 + a^2y^2 = a^2(a^2 - c^2)$$

Siden  $b^2 = c^2 - a^2$ , kan dette skrives

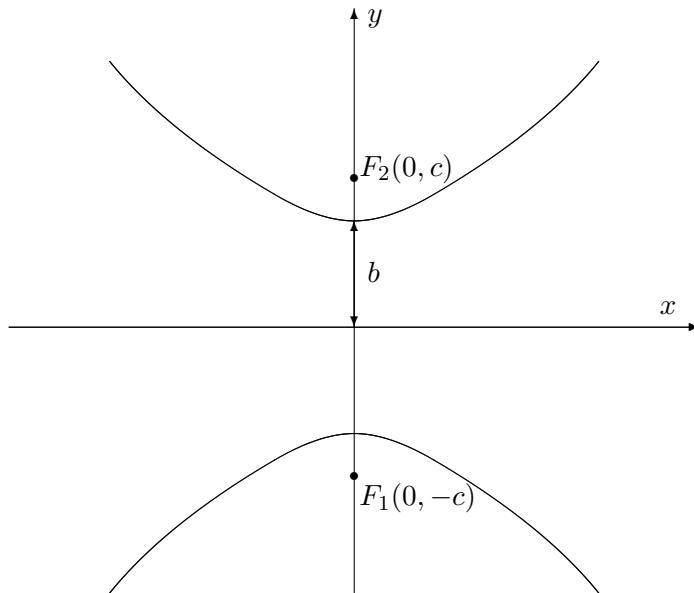
$$-b^2x^2 + a^2y^2 = -a^2b^2$$

Til slutt deler vi med  $-a^2b^2$  og får

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

Vi har dermed vist at alle punktene på hyperbelen tilfredsstiller ligningen  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ , og det viser seg også at de punktene som tilfredsstiller denne ligningen, faktisk ligger på hyperbelen (kvadreringene våre har altså ikke ført til falske løsninger).

Dersom vi plasserer minusstegnet foran  $x$ -leddet istedenfor  $y$ -leddet, altså  $\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$ , så fremstiller ligningen fortsatt en hyperbel, men nå har  $x$ - og  $y$ -aksen byttet roller. Det er  $b$  som er halvaksen, brennpunktene ligger på  $y$ -aksen i posisjonene  $(0, -c)$  og  $(0, c)$ , og hyperbelen “ligger” istedenfor å stå, slik figur 20 viser.



Figur 20: Hyperbel med ligning  $\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$

La oss oppsummere resultatene våre så langt.

**Setning 3.6.5** *Ligningen*

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

fremstiller en hyperbel med halvakse  $a$  og med brennpunkter i  $(-c, 0)$  og  $(c, 0)$  der  $c = \sqrt{a^2 + b^2}$ . Ligningen

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$$

fremstiller en hyperbel med halvakse  $b$  og med brennpunkter i  $(0, -c)$  og  $(0, c)$ .

Også hyperbler kan ha sentrum andre steder enn i origo. Ligningene

$$\frac{(x - m)^2}{a^2} - \frac{(y - n)^2}{b^2} = 1$$

og

$$\frac{(y - n)^2}{b^2} - \frac{(x - m)^2}{a^2} = 1$$

fremstiller hyperbler med sentrum i  $(m, n)$ .

**Eksempel 3:** Vi skal vise at ligningen

$$-3x^2 + 4y^2 + 6x + 32y + 49 = 0$$

fremstiller en hyperbel, og finne sentrum og brennpunkter.

Akkurat som for ellipsen og parabelen fullfører vi kvadratene og forsøker å bringe ligningen over på en av standardformene ovenfor. Vi har:

$$\begin{aligned} -3x^2 + 4y^2 + 6x + 32y + 49 &= -3x^2 + 6x + 4y^2 + 32y + 49 = \\ -3(x^2 - 2x) + 4(y^2 + 8y) + 49 &= -3(x^2 - 2x + 1) + 3 + 4(y^2 + 8y + 16) - 64 + 49 = \\ &= -3(x - 1)^2 + 4(y + 4)^2 - 12 \end{aligned}$$

Ligningen vår kan altså skrives

$$-3(x - 1)^2 + 4(y + 4)^2 = 12$$

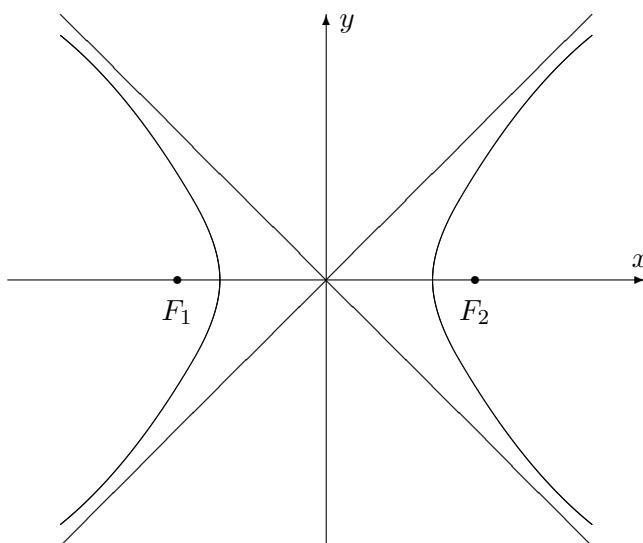
og deler vi på 12, får vi

$$-\frac{(x - 1)^2}{4} + \frac{(y + 4)^2}{3} = 1$$

Dette er det samme som

$$\frac{(y + 4)^2}{(\sqrt{3})^2} - \frac{(x - 1)^2}{2^2} = 1$$

Dette er ligningen til en hyperbel med sentrum i  $(1, -4)$ , med  $a = 2$  og  $b = \sqrt{3}$ . Brennvidden er  $c = \sqrt{a^2 + b^2} = \sqrt{2^2 + (\sqrt{3})^2} = \sqrt{7}$ . Siden aksen til parabelen er parallel med  $y$ -aksen, ligger brennpunktene i  $(1, -4 - \sqrt{7})$  og  $(1, -4 + \sqrt{7})$ . ♣



Figur 21: Hyperbel med asymptoter

Asymptoter er et viktig hjelpemiddel når vi skal tegne hyperbler (se figur 21). Det viser seg at hyperblene  $\frac{(x-m)^2}{a^2} - \frac{(y-n)^2}{b^2} = 1$  og  $\frac{(y-n)^2}{b^2} - \frac{(x-m)^2}{a^2} = 1$  har de samme asymptotene.

**Setning 3.6.6** Hyperblene  $\frac{(x-m)^2}{a^2} - \frac{(y-n)^2}{b^2} = 1$  og  $\frac{(y-n)^2}{b^2} - \frac{(x-m)^2}{a^2} = 1$  har asymptotene

$$y - n = \pm \frac{b}{a}(x - m)$$

når  $x \rightarrow \pm\infty$ .

*Bevis:* For at ikke regningene skal bli altfor uoversiktelig, nøyer vi oss med å vise at  $y = \frac{b}{a}x$  er en asymptote for hyperbelen  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$  når  $x \rightarrow \infty$ :

Løser vi ligningen  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$  med hensyn på  $y$ , får vi

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

Siden vi åpenbart er interessert i den positive delen, må vi vise at

$$\lim_{x \rightarrow \infty} \left( \frac{b}{a} \sqrt{x^2 - a^2} - \frac{b}{a}x \right) = 0$$

Bruker vi trikset med å gange med den konjugerte over og under brøkstreken, får vi

$$\begin{aligned} \lim_{x \rightarrow \infty} \left( \frac{b}{a} \sqrt{x^2 - a^2} - \frac{b}{a}x \right) &= \frac{b}{a} \lim_{x \rightarrow \infty} \left( \sqrt{x^2 - a^2} - x \right) = \\ &= \frac{b}{a} \lim_{x \rightarrow \infty} \frac{\left( \sqrt{x^2 - a^2} - x \right) \left( \sqrt{x^2 - a^2} + x \right)}{\sqrt{x^2 - a^2} + x} = \\ &= \frac{b}{a} \lim_{x \rightarrow \infty} \frac{-a^2}{\sqrt{x^2 - a^2} + x} = 0 \end{aligned}$$

□

Vi tar med et eksempel.

**Eksempel 4:** Finn asymptotene til hyperbelen

$$\frac{(x-3)^2}{3^2} - \frac{(y+4)^2}{4^2} = 1$$

Formelen ovenfor sier at asymptotene er

$$y + 4 = \pm \frac{4}{3}(x - 3)$$

Tar vi de to fortumnene hver for seg, får vi

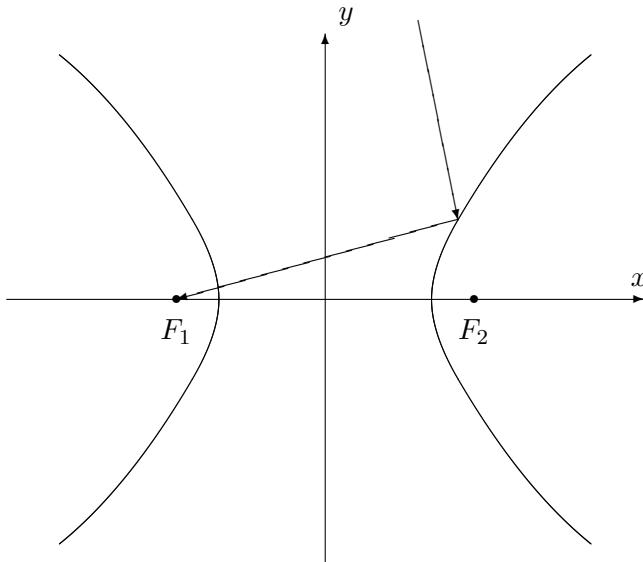
$$y = \frac{4}{3}x - 8$$

og

$$y = -\frac{4}{3}x$$



Akkurat som de andre kjeglesnittene har hyperbelen en refleksjonsegenskap. Den er ikke så mye brukt som refleksjonsegenskapene til parabelen og ellipsen, men for ordens skyld tar vi den med (uten bevis).



Figur 22: Refleksjonsegenskapen for hyperbler

**Setning 3.6.7** En stråle som kommer fra utsiden av en hyperbel med retning mot det ene brennpunktet, reflekteres i retning av det andre brennpunktet (se figur 22).  $\square$

La oss helt til slutt ta med et ord om parametrisering. Hyperbelen  $\frac{(x-m)^2}{a^2} - \frac{(y-n)^2}{b^2} = 1$  kan parametriseres ved

$$\mathbf{r}(t) = (m + a \cosh t) \mathbf{i} + (n + b \sinh t) \mathbf{j}$$

der

$$\cosh t = \frac{e^t + e^{-t}}{2}$$

og

$$\sinh t = \frac{e^t - e^{-t}}{2}$$

er henholdsvis hyperbolsk sinus og cosinus (se seksjon 7.7 i *Kalkulus*). Du kan sjekke dette ved å bruke at

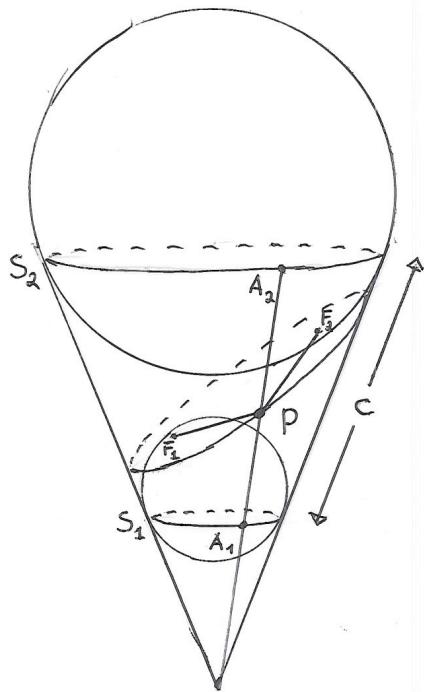
$$\cosh^2 t - \sinh^2 t = 1$$

### Dandelins iskrembevis

Helt i begynnelsen av seksjonen (figur 1) så vi hvordan kjeglesnitt fremkommer når vi snitter en kjegle med et plan. I dette avslutningsavsnittet skal

vi se hvordan vi kan knytte denne romgeometriske beskrivelsen til de plan-geometriske definisjonene vi ellers har brukt. Vi skal gjennomføre beviset for ellipser, men det finnes tilsvarende argumenter for hyperbler og parabler (se oppgavene). Beviset stammer fra den belgiske matematikeren Germinal P. Dandelin (1794-1847) og er et sjeldent eksempel på at det finnes elegante og naturlige geometriske argumenter som ikke ble funnet av de gamle grekerne!

Før vi begynner, er det nyttig å være klar over følgende lille kjensgjerning: Dersom du har et punkt  $P$  utenfor en kule, er det uendelig mange linjer som går gjennom  $P$  og tangerer kulen. Uansett hvilken av disse linjene du velger, vil avstanden fra  $P$  til tangeringspunktet alltid være den samme. Den enkleste måten å se dette på geometrisk, er nok å tenke seg at punktet ligger rett over toppen av kulen som polarstjernen over Nordpolen — da vil alle tangenter berøre jordkloden langs samme breddegrad.



Figur 23: Dandelin's iskrembevis

Figur 23 viser utgangspunktet for Dandelin's bevis. Vi kutter over en kjegle med et plan som er mindre bratt enn sidekanten til kjeglen. Inni i kjeglen legger vi to kuler som "hviler" i kjeglen slik en perfekt iskremkule vil hvile i en kjegleformet kjeks — én som tangerer planet fra undersiden og én som tangerer planet fra oversiden (se figuren). Disse to kulene tangerer overskjæringsplanet i to punkter  $F_1$  og  $F_2$ , og vi skal vise at snittkurven til kjeglen og planet er en ellipse med brennpunkter i  $F_1$  og  $F_2$ . Ifølge den

geometriske definisjonen av en ellipse er det nok å vise at summen av avstandenene  $|PF_1|$  og  $|PF_2|$  er den samme uansett hvilket punkt  $P$  vi velger på snittkurven.

Observer at de to kulene tangerer kjeglen i to horisontale sirkler  $S_1$  og  $S_2$ , og la  $c$  være avstanden mellom sirklene målt langs kjeglekanten. Vi skal vise at  $|PF_1| + |PF_2| = c$ .

Trekk linjen gjennom  $P$  og bunnpunktet  $O$  i kjeglen. Denne linjen skjærer sirklene  $S_1$  og  $S_2$  i punktene  $A_1$  og  $A_2$ . Vi ser nå at  $|PF_1| = |PA_1|$  siden begge er tangenter til den nederste kullen. Tilsvarende er  $|PF_2| = |PA_2|$  siden begge er tangenter til den øverste kullen. Men dermed er  $|PF_1| + |PF_2| = |PA_1| + |PA_2| = c$ , og beviset er fullført.

Som allerede nevnt, finnes det tilsvarende bevis for hyperbler og parabler. Du finner noen hint i oppgave 16 og 17.

### Oppgaver til seksjon 3.6

I oppgave 1 til 7 skal du undersøke om ligningen fremstiller et kjeglesnitt. Bestem i så fall hva slags kjeglesnitt, og finn sentrum, halvakser, brennpunkter, toppunkt, asymptoter der det er aktuelt. Lag en skisse av kjeglesnittet.

**1.**  $4x^2 + 9y^2 + 32x - 18y + 37 = 0$

**2.**  $y^2 - 4x - 2y - 7 = 0$

**3.**  $x^2 - y^2 - 2x + 4y - 7 = 0$

**4.**  $16y^2 - 9x^2 + 32y + 54x - 209 = 0$

**5.**  $3x^2 + 5y^2 + 6x - 20y + 8 = 0$

**6.**  $x^2 + 4x + 2y - 4 = 0$

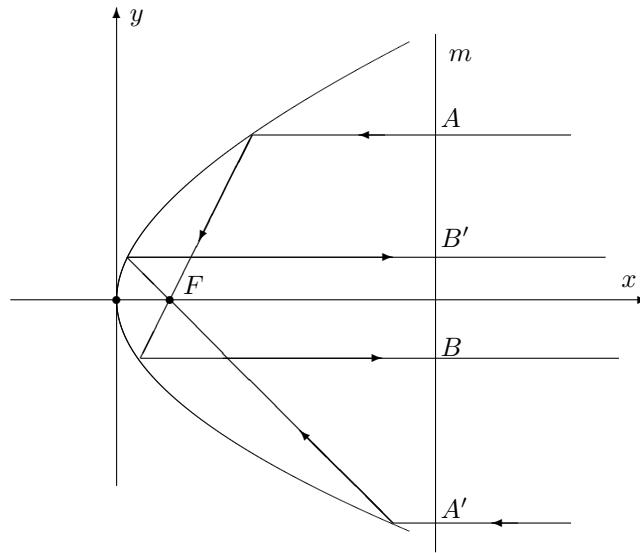
**7.**  $3x^2 + y^2 - 6x + 4y + 16 = 0$

**8.** Vis at ligningen til tangenten i et punkt  $(x_0, y_0)$  på ellipsen  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$  kan skrives  $\frac{xx_0}{a^2} + \frac{yy_0}{b^2} = 1$ .

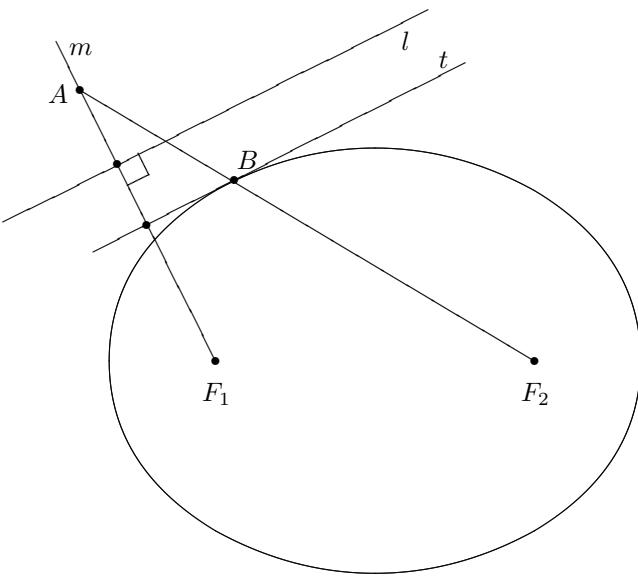
**9.** Vis at tangenten til parabelen  $y^2 = 4ax$  i punktet  $(x_0, y_0)$  skjærer  $x$ -aksen i punktet  $(-x_0, 0)$ .

**10.** Et punkt  $P$  ligger “inni” en parabel. Finn den korteste veien fra  $P$ , via parabelen, til brennpunktet  $F$ .

**11.** Figuren viser en parabel og en linje  $m$  som er parallel med styrelinjen. To lysstråler kommer inn parallelt med parabelaksen og brytes når de treffer parabelen. Den ene strålen skjærer linjen  $m$  i punktene  $A$  og  $B$ , den andre i punktene  $A'$  og  $B'$ . Bruker lyset kortest tid fra  $A$  til  $B$  eller fra  $A'$  til  $B'$ ?



**12.** Figuren viser en linje  $l$  og en ellips med brennpunkter  $F_1$  og  $F_2$ , og store halvaksse  $a$ . Vi skal nå beskrive hvordan du kan finne en tangent til ellipsen som er parallel med  $l$ .



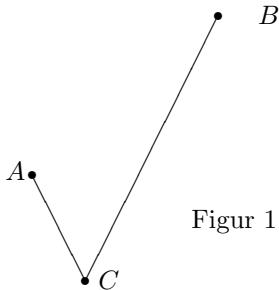
Trekk linjen  $m$  som går gjennom  $F_1$  og står normalt på  $l$ . Sett av et punkt  $A$  på  $m$  som har avstand  $2a$  til  $F_2$ . Trekk midtnormalen  $t$  på linjestykket  $F_1A$ . Da er  $t$  parallel med  $l$  og tangent til ellipsen.

I resten av oppgaven skal vi vise at denne metoden er riktig.

- La  $B$  være skjæringspunktet mellom ellipsen og linjestykket  $F_2A$ . Forklar hvorfor  $|AB| = |BF_1|$ .

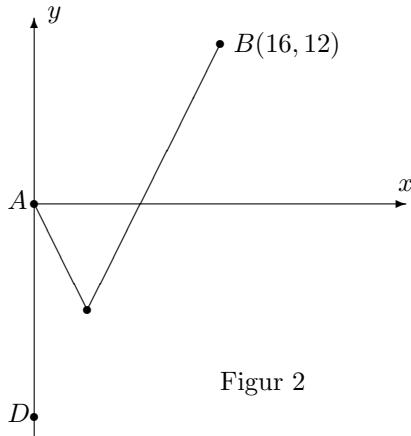
- b) Vis at  $B$  ligger på  $t$ .
- c) La  $C$  være et annet punkt på  $t$ . Forklar hvorfor  $|F_2C| + |CF_1| = |F_2C| + |CA| > 2a$ .
- d) Bruk resultatet i c) til å vise at  $t$  tangerer ellipsen i  $B$ .

**13.** Endene til et 34 meter langt tau er festet i to punkter  $A$  og  $B$ . Avstanden fra  $A$  til  $B$  er 20 meter. Et lodd er festet til et punkt  $C$  på tauet. Loddet trekker tauet stramt slik at det danner to rette linjestykker  $AC$  og  $CB$  (se figur 1)



Figur 1

- a) Forklar at punktet  $C$  ligger på en ellipse med brennpunkter i  $A$  og  $B$  og med store halvaksen 17 meter. Hva er den andre halvaksen?



Figur 2

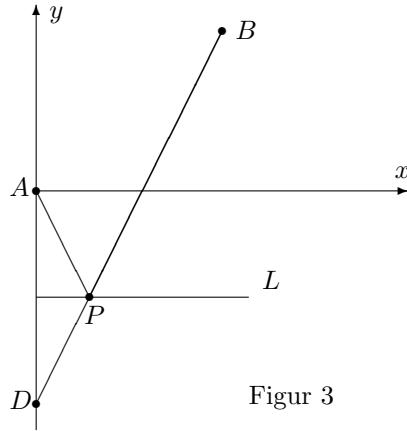
Hvis loddet får gli fritt langs tauet, vil det falle til ro i det laveste punktet på ellipsen. I resten av oppgaven skal vi arbeide for å finne dette punktet. I figur 2 har vi lagt punktene inn i et koordinatsystem slik at  $A$  har koordinatene  $(0,0)$  og  $B$  har koordinatene  $(16, 12)$ .

- b) Punktet  $D$  ligger på den negative  $y$ -aksen en taulengde ( $= 34$  meter) fra  $B$ . Finn koordinatene til  $D$ .

Linjen  $L$  er parallel med  $x$ -aksen og ligger like langt fra  $A$  som fra  $D$ . Denne linjen skjærer linjestykket  $DB$  i et punkt  $P$  (se figur 3 nedenfor).

- c) Vis at  $P$  ligger på ellipsen.
- d) Vis at ingen andre punkter på  $L$  ligger på ellipsen. Hvorfor betyr dette at  $P$  er det laveste punktet på ellipsen?

- e) Finn koordinatene til  $P$ .

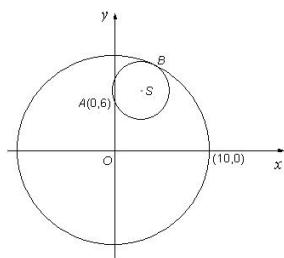


Figur 3

**14.**  $O$  og  $A$  er to punkter i planet med koordinater henholdsvis  $(0, 0)$  og  $(0, 6)$ . Et kjeglesnitt består av alle de punkter  $X$  med koordinater  $(x, y)$  slik at  $|OX| + |AX| = 10$ .

- a) Hva slags kjeglesnitt er dette? Finn ligningen til kjeglesnittet og tegn figur.

Figuren nedenfor viser en sirkel  $C_1$  med sentrum i  $O$  og radius 10. En mindre sirkel  $C_2$  går gjennom  $A$  og tangerer  $C_1$  i punktet  $B$ .



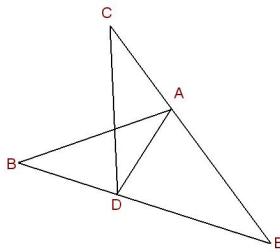
- b) Vis at sentrum  $S$  i  $C_2$  ligger på kjeglesnittet i punkt a).

Et ingeniørfirma planlegger en undersjøisk rørledning. Røret skal ha indre radius 10 cm, og det skal blant annet inneholde en elektrisk kabel med radius 3 cm. Figuren ovenfor kan oppfattes som et tverrsnitt gjennom røret og kabelen. På grunn av strømtilførselen må ytterkanten av kabelen gå gjennom punkt  $A$ , og for at kabelen skal være minst mulig i veien for de andre installasjonene i røret, ønsker man at den skal tangere innsiden av røret.

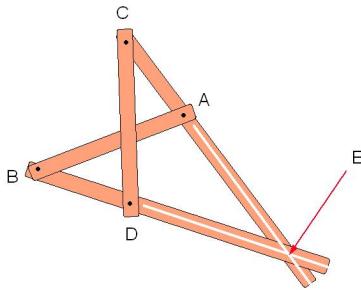
- c) Finn koordinatene til midtpunktet  $S$  i kabelen. Finn også koordinatene til punktet  $B$  der kabelen skal festes til røret.

**15.** I denne oppgaven er  $a$  og  $c$  to positive tall, og  $c > a$ .

- a)  $A$  og  $B$  er to punkter i planet. Et kjeglesnitt består av alle punkter  $X$  slik at  $|AX| - |BX| = \pm 2a$ . Beskriv kjeglesnittet
- b) På figuren nedenfor er  $|AB| = |CD| = 2c$  og  $|AC| = |BD| = 2a$ . Forklar at  $\angle BDA = \angle CAD$ . Forklar også at  $|DE| = |AE|$ , og vis at punktet  $E$  ligger på kjeglesnittet i spørsmål a).



Den nederlandske matematikeren Frans van Schooten konstruerte på 1600-tallet et apparat til å tegne hyperbler (se figuren nedenfor). Instrumentet er laget slik at  $|AB| = |CD| = 2c$  og  $|AC| = |BD| = 2a$ . I punktene  $A, B, C, D$  er stengene festet med en nagle slik at de kan rotere i forhold til hverandre. Punktene  $A$  og  $B$  er festet til papiret mens en penn i  $E$  tegner en kurve når man vrir på instrumentet.



- c) Forklar at instrumentet tegner opp (deler av) en hyperbel.

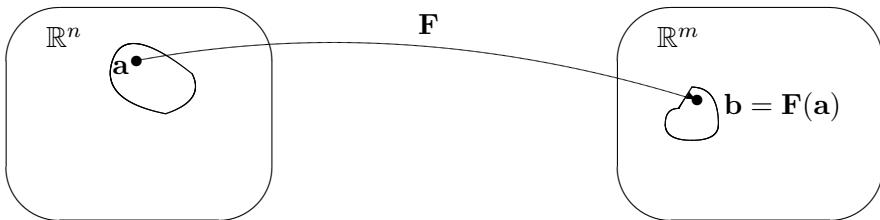
(Takk til Tor Andersen for tegningen av van Schootens instrument).

**16.** Det finnes en variant av Dandelinis iskrembevis for hyperbler. I dette tilfellet arbeider man også med to kuler — den ene i den øvre delen av dobbeltkjeglen, den andre i den nedre delen. Gjennomfør resonnementet.

**17.** Gjennomfør Dandelinis iskrembevis for parabler. I dette tilfellet er det bare én kule å arbeide med. Dennekulen tangerer kjeglen i en sirkel  $S_1$ . Styrelinjen til parabelen er skjæringlinjen mellom overskjæringsplanet (som parabelen ligger i) og planet til  $S_1$ .

### 3.7 Grafisk fremstilling av skalarfelt

I kapittel 2 studerte vi funksjoner fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ . Generelt er det ikke mulig å fremstille slike funksjoner grafisk på en realistisk måte, og vi nøyde oss derfor med stiliserte figurer av typen nedenfor.

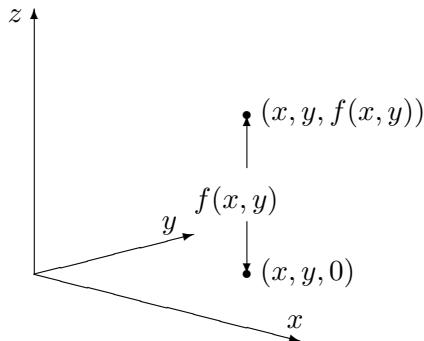
Figur 1: Funksjon  $\mathbf{F}$  fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$ 

Det finnes imidlertid tilfeller der man kan gi gode grafiske fremstillinger av funksjoner av flere variable, og det er dette som er temaet for denne og den neste seksjonen. Før vi setter igang for alvor, tar vi med en liten forsvarstale for hvorfor det fortsatt er lurt å lære seg å tegne grafer for hånd til tross for det finnes både lommeregnere og dataprogrammer (slik som MATLAB) som gir flotte grafiske fremstillinger. Det er tre momenter jeg vil trekke frem:

- En tredimensjonal graf kan være vanskelig å tolke, spesielt når deler av grafen skjuler seg bak andre deler. Det kan også være vanskelig å finne det riktige vinduet; bruker man gal målestokk eller fokuserer på feil sted, risikerer man å gå glipp av de interessante detaljene. Lager du grafene for hånd, blir du ofte ledet til stedene der det interessante skjer. Det viser seg også ofte at en litt karikert håndtegning er lettere å forstå enn en eksakt datautskrift — akkurat som en god karikaturtegning ofte sier mer om en politiker enn det et offisielt fotografi gjør.
- Ofte er vi ikke bare interessert i å vite *hvordan* en funksjon ser ut — vi vil også vite *hvorfor* den ser ut som den gjør. Når vi analyserer en funksjon matematisk, får vi en helt annen forståelse av hvorfor bunnene og toppene ligger der de gjør, enn når vi bare studerer et bilde på en dataskjerm.
- Av og til vil vi lage en graf med et spesielt utseende eller ta en graf vi allerede har, og modifisere den i en spesiell retning. Man kan gjøre dette med dataverktøy etter prøve-og-feile-metoden, men det tar ofte lang tid og minner om å lete etter en nål i en høystakk. Skjønner man matematikken som ligger bak grafene, er det mye lettere å se hva som trengs.

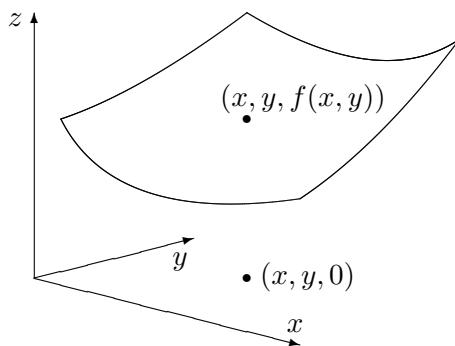
La meg understreke at punktene ovenfor ikke betyr at dataverktøy er unyttig når man skal fremstille funksjoner grafisk — tvert i mot finnes det mange funksjoner som er så kompliserte at de *bare* kan fremstilles grafisk ved hjelp av slike hjelpe middel. Det jeg vil argumentere for, er at man ofte kommer lengst når man kombinerer datamaskinens fremstilling med en matematisk analyse.

La oss komme igang! Vi begynner med å minne om hvordan vi grafisk kan fremstille funksjoner  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , altså skalarfelt av to variable (husk seksjon 2.1). For å unngå altfor mange indekser, skal vi kalle variablene  $x$  og  $y$  istedenfor  $x_1$  og  $x_2$ , og vi skal bruke  $z$  som en betegnelse på funksjonsverdien. Vi ser altså på funksjoner  $z = f(x, y)$ . For å tegne funksjonsgrafen lager vi først et tre-dimensjonalt koordinatsystem som vist på figur 2.



Figur 2: Plotting av skalarfelt

Gitt variabelverdier  $x$  og  $y$ , finner vi punktet  $(x, y, 0)$  i  $xy$ -planet. Vi flytter oss nå loddrett (dvs. parallelt med  $z$ -aksen) til vi finner punktet  $(x, y, f(x, y))$ . Dette er det første punktet på funksjonsgrafen vår. Gjentar vi denne prosedyren for stadig flere variabelverdier  $(x, y)$ , vokser grafen etterhvert frem som en flate i rommet (se figur 3).



Figur 3: Grafisk fremstilling av skalarfelt

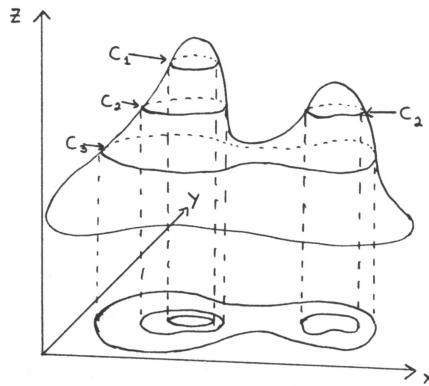
Selv om denne prosedyren på en grei måte forklarer hva grafen til et skalarfelt er, så er den i praksis ubruklig som en oppskrift for hvordan man tegner grafen. Prøver du den, selv på en enkel funksjon, oppdager du fort at du helt mister romfølelsen i bildet. Vi må derfor finne frem til andre og mer effektive måter å tegne grafer på.

Et godt hjelpemiddel er nivåkurvene til funksjonen. For hvert tall  $c$ ,

består *nivåkurven*

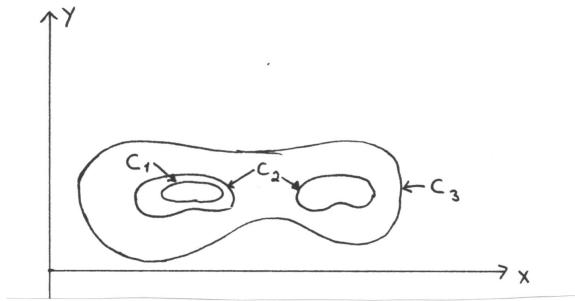
$$N_c = \{(x, y) : f(x, y) = c\}$$

av de punktene i  $xy$ -planet der funksjonen har verdien  $c$ . Figur 4 viser hvordan nivåkurvene fremkommer — vi tar utgangspunkt i de punktene på grafen hvor funksjonsverdien er hhv.  $c_1$ ,  $c_2$  og  $c_3$ , og projiserer dem ned i  $xy$ -planet.



Figur 4: Fremstilling av nivåkurver

På figur 5 har vi tegnet opp disse punktene i et todimensjonalt koordinatsystem. Legg merke til at én av nivåkurvene (den som tilsvarer verdien  $c_2$ ) består av to deler — en for hver “fjelltopp” på grafen.



Figur 5: Nivåkurver

På figur 4 og 5 startet vi med funksjonsgrafen og tegnet inn nivåkurvene etterpå. I praksis går vi som regel den andre veien — vi starter med nivåkurvene og vil bruke dem som et utgangspunkt for å tegne grafen. Her er et eksempel:

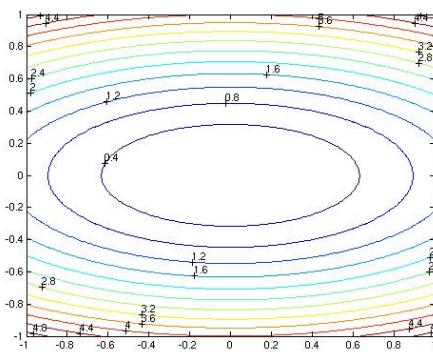
**Eksempel 1:** Vi skal skissere grafen til funksjonen  $f(x, y) = x^2 + 4y^2$ . Nivåkurvene blir

$$x^2 + 4y^2 = c.$$

Er  $c$  negativ, finnes det ingen punkter som oppfyller denne ligningen. Er  $c$  positiv, får vi ellipsene

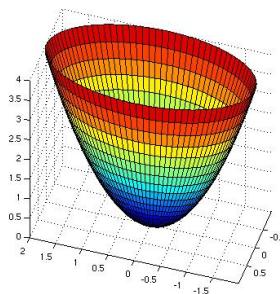
$$\frac{x^2}{(\sqrt{c})^2} + \frac{y^2}{\left(\frac{\sqrt{c}}{2}\right)^2} = 1.$$

Figur 6 viser noen av nivåkurvene (de tilhørende  $c$ -verdiene står på). Nivåkurvene til  $f$  er altså ellipser med samme form som ligger tettere og tettere dess større  $c$  blir.



Figur 6: Nivåkurver til  $f(x, y) = x^2 + 4y^2$

Vi kan tenke oss at flaten består av elliptiske ringer lagt oppå hverandre. For å få et bedre inntrykk av hvordan disse ringene vokser når vi går oppover, kan det være lurt å se på de såkalte *konturene* til flaten, dvs. skjæringskurvene mellom flaten og plan parallelle med henholdsvis  $xz$ -planet og  $yz$ -planet. Setter vi for eksempel  $y = 0$  i uttrykket  $z = x^2 + 4y^2$ , får vi ligningen  $z = x^2$  for skjæringskurven mellom flaten og  $xz$ -planet. Dette betyr at ringene vokser på en slik måte at de passer inn i en parabel. Setter vi isteden  $x = 0$ , får vi uttrykket  $z = 4y^2$  som gir skjæringen mellom flaten og  $yz$ -planen. Dette er også en parabel, men en som er "slankere" enn den forrige.



Figur 7: Flaten  $z = x^2 + 4y^2$

Kombinerer vi den informasjonen vi nå har, får vi et ganske godt bilde av flaten – den er pokalformet med et litt avlangt tverrsnitt. Figur 7 viser en tredimensjonal tegning med nivåkurver.



La oss nå se på et eksempel hvor det lønner seg å modifisere teknikken ovenfor en smule.

**Eksempel 2:** Vi skal skisser grafen til funksjonen

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

Nivåkurvene til denne funksjonen er gitt ved

$$x^2 - y^2 = c$$

Disse kurvene ser litt forskjellige ut avhengig av om  $c$  er positiv, negativ eller 0. Er  $c$  positiv, får vi “stående” hyperbler

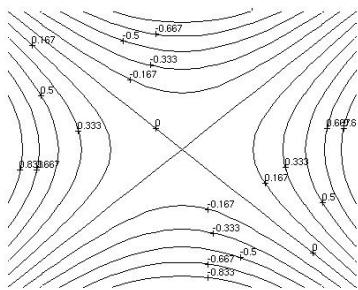
$$\frac{x^2}{(\sqrt{c})^2} - \frac{y^2}{(\sqrt{c})^2} = 1$$

Er  $c$  negativ, får vi “liggende” hyperbler

$$\frac{y^2}{(\sqrt{-c})^2} - \frac{x^2}{(\sqrt{-c})^2} = 1$$

Er  $c = 0$ , får vi ligningen  $x^2 - y^2 = 0$ , som gir oss de to rette linjene  $y = \pm x$ . Noen av nivåkurvene er vist på figur 8 (de tilhørende  $c$ -verdiene står på).

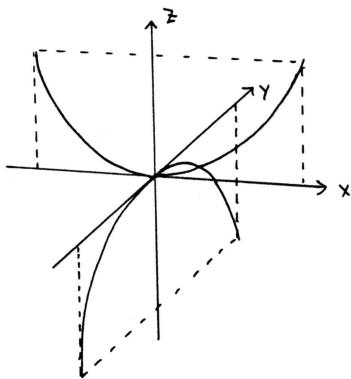
Nivåkurvene gir oss et hovedinntrykk av funksjonen. Starter vi i origo, vokser funksjonen hvis vi går langs  $x$ -aksen (uansett om vi går mot høyre eller venstre), mens den synker når vi går langs  $y$ -aksen. Noe godt utgangspunkt for å tegne funksjonsgrafen har vi allikevel ikke så langt.



Figur 8: Nivåkurver for  $f(x, y) = x^2 - y^2$

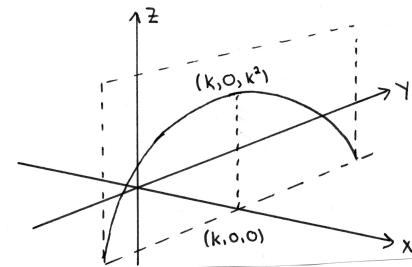
La oss bruke trikset fra forrige gang med å sette  $x$  og  $y$  lik 0 etter tur for å finne konturene til grafen. Vi får  $z = x^2$  og  $z = -y^2$  som er to parabler – den

ene vender oppover og den andre nedover. Tegner vi begge disse parablene i et tre-dimensjonalt aksekors, får vi figur 9 der det ser ut som om den ene parabelen henger på tvers over den andre.

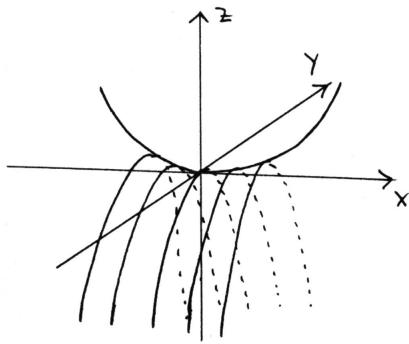


Figur 9: Skjæring med  $xz$ - og  $yz$ -planet

For å skaffe oss et enda bedre inntrykk av grafen, kan vi velge å sette  $x$  lik en annen konstant  $k$  for å se hvordan grafen skjærer planet  $x = k$ . Da får vi parabelen  $z = -y^2 + k^2$  (se figur 10).

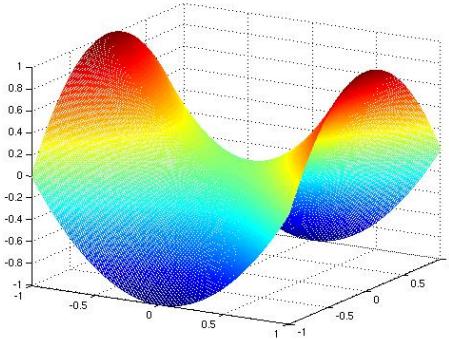


Figur 10: Skjæring med planet  $x = k$



Figur 11: Skisse av flaten  $z = x^2 - y^2$

Setter vi sammen den informasjonen vi nå har, får vi bildet i figur 11 der det ser ut som en serie av parabler er hengt opp på en parabelformet klessnor.



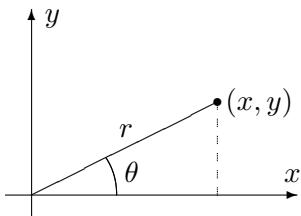
Figur 12. Flaten  $z = x^2 - y^2$

Tar vi utgangspunkt i denne forestillingen, er det lett å lage en tegning som i figur 12. Flaten ser ut omtrent som en sal (på en hest). ♣

Ved å kombinere teknikkene i eksemplene ovenfor kan du skissere de fleste funksjonsgrafer. I kapittel 5 vil du i tillegg lære teknikker for å finne maksimums- og minimumspunkter. Men det er ikke bare matematiske knep som er nyttige når man skal tegne en graf, det kan også være lurt å tenke litt på det tegnetekniske. Ofte vil du finne at en flate blir enklere å tegne dersom du skifter synsvinkel og ser den fra en annen kant.

### Polarkoordinater

Når vi skal angi posisjonen til et punkt i planet, er det vanligste å oppgi  $x$ - og  $y$ -koordinaten. I en del sammehenger er det imidlertid enklere og nyttigere å bruke *polarkoordinater*  $(r, \theta)$ . Polarkoordinatene fungerer på akkurat samme måte her som i teorien for komplekse tall (se figur 13).



Figur 13: Polarkoordinater

For å finne polarkoordinatene til punktet  $(x, y)$  regner vi først ut

$$r = \sqrt{x^2 + y^2}.$$

Deretter regner vi ut

$$\sin \theta = \frac{y}{r}.$$

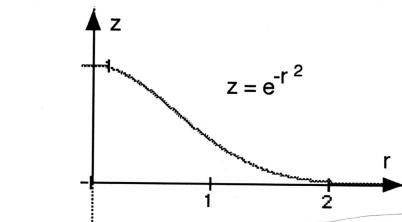
(du kan godt regne ut  $\cos \theta = \frac{x}{r}$  isteden). Det er to vinkler i første omløp med samme sinus, men ved å se på hvilken kvadrant punktet  $(x, y)$  ligger i, er det ikke vanskelig å plukke ut den riktige vinkelen.

Det hender også at vi må gå den andre veien – at vi kjenner polarkoordinatene  $r$  og  $\theta$ , og ønsker å finne  $x$  og  $y$ . Dette er lettere – vi observerer bare at

$$\begin{aligned} x &= r \cos \theta \\ y &= r \sin \theta. \end{aligned}$$

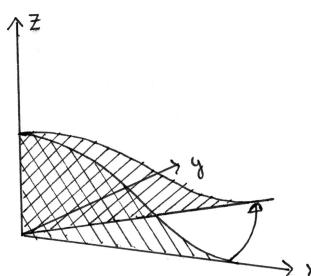
Siden vi kan angi punkter i planet ved hjelp av polarkoordinater  $(r, \theta)$  istedenfor kartesiske koordinater  $(x, y)$ , kan vi også beskrive funksjoner av to variable ved hjelp av polarkoordinater  $z = g(r, \theta)$  istedenfor kartesiske koordinater  $z = f(x, y)$ . Ofte kan det være nyttig å skrive om en funksjon til polarkoordinater for å få et bedre inntrykk av grafen.

**Eksempel 3:** Hvis vi skriver om  $z = e^{-(x^2+y^2)}$  til polarkoordinater, får vi  $z = e^{-r^2}$ . Tegner vi  $z = e^{-r^2}$  som en funksjon av én variabel, får vi grafen i figur 14.

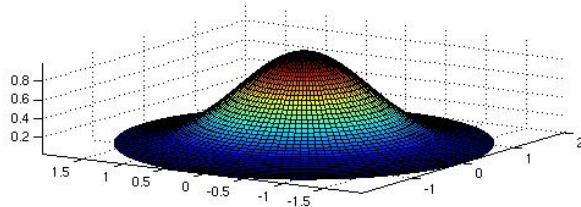


Figur 14: Grafen til  $z = e^{-r^2}$ .

Grafen til funksjonen  $z = e^{-(x^2+y^2)}$  får vi ved å rottere denne grafen om  $z$ -aksen. Figur 15 viser rotasjonen, og figur 16 viser resultatet.



Figur 15: Rotasjon av grafen  $z = e^{-r^2}$ .

Figur 16: Grafen  $z = e^{-(x^2+y^2)}$ .

Vi tar med et eksempel til:

**Eksempel 4:** Skriver vi om funksjonen  $z = x^2 - y^2$  i eksempel 2 til polar-kordinater, får vi

$$z = x^2 - y^2 = (r \cos \theta)^2 - (r \sin \theta)^2 = r^2(\cos^2 \theta - \sin^2 \theta) = r^2 \cos 2\theta.$$

Dette betyr at hvis vi holder vinkelen  $\theta$  konstant og varierer avstanden  $r$ , så følger  $z$  en parabelbue  $z = r^2 \cos 2\theta$ . Fortegnet til  $\cos 2\theta$  avgjør om parabelen vokser oppover eller nedover når  $r$  øker, og størrelsen til  $|\cos 2\theta|$  avgjør hvor rask denne veksten er. Forsøk å lage en skisse av grafen ut i fra den informasjonen du nå har.



### Funksjoner av tre variable

Hittil har vi holdt oss til funksjoner av to variable. Det er ikke mulig å tegne grafen til funksjoner av flere enn to variable på samme måte, men mye av den intuisjonen vi får fra det todimensjonale tilfellet, kan overføres til funksjoner av flere variable. For funksjoner  $u = f(x, y, z)$  av tre variable kan vi få en viss informasjon ved å se på nivåflater. Gitt et tall  $c$ , er den tilhørende nivåflaten gitt ved

$$N_c = \{(x, y, z) : f(x, y, z) = c\}$$

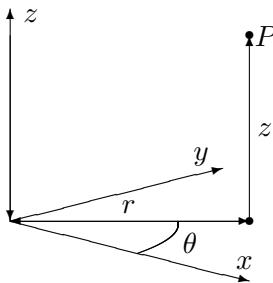
Disse mengdene kalles nivåflater fordi de ofte danner flater i rommet. Tar vi funksjonen

$$f(x, y, z) = x^2 + y^2 + z^2$$

som et eksempel, ser vi at nivåflaten  $N_c$  er et kuleskall med radius  $\sqrt{c}$  når  $c > 0$ . Når  $c = 0$ , består nivåflaten bare av punktet  $(0, 0, 0)$ , og når  $c < 0$ , inneholder ikke  $N_c$  noen punkter i det hele tatt. Dette gir oss et visst inntrykk av hvordan funksjonen  $f$  er — den er null i origo, konstant på kuleskall og

har tettere og tettere nivåflater dess lengre ut vi kommer (dvs. den vokser raskere og raskere),

Også for funksjoner av tre variable kan det ofte lønne seg å skrive om til andre koordinatsystemer. Vi skal se raskt på to slike koordinatsystemer – sylinderkoordinater og kulekoordinater. Figur 17 viser grunnideen for *sylinderkoordinater*; vi angir posisjonen til punktet  $P$  ved hjelp av de tre størrelsene  $r, \theta$  og  $z$ , der  $z$  er  $P$ 's høyde over  $xy$ -planet, mens  $r$  og  $\theta$  er polarkoordinatene til  $P$ 's projeksjon ned i  $xy$ -planet. Sylinderkoordinater er altså nært beslektet med polarkoordinater — vi har bare hektet på en tredje koordinat  $z$  for å kunne beskrive punktets høyde.



Figur 17: Sylinderkoordinater

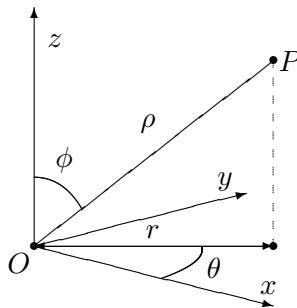
**Eksempel 5:** Vi skriver funksjonen

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

ved hjelp av sylinderkoordinater. Siden  $x^2 + y^2 = r^2$ , får vi

$$u = r^2 e^{-z}$$

Dette gir oss et visst inntrykk av hvordan  $f$  oppfører seg; så lenge vi holder oss i en fast høyde, vokser funksjonen proporsjonalt med kvadratet av avstanden  $r$  til  $z$ -aksen, men når vi forflytter oss rett oppover, avtar funksjonen proporsjonalt med  $e^{-z}$ .  $\square$



Figur 18: Kulekoordinater

I kulekoordinater beskrives posisjonen til et punkt  $P$  ved hjelp av en lengde  $\rho$  og to vinkler  $\theta$  og  $\phi$ . Figur 18 viser ideen:  $\rho$  er avstanden fra  $P$  til origo,  $\phi$  er vinkelen mellom  $z$ -aksen og vektoren  $\overrightarrow{OP}$ , og  $\theta$  er den samme vinkelen som for sylinderkoordinatene, nemlig vinkelen mellom  $x$ -aksen og projeksjonen  $\overrightarrow{OP'}$  av  $\overrightarrow{OP}$  ned i  $xy$ -planet. Vinkelen  $\phi$  ligger mellom 0 og  $\pi$ , mens  $\theta$  ligger mellom 0 og  $2\pi$ .

For å uttrykke  $x$ ,  $y$  og  $z$  ved hjelp av  $\rho$ ,  $\phi$  og  $\theta$ , observerer vi først at

$$z = \rho \cos \phi.$$

Vi ser også at  $|\overrightarrow{OP'}| = |\overrightarrow{OP}| \sin \phi = \rho \sin \phi$ . Dette betyr at

$$\begin{aligned} x &= |\overrightarrow{OP'}| \cos \theta = \rho \cos \theta \sin \phi \\ y &= |\overrightarrow{OP'}| \sin \theta = \rho \sin \theta \sin \phi \end{aligned}$$

La oss se hvordan disse formlene brukes i praksis:

**Eksempel 6:** Vi skal skrive om

$$u = x^2 + y^2 - z^2$$

til kulekoordinater. Vi ser at

$$\begin{aligned} x^2 + y^2 &= (\rho \cos \theta \sin \phi)^2 + (\rho \sin \theta \sin \phi)^2 = \rho^2 \sin^2 \phi \\ z^2 &= \rho^2 \cos^2 \phi \end{aligned}$$

så

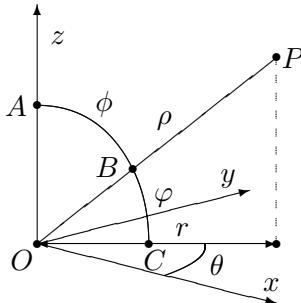
$$u = \rho^2 \sin^2 \phi - \rho^2 \cos^2 \phi = -\rho^2 \cos 2\phi.$$

Dette betyr at  $u$  er uavhengig av vinkelen  $\theta$ . Holder vi vinkelen  $\phi$  konstant, vokser eller avtar  $u$  proporsjonalt med  $\rho^2$ . Om  $u$  er positiv eller negativ avhenger av størrelsen på  $\phi$ ;  $u$  er positiv for  $\phi \in \left(\frac{\pi}{4}, \frac{3\pi}{4}\right)$  og negativ for  $\phi \in \left(0, \frac{\pi}{4}\right) \cup \left(\frac{3\pi}{4}, \pi\right)$ .

**Advarsel:** Det er litt forskjellige konvensjoner ute og går når det gjelder vinkelen  $\phi$  i kulekoordinater. Vi har gjort det valget som er vanlig i matematikkbøker, nemlig at  $\phi$  er vinkelen mellom  $z$ -aksen og radius vektor  $\overrightarrow{OP}$ , mens det i enkelte andre fag er vanlig å bruke vinkelen mellom  $\overrightarrow{OP}$  og  $xy$ -planet isteden. Figur 19 viser de to vinklene (vi har kalt dem  $\phi$  og  $\varphi$  som bare er to forskjellige utforminger av den greske bokstaven "phi"). "Vår" vinkel  $\phi$  er markert av buen fra  $A$  til  $B$ , mens den "alternative" vinkelen  $\varphi$  er markert av buen fra  $B$  til  $C$ . Vi ser at  $\phi$  og  $\varphi$  er komplementvinkler, dvs. at  $\phi + \varphi = \frac{\pi}{2}$ . Det betyr at

$$\sin \phi = \cos \varphi \quad \text{og} \quad \cos \phi = \sin \varphi$$

Bytter du fra én konvensjon til den andre, må du altså bytte ut alle sinus'er med cosinus'er og omvendt (men bare dem som gjelder  $\phi$ , den andre vinkelen  $\theta$  er lik i begge tilfeller!)



Figur 19: De to variantene av vinkelen “phi” i kulekoordinater

Du kommer til å få bruk for sylinder- og kulekoordinater for alvor når du kommer til integraler av funksjoner av tre variable. I et slikt integral skal du integrere en funksjon over et område i rommet, og hvis funksjonen og/eller området er lettere å beskrive i sylinder- eller kulekoordinater, lønner det seg ofte å bytte til disse (omtrent som du bytter variabel i et vanlig integral).

### Nivåflater og tangentplan

La oss avslutte denne seksjonen med to litt mer teoretiske temaer. Vi har allerede definert nivåkurver for funksjoner av to variable og nivåflater for funksjoner av tre variable. Den generelle definisjonen ser slik ut:

**Definisjon 3.7.1** *Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable, og at  $c \in \mathbb{R}$  er et tall. Da kalles mengden*

$$N_c = \{\mathbf{x} \in A \mid f(\mathbf{x}) = c\}$$

*en nivåflate for  $f$ .*

Vi vet allerede at gradienten  $\nabla f(\mathbf{a})$  peker i den retningen hvor funksjonen  $f$  vokser raskest i punktet  $\mathbf{a}$ . Ikke overraskende står denne retningen alltid normalt på nivåflaten  $N_c$  i følgende forstand: Dersom  $\mathbf{r}$  er en deriverbar kurve som ligger på nivåflaten (*dvs.*  $f(\mathbf{r}(t)) = c$  for alle  $t$ ), og  $\mathbf{r}$  er i punktet  $\mathbf{a}$  ved tiden  $t_0$ , så er

$$\nabla f(\mathbf{a}) \cdot \mathbf{r}'(t_0) = 0$$

*dvs. tangentvektoren til kurven i punktet  $\mathbf{a}$  står normalt på gradienten  $\nabla f(\mathbf{a})$  i punktet.*

*Bevis:* Vi bruker kjerneregelen i versjon 3.2.1. Siden  $\mathbf{r}$  ligger på nivåflaten  $N_c$  er funksjonen  $u(t) = f(\mathbf{r}(t))$  konstant lik  $c$ , og har derfor derivert 0. Bruker vi kjerneregelen, får vi imidlertid

$$u'(t_0) = \nabla f(\mathbf{r}(t_0)) \cdot \mathbf{r}'(t_0) = \nabla f(\mathbf{a}) \cdot \mathbf{r}'(t_0)$$

Følgelig er  $\nabla f(\mathbf{a}) \cdot \mathbf{r}'(t_0) = 0$ , og setningen er bevist.  $\square$

Det neste temaet vi skal se på, er tangentplan. Vi begynner med en funksjon  $z = f(x, y)$  av to variable og tenker oss at vi har et punkt  $\mathbf{b} = (x_0, y_0, f(x_0, y_0))$  på funksjonsgrafen. Vi tenker oss også at vi har en stor, bevegelig, plan flate som vi skrur fast til funksjonsgrafen i punktet  $\mathbf{b}$ . Etterhvert som skruen fastner, får flaten mindre og mindre rom til å bevege seg, og til slutt sitter den helt fast inntil flaten. Det planet som flaten nå definerer, er *tangentplanet* til  $f$  i punktet  $\mathbf{b}$ .

La oss se om vi kan finne en mer matematisk beskrivelse av tangentplanet. For å finne ligningen til et plan (se seksjon 1.4) trenger vi to ting; et punkt i planet og en normalvektor. Vi vet at tangentplanet vårt går gjennom  $\mathbf{b}$ , så alt vi trenger er en normalvektor. Det er flere måter å finne en normalvektor på, og vi skal først bruke et lite triks som bygger på det vi nettopp har lært om grader.

De punktene  $(x, y, z)$  som ligger på funksjonsgrafen tilfredsstiller ligningen  $z = f(x, y)$  eller, om vi flytter litt rundt på leddene,

$$z - f(x, y) = 0$$

Lar vi  $g(x, y, z) = z - f(x, y)$  være en funksjon av tre variable, ser vi at funksjonsgrafen vår rett og slett er nivåflaten  $N_0$  til funksjonen  $g$ . Siden gradienten til  $g$  står normalt på flaten, må den også stå normalt på tangentplanet. Alt vi behøver å gjøre for å finne en normalvektor, er derfor å regne ut gradienten til  $g(x, y, z) = z - f(x, y)$  i punktet  $\mathbf{b} = (x_0, y_0, f(x_0, y_0))$ . Generelt har vi

$$\frac{\partial g}{\partial x}(x, y, z) = -\frac{\partial f}{\partial x}(x, y), \quad \frac{\partial g}{\partial y}(x, y, z) = -\frac{\partial f}{\partial y}(x, y), \quad \frac{\partial g}{\partial z}(x, y, z) = 1$$

som gir

$$\nabla g(x, y, z) = -\frac{\partial f}{\partial x}(x, y) \mathbf{i} - \frac{\partial f}{\partial y}(x, y) \mathbf{j} + \mathbf{k}$$

En normalvektor i punktet  $\mathbf{b} = (x_0, y_0, f(x_0, y_0))$  er dermed

$$\mathbf{n} = \nabla g(x_0, y_0, f(x_0, y_0)) = -\frac{\partial f}{\partial x}(x_0, y_0) \mathbf{i} - \frac{\partial f}{\partial y}(x_0, y_0) \mathbf{j} + \mathbf{k}$$

Ligningen for planet gjennom  $\mathbf{b}$  med normalvektor  $\mathbf{n}$  er

$$0 = \mathbf{n} \cdot (\mathbf{x} - \mathbf{b}) =$$

$$\left(-\frac{\partial f}{\partial x}(x_0, y_0) \mathbf{i} - \frac{\partial f}{\partial y}(x_0, y_0) \mathbf{j} + \mathbf{k}\right) \cdot ((x - x_0) \mathbf{i} + (y - y_0) \mathbf{j} + (z - f(x_0, y_0)) \mathbf{k})$$

som etter litt opprydning gir

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

Vi har dermed kommet frem til en fornuftig definisjon av tangentplanet:

**Definisjon 3.7.3** Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av to variable og at  $f$  er deriverbar i punktet  $(x_0, y_0)$ . Tangentplanet til  $f$  i punktet  $(x_0, y_0, f(x_0, y_0))$  er da definert ved ligningen

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0)$$

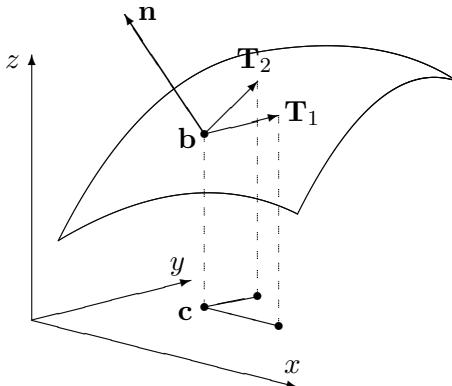
Normalretningen i punktet er gitt ved vektoren

$$\mathbf{n} = -\frac{\partial f}{\partial x}(x_0, y_0) \mathbf{i} - \frac{\partial f}{\partial y}(x_0, y_0) \mathbf{j} + \mathbf{k}$$

Figur 20 viser en annen måte vi kan finne frem til normalvektoren  $\mathbf{n}$  på. Vi tenker oss at vi starter i punktet  $\mathbf{b} = (x_0, y_0, f(x_0, y_0))$  på funksjonsgrafen, og at vi går på tangentplanet med konstant  $y$ -koordinat (“skyggen” vår i  $xy$ -planet starter dermed i punktet  $\mathbf{c} = (x_0, y_0, 0)$  og beveger seg parallelt med  $x$ -aksen). Siden stigningstallet i  $x$ -retning er  $\frac{\partial f}{\partial x}$ , vil vi ha steget  $\frac{\partial f}{\partial x}(x_0, y_0)$  enheter i  $z$ -retning når vi har gått en enhet i  $x$ -retning. Dette gir oss vektoren  $\mathbf{T}_1 = (1, 0, \frac{\partial f}{\partial x}(x_0, y_0))$  (se figur 20). Beveger vi oss på tilsvarende måte i  $y$ -retning, får vi vektoren  $\mathbf{T}_2 = (0, 1, \frac{\partial f}{\partial y}(x_0, y_0))$  på figur 19. Både  $\mathbf{T}_1$  og  $\mathbf{T}_2$  er tangentvektorer til flaten, og vektorproduktet deres må derfor være en normalvektor. Regner du ut dette produktet, ser du at

$$\mathbf{n} = \mathbf{T}_1 \times \mathbf{T}_2$$

der  $\mathbf{n}$  er som ovenfor.



Figur 20: Normalvektoren  $\mathbf{n}$  og tangentvektorene  $\mathbf{T}_1$  og  $\mathbf{T}_2$

Det er på tide med et eksempel.

**Eksempel 7.** Vi skal finne en normalvektor for funksjonen  $f(x, y) = x^3y^2$  i punktet  $(2, -1)$ . Vi skal dessuten finne ligningen til tangentplanet i punktet.

Først trenger vi de partielle deriverte:

$$\frac{\partial f}{\partial x}(x, y) = 3x^2y^2 \implies \frac{\partial f}{\partial x}(2, -1) = 12$$

$$\frac{\partial f}{\partial y}(x, y) = 2x^3y \implies \frac{\partial f}{\partial y}(2, -1) = -16$$

Setter vi inn i formelen for normalvektor i definisjon 3.7.3, får vi

$$\mathbf{n} = -\frac{\partial f}{\partial x}(x_0, y_0)\mathbf{i} - \frac{\partial f}{\partial y}(x_0, y_0)\mathbf{j} + \mathbf{k} = -12\mathbf{i} + 16\mathbf{j} + \mathbf{k}$$

For å finne ligningen for tangentplanet, kan vi f.eks. bruke den øverste ligningen i definisjon 3.7.3.

$$z = 8 + 12(x - 2) - 16(y + 1)$$

(der vi også har brukt at  $f(2, -1) = 8$ ). Denne formelen kan også skrives

$$12x - 16y - z = 32$$



La oss gjøre en observasjon til. Innfører vi notasjonen  $\mathbf{a} = (x_0, y_0)$ ,  $\mathbf{r} = (x, y)$ , kan ligningen for tangenplanet skrives

$$z = f(x_0, y_0) + \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{r} - \mathbf{a})$$

Dette minner om formelen for lineariseringen til  $f$  (se definisjon 2.9.4). For et skalarfelt er Jacobi-matrisen det samme som gradienten, og lineariseringen er dermed gitt ved

$$T_{\mathbf{a}}(\mathbf{r}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{r} - \mathbf{a})$$

Altså er tangentplanet i  $\mathbf{a}$  rett og slett grafen til lineariseringen i  $\mathbf{a}$ . Dette er ikke så rart — lineariseringen er den affinavbildningen som tilnærmer  $f$  best i nærheten av  $\mathbf{a}$ , og tangentplanet er det planet som tilnærmer grafen til  $f$  best i nærheten av  $\mathbf{a}$ . Det er ikke vanskelig å vise at en funksjon  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  er en affinavbildning hvis og bare hvis grafen er et plan, og dermed er ringen sluttet.

Definisjon 3.7.3 har en generalisering til høyere dimensjoner som vi tar med for ordens skyld.

**Definisjon 3.7.4** Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable og at  $f$  er deriverbar i punktet  $\mathbf{a}$ . Normalretningen til funksjonsgrafen i punktet  $(\mathbf{a}, f(\mathbf{a}))$  er gitt ved vektoren

$$\mathbf{n} = \left( -\frac{\partial f}{\partial x_1}(\mathbf{a}), -\frac{\partial f}{\partial x_2}(\mathbf{a}), \dots, -\frac{\partial f}{\partial x_n}(\mathbf{a}), 1 \right)$$

Tangentplanet til  $f$  i punktet  $(\mathbf{a}, f(\mathbf{a}))$  består av de punktene  $(\mathbf{x}, z) \in \mathbb{R}^{n+1}$  som tilfredsstiller ligningen:

$$z = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$$

### MATLAB-kommentarer

MATLAB er et utmerket hjelpemiddel for å fremstille grafer til funksjoner  $z = f(x, y)$ . For å tegne en slik graf med MATLAB må du først definere et rutenett (et “grid”) i  $xy$ -planet. Deretter må du få MATLAB til å regne ut funksjonsverdiene i alle hjørnene i rutenettet, og til slutt må du få MATLAB til å tegne resultatet (MATLAB vil da plotte alle punkter av typen  $(x, y, f(x, y))$  der  $(x, y)$  er et hjørne i rutenettet, og så forbinde alle nabopunkter med rette streker).

**Eksempel 8:** Vi skal tegne grafen til  $f(x, y) = x^3 - 4y^2$  over rektangelet  $x \in [-3, 3]$ ,  $y \in [-5, 5]$ .

Vi lager først en oppdeling av de to intervallene vi er interessert i, ved å skrive

```
>> r=-3:0.1:3;
>> s=-5:0.1:5;
```

(husk semikolon etter kommandoene, ellers vil du få lange tallremser som output!) Her har vi valgt å dele opp begge intervallene i skritt med lengde 0.1, men du kan godt velge en finere eller grovere oppdeling. Det kan være lurt å prøve en skikkelig grov oppdeling (f.eks. skritt lengde 0.5) en gang slik at du virkelig ser hvordan MATLAB tegner grafer.

Neste skritt er å lage et rutenett av oppdelingene våre. Dette gjør vi med kommandoen

```
>> [x,y]=meshgrid(r,s);
```

Vi kan nå definere funksjonen:

```
>> z=x.^3-4*y.^2;
```

(husk å bruke .-versjonene av de algebraiske operasjonene!) Dermed er vi klare til selve plottingen som utføres av kommandoen

```
>> mesh(x,y,z)
```

Grafen kommer opp i et eget vindu på vanlig måte. Husk at du kan dreie på grafen ved først å klikke på ikonet som symboliserer dreining, og så dra grafen rundt med musa. Bruker du kommandoen `surf(x,y,z)` istedenfor `mesh(x,y,z)`, vil MATLAB tegne grafen med fargelegging av hvert rutelement. Det er ofte klargjørende når grafen varierer mye.

Vil du se nivåkurvene istedenfor grafen, bytter du ut den siste kommando'en ovenfor med

```
>> contour(x,y,z)
```

Når du bruker `contour` på denne måten, velger MATLAB selv hvilke nivåkurver den skal tegne. Siden MATLAB ikke er veldig dyktig til å finne de mest interessante nivåkurvene, hender det at du må hjelpe til. Dersom vektoren  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  er lagt inn, vil kommandoen

```
>> contour(x,y,z,v)
```

tvinge MATLAB til å tegne nivåkurvene med verdier  $v_1, v_2, \dots, v_n$ . Vil du bare regulere *antall* nivåkurver, men ikke spesifisere verdiene, kan du bruke denne kommando'en

```
>> contour(x,y,z,n)
```

som får MATLAB til å tegne opp  $n$  nivåkurver. Med kommandoen `clabel` får du MATLAB til å skrive nivået til nivåkurvene på grafen. Prøv

```
>> clabel(contour(x,y,z,12))
```

MATLAB vil normalt tegne nivåkurvene i forskjellige farger. Dette er nyttig på skjermen, men kan være mindre praktisk dersom du ønsker å lime figuren inn i et svart-hvitt dokument. Skriver du

```
>> contour(x,y,z,8,'k')
```

får du 8 nivåkurver tegnet i svart ('k' er symbolet for svart farge). Ønsker du at MATLAB skal tegne nivåkurvene og grafen i samme plot, bruker du kommando'en

```
>> meshc(x,y,z)
```

Det finnes mange andre kommandoer du også kan bruke (og mange flere måter å modifisere kommandoene ovenfor på!). Skriv

```
>> help graph3d
```

for å få en oversikt.

### Oppgaver til seksjon 3.7

**1.** Finn nivåkurvene til funksjonen. Tegn nok av dem til at du kan danne deg et bilde av funksjonsgrafen

$$\begin{array}{ll} \text{a)} & f(x, y) = 4x^2 + 3y^2 \\ \text{b)} & f(x, y) = \frac{1}{x^2 - y^2} \\ \text{c)} & f(x, y) = e^{x^2 - y^2} \end{array} \quad \begin{array}{ll} \text{d)} & f(x, y) = e^{x^2 - y} \\ \text{e)} & f(x, y) = \frac{x}{x^2 + y^2} \end{array}$$

**2.** Skisser grafen til funksjonen og sammenlign resultatet med det du får når du bruker MATLAB.

$$\begin{array}{ll} \text{a)} & f(x, y) = 2x^2 + y^2 \\ \text{b)} & f(x, y) = y^2 - x \\ \text{c)} & f(x, y) = \sin(x^2 + y^2) \end{array} \quad \begin{array}{ll} \text{d)} & f(x, y) = x^2 - 4y^2 \\ \text{e)} & f(x, y) = \ln(xy) \end{array}$$

**3.** Skriv om funksjonen til polarkoordinater. Skisser grafen og sammenlign resultatet med det du får når du bruker MATLAB.

$$\begin{array}{ll} \text{a)} & f(x, y) = \frac{1}{\sqrt{x^2 + y^2}} \\ \text{b)} & f(x, y) = \frac{x}{x^2 + y^2} \\ \text{c)} & f(x, y) = \frac{y}{x} \end{array} \quad \begin{array}{ll} \text{d)} & f(x, y) = x^2 - 4y^2 \\ \text{e)} & f(x, y) = e^{xy} \end{array}$$

**4.** Skriv om funksjonen til både sylinder- og kulekoordinater. Avgjør hva du synes er mest informativt i hvert enkelt tilfelle.

$$\begin{array}{ll} \text{a)} & f(x, y, z) = (x^2 + y^2)e^{-z^2} \\ \text{b)} & f(x, y, z) = \frac{1}{x^2 + y^2 + z^2} \\ \text{c)} & f(x, y, z) = \frac{x^2 + y^2}{z^2} \end{array} \quad \begin{array}{ll} \text{d)} & f(x, y, z) = \frac{x^2 - y^2}{z} \\ \text{e)} & f(x, y, z) = z \cdot \arctan\left(\frac{y}{x}\right) \\ \text{f)} & f(x, y, z) = x^2 + y^2 - 2z^2 \end{array}$$

**5.** Finn en ligning for tangentplanet til funksjonen i det angitte punktet.

$$\begin{array}{ll} \text{a)} & f(x, y) = x^2y \quad \text{i punktet } (1, -2) \\ \text{b)} & f(x, y) = xe^{-xy} \quad \text{i punktet } (1, 0) \\ \text{c)} & f(x, y) = x^2y - xy^2 \quad \text{i punktet } (2, -2) \end{array}$$

**6.** I eksempel 3 i seksjon 2.2 studerte vi funksjonen

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & \text{for } (x, y) \neq (0, 0) \\ 0 & \text{når } (x, y) = 0 \end{cases}$$

som et eksempel på en funksjon som "oppfører seg kontinuerlig" når vi nærmer oss origo langs rette linjer, men ikke når vi følger parabelen  $y = x^2$ . Bruk MATLAB til å tegne grafen til funksjonen, og studer fenomenet ved å dreie på grafen. Tegn også inn kurven  $\mathbf{r}(t) = t\mathbf{i} + t^2\mathbf{j} + \frac{1}{2}\mathbf{k}$  i samme figur. Tegn til slutt konturkurvene til funksjonen.

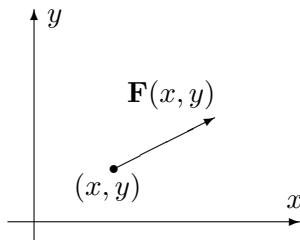
**7.** I oppgave 2.5.4 studerte vi funksjonen

$$f(x, y) = \begin{cases} \frac{x^3 y - x y^3}{(x^2 + y^2)^2} & \text{for } (x, y) \neq (0, 0) \\ 0 & \text{når } (x, y) = 0 \end{cases}$$

som et eksempel på en funksjon der de blandede annenordens partiellderiverte er forskjellige. Bruk MATLAB til å tegne grafen og konturkurvene til funksjonen, og prøv å forstå fra figuren hvorfor  $\frac{\partial^2 f}{\partial y \partial x}(0, 0)$  er negativ mens  $\frac{\partial^2 f}{\partial x \partial y}(0, 0)$  er positiv.

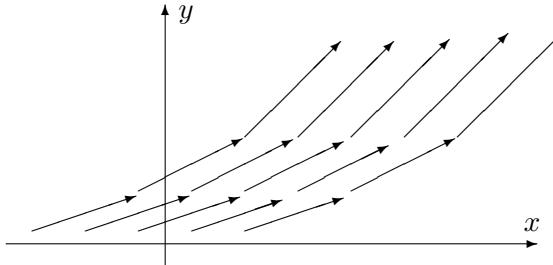
### 3.8 Grafisk fremstilling av vektorfelt

I denne seksjonen skal vi se på grafisk fremstilling av vektorfelt i to og tre dimensjoner, altså av funksjoner  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  og  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ . Det er lettest å forstå disse fremstillingene hvis man tenker litt praktisk. Dersom  $\mathbf{F}(x, y)$  angir vindretningen i punktet  $(x, y)$ , er det naturlig å illustrere dette som vist på figur 1; vi plasserer rett og slett vektoren  $\mathbf{F}(x, y)$  med startpunkt  $(x, y)$ .



Figur 1: Vektorfeltet  $\mathbf{F}(x, y)$

Gjør vi det samme i mange punkter, får vi en oversikt over hvordan vindfeltet er (se figur 2).



Figur 2: Vektorfeltet  $\mathbf{F}(x, y)$

Det går an å lage slike fremstillinger for hånd, men det er tidkrevende og gir ikke noen spesiell innsikt i hvordan vektorfeltet oppfører seg. MATLAB har en praktisk kommando som heter **quiver** for å lage slike diagrammer. Vi skal se litt på hvordan kommandoen fungerer. Dersom du har lagt inn to  $m \times n$  matriser  $u$  og  $v$ , vil kommandoen

```
>> quiver(u,v)
```

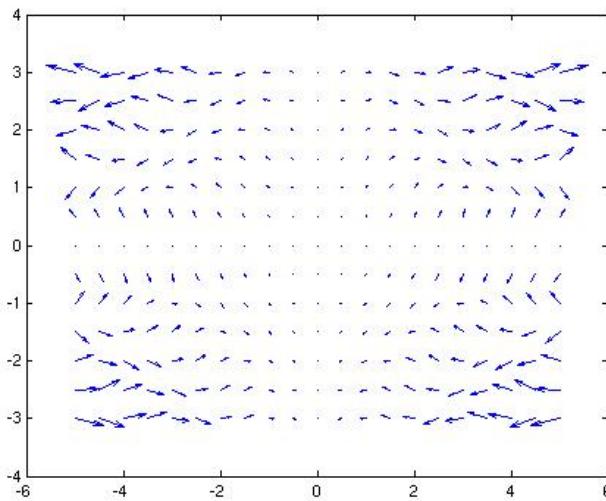
få MATLAB til å produsere et bilde der det ut fra hvert punkt med heltallige koordinater  $(i, j)$  (der  $1 \leq i \leq m$  og  $1 \leq j \leq n$ ) går en vektor som er proporsjonal med  $(u_{ij}, v_{ij})$ . Vektoren er *proporsjonal med*  $(u_{ij}, v_{ij})$  (og ikke *lik*  $(u_{ij}, v_{ij})$  som man kanskje skulle tro) siden MATLAB automatisk skalerer vektorene slik at de ikke overlapper. Alle vektorene skaleres likt slik at det innbyrdes størrelsesforholdet er korrekt. MATLAB lager altså et rutenett

med  $m \times n$  ruter, der hver rute har sidekant 1, og der det fra hvert hjørne går en vektor spesifisert av de korresponderende elementene i  $u$  og  $v$ .

Ofte ønsker vi selv å spesifisere hvilket rutenett vektorene våre skal starte i (og ikke bruke "standardnettet" ovenfor). Dersom  $\mathbf{x}$  og  $\mathbf{y}$  er vektorer med hhv.  $m$  og  $n$  komponenter, kan vi bruke kommandoen

```
>> quiver(x,y,u,v)
```

Dette får MATLAB til å lage en figur der det ut fra hvert punkt av typen  $(x_i, y_j)$  går en vektor parallel med  $(u_{ij}, v_{ij})$  (se figur 3).



Figur 3: Vektorfelt fremstilt med `quiver`.

**Eksempel 1:** Vi skal lage en fremstilling av vektorfeltet

$$\mathbf{F}(x, y) = xy \mathbf{i} + x \sin(xy) \mathbf{j}$$

over mengden  $-5 \leq x \leq 5$ ,  $-3 \leq y \leq 3$ .

Vi starter med å lage rutenettet. Vi bør ikke lage oppdelingen *for fin*, for da blir det vanskelig å se vektorene. Vi velger en rutelengde på 0.5:

```
>> x=-5:0.5:5;
>> y=-3:0.5:3;
```

Vi lager så et rutenett av  $\mathbf{x}$  og  $\mathbf{y}$ .

```
>> [x,y]=meshgrid(x,y);
```

Nå kan vi legge inn vektorfeltet:

```
>> u=x.*y;
>> v=x.*sin(x.*y);
```

Dermed er vi klare til å lage figuren

```
>> quiver(x,y,u,v)
```

MATLAB svarer med figur 3 (se ovenfor). 

Vi kan fremstille vektorfelt  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  på samme måte, men resultatene er ofte vanskeligere å tolke fordi det ikke er så lett å få romfølelse i bildet. MATLAB har en kommando `quiver3` som fungerer på samme måte som `quiver`, men som tar tre variable der `quiver` tar to.

La oss se på en kommando til. Vi tenker oss at vektorfeltet  $\mathbf{F}$  angir hastigheten til noe som strømmer, f.eks. at  $\mathbf{F}(x, y)$  angir strømningshastigheten til havet i punktet  $(x, y)$ . Anta at en partikkkel som flyter med strømmen er i punktet  $\mathbf{r}(t) = (x(t), y(t))$  ved tiden  $t$ ; da er  $\mathbf{v}(t) = \mathbf{r}'(t) = \mathbf{F}(x(t), y(t))$ . Dette betyr at banen som partikkelen følger, hele tiden har vektorfeltet som tangent. Vi kaller slike baner for strømningslinjer.

**Eksempel 2:** La oss bruke MATLAB til å finne en strømningslinje for vektorfeltet i eksempel 1. Vi må først legge inn  $x$ ,  $y$ ,  $u$  og  $v$  på samme måte som ovenfor. I dette tilfellet kan det imidlertid være lurt å bruke et rutenett som er mindre grovt, så vi starter med sekvensen

```
>> x=-5:0.05:5;
>> y=-3:0.05:3;
>> [x,y]=meshgrid(x,y);
>> u=x.*y;
>> v=x.*sin(x.*y);
```

For å lage strømningslinjen som starter i punktet  $(1, -1)$  skriver vi nå

```
>> streamline(x,y,u,v,1,-1)
```

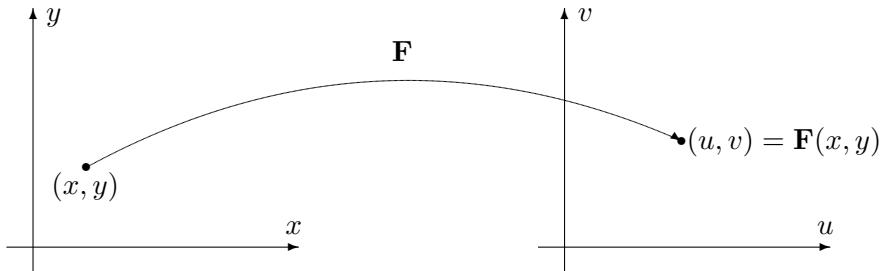
Hvis du taster inn den siste kommandoen på nytt med et annet startpunkt, f.eks.

```
>> streamline(x,y,u,v,0.5,1)
```

tegner MATLAB en ny strømningslinje i det samme figurvinduet. Det er også instruktivt å tegne inn vektorfeltet og strømningslinjene i samme figur (prøv!). 

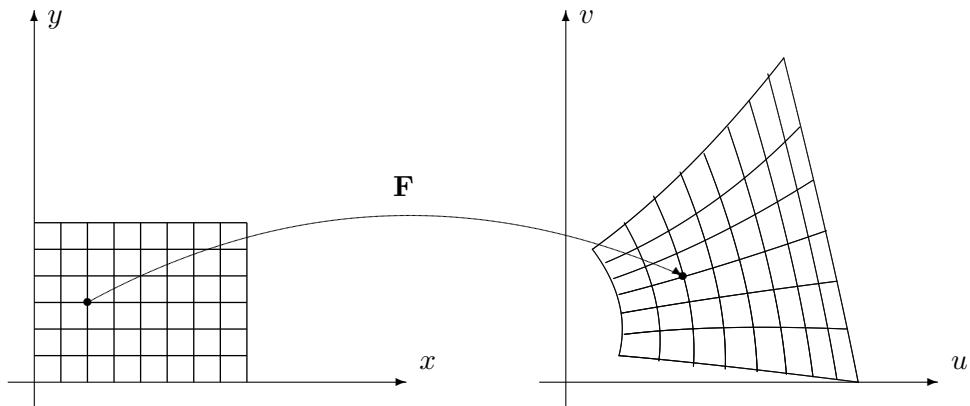
De grafiske fremstillingene vi hittil har sett på, er naturlige i situasjoner der man tenker på  $\mathbf{F}(x, y)$  som en vektor som starter i punktet  $(x, y)$ . I andre

situasjoner er det mer naturlig å tenke på  $\mathbf{F}$  som en avbildning som sender punkter  $(x, y)$  i  $\mathbb{R}^2$  til nye punkter  $(u, v)$  i  $\mathbb{R}^2$  (se figur 4).



Figur 4:  $\mathbf{F}$  avbilder  $(x, y)$  på  $(u, v)$ .

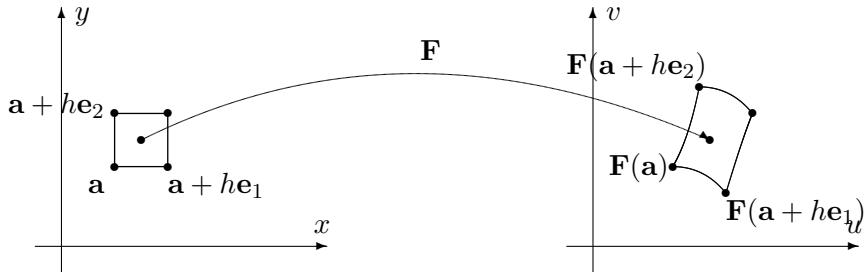
For å få et godt bilde av funksjonen, må vi se hvordan den virker på flere punkter. I figur 5 har vi tegnet opp hvordan  $\mathbf{F}$  virker på punktene i et rutenett.



Figur 5: Bildet av et rutenett under  $\mathbf{F}$ .

Figur 5 ovenfor minner om figur 1 i seksjon 2.9. Det er samme situasjon som er avbildet, den eneste forskjellen er at i seksjon 2.9 er  $\mathbf{F}$  en affinavbildning, mens den her er en generell avbildning. Affinavbildninger avbilder parallelle linjer på parallelle linjer, og bildet av rutnettet består av parallelogrammer av samme størrelse. I det generelle tilfellet avbildes rette linjer på krumme kurver, og bildene av kvadratene har varierende størrelse.

I figur 6 har vi tegnet opp hvordan  $\mathbf{F}$  virker på et lite kvadrat med sidekant  $h$ . Vi ønsker å estimere hvor stort bildet av kvadratet er sammenlignet med kvadratet. Arealet av kvadratet er åpenbart  $h^2$ . Arealet av bildet er tilnærmet lik parallelogrammet utspent av vektorene  $\mathbf{F}(\mathbf{a} + h\mathbf{e}_1) - \mathbf{F}(\mathbf{a})$  og  $\mathbf{F}(\mathbf{a} + h\mathbf{e}_2) - \mathbf{F}(\mathbf{a})$ . Siden  $h$  er liten, er disse størrelsene tilnærmet lik



Figur 6: Avbildning av et kvadrat

henholdsvis  $\frac{\partial \mathbf{F}}{\partial x}(\mathbf{a})h$  og  $\frac{\partial \mathbf{F}}{\partial y}(\mathbf{a})h$ . Arealet utspent av disse størrelsene er

$$\left| \begin{vmatrix} \frac{\partial F_1}{\partial x}(\mathbf{a}) & \frac{\partial F_1}{\partial y}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x}(\mathbf{a}) & \frac{\partial F_2}{\partial y}(\mathbf{a}) \end{vmatrix} \right| h^2 = |\det(\mathbf{F}'(\mathbf{a}))| h^2$$

Dette viser at (tallverdien til) determinanten til Jacobi-matrissen  $\mathbf{F}'(\mathbf{a})$  er et godt mål på hvor mye  $\mathbf{F}$  forstørrer arealer. Legg merke til at denne størrelsen normalt avhenger av  $\mathbf{a}$ , og derfor varierer fra sted til sted.

Som vi observerte i seksjon 2.9, kan vi også finne frem til dette på en litt annen måte. I nærheten av  $\mathbf{a}$  er lineariseringen  $T_{\mathbf{a}}\mathbf{F}$  en svært god tilnærming til  $\mathbf{F}$ . Dette er en affinavbildning med matrise

$$\mathbf{F}'(\mathbf{a}) = \begin{pmatrix} \frac{\partial F_1}{\partial x}(\mathbf{a}) & \frac{\partial F_1}{\partial y}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x}(\mathbf{a}) & \frac{\partial F_2}{\partial y}(\mathbf{a}) \end{pmatrix}$$

og har derfor forstørrelsesfaktor  $|\det(\mathbf{F}'(\mathbf{a}))|$ .

Determinanten til Jacobi-matrissen kalles *Jacobi-determinanten* og spiller en viktig rolle når vi skal skifte variabel i integraler av flere variable. Det skyldes nettopp at den er et mål på den lokale forstørrelsesfaktoren til en avbildning.

### Oppgaver til seksjon 3.8

**1.** Bruk kommandoen `quiver` til å tegne vektorfeltet. Tegn også inn noen strømningslinjer.

- a)  $\mathbf{F}(x, y) = \cos x \mathbf{i} + \sin x \mathbf{j}$
- b)  $\mathbf{F}(x, y) = -\frac{x}{(x^2+y^2)^{\frac{3}{2}}} \mathbf{i} - \frac{y}{(x^2+y^2)^{\frac{3}{2}}} \mathbf{j}$
- c)  $\mathbf{F}(x, y) = \left( \frac{1-x}{(x-1)^2+y^2} + \frac{1+x}{(x+1)^2+y^2} \right) \mathbf{i} + \left( \frac{-y}{(x-1)^2+y^2} + \frac{y}{(x+1)^2+y^2} \right) \mathbf{j}$

**2.** I denne oppgaven skal vi se nærmere på vektorfeltet

$$\mathbf{F}(x, y) = -\frac{y}{x^2+y^2} \mathbf{i} + \frac{x}{x^2+y^2} \mathbf{j}$$

i eksempel 3 i seksjon 3.5

- a) Bruk kommandoen `quiver` til å tegne vektorfeltet. Bruk en forholdsvis grov oppdeling på aksene.
- b) Forklar at strømningslinjene til  $\mathbf{F}$  er sirkler med sentrum i origo.
- c) Tegn strømningslinjen som starter i punktet  $(1, 0)$  på samme figur som vektorfeltet i a). Sammenlign resultatet med b).
- d) Gjenta punkt a) og c) med mye finere oppdeling av aksene. Hva skjer med strømningslinjene?

**3.** I denne oppgaven skal vi bruke MATLAB til å eksperimentere litt med avbildninger slik som demonstrert i figur 5. Vi har altså en funksjon  $(u, v) = \mathbf{F}(x, y)$  og vil se hvordan den avbilder et rutenett i  $xy$ -planet. For å slippe å lage et for omfattende MATLAB-program, skal vi nøye oss med en tilnærming der MATLAB tegner opp hjørnene i det fordreide rutenettet og forbinder dem med rette streker (som i de fleste eksemplene burde vært buede kurver). Her er programmet i det tilfellet  $u = 3x - y$  og  $v = x + 2y$ :

```
>> r=-2:0.25:2; %lager oppdeling av x-aksen
>> s=-2:0.25:2; %lager oppdeling av y-aksen
>> [x,y]=meshgrid(r,s); %lager rutenett av oppdelingene
>> u=3.*x-y; %regner ut u av alle hjørnene i rutenettet
>> v=x+2.*y; %regner ut v av alle hjørnene i rutenettet
>> plot(u,v,u',v') %tegner opp bildet av alle hjørnene i
%rutenettet og forbinder dem med rette streker
%Den første delen av kommandoene (dvs. plot(u,v))
%tegner opp strekene mellom "loddrette naboer", mens den
%andre delen (dvs. plot(u',v')) tegner opp strekene mellom
%"vannrette naboer".
```

- a) Kjør programmet ovenfor med de angitte funksjonene  $u = 3x - y$  og  $v = x + 2y$ . Beskriv rutenettet du ser (avbildningen  $\mathbf{F}(x, y) = \begin{pmatrix} 3x - y \\ x + 2y \end{pmatrix}$  er lineær).
- b) Kjør programmet på nytt, men la  $u = x \cos y$ ,  $v = x \sin y$ , og bruk en oppdeling slik at  $0 \leq x \leq 5$ ,  $0 \leq y \leq 2\pi$ . Beskriv rutenettet, og forklar sammenhengen med polarkoordinater.
- c) Kjør programmet igjen med  $u = \sqrt{\frac{x}{y}}$  og  $v = \sqrt{xy}$ . Velg en oppdeling slik at  $x \geq 0$  og  $y \geq 0$ . Beskriv rutenettet.

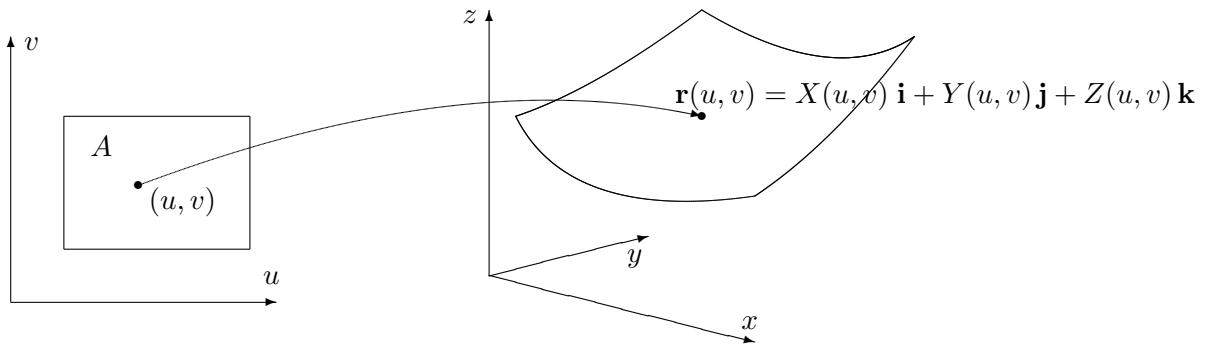
### 3.9 Parametriserte flater

De flatene vi hittil har studert, har enten vært gitt ved funksjonsgrafer  $z = f(x, y)$  eller ved ligninger  $g(x, y, z) = 0$ . Vi skal nå se på en mer generell måte å fremstille flater på. Den minner om parameterfremstilling av kurver og kalles da også parameterfremstilling av flater. En slik fremstilling er lett

og slett gitt ved et område  $A$  i  $\mathbb{R}^2$  og en kontinuerlig funksjon  $\mathbf{r} : A \rightarrow \mathbb{R}^3$ . Det er ofte praktisk å skrive funksjonen på denne måten

$$\mathbf{r}(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j} + Z(u, v)\mathbf{k}, \quad \text{der } (u, v) \in A$$

Figur 1 viser hvordan en slik funksjon avbilder området  $A$  opp på en flate. Vi kaller  $\mathbf{r}$  en *parametrisering* av flaten.

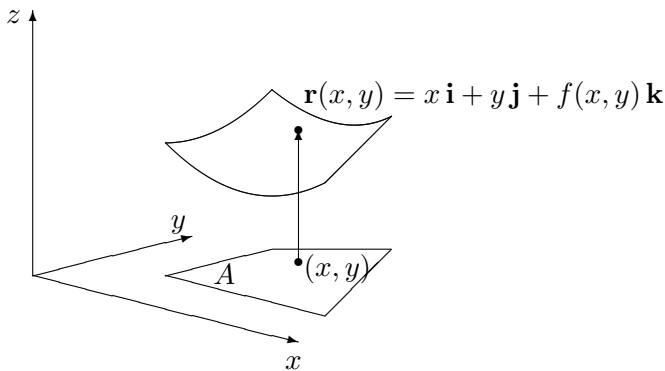


Figur 1: Parametrisert flate

Akkurat som samme kurve kan parametriseres på mange måter, kan også en flate parametriseres på forskjellig vis. Vær også oppmerksom på at ikke alle funksjoner  $\mathbf{r}$  av typen ovenfor gir opphav til en flate, det kan f.eks. hende at vi bare får et punkt eller en kurve. Dette avhenger både av området  $A$  og av funksjonene  $X, Y, Z$ , men er ikke noe du behøver å bekymre deg om for øyeblikket.

Det enkleste tilfellet er der området  $A$  ligger i  $(x, y)$ -planet og løftes opp på flaten av en funksjon  $z = f(x, y)$  (se figur 2). I så fall er parametriseringen  $\mathbf{r}$  gitt ved

$$\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + f(x, y)\mathbf{k}$$



Figur 2: Parametrisering av typen  $\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + f(x, y)\mathbf{k}$

Parametriseringer av denne typen er enkle å forstå, men de er ikke alltid de beste å arbeide med. Den øvre delen av en kuleflate vil f.eks. få parametriseringen

$$\mathbf{r}(x, y) = x \mathbf{i} + y \mathbf{j} + \sqrt{R^2 - x^2 - y^2} \mathbf{k}$$

der  $x^2 + y^2 \leq R^2$ . Dette uttrykket er tungt å arbeide med pga. kvadratroten, og det har også den ulempen at det bare beskriver halve kuleflatene — ønsker vi også å få med den nedre halvkulen, må vi bruke en parametrisering til:

$$\mathbf{s}(x, y) = x \mathbf{i} + y \mathbf{j} - \sqrt{R^2 - x^2 - y^2} \mathbf{k}$$

For de fleste formål er det bedre å beskrive kuleflatene ved hjelp av en annen parametrisering. Tar vi utgangspunkt i kulekoordinatene  $\rho, \phi, \theta$ , ser vi at alle punktene på kulen fremkommer om vi setter  $\rho = R$  og lar  $\phi$  og  $\theta$  løpe over intervallene  $[0, \pi]$  og  $[0, 2\pi]$ . For polarkoordinater har vi den generelle sammenhengen  $x = \rho \sin \phi \cos \theta, y = \rho \sin \phi \sin \theta, z = \rho \cos \phi$  (se seksjon 3.9). Det betyr at punktene på kuleflatene kan fremstilles ved

$$\mathbf{r}(\phi, \theta) = R \sin \phi \cos \theta \mathbf{i} + R \sin \phi \sin \theta \mathbf{j} + R \cos \phi \mathbf{k},$$

der  $0 \leq \phi \leq \pi, 0 \leq \theta < 2\pi$ . Dette gir oss en parametrisering av hele kuleflatene på en gang, og det uten at vi må ty til ubehagelige kvadratrøtter. Legg forøvrig merke til at “nordpolen” og “sydpolen” på kulen blir representert flere ganger; ethvert punkt av typen  $(0, \theta)$  havner på nordpolen (i den forstand at  $\mathbf{r}(0, \theta) = (0, 0, R)$ ), og hvert punkt av typen  $(\pi, \theta)$  havner på tilsvarende vis på sydpolen. Slike “multiple representasjoner” er umulig å unngå, men de skaper innimellom litt problemer for teorien.

La oss se på en enkel parametrisering til. Vi ønsker å parametrisere kjegleflatene  $z = \sqrt{x^2 + y^2}$  (lag en figur). Det kan vi gjøre lett og slett ved

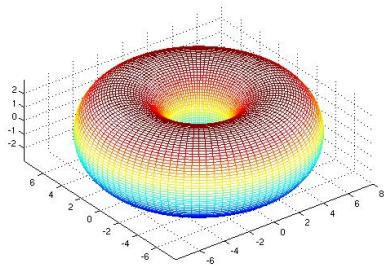
$$\mathbf{r}(x, y) = x \mathbf{i} + y \mathbf{j} + \sqrt{x^2 + y^2} \mathbf{k}$$

men dette gir oss en kvadratrot som er ubehagelig i noen sammenhenger. Det kan derfor være lurer å bruke cylinderkoordinater siden vi da får  $z = \sqrt{x^2 + y^2} = r$ . Dette gir parametriseringen

$$\mathbf{r}(r, \theta) = r \cos \theta \mathbf{i} + r \sin \theta \mathbf{j} + r \mathbf{k}$$

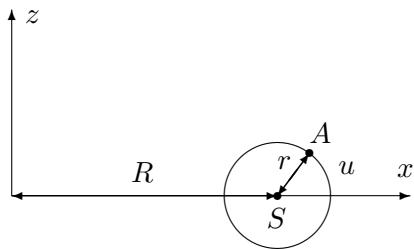
der  $0 \leq \theta < 2\pi, r \geq 0$ .

La oss forsøke oss på en litt vanskeligere parametrisering. Vi skal parametrisere en *torus* (dvs. overflaten til en smultring). Figur 3 viser en torus fremstilt ved hjelp av MATLAB.

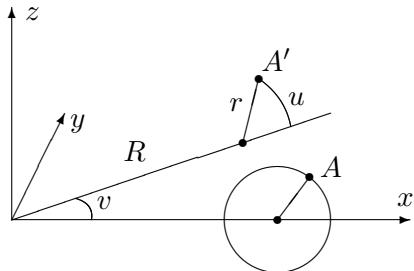


Figur 3: En torus

Vi tenker oss at flaten fremkommer når vi roterer sirkelen på figur 4 én gang rundt  $z$ -aksen. Legg merke til at sirkelen vi dreier rundt har radius  $r$  og at sentrum  $S$  i denne sirkelen ligger en avstand  $R$  fra origo. På sirkelen har vi merket av et punkt  $A$  slik at vinkelen mellom  $\overrightarrow{SA}$  og  $x$ -aksen er  $u$ .

Figur 4: Torus som omdreiningslegeme om  $z$ -aksen

Et typisk punkt  $A'$  på flaten fremkommer altså når vi dreier et punkt som  $A$  på figuren en vinkel  $v$  om  $z$ -aksen.

Figur 5: Punktet  $A$  roteres til  $A'$ 

Punktet  $A'$  på flaten kan dermed beskrives av to vinkler; vinkelen  $u$  som forteller oss hvor på den opprinnelige sirkelen  $A$  ligger, og vinkelen  $v$

som forteller oss hvor mye vi dreier (se figur 5). Litt trigonometri viser at koordinatene til det dreide punktet  $A'$  er

$$x = (R + r \cos u) \cos v, \quad y = (R + r \cos u) \sin v, \quad z = r \sin u$$

Dermed får vi parametriseringen

$$\mathbf{r}(u, v) = (R + r \cos u) \cos v \mathbf{i} + (R + r \cos u) \sin v \mathbf{j} + r \sin u \mathbf{k},$$

der  $u, v \in [0, 2\pi)$ .

Tidligere i dette kapitlet har vi sett hvordan vi kan integrere funksjoner langs en parametrisert kurve, og det er naturlig å spørre om vi kan integrere funksjoner over en parametrisert flate på tilsvarende måte. Det kan vi, men før vi ser på denne teorien, må vi lære litt om dobbeltintegraler. Det er temaet for kapittel 6.

### Parametriserte flater i MATLAB

Det er lett å bruke MATLAB til å tegne parametriserte flater. Fremgangsmåten er en variant av den vi bruker når vi tegner grafen til en funksjon av to variable, og burde fremgå av følgende eksempel.

**Eksempel 1:** Vi skal bruke MATLAB til å tegne den parametriserte flaten

$$\mathbf{r}(u, v) = (u + v) \mathbf{i} + (u - v) \mathbf{j} + uv \mathbf{k}, \text{ der } -5 \leq u, v \leq 5$$

Vi taster da:

```
>> u=-5:.1:5;
>> v=-5:.1:5;
>> [U,V]=meshgrid(u,v);
>> x=U+V;
>> y=U-V;
>> z=U.*V;
>> surf(x,y,z)
```

og MATLAB tegner flaten.



I det neste eksemplet lar vi MATLAB tegne en kuleflate med radius 3 og sentrum i origo.

**Eksempel 2:** For å slippe for lange variabelnavn, bruker vi  $u$  for den variablene som i kulekoordinater vanligvis heter  $\phi$  og  $v$  for den som vanligvis heter  $\theta$ . Kommandoen `linspace(a,b,n)` gir oss en vektor med  $n$  komponenter som begynner med  $a$  og ender med  $b$ .

```
>> u=linspace(0,pi,100);
>> v=linspace(0,2*pi,200);
>> [U,V]=meshgrid(u,v);
>> x=3*sin(U).*cos(V);
>> y=3*sin(U).*sin(V);
>> z=3*cos(U);
>> surf(x,y,z)
>> axis('equal')
```

Husk kommandoen `>> axis('equal')` som gir samme målestokk langs alle akser — uten den ser kulen ut som en ellipsoide.



### Oppgaver til seksjon 3.9

- 1.** Finn to parametriseringer av paraboloiden  $z = x^2 + y^2$ , én ved hjelp av vanlige koordinater  $(x, y)$  og én ved hjelp av polarkoordinater  $(r, \theta)$ .
- 2.** Finn en parametrisering av den delen av kuleflaten  $x^2 + y^2 + z^2 = 4$  som ligger i første oktant (dvs. i området der  $x \geq 0, y \geq 0, z \geq 0$ ).
- 3.** Finn en parametrisering av den delen av cylinderflaten  $x^2 + y^2 = 1$  som ligger mellom  $z = 0$  og  $z = 1$ .
- 4.** Finn en parametrisering av sylinderen  $y^2 + z^2 = 9$ .
- 5.** Finn en parametrisering av kjegleflaten  $x = \sqrt{y^2 + z^2}$ .
- 6.** Finn en parametrisering av ellipsoiden

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

- 7.** En flate er parametrisert ved

$$\mathbf{r}(u, v) = u \mathbf{i} + 2 \cos v \mathbf{j} + 2 \sin v \mathbf{k}, \quad \text{der } 0 \leq u \leq 2, 0 \leq v < 2\pi$$

Beskriv flaten med ord.

- 8.** Finn en parametrisering av den delen av kuleflaten  $x^2 + y^2 + z^2 = 4$  som ligger over  $xy$ -planet og inni kjeglen  $z^2 = 3(x^2 + y^2)$ .
- 9.** Bruk MATLAB til å lage en tegning av den delen av kulen  $x^2 + y^2 + z^2 = 4$  som ligger i første oktant (dvs. området der  $x, y, z \geq 0$ ).
- 10.** Bruk MATLAB til å lage en tegning av flaten i oppgave 7.
- 11.** Bruk MATLAB til å lage en tegning av flaten parametrisert ved

$$\mathbf{r}(u, v) = uv^2 \mathbf{i} + u \mathbf{j} + \sin(uv) \mathbf{k}, \quad \text{der } -1 \leq u \leq 1, 0 \leq v \leq 3$$

**12.** Bruk MATLAB til å lage en tegning av sylinderen  $x^2 + y^2 = 9$ , når  $0 \leq z \leq 2$ .

**12.** Bruk MATLAB til å lage en tegning av ellipsoiden i oppgave 6. Bruk samme målestokk på alle akser.

**13.** Bruk MATLAB til å lage en tegning av en torus der  $r = 3$  og  $R = 5$ .

## Kapittel 4

# Lineær algebra i $\mathbb{R}^n$

I dette kapitlet skal vi se nærmere på en del spørsmål om vektorer og matriser som vi så vidt har vært borti tidligere, men som vi hittil ikke har gitt noen skikkelig svar på — spørsmål som: når er en matrise inverterbar og hvordan regner vi ut den inverse matrisen, hva er determinanter og hvordan regner vi dem ut på en effektiv måte, hvordan finner vi egenverdier og egenvektorer og hva kan de brukes til? Disse spørsmålene er viktige dels fordi mange praktiske problemer kan formuleres ved hjelp av vektorer og matriser, og dels fordi mange problemer i andre deler av matematikken kan løses ved hjelp av matrisemetoder. Vi skal se eksempler på begge fenomenene i dette og de neste kapitlene.

Den videregåede teorien for vektorer og matriser (og deres generaliseringer) kalles *lineær algebra* og blir etter hvert ganske abstrakt, men den har et ytterst konkret utgangspunkt, nemlig de lineære ligningssystemene du lærte å løse på ungdomsskolen, slike som

$$2x - y = 3$$

$$x + 3y = 4$$

Senere har du truffet slike ligningssystemer i mange sammenhenger, ofte med flere ligninger og flere ukjente. Et typisk eksempel er delbrøkoppspalting der vi bruker lineære ligningssystemer til å finne konstantene i oppspaltingen.

Det viser seg at nøkkelen til alle de spørsmålene vi startet med, ligger i et nærmere studium av slike lineære ligningssystemer (dvs. ligningssystemer der de ukjente bare opptrer i første potens og ikke inni mer kompliserte funksjoner). Kanskje kan et slikt studium virke unødvendig siden vi vet hvordan vi skal løse disse ligningssystemene i praksis, men det viser seg at det er mye å tjene på å gå systematisk til verks og studere litt mer teoretiske spørsmål som: når har et lineært ligningssystem løsninger, og hvor mange løsninger har det i så fall? Etter hvert som kapitlet skrider frem, vil du se hvordan kunnskap om lineære ligningssystemer kaster lys over spørsmålene vi startet med. Du vil også se hvordan systematiske løsningsmetoder for lineære

ligninger kan brukes til å utvikle raske metoder for å invertere matriser og regne ut determinater.

Som allerede antydet er lineær algebra en av de delene av matematikken som har størst umiddelbar nytte — det viser seg at svært mange sentrale problemstillinger i matematikk, fysikk, informatikk, statistikk, økonomi og andre fag kan formuleres (i hvert fall tilnærmet) ved hjelp av lineære ligningssystemer og løses ved hjelp av lineær algebra. Disse ligningssystemene består ofte av mange tusen ligninger og ukjente, og de kan bare løses ved hjelp av datamaskiner. Når man skal programmere en datamaskin til å løse slike problemer, kan man ikke stole på snarveier og smarte triks — man må ha systematiske metoder som alltid fungerer. Vi skal begynne med å se på en metode som kalles *Gauss-eliminasjon* eller *Gauss-Jordan-eliminasjon*.

## 4.1 Noen eksempler på Gauss-eliminasjon

La oss introdusere metoden gjennom et enkelt eksempel. Vi skal løse ligningssystemet

$$\begin{aligned} 2y + z &= -1 \\ 3x + 5y + z &= 2 \\ x + 2y + z &= 1 \end{aligned} \tag{4.1.1}$$

ved å omforme det til stadig enklere ligningssystemer som har nøyaktig de samme løsningene (som vi skal se senere, kan et lineært ligningssystem ha én, ingen eller uendelig mange løsninger). Til slutt sitter vi igjen med et ligningssystem som er så enkelt at vi kan løse det med hoderegning. Ideen bak metoden er å *eliminere* (fjerne) ukjente fra så mange ligninger som mulig — først eliminerer vi den første variablen  $x$  fra alle ligninger unntatt den øverste, deretter eliminerer vi den andre variablen  $y$  fra alle ligninger unntatt de to øverste osv. Det ligningssystemet vi sitter igjen med til slutt, kan løses nedenfra — først finner vi verdien til den siste variablen fra den nederste ligningen, deretter bruker vi denne verdien til å finne verdien til variablen foran osv. Dette kan høres komplisert ut, men når vi bruker metoden på ligningssystemet ovenfor, vil du fort se hvordan den virker.

Siden vi skal eliminere  $x$  fra alle ligningene unntatt den øverste, passer det dårlig at den øverste ligningen mangler  $x$ -ledd. Vi begynner derfor med å bytte om den første og siste ligningen:

$$\begin{aligned} x + 2y + z &= 1 \\ 3x + 5y + z &= 2 \\ 2y + z &= -1 \end{aligned} \tag{4.1.2}$$

Det nye ligningssystemet har selvfølgelig akkurat de samme løsningene som det gamle; rekkefølgen av ligningene spiller ingen rolle. Neste skritt i metoden er å bruke  $x$ -leddet i den øverste ligningen til å eliminere  $x$ -leddet i de andre. Ganger vi den første ligningen med  $-3$ , får vi  $-3x - 6y - 3z = -3$ , og legger vi denne ligningen til ligning nummer to i (4.1.2), får vi dette ligningssystemet:

$$\begin{aligned} x + 2y + z &= 1 \\ -y - 2z &= -1 \\ 2y + z &= -1 \end{aligned} \tag{4.1.3}$$

Dette ligningssystemet må ha nøyaktig de samme løsningene som det foregående (vær sikker på at du skjønner hvorfor!). Normalt ville vi nå fortsette med å eliminere  $x$ -en i den nederste linjen, men siden den allerede er borte, kan vi konsentrere oss om  $y$ -ene isteden. Vi skal bruke  $y$ -leddet i den nest øverste ligningen til å eliminere  $y$ -leddet i den nederste ligningen. Multipliserer vi den nest øverste ligningen med  $2$ , får vi  $-2y - 4z = -2$ , og legger vi dette til den nederste ligningen, får vi

$$\begin{aligned} x + 2y + z &= 1 \\ -y - 2z &= -1 \\ -3z &= -3 \end{aligned} \tag{4.1.4}$$

Igjen ser vi at dette ligningssystemet har akkurat de samme løsningene som det foregående. Nå er vi nesten ferdige, men vi kan forenkle ligningene ytterligere ved å gange den midterste med  $-1$  og den nederste med  $-\frac{1}{3}$ :

$$\begin{aligned} x + 2y + z &= 1 \\ y + 2z &= 1 \\ z &= 1 \end{aligned} \tag{4.1.5}$$

Dette ligningssystemet er enkelt å løse; fra den nederste ligningen ser vi at  $z = 1$ , setter vi dette inn i den midterste, får vi at  $y = -1$ , og setter vi disse verdiene for  $y$  og  $z$  inn i den øverste ligningen, ser vi at  $x = 2$ . Siden alle ligningssystemene våre har de samme løsningene, betyr dette at  $x = 2$ ,  $y = -1$ ,  $z = 1$  også er (den eneste) løsningen til det opprinnelige ligningssystemet (4.1.1).

**Bemerkning:** Det finnes raskere måter å løse ligningssystemet (4.1.1) på. Sammenligner vi den første og den siste ligningen, ser vi med en gang at  $x = 2$ , og deretter er det ikke vanskelig å finne  $y$  og  $z$ . Slike snarveier er viktige når vi løser ligningssystemer for hånd, men det er ikke det som er poenget for oss nå — vi er på jakt etter systematiske metoder som fungerer

for alle lineære ligningssystemer.

Av åpenbare grunner sier vi at ligningssystemet (4.1.5) er på *trappeform*, og målet for metoden vi er iferd med å beskrive, er å føre et hvilket som helst lineært ligningssystem over på trappeform. Legg merke til at vi i prosessen ovenfor har brukt tre operasjonstyper:

- (i) Bytte om to ligninger i ligningssystemet
- (ii) Gange en ligning i systemet med et tall forskjellig fra 0
- (iii) Velge én av ligningene i systemet og erstatte den med det vi får når vi legger til et multiplum av en av de andre ligningene

Det viser seg at alle lineære ligningssystemer kan føres over på trappeform ved å bruke disse tre operasjonene.

Når vi bruker operasjonene ovenfor, endrer vi ikke løsningene til ligningssystemet — vi mister ikke løsninger og pådrar oss heller ikke nye, ”falske” løsninger. Det betyr at løsningene til det enkle trappeformede systemet vi ender opp med, er de samme som løsningene til det mer kompliserte systemet vi startet med. Det er lettere å forstå hvorfor dette er viktig dersom vi ser på et ligningssystem med uendelig mange løsninger.

Vi skal se på ligningssystemet

$$\begin{aligned} x + 2y + z - u &= 3 \\ -x - y - 4z + 2u &= -1 \\ 2x + 5y - z &= 9 \\ x + 7z - 5u &= -3 \end{aligned}$$

Vi bruker først  $x$ -leddet i den øverste ligningen til å kvitte oss med  $x$ -leddene i de andre ligningene. Legger vi den øverste linjen til den andre, får vi

$$\begin{aligned} x + 2y + z - u &= 3 \\ y - 3z + u &= 2 \\ 2x + 5y - z &= 9 \\ x + 7z - 5u &= -3 \end{aligned}$$

For å bli kvitt  $x$ -leddet i ligning nummer tre, ganger vi først den øverste ligningen med -2 og får  $-2x - 4y - 2z + 2u = -6$ . Legger vi dette til den tredje ligningen, får vi

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

For å kvitte oss med  $x$ -leddet i den nederste linjen, ganger vi den øverste linjen med  $-1$  og får  $-x - 2y - z + u = -3$ . Legger vi dette til den nederste linjen, får vi

$$\begin{aligned} x + 2y + z - u &= 3 \\ y - 3z + u &= 2 \\ y - 3z + 2u &= 3 \\ -2y + 6z - 4u &= -6 \end{aligned}$$

Nå har vi kvittet oss med alle de  $x$ -leddene vi ønsket å fjerne. Neste trinn på programmet er å bruke  $y$ -leddet i ligning nummer to til å eliminere  $y$ -leddet i alle ligningene nedenfor. Ganger vi linje nummer to med  $-1$ , får vi  $-y + 3z - u = -2$ , og legger vi dette til linje nummer tre, får vi

$$\begin{aligned} x + 2y + z - u &= 3 \\ y - 3z + u &= 2 \\ u &= 1 \\ -2y + 6z - 4u &= -6 \end{aligned}$$

For å bli kvitt  $y$ -leddet i den nederste ligningen, ganger vi først ligning nummer to med  $2$  og får  $2y - 6z + 2u = 4$ . Legger vi dette til den nederste linjen, får vi

$$\begin{aligned} x + 2y + z - u &= 3 \\ y - 3z + u &= 2 \\ u &= 1 \\ -2u &= -2 \end{aligned}$$

Ifølge systemet vårt skulle vi nå ha brukt  $z$ -leddet i den tredje ligningen til å kvitte oss med  $z$ -leddet i den fjerde, men det er ikke noe  $z$ -ledd i tredje ligning, og vi kan heller ikke skaffe oss noe ved å bytte om på ligning 3 og 4. Vi går derfor videre til den neste variabelen  $u$ , og bruker  $u$ -leddet i den tredje ligningen til å eliminere  $u$ -leddet i fjerde. Ganger vi ligning 3 med  $2$ , får vi  $2u = 2$ , og legger vi dette til ligning 4, får vi

$$\begin{aligned} x + 2y + z - u &= 3 \\ y - 3z + u &= 2 \\ u &= 1 \\ 0 &= 0 \end{aligned}$$

Vi har nå fått ligningssystemet over på trappeform, men trappen er litt mindre regelmessig enn i stad siden ikke alle trappetrinnene er like lange. Dette er helt greit og ikke til å unngå i mange tilfeller.

La oss nå prøve å løse ligningssystemet ovenfor. Den siste ligningen er alltid oppfylt og kan bare glemmes. Den nest nederste ligningen forteller oss at  $u = 1$ . Ligningen over gir oss ikke noe krav på  $z$ , men forteller oss at vi kan regne ut  $y$  når vi kjenner  $z$ . Dette betyr at vi kan velge  $z$  helt fritt, men når valget er gjort, må vi la  $y = 2 + 3z - u = 1 + 3z$  (husk at  $u = 1$ ). Den øverste ligningen lar oss på tilsvarende måte regne ut  $x$  når  $z$  er valgt — vi får  $x = 3 - 2y - z + u = 3 - 2(1 + 3z) - z + 1 = 2 - 7z$ . Dette betyr at ligningen har uendelig mange løsninger, nemlig

$$\begin{aligned}x &= 2 - 7z \\y &= 1 + 3z \\z &= z \\u &= 1\end{aligned}$$

der  $z$  er et fritt valgt tall. Det er instruktivt å sette disse uttrykkene inn i det opprinnelige ligningssystemet og se at de passer.

Når man arbeider med ligningssystemer med uendelig mange løsninger, er det lett å forstå fordelene ved bare å bruke operasjoner som ikke endrer løsningsmengden til ligningssystemet — i systemet ovenfor hadde det ikke vært enkelt å holde styr på hvilke falske løsninger vi hadde pådratt oss og hvilke ekte vi hadde mistet.

Vi tar med enda et eksempel som viser hva som skjer når ligningssystemet *ikke* har løsninger. Vi starter med ligningssystemet

$$\begin{aligned}x - y + z - u &= 1 \\2x - 2y - z + u &= 0 \\-x + 2y + u &= 2 \\2x - y + u &= 2\end{aligned}$$

og bruker på vanlig måte  $x$ -leddet i den første ligningen til å eliminere  $x$ -leddene i de andre ligningene. Ganger vi den øverste ligningen med -2 og legger resultatet til den nest øverste, får vi

$$\begin{aligned}x - y + z - u &= 1 \\-3z + 3u &= -2 \\-x + 2y + u &= 2 \\2x - y + u &= 2\end{aligned}$$

Legger vi den øverste linjen til den tredje, får vi

$$\begin{aligned}x - y + z - u &= 1 \\-3z + 3u &= -2 \\y + z &= 3 \\2x - y + u &= 2\end{aligned}$$

Til slutt ganger vi den øverste linjen med -2 og legger resultatet til den nederste:

$$\begin{aligned} x - y + z - u &= 1 \\ -3z + 3u &= -2 \\ y + z &= 3 \\ y - 2z + 3u &= 0 \end{aligned}$$

Neste post på programmet er vanligvis at vi bruker  $y$ -leddet i den andre ligningen til å kvitte oss med  $y$ -leddene nedenfor, men i dette tilfellet er det ikke noe  $y$ -ledd i den andre ligningen. Vi bytter derfor om på ligning 2 og 3:

$$\begin{aligned} x - y + z - u &= 1 \\ y + z &= 3 \\ -3z + 3u &= -2 \\ y - 2z + 3u &= 0 \end{aligned}$$

Ganger vi den andre linjen med -1 og legger resultatet til den nederste linjen, får vi

$$\begin{aligned} x - y + z - u &= 1 \\ y + z &= 3 \\ -3z + 3u &= -2 \\ -3z + 3u &= -3 \end{aligned}$$

Det gjenstår å bruke  $z$ -leddet i tredje linje til å eliminere  $z$ -leddet i den nederste linjen. Ganger vi den tredje linjen med -1 og legger resultatet til den nederste, får vi

$$\begin{aligned} x - y + z - u &= 1 \\ y + z &= 3 \\ -3z + 3u &= -2 \\ 0 &= -1 \end{aligned}$$

Dermed har vi fått ligningssystemet på trappeform. Siden den nederste linjen er umulig å oppfylle (ingen valg av  $x, y, z, u$  kan få 0 til å bli lik -1), har systemet ingen løsning. Det betyr at det opprinnelige systemet heller ikke har noen løsning.

Det viser seg at vi nå har sett alle de mulighetene som finnes — et lineært ligningssystem kan ha enten én, uendelig mange eller ingen løsninger. Dette kan virke underlig, men det er ikke vanskelig å få et geometrisk innblikk

i det som foregår. La oss se på et ligningssystem med tre ligninger og tre ukjente:

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_1 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ a_{31}x + a_{32}y + a_{33}z &= b_3 \end{aligned}$$

Hver av disse ligningene beskriver et plan — la oss kalle dem *I*, *II* og *III*. En løsning til ligningssystemet vil være (koordinatene til) et punkt som ligger i alle tre planene. “Normalt” vil plan *I* og *II* skjære hverandre langs en rett linje *m* (unntakene er hvis *I* og *II* er parallelle eller sammenfallende). Denne linjen *m* vil normalt skjære det tredje planet i ett eneste punkt, og koordinatene til dette punktet gir oss da den eneste løsningen til ligningssystemet. Det finnes imidlertid unntak; dersom skjæringslinjen *m* ligger i planet *III*, vil alle punktene på denne linjen gi oss en løsning — i dette tilfellet har vi altså uendelig mange løsninger. En tredje mulighet er at linjen *m* ikke skjærer planet *III* i det hele tatt — i så fall har ligningssystemet ingen løsning.

I den neste seksjonen skal vi se grundigere på metoden ovenfor, men før vi begynner, kan det være lurt å gjøre en enkel observasjon. I alle regningene våre har variablene  $x, y, z, u$  spilt en underordnet rolle — de har bare vært med som en slags “plassholdere” som har fortalt oss hvilke koeffisienter som hører sammen. Vi kan spare skrivearbeid og tjene oversikt ved å fjerne variablene og isteden organisere koeffisientene i en matrise. Vårt første ligningssystem

$$\begin{aligned} 2y + z &= -1 \\ 3x + 5y + z &= 2 \\ x + 2y + z &= 1 \end{aligned}$$

kan f.eks. ertattes med matrisen

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

Legg merke til hvordan *B* er laget: Elementene i den første søylen er koeffisienten til  $x$  i ligningssystemet, elementene i den andre søylen er koeffisientene til  $y$ , elementene i den tredje søylen er koeffisientene til  $z$ , mens den fjerde søylen består av konstantene på høyresiden av ligningssystemet. Vi kaller *B* den *utvidede matrisen* til ligningssystemet (4.1.1) (*utvidet* fordi vi ikke bare har med koeffisientene, men også konstantene på høyresiden). Vi kan nå utføre akkurat de samme operasjonene på denne matrisen som vi i stad

utførte på ligningssystemet. Vi får da denne sekvensen

$$\begin{aligned} \left( \begin{array}{cccc} 0 & 2 & 1 & -1 \\ 3 & 5 & 1 & 2 \\ 1 & 2 & 1 & 1 \end{array} \right) &\Rightarrow \left( \begin{array}{cccc} 1 & 2 & 1 & 1 \\ 3 & 5 & 1 & 2 \\ 0 & 2 & 1 & -1 \end{array} \right) \Rightarrow \left( \begin{array}{cccc} 1 & 2 & 1 & 1 \\ 0 & -1 & -2 & -1 \\ 0 & 2 & 1 & -1 \end{array} \right) \Rightarrow \\ &\Rightarrow \left( \begin{array}{cccc} 1 & 2 & 1 & 1 \\ 0 & -1 & -2 & -1 \\ 0 & 0 & -3 & -3 \end{array} \right) \Rightarrow \left( \begin{array}{cccc} 1 & 2 & 1 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{array} \right) \end{aligned}$$

Når vi nå har fått matrisen på trappeform, kan vi gå tilbake til ligningssystemet ved å sette inn variablene igjen:

$$\begin{aligned} x + 2y + z &= 1 \\ y + 2z &= 1 \\ z &= 1 \end{aligned}$$

Siden matrisenotasjonen er mer oversiktlig enn ligningsnotasjonen, skal vi stort sett holde oss til den i fortsettelsen.

**Bemerkning:** Som vi allerede har nevnt, må man i praksis ofte løse store ligningssystemer med tusenvis av ligninger og ukjente. Dette må selvfølgelig gjøres med datamaskin, men selv for en datamaskin kan regneoppgavene bli uoverkommelige dersom vi ikke bruker effektive metoder. Et godt mål på effektiviteten til en metode er å telle hvor mange regneoperasjoner den krever. Når vi løser et ligningssystem ved hjelp av Gauss-eliminasjon, er det to trinn som krever utregninger — å gange en ligning/rad med et tall, og å subtrahere et multiplum av én ligning/rad fra en annen. Begge disse trinnene krever flere utregninger siden vi må behandle alle ikke-null koeffisienter i ligningen/raden. Teller vi opp alle de regneoperasjonene vi trenger for å løse et ligningssystem med  $n$  ligninger og  $n$  ukjente ved hjelp av Gauss-eliminasjon, får vi omtrent  $\frac{n^3}{3}$ . Er  $n = 1000$ , trenger vi altså omtrent  $3.3 \cdot 10^8$  operasjoner! Det viser seg at det er mulig å senke dette tallet noe ved å benytte andre metoder, men disse metodene er ofte så tungvinne å bruke at de bare anvendes når man er helt på grensen av det maskinene kan greie.

I anvendt lineær algebra er effektivitetsberegninger av denne typen svært viktige, men vi har så mye annet å konsentrere oss om i dette kompendiet at vi skal nøyne oss med noen få råd her og der om hvordan man bør gå frem — eller helst: hvordan man *ikke* bør frem!

## Oppgaver til seksjon 4.1

1. Finn alle løsningene til ligningssystemet

$$\begin{aligned} x + 2y - z &= 3 \\ 2x + 3y - 3z &= -1 \\ -x + 2y + 3z &= 1 \end{aligned}$$

**2.** Finn alle løsningene til ligningssystemet

$$\begin{aligned}x - y + 2z &= 3 \\2x - 2y &= 4 \\-3x + 2y + z &= 0\end{aligned}$$

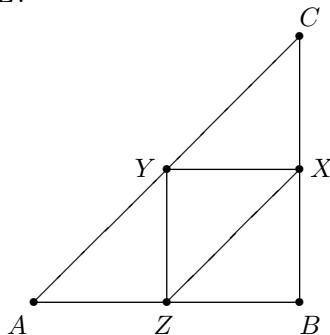
**3.** Finn alle løsningene til ligningssystemet

$$\begin{aligned}2x - 4y + 6z &= -2 \\-3x + 2y - z &= 8 \\x - 6y + 11z &= 4\end{aligned}$$

**4.** Finn alle løsningene til ligningssystemet

$$\begin{aligned}x - 2y + 3z &= 1 \\-x + y - 2z &= 0 \\-3x + 5y - 8z &= 2\end{aligned}$$

**5.** Figuren viser et nettverk av ledninger. Temperaturen i hvert av punktene  $X$ ,  $Y$  og  $Z$  er lik gjennomsnittet av temperaturen i nabopunktene (dvs. punktene som de er forbundet til ved hjelp av en ledning). Anta at temperaturen i hjørnepunktene  $A$ ,  $B$  og  $C$  er henholdsvis  $a$ ,  $b$  og  $c$ . Finn temperaturene  $x$ ,  $y$  og  $z$  i punktene  $X$ ,  $Y$  og  $Z$ .



**6.** I et game i tennis må du vinne med minst to poeng. Når man kommer til stillingen 40-40, vil spilleren som vinner den neste ballvekslingen få "fordel". Vinner hun også den neste ballvekslingen, vinner hun gamet, hvis ikke går stillingen tilbake til "like". Spilleren som vinner den neste ballvekslingen vil så få fordel, osv. I denne delen av spillet er det altså tre mulige stillinger: "like", "fordel spiller A" og "fordel spiller B". I tennis lønner det seg å serve, og vi antar at i det gamet vi ser på, er det spiller  $A$  som har serven og dermed har 60% sjanse for å vinne en ballveksling.

Vi lar  $x$  være spiller  $A$ s sannsynlighet for å vinne gamet dersom stillingen er like,  $y$  hennes sannsynlighet for å vinne gamet dersom hun har fordel og  $z$  hennes sannsynlighet for å vinne gamet dersom motstanderen har fordel.

a) Forklar at

$$\begin{aligned}x &= 0.6y + 0.4z \\y &= 0.4x + 0.6 \\z &= 0.6x\end{aligned}$$

b) Finn  $x$ ,  $y$  og  $z$ .

## 4.2 Trappeform

Vi skal nå ta en nærmere kikk på prosedyren vi introduserte i forrige seksjon. For oversiktens skyld skal vi stort sett arbeide med matriser og ikke ligningssystemer, så vi antar at vi starter med en matrise

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

Med en *radoperasjon* på  $A$  mener vi én av følgende operasjoner:

- (i) Bytte om to rader i  $A$
- (ii) Gange en av radene i  $A$  med et tall forskjellig fra 0
- (iii) Velge én av radene i  $A$  og erstatte den med det vi får når vi legger til et multiplum av en av de *andre* radene.

Legg merke til at dette er nøyaktig de samme operasjonene som vi brukte på ligningssystemene i forrige seksjon.

**Definisjon 4.2.1** Vi sier at to  $m \times n$ -matriser  $A, B$  er radekvivalente der som det finnes en sekvens av radoperasjoner som forvandler  $A$  til  $B$ . Vi skriver  $A \sim B$  når  $A$  og  $B$  er radekvivalente.

**Eksempel 1:** Matrisene

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

og

$$B = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

er radekvivalente siden  $A$  kan forvandles til  $B$  gjennom denne sekvensen av radoperasjoner (radene i matrisen kalles  $I$ ,  $II$  og  $III$ , og symbolene over  $\sim$ -tegnene antyder hvilke operasjoner vi bruker):

$$\begin{aligned} A &= \begin{pmatrix} 2 & 4 & -2 \\ 1 & 2 & 0 \\ -1 & 3 & 1 \end{pmatrix} \xrightarrow{\frac{1}{2}I} \begin{pmatrix} 1 & 2 & -1 \\ 1 & 2 & 0 \\ -1 & 3 & 1 \end{pmatrix} \xrightarrow{II + (-1)I} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ -1 & 3 & 1 \end{pmatrix} \xrightarrow{III + I} \\ &\quad \xrightarrow{III \sim I} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ 0 & 5 & 0 \end{pmatrix} \xrightarrow{II \leftrightarrow III} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{\frac{1}{5}II} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = B \quad \clubsuit \end{aligned}$$

Det er lett å se at dersom vi kan forvandle  $A$  til  $B$  ved hjelp av radoperasjoner, så kan vi også bruke radoperasjoner til å forvandle  $B$  til  $A$ . Alt vi behøver å gjøre, er å “reversere” de operasjonene som forvandlet  $A$  til  $B$  — der vi tidligere ganget en rad med 3, ganger vi den nå med  $\frac{1}{3}$ ; der vi tidligere la til 7 ganger en rad, legger vi nå til -7 ganger den samme raden, og der vi tidligere byttet om to rader, bytter vi dem nå tilbake igjen. I tillegg må vi passe på å begynne bakfra; vi begynner med å reversere den *siste* operasjonen som var med på å forvandle  $A$  til  $B$ . Det neste eksemplet viser hvordan dette foregår i praksis.

**Eksempel 2:** I forrige eksempel forvandlet vi matrisen

$$A = \begin{pmatrix} 2 & 4 & -2 \\ 1 & 2 & 0 \\ -1 & 3 & 1 \end{pmatrix} \quad \text{til} \quad B = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

og vi skal nå vise hvordan vi kan reversere denne prosedyren slik at  $B$  blir forvandlet til  $A$ . Siden den siste operasjonen vi brukte da vi forvandlet  $A$  til  $B$ , var å gange rad 2 med  $\frac{1}{5}$ , gjør vi nå det omvendte — vi ganger rad 2 med 5:

$$B = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{5II} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Den nest siste operasjonen vi brukte, var å bytte om rad 2 og 3, så nå bytter vi tilbake igjen:

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{III \leftrightarrow II} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ 0 & 5 & 0 \end{pmatrix}$$

Operasjonen før dette var å legge rad 1 til rad 3, så nå trekker vi isteden rad 1 fra rad 3 (for å holde oss i vår offisielle språkbruk, burde vi heller si at vi ganger rad 1 med  $-1$  og legger resultatet til rad 3):

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ 0 & 5 & 0 \end{pmatrix} \xrightarrow{III + (-1)I} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ -1 & 3 & 1 \end{pmatrix}$$

Den neste operasjonen vi må reversere, er å legge  $-1$  ganger rad 1 til rad 2. Vi legger derfor rad 1 til rad 2:

$$\begin{pmatrix} 1 & 2 & -1 \\ 0 & 0 & 1 \\ -1 & 3 & 1 \end{pmatrix} \xrightarrow{II + I} \begin{pmatrix} 1 & 2 & -1 \\ 1 & 2 & 0 \\ -1 & 3 & 1 \end{pmatrix}$$

Det gjenstår nå bare å reverse en operasjon. Den var å gange rad 1 med  $\frac{1}{2}$ , så nå ganger vi rad 1 med 2:

$$\left( \begin{array}{ccc} 1 & 2 & -1 \\ 1 & 2 & 0 \\ -1 & 3 & 1 \end{array} \right) \xrightarrow{III+(-1)I} \left( \begin{array}{ccc} 2 & 4 & -2 \\ 1 & 2 & 0 \\ -1 & 3 & 1 \end{array} \right) = A$$

Dermed har vi reversert hele prosedyren og gjort  $B$  om til  $A$ . ♣

Som antydet i forrige seksjon, er hensikten med radoperasjoner å forvandle en vilkårlig matrise til en matrise på trappeform. Aller først må vi definere hva dette betyr:

**Definisjon 4.2.2** En matrise er på trappeform dersom:

- (i) Enhver rad består enten bare av nuller, eller så er det første ikke-null elementet et ett-tall.
- (ii) Enhver rad som ikke bare består av nuller, begynner med minst én null mer enn raden over.

En matrise på trappeform blir også kalt en trappematrise.

Legg merke til at dersom en trappematrise har rader som bare består av nuller, må disse være samlet nederst i matrisen (det følger fra punkt (ii) ovenfor).

**Eksempel 3:** Følgende matriser er på trappeform:

$$\left( \begin{array}{cccc} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right), \quad \left( \begin{array}{ccccc} 0 & 1 & 2 & -1 & 2 \\ 0 & 0 & 1 & 3 & 4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right), \quad \left( \begin{array}{ccccc} 1 & 3 & 0 & 0 & 2 \\ 0 & 1 & 0 & 7 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{array} \right)$$

Disse matrisene er ikke på trappeform (forklar hvorfor):

$$\left( \begin{array}{cccc} 2 & 2 & 0 & -1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{array} \right), \quad \left( \begin{array}{ccccc} 0 & 1 & 2 & -1 & 2 \\ 0 & 1 & 1 & 3 & 4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right), \quad \left( \begin{array}{ccccc} 0 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 7 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{array} \right)$$



For å kunne beskrive matriser på trappeform trenger vi litt terminologi. Det første ikke-null elementet i en rad (det er nødvendigvis et ett-tall) kalles et *pivotelement*, og søylen det står i, kalles en *pivotsøyle* (se illustrasjonen

nedenfor). Pivotelementer og pivotsøyler kommer til å spille en sentral rolle i teorien i dette kapitlet.

The diagram shows a 3x4 matrix with arrows indicating pivot elements and pivot rows. The matrix is:

$$\begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Two arrows point to the first column, with one arrow pointing to the top-left element (1) and another pointing to the second element (0) in the same row. These are labeled "pivotelementer". Two arrows point to the second row, with one arrow pointing to the second element (0) and another pointing to the fourth element (4). These are labeled "pivotsøyler".

Vi har allerede nevnt flere ganger at poenget med radoperasjoner er å forvandle en gitt matrise til en matrise på trappeform. Det neste resultatet forteller oss at dette alltid er mulig.

**Setning 4.2.3** *Enhver matrise er radekvivalent med en matrise på trappeform.*

*Bevis:* Dette beviset er bare en systematisering av den metoden vi har brukt i eksemplene. Først ser vi på søylene i matrisen og plukker ut den første som ikke bare består av nuller. Ved eventuelt å bytte om to rader kan vi sørge for at det øverste elementet i denne søylen ikke er null, og ved å gange den øverste raden med et passende tall, kan vi gjøre om dette elementet til et ett-tall. Ved hjelp av dette ett-tallet kan vi nå eliminere alle ikke-null elementer i søylen under — vi bare ganger den øverste linjen med et passe tall og adderer resultatet til den raden vi arbeider med. Når dette arbeidet er ferdig, står ett-tallet alene igjen i søylen og er blitt vårt første pivotelement. Vi “glemmer” nå den øverste raden i matrisen og arbeider bare med resten (så når vi nå sier “restmatrisen” mener vi det som står igjen etter at vi har fjernet første rad).

Vi begynner nå prosedyren på nytt med å lete oss frem til den første søylen i “restmatrisen” som ikke bare består av nuller. Ved eventuelt å bytte om rader sikrer vi oss at det øverste elementet ikke er 0, og ved å gange raden med et passende tall, forvandler vi elementet til et ett-tall. På vanlig måte bruker vi dette ett-tallet til å eliminere elementene i søylen nedenfor. Nå har vi funnet vårt andre pivotelement, så vi “glemmer” raden det står i, og gjentar prosedyren på radene nedenfor.

Denne prosedyren fortsetter inntil én av to ting skjer. Enten er det ikke flere rader igjen (og da er vi ferdige fordi det står et pivotelement i hver eneste rad), eller så består alle søylene i den “restmatrisen” vi ser på, av bare nuller (og da er vi ferdige fordi vi har pivotelementer i alle de radene som ikke bare består av nuller).  $\square$

**Bemerkning:** Når vi bruker radoperasjoner til å bringe en matrise  $A$  over på trappeform, sier vi at vi *radreduserer*  $A$ . Det er greit å være klar over

at matriser kan radreduseres på forskjellige vis — når du bruker metoden, har du ofte et valg mellom flere muligheter (f.eks. hvilke rader du vil bytte om på), og det endelige resultatet vil ofte avhenge av hvilke valg du gjør underveis. Det kan derfor godt hende at fasitsvaret på en oppgave er forskjellig fra ditt svar, selv om ditt også er riktig. Det viser seg imidlertid at alle trappeformene til en matrise har de samme pivotsøylene (og dermed de samme pivotelementene), og dette kan du bruke som en sjekk på at svaret ditt er rimelig.

### Ligningssystemer på trappeform

Vi skal nå vende tilbake til lineære ligningssystemer og se dem i lys av det vi nettopp har lært. Vi starter med et generelt, lineært ligningssystem med  $n$  ligninger og  $m$  ukjente (vi insisterer altså ikke på at det skal være like mange ligninger som ukjente — det finnes en del problemstillinger der det er aktuelt å se på både “for få” og “for mange” ligninger):

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Vi tar nå den utvidede matrisen

$$B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

og bringer den på trappeform ved hjelp av radoperasjoner. Den matrisen  $C$  vi da får, tilhører et ligningssystem med de samme løsningene som det opprinnelige systemet. I dette systemet vil noen variable (“ukjente”) korrespondere til pivotsøyler, og disse kaller vi *basisvariable*, mens de andre kalles *frie variable*. La oss se på et eksempel:

**Eksempel 4:** Vi starter med ligningssystemet

$$\begin{aligned} x + 2y + z &= 1 \\ -x + y - z &= 0 \\ x + 5y + z &= 2 \end{aligned}$$

Den utvidede matrisen er

$$B = \begin{pmatrix} 1 & 2 & 1 & 1 \\ -1 & 1 & -1 & 0 \\ 1 & 5 & 1 & 2 \end{pmatrix}$$

Radreduserer vi  $B$ , får vi:

$$B = \begin{pmatrix} 1 & 2 & 1 & 1 \\ -1 & 1 & -1 & 0 \\ 1 & 5 & 1 & 2 \end{pmatrix} \xrightarrow{II+I} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 3 & 0 & 1 \\ 1 & 5 & 1 & 2 \end{pmatrix} \xrightarrow{III+(-1)I} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 3 & 0 & 1 \\ 0 & 3 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{III-II} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\frac{1}{3}II} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{pmatrix} = C$$

I denne trappematrisen  $C$  er øylen 1 og 2 pivotsøyler. Øylen 1 og 2 tilsvarer variablene  $x$  og  $y$ , så disse er basisvariable, mens  $z$  (som tilsvarer øyle 3) er en fri variabel. Vi ser at  $C$  er den utvidede matrisen til systemet

$$\begin{aligned} x + 2y + z &= 1 \\ y &= \frac{1}{3} \\ 0 &= 0 \end{aligned}$$

Den nederste ligningen er alltid oppfylt og kan neglisjeres. De to andre ligningene legger ingen føringer på den *frie variablen*  $z$  som kan velges fritt, og når den er valgt, kan vi regne ut verdiene til basisvariablene  $x$  og  $y$ . Vi får  $y = \frac{1}{3}$  og  $x = 1 - 2y - z = \frac{1}{3} - z$  (at  $y$  er uavhengig av  $z$  er en tilfeldighet).

La oss nå forandre eksemplet litt. Dersom vi endrer konstantleddet i den tredje ligningen i det opprinnelige problemet fra 2 til 1, får vi ligningssystemet

$$\begin{aligned} x + 2y + z &= 1 \\ -x + y - z &= 0 \\ x + 5y + z &= 1 \end{aligned}$$

med utvidet matrise

$$B' = \begin{pmatrix} 1 & 2 & 1 & 1 \\ -1 & 1 & -1 & 0 \\ 1 & 5 & 1 & 1 \end{pmatrix}$$

Radreduserer vi denne matrisen, ender vi opp med trappematrisen

$$C' = \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Nå har vi fått et pivotelement i siste øyle, og det har dramatiske konsekvenser. Det korresponderende ligningssystemet er nemlig

$$\begin{aligned} x + 2y + z &= 1 \\ y &= \frac{1}{3} \\ 0 &= 1 \end{aligned}$$

og her ser vi at den siste ligningen aldri kan oppfylles. Ligningssystemet har derfor ingen løsninger.



Mønsteret du ser i eksemplet ovenfor er helt generelt:

**Setning 4.2.4** *Anta at den utvidede matrisen til et lineært ligningssystem kan radreduseres til trappematrissen  $C$ . Da gjelder:*

- (i) *Dersom den siste søylen i  $C$  er en pivotsøyle, har ligningssystemet ingen løsninger.*

*Dersom den siste søylen ikke er en pivotsøyle, har vi videre:*

- (ii) *Dersom alle de andre søylene i  $C$  er pivotsøyler, har ligningssystemet nøyaktig én løsning.*
- (iii) *Dersom minst én av de andre søylene ikke er en pivotsøyle, har ligningssystemet uendelig mange løsninger.*

*Bevis:* Siden det opprinnelige ligningssystemet og det som hører til  $C$ , har nøyaktig de samme løsningene, kan vi konsentrere oss om ligningssystemet til  $C$ . Anta først at den siste søylen i  $C$  er en pivotsøyle. Da inneholder ligningssystemet en ligning av formen  $0 = 1$ , og har derfor ingen løsninger.

Anta så at den siste søylen i  $C$  ikke er en pivotsøyle. Dersom det finnes andre søyler som ikke er pivotsøyler, har systemet frie variable. Gi disse hvilke som helst verdier du ønsker. Du kan nå regne ut verdien til de andre variablene ved å begynne nedenfra med den nederste (ikke-trivuelle) ligningen. Denne ligningen gir deg verdien til den siste av basisvariablene. Gå nå videre til ligningen over og regn ut den tilhørende basisvariablen. Fortsett oppover i ligningssystemet til du har regnet ut alle basisvariablene. Det viser at for hvert valg av frie variable, har ligningssystemet nøyaktig én løsning. Finnes det frie variable, har derfor ligningssettet uendelig mange løsninger. Finnes det ikke frie variable (dvs. at alle søyler unntatt den siste er pivotsøyler), må ligningssettet ha nøyaktig én løsning.  $\square$

Det er ofte viktig å vite når et ligningssystem har nøyaktig én løsning (vi kaller dette en *entydig* løsning), og vi tar derfor med dette som et separat resultat:

**Korollar 4.2.5** *Anta at den utvidede matrisen til et lineært ligningssystem kan radreduseres til trappematrissen  $C$ . Da har ligningssystemet en entydig løsning hvis og bare hvis alle søyler i  $C$  unntatt den siste, er pivotsøyler.*

*Bevis:* Dette er bare en omskrivning av punkt (ii) i setningen ovenfor.  $\square$

Matrisen nedenfor viser en typisk trappematrise som tilsvarer et ligningsystem med entydig løsning:

$$C = \begin{pmatrix} 1 & -2 & 3 & 0 & 4 & 2 \\ 0 & 1 & 1 & 3 & -2 & 6 \\ 0 & 0 & 1 & -2 & 3 & \pi \\ 0 & 0 & 0 & 1 & 3 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Vi ser at pivotelementene begynner øverst i venstre hjørne og fortsetter nedover diagonalen inntil de når den nest siste søylen. Under dette nivået kan det godt være noen rader med bare nuller.

### Ligningssystemer med samme venstreside

Både i teori og praksis hender det ofte at vi har behov for å løse “det samme” ligningssystemet gang på gang med forskjellig høyreside. Mer presist betyr dette at vi ønsker å løse systemet

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

mange ganger for de samme koeffisientene  $a_{ij}$ , men for forskjellige  $b_i$ . Et spørsmål som da dukker opp, er når det er mulig å løse ligningssystemet for alle valg av  $b_1, b_2, \dots, b_m$ . For å løse dette problemet ser vi på *matrisen*

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

til ligningssystemet (ikke bland denne sammen med den *utvidede matrisen*)

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

der  $b$ -ene er med!) Vi radreduserer så  $A$  til en trappematrise  $D$ . Det viser seg at ligningssystemet vårt har en løsning for alle valg av  $b_1, b_2, \dots, b_m$  dersom alle radene i  $D$  inneholder et pivotelement.

**Setning 4.2.6** *Anta at*

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

*kan radreduseres til trappematrisen D. Da har ligningssystemet*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

*en løsning for alle valg av  $b_1, b_2, \dots, b_m$  hvis og bare hvis alle radene i D inneholder pivotelementer.*

*Bevis:* Anta først at D har pivotelementer i alle rader, og velg  $b_1, b_2, \dots, b_m$ . La B være den utvidede matrisen

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

Vi vet at A lar seg forvandle til D gjennom en sekvens av radoperasjoner, og vi lar nå C være den matrisen vi får når vi lar B gjennomgå den samme sekvensen av operasjoner. Da er

$$C = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} & \tilde{b}_1 \\ d_{21} & d_{22} & \cdots & d_{2n} & \tilde{b}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} & \tilde{b}_m \end{pmatrix}$$

der  $d_{ij}$  er elementene i D og  $\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_m$  er resultatet av å bruke disse radoperasjonene på  $b_1, b_2, \dots, b_m$ . Siden D har et pivotelement i hver rad, kan ikke den siste søylen i C være en pivotsøyle, og følgelig har ligningssystemet minst en løsning.

Anta nå omvendt at D mangler pivotelement i noen av radene, og la rad nummer  $j$  være den første av disse. Hvis vi kan finne  $b_1, b_2, \dots, b_m$  slik at  $\tilde{b}_j$  er lik 1, vil C få et pivotelement i siste søyle, og ligningssystemet vil da ikke ha en løsning. Siden radoperasjonene er reverserbare, er det ikke vanskelig å finne et slikt sett med b-er. Vi velger oss rett og slett et sett  $\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_m$

med den egenskapen vi ønsker oss ved å la  $\tilde{b}_j = 1$ , mens  $\tilde{b}_i = 0$  for alle andre indeksene  $i$ . Så danner vi matrisen

$$C = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} & \tilde{b}_1 \\ d_{21} & d_{22} & \dots & d_{2n} & \tilde{b}_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_{m1} & d_{m2} & \dots & d_{mn} & \tilde{b}_m \end{pmatrix}$$

Legg merke til at  $C$  er på trappeform og at  $\tilde{b}_j$  er et pivotelement i siste rad. Nå bruker vi de omvendte av de operasjonene som førte  $A$  til  $B$  til å føre  $C$  tilbake til en matrise på formen

$$B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

(husk reverseringsprosedyren i eksempel 2). Ligningssystemet til  $B$  kan ikke ha en løsning siden vi kan bruke de opprinnelige radoperasjonene til å forvandle  $B$  til  $C$ , og  $C$  har et pivotelement i siste rad.  $\square$

Legg merke til at dersom  $m > n$  (dvs. at matrisen  $A$  har flere rader enn søyler), så er det ikke plass til et pivotelement i hver rad (husk at pivotelementene må flytte seg minst ett skritt mot høyre for hver rad), og ligningssystemet kan derfor ikke ha løsninger for alle  $b_1, b_2, \dots, b_m$ .

La oss også se på betingelsen for at ligningssystemet har en *entydig* løsning for alle valg av  $b_1, b_2, \dots, b_m$ :

**Korollar 4.2.7** *Anta at*

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

*kan redreses til trappematrisen  $D$ . Da har ligningssystemet*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{aligned}$$

*en entydig løsning for alle valg av  $b_1, b_2, \dots, b_m$  hvis og bare hvis alle raderne og alle søylene i  $D$  inneholder pivotelementer. Dette betyr at  $D$  er en kvadratisk matrise med pivotelementer på diagonalen.*

*Bevis:* Vi vet fra setningen ovenfor at det må være pivotelementer i alle rader dersom vi skal ha løsninger for alle valg av  $b_1, b_2, \dots, b_m$ . Vi vet også at for å få entydige løsninger, må vi ha pivotelementer i alle søyler. Skal vi få plass til pivotelementer i alle rader og alle søyler, må  $D$  være en kvadratisk matrise med pivotelementer på diagonalen (prøv deg frem, så vil du se).  $\square$

**Bemerkning:** Ifølge resultatet ovenfor er det bare mulig å ha entydige løsninger for alle  $b_1, b_2, \dots, b_m$  dersom  $D$  — og dermed  $A$  — er en kvadratisk matrise, dvs. at vi har like mange ligninger som ukjente. Dette er en viktig observasjon som vi skal møte igjen i ulike forkledninger senere i kapitlet.

Det er på tide med et eksempel:

**Eksempel 5:** Vi skal undersøke om ligningssystemet

$$\begin{aligned} 3x + y - 2z &= b_1 \\ -x + 2y - z &= b_2 \\ x + z &= b_3 \end{aligned}$$

har en løsning for alle valg av  $b_1, b_2, b_3$ . Matrisen til systemet er

$$\left( \begin{array}{ccc} 3 & 1 & -2 \\ -1 & 2 & -1 \\ 1 & 0 & 1 \end{array} \right)$$

Bruker vi radoperasjoner på denne matrisen, får vi (vi tillater oss å ta flere operasjoner i slengen slik at ikke utledningen skal bli for lang):

$$\begin{aligned} \left( \begin{array}{ccc} 3 & 1 & -2 \\ -1 & 2 & -1 \\ 1 & 0 & 1 \end{array} \right) &\stackrel{I \leftrightarrow II}{\sim} \left( \begin{array}{ccc} -1 & 2 & -1 \\ 3 & 1 & -2 \\ 1 & 0 & 1 \end{array} \right) \stackrel{III+3I}{\sim} \left( \begin{array}{ccc} -1 & 2 & -1 \\ 0 & 7 & -5 \\ 1 & 0 & 1 \end{array} \right) \\ &\stackrel{\frac{1}{7}II}{\sim} \left( \begin{array}{ccc} -1 & 2 & -1 \\ 0 & 1 & -\frac{5}{7} \\ 1 & 0 & 1 \end{array} \right) \stackrel{III+(-2)II}{\sim} \left( \begin{array}{ccc} -1 & 2 & -1 \\ 0 & 1 & -\frac{5}{7} \\ 0 & 0 & \frac{10}{7} \end{array} \right) \stackrel{\frac{(-1)I}{10}}{\sim} \left( \begin{array}{ccc} 1 & -2 & 1 \\ 0 & 1 & -\frac{5}{7} \\ 0 & 0 & 1 \end{array} \right) \end{aligned}$$

Siden den siste matrisen har et pivotelement i hver rad, har ligningssystemet en løsning for alle valg av  $b_1, b_2, b_3$ . Siden det også er et pivotelement i hver søyle, er denne løsningen entydig. ♣

### Oppgaver til seksjon 4.2

1. Avgjør om matrisene er på trappeform:

$$A = \left( \begin{array}{ccc} 1 & -2 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{array} \right), B = \left( \begin{array}{ccc} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right), C = \left( \begin{array}{ccc} 1 & 0 & 2 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{array} \right)$$

$$D = \begin{pmatrix} 1 & -2 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, E = \begin{pmatrix} 1 & 3 & -2 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}, F = \begin{pmatrix} 1 & -2 & 0 & -1 \\ 0 & 1 & 2 & 1 \\ 0 & 1 & 4 & -3 \end{pmatrix}$$

**2.** Reduser matrisen til trappeform:

a)  $A = \begin{pmatrix} 1 & 2 & 1 \\ -1 & -3 & 2 \\ 2 & 1 & 2 \end{pmatrix}$

b)  $B = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 2 \\ -2 & 1 & 3 \end{pmatrix}$

c)  $C = \begin{pmatrix} 1 & 1 & -2 & 3 \\ 2 & 1 & 3 & 0 \\ -1 & 0 & -5 & 2 \end{pmatrix}$

d)  $D = \begin{pmatrix} 1 & 2 & 1 & 1 \\ -1 & 3 & -1 & 2 \\ 1 & 12 & 1 & 7 \end{pmatrix}$

**3.** Løs ligningssystemet ved å radredusere den utvidede matrisen.

$$\begin{aligned} x - y + 2z &= 1 \\ 2x + y + z &= 1 \\ -2x - y + z &= 0 \end{aligned}$$

**4.** Løs ligningssystemet ved å radredusere den utvidede matrisen.

$$\begin{aligned} 3x - 4y + z &= 2 \\ x - 2y &= 1 \\ -2x + 2y - z &= -1 \end{aligned}$$

**5.** Løs ligningssystemet ved å radredusere den utvidede matrisen.

$$\begin{aligned} x + 2y + z &= 2 \\ 2x - 4y + 3z &= 1 \\ 3x - 2y + 4z &= 6 \end{aligned}$$

**6.** Løs ligningssystemet ved å radredusere den utvidede matrisen.

$$\begin{aligned} x + 3y - z + 3u &= 4 \\ x + 2y - 2z + 3u &= 0 \\ 2x + 2y - 5z + 5u &= 1 \end{aligned}$$

**7.** Avgjør om ligningssystemet har en løsning for alle valg av  $b_1, b_2, b_3$ :

$$\begin{aligned} 2x + 4y - 4z &= b_1 \\ 2x - y + 3z &= b_2 \\ x - y + 2z &= b_3 \end{aligned}$$

**8.** Avgjør om ligningssystemet har en løsning for alle valg av  $b_1, b_2, b_3$ :

$$\begin{aligned}x + 2y - 3z &= b_1 \\-2x - 4y + z &= b_2 \\x + 2y + z &= b_3\end{aligned}$$

**9.** (Eksamens i MAT 1110, 14/6 2004, litt tilpasset).

a) Reduser matrisen til trappeform:

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

b) Løs ligningssystemet

$$\begin{aligned}x + 2y + 2u &= 5 \\y + z + u &= 3 \\-2y + z + u &= 0 \\x + 2y + z + 3u &= 7\end{aligned}$$

**10.** Et bilutleiefirma har kontor i tre byer  $A$ ,  $B$  og  $C$ . Av de bilene som leies i  $A$ , blir 60% returnert i  $A$ , 30% i  $B$  og 10% i  $C$ . Av de bilene som leies i  $B$ , blir 30% returnert i  $A$ , 50% i  $B$  og 20% i  $C$ . Av de bilene som leies i  $C$ , blir 60% returnert i  $A$ , 10% i  $B$  og 30% i  $C$ . Bilfirmaet har totalt 120 biler. Hvordan skal det fordele disse bilene i  $A$ ,  $B$  og  $C$  slik at det i hver by returneres like mange biler som det leies ut?

### 4.3 Redusert trappeform

Når vi omformer en matrise til trappeform, sørger vi for at elementene *under* pivotelementene alltid er null, men vi bryr oss ikke om elementene *over* pivotelementene. For noen formål lønner det seg å sørge for at disse elementene også er null. Vi sier da at matrisen er på *redusert trappeform*. Her er den presise definisjonen:

**Definisjon 4.3.1** Vi sier at en matrise er på redusert trappeform dersom den er på trappeform og alle elementene i pivotsøylene, unntatt pivotelementene, er 0.

**Eksempel 1:** Disse matrisene er på redusert trappeform:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 0 & 0 & 5 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Disse matrisene er på trappeform, men *ikke* på redusert trappeform:

$$\begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -3 & 2 \\ 0 & 1 & 4 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 3 & 3 & 5 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

♣

Alle matriser kan omformes til redusert trappeform ved hjelp av radoperasjoner. Vi radreduserer dem først til vanlig trappeform, og bruker deretter pivotelementene til å skaffe oss de resterende nullene i pivotsøylene. Det er lurest å begynne bakfra med de pivotelementene som står lengst til høyre. Her er et eksempel:

**Eksempel 2:** Vi starter med en matrise som er på vanlig trappeform

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

Deretter tar vi utgangspunkt i pivotelementet nederst til høyre, og bruker det til å skaffe oss nuller i posisjonene over:

$$\begin{pmatrix} 1 & 2 & -3 & 4 & 2 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\sim II+(-3)III} \begin{pmatrix} 1 & 2 & -3 & 4 & 2 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{\sim I+(-2)III} \begin{pmatrix} 1 & 2 & -3 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Dermed har vi ordnet opp i den bakersteøylen, og vi går nå mot venstre på jakt etter neste pivotsøyle. Det er øyle nummer 3, og vi bruker nå pivotelementet her til å skaffe flere nuller:

$$\begin{pmatrix} 1 & 2 & -3 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\sim I+3II} \begin{pmatrix} 1 & 2 & 0 & 10 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Dermed er matrisen brukt på redusert trappeform.

♣

Poenget med å starte prosessen bakfra er at vi slipper unna mye regnearbeid fordi de fleste tallene vi adderer er 0.

**Setning 4.3.2** *Enhver matrise er radekvivalent med en matrise på redusert trappeform.*

*Bevis:* Vi vet allerede at matrisen er radekvivalent med en matrise på vanlig trappeform, så alt vi trenger, er å vise er at enhver matrise på vanlig trappeform er ekvivalent med en matrise på redusert trappeform. Dette følger fra prosedyren vi har beskrevet ovenfor. Det eneste som kunne gått galt med denne prosedyren, var hvis noen av de operasjonene vi gjorde underveis, ”ødela” nuller vi allerede hadde skaffet oss på et tidligere tidspunkt, men det er lett å sjekke at det ikke skjer.  $\square$

**Bemerkning:** Det viser seg at den reduserte trappeformen til en matrise er entydig bestemt — uansett hvilken sekvens av radoperasjoner du bruker for å bringe en matrise på redusert trappeform, blir sluttresultatet det samme. Vi skal derfor av og til snakke om ”den reduserte trappeformen” i bestemt form.

Det neste resultatet skal vi ofte få bruk for. Husk at korollar 4.2.7 forteller oss at ligningssystemet

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

bare kan ha en entydig løsning for alle valg av  $b_1, b_2, \dots, b_m$  dersom  $m = n$ , dvs. dersom vi har like mange ligninger som ukjente.

#### Setning 4.3.3 Ligningssystemet

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

har en entydig løsning for alle valg av  $b_1, b_2, \dots, b_n$  hvis og bare hvis den tilhørende matrisen

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

er radekvivalent med identitetsmatrisen

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

*Bevis:* Dersom  $A$  er radekvivalent med identitetsmatrisen, følger det umiddelbart fra korollar 4.2.7 at ligningssystemet alltid har en entydig løsning. På den annen side: Dersom ligningssystemet alltid har en entydig løsning, vet vi fra korollar 4.2.7 at  $A$  er radekvivalent med en trappematrise der alle pivotelementene ligger på diagonalen. Den reduserte trappeformen til en slik matrise er identitetsmatrisen (hvorfor?), og følgelig er  $A$  radekvivalent med identitetsmatrisen.  $\square$

### Redusert trappeform i MATLAB

Det er ganske kjedelig å føre store matriser over på trappeform for hånd, men heldigvis finnes det hjelpefunksjoner. Dersom du har tastet inn en matrise  $A$  i MATLAB, vil kommandoen `>> rref(A)` få MATLAB til å regne ut den reduserte trappeformen til  $A$ . Her er et eksempel på en kjøring:

```
>> A=[2 -1 4 5 6
6 -1 3 2 1
-2 3 1 0 5];
>> B=rref(A)

B =

```

1.0000	0	0	-0.4286	-0.7500
0	1.0000	0	-0.7143	0.5000
0	0	1.0000	1.2857	2.000

Kommandonavnet `rref` kan virke litt mystisk, men det skyldes rett og slett at redusert trappeform heter *reduced row echelon form* på engelsk. Vanlig trappeform heter *row echelon form*.

La oss se hvordan vi kan bruke kommandoen `rref` til å løse et lignings-system.

**Eksempel 3:** Vi skal løse ligningssystemet

$$\begin{aligned} 2x - y + 3z + u + v &= 2 \\ 3x + y - z + 2u - v &= 3 \\ -x - 2y + 4z + u + 2v &= 4 \end{aligned}$$

Den utvidede matrisen er

$$B = \begin{pmatrix} 2 & -1 & 3 & 1 & 1 & 2 \\ 3 & 1 & -1 & 2 & -1 & 3 \\ -1 & -2 & 4 & 1 & 2 & 4 \end{pmatrix}$$

og putter vi denne inn i MATLAB og bruker `rref`, får vi

```
>> B=[2 -1 3 1 1 2
      3 1 -1 2 -1 3
      -1 -2 4 1 2 4];
>> C=rref(B)

C =

1.0000      0      0.4000      0      0      -0.5000
0      1.0000     -2.2000      0     -1.0000     -0.5000
0          0          0      1.0000      0      2.5000
```

Den reduserte trappeformen er altså

$$C = \begin{pmatrix} 1 & 0 & 0.4 & 0 & 0 & -0.5 \\ 0 & 1 & -2.2 & 0 & -1 & -0.5 \\ 0 & 0 & 0 & 1 & 0 & 2.5 \end{pmatrix}$$

Vi ser at pivotsøylene er søyle 1, 2 og 4, og at de frie variablene er  $z$  (som korresponderer til søyle 3) og  $v$  (som korresponderer til søyle 5). Vi kan derfor velge  $z$  og  $v$  fritt og løse for de andre variablene. Det er lettest å gjøre dette hvis vi først skriver opp ligningssystemet til  $C$ :

$$\begin{aligned} x + 0.4z &= -0.5 \\ y - 2.2z - v &= -0.5 \\ u &= 2.5 \end{aligned}$$

Vi ser at løsningene er gitt ved

$$x = -0.5 - 0.4z$$

$$y = -0.5 + 2.2z + v$$

$$z = z$$

$$u = 2.5$$

$$v = v$$

der  $z$  og  $v$  kan velges fritt.



### Oppgaver til seksjon 4.3

1. Avgjør om matrisene er på redusert trappeform:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

$$D = \begin{pmatrix} 1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, F = \begin{pmatrix} 1 & 2 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 7 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

2. Omform matrisene til redusert trappeform:

a)  $A = \begin{pmatrix} 1 & 2 \\ 3 & 5 \end{pmatrix}$

b)  $B = \begin{pmatrix} 2 & 1 & 3 & 4 \\ 1 & 1 & 0 & 3 \end{pmatrix}$

c)  $C = \begin{pmatrix} 1 & 1 & 1 & 2 \\ -1 & 2 & -1 & 3 \\ 2 & 2 & 1 & 1 \end{pmatrix}$

d)  $D = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ -2 & -4 & 5 \\ -1 & -2 & 8 \end{pmatrix}$

3. Bruk MATLAB til å omforme disse matrisene til redusert trappeform:

a)  $A = \begin{pmatrix} 1 & 3 & -2 & 2 & 3 \\ 0 & 2 & -3 & 4 & 6 \\ -2 & 3 & 1 & -4 & 5 \\ 0 & -2 & -1 & -2 & 8 \\ 2 & 3 & 1 & 0 & -1 \end{pmatrix}$

b)  $B = \begin{pmatrix} 0.25 & 0.5 & 1.5 & 0.75 \\ 1 & 0.55 & 0.7 & 0.25 \\ -0.25 & 3 & 0.75 & -0.1 \end{pmatrix}$

c)  $C = \begin{pmatrix} 1 & 2 & 0.5 & 3 & -1 & 2 \\ 2 & 0.5 & -1 & -1 & 2 & 3 \\ 1 & 1 & 1 & 2 & 3 & 4 \\ 2 & -1 & 2 & 3 & -2 & 0 \end{pmatrix}$

4. Avgjør om ligningssystemet har en entydig løsning for alle valg av  $b_1, b_2, b_3$ . Bruk gjerne MATLAB som hjelpemiddel.

$$\begin{aligned} x + 2y + z &= b_1 \\ 2x + 4y + 3z &= b_2 \\ -x + 3y + 2z &= b_3 \end{aligned}$$

**5.** Avgjør om ligningssystemet har en entydig løsning for alle valg av  $b_1, b_2, b_3$ . Bruk gjerne MATLAB som hjelphemiddel.

$$\begin{aligned} 2x - y + z &= b_1 \\ -x + 3y + 2z &= b_2 \\ 3x - 4y - z &= b_3 \end{aligned}$$

**6.** Finn alle løsningene til ligningssystemet. Bruk først MATLAB til å skrive den utvidede matrisen på redusert trappeform.

$$\begin{aligned} 2x - y + z + 3u &= -4 \\ -x + 2y + 4z + 3u &= 2 \\ -2x + y + 3z - 4u &= -1 \end{aligned}$$

**7.** Finn alle løsningene til ligningssystemet. Bruk først MATLAB til å skrive den utvidede matrisen på redusert trappeform.

$$\begin{aligned} x + y - z + 2u - v &= 1 \\ -2x - 2y + z - u + v &= 2 \\ 3x + 3y - 2u + 2v &= 1 \end{aligned}$$

## 4.4 Matriseligninger

Dersom vi starter med en  $m \times n$ -matrise

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

og en søylevektor

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

kan vi regne ut en søylevektor

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

ved å ta produktet

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

Vi kan også snu problemstillingen på hodet: Dersom vi starter med  $A$  og  $\mathbf{b}$ , ønsker vi å finne en vektor  $\mathbf{x}$  slik at

$$A\mathbf{x} = \mathbf{b} \quad (4.4.1)$$

Vi kaller dette en *matriseligning*. Dersom vi skriver ut ligning (4.4.1) på komponentform, får vi

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \quad (4.4.2)$$

Det å løse matriseligningen (4.4.1) er altså det samme som å løse ligningsystemet

$$\begin{array}{lcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array}$$

All den kunnskapen vi har skaffet oss om lineære ligningssystemer, kan vi nå overføre til matriseligninger. Først litt notasjon — vi skal skrive

$$B = (A, \mathbf{b})$$

for den *utvidede matrisen*

$$B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

La oss nå oversette setning 4.2.4 til matrisespråk.

**Setning 4.4.1** La  $B = (A, \mathbf{b})$  være den utvidede matrisen til matriseligningen

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

og anta at  $B$  kan radreduseres til trappematrisen  $C$ . Da gjelder:

- (i) Dersom den siste søylen i  $C$  er en pivotsøyle, har matriseligningen ingen løsninger.

Dersom den siste søylen ikke er en pivotsøyle, har vi videre:

- (ii) Dersom alle de andre søylene i  $C$  er pivotsøyler, har matriseligningen nøyaktig én løsning.
- (iii) Dersom minst én av de andre søylene ikke er en pivotsøyle, har matriseligningen uendelig mange løsninger.

*Bevis:* Dette er bare en omformulering av setning 4.2.4.  $\square$

**Eksempel 1:** Finn alle løsninger til matriseligningen  $A\mathbf{x} = \mathbf{b}$  når

$$A = \begin{pmatrix} 2 & -1 & 3 & 3 & 4 \\ 3 & 0 & 2 & 1 & -1 \\ -1 & 2 & -4 & -5 & 0 \end{pmatrix} \quad \text{og} \quad \mathbf{b} = \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}$$

Vi bruker først radoperasjoner på den utvidede matrisen

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

Dersom vi velger å bruke MATLAB, får vi

```
B=[2 -1 3 3 4 1
3 0 2 1 -1 4
-1 2 -4 -5 0 1];
>> C=rref(B)
C =
1.0000      0      0.6667      0.3333      0      1.2963
0      1.0000     -1.6667     -2.3333      0      1.1481
0          0          0          0      1.0000     -0.1111
```

$B$  er altså radekvivalent med matrisen

$$C = \begin{pmatrix} 1 & 0 & 0.6667 & 0.3333 & 0 & 1.2963 \\ 0 & 1 & -1.6667 & -2.3333 & 0 & 1.1481 \\ 0 & 0 & 0 & 0 & 1 & -0.1111 \end{pmatrix}$$

Vi ser at den siste søylen ikke er en pivotsøyle, så ligningen har løsninger. Vi ser også at søyler 3 og 4 ikke er pivotsøyler, så ligningen har uendelig mange løsninger. Variablene  $x_3$  og  $x_4$  er frie, og velger vi verdier for disse, kan vi regne ut de andre variablene:

$$x_1 = 1.2963 - 0.6667x_3 - 0.3333x_4$$

$$x_2 = 1.1481 + 1.6667x_3 + 2.3333x_4$$

$$x_5 = -0.1111$$

Skriver vi løsningen på vektorform, har vi dermed

$$\begin{aligned}\mathbf{x} &= \begin{pmatrix} 1.2963 - 0.6667x_3 - 0.3333x_4 \\ 1.1481 + 1.6667x_3 + 2.3333x_4 \\ x_3 \\ x_4 \\ -0.1111 \end{pmatrix} = \\ &= \begin{pmatrix} 1.2963 \\ 1.1481 \\ 0 \\ 0 \\ -0.1111 \end{pmatrix} + x_3 \begin{pmatrix} -0.6667 \\ 1.6667 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -0.3333 \\ 2.3333 \\ 0 \\ 1 \\ 0 \end{pmatrix}\end{aligned}$$

Denne skrivemåten gir en god oversikt over løsningene. ♣

La oss nå undersøke når matriseligningen  $A\mathbf{x} = \mathbf{b}$  har en løsning for alle vektorer  $\mathbf{b}$ . Svaret ligger i setning 4.2.6 og korollar 4.2.7:

**Setning 4.4.2** *Anta at matrisen  $A$  er radekvivalent med trappematrisen  $D$ . Da har ligningen*

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

*løsning for alle vektorer  $\mathbf{b} \in \mathbb{R}^m$  hvis og bare hvis alle radene i  $D$  inneholder et pivotelement. Løsningen er entydig dersom også alle søylene i  $D$  inneholder et pivotelement — dette betyr at  $A$  er en kvadratisk matrise som er radekvivalent med identitetsmatrisen.*

*Bevis:* Som allerede nevnt er dette bare en omskrivning av setning 4.2.6 og korollar 4.2.7 til matrisespråk. □

### Homogene ligninger

En matriseligning  $A\mathbf{x} = \mathbf{b}$  kan ha ingen løsninger for noen verdier av  $\mathbf{b}$  og én eller uendelig mange løsninger for andre verdier av  $\mathbf{b}$ . For å forstå sammenhengen er det lurt å ta utgangspunkt i *homogene* ligninger, dvs. ligninger av typen  $A\mathbf{x} = \mathbf{0}$  der høyresiden er null. En homogen ligning har alltid løsningen  $\mathbf{x} = \mathbf{0}$ , så spørsmålet er om dette er den eneste løsningen, eller om det finnes uendelig mange andre. Fra setning 4.4.1 får vi:

**Korollar 4.4.3** *Anta at matrisen  $A$  har trappeform  $D$ . Dersom alle søylene i  $D$  er pivotsøyler, har den homogene ligningen  $A\mathbf{x} = \mathbf{0}$  bare løsningen  $\mathbf{x} = \mathbf{0}$ . Dersom  $D$  har søyler som ikke er pivotsøyler, har ligningen uendelig*

mange løsninger. Dersom  $A$  er en kvadratisk  $n \times n$ -matrise, betyr dette at ligningen  $A\mathbf{x} = \mathbf{0}$  har  $\mathbf{0}$  som eneste løsning hvis og bare  $A$  er radekvivalent med  $I_n$ .

*Bevis:* Den generelle delen av korollaret er følger direkte fra setning 4.4.1. Tillegget om kvadratiske matriser følger fordi en kvadratisk matrise har pivotelementer i alle søyler hvis og bare hvis alle pivotelementene står på diagonalen, og det er det samme som at den kan radresuseres til  $I_n$ .  $\square$

Den neste setningen gir oss sammenhengen mellom løsningene av homogene og inhomogene ligninger. Har du studert differens- eller differensial-ligninger, vil du ha sett lignende resultater før. Indeksene  $p$  og  $h$  på vektorene  $\mathbf{x}_p$  og  $\mathbf{x}_h$  står for henholdsvis *partikulær* og *homogen*.

**Setning 4.4.4** *Anta  $\mathbf{x}_p$  er en løsning av matriseligningen  $A\mathbf{x} = \mathbf{b}$ . De andre løsningene er da vektorene på formen*

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$$

*der  $\mathbf{x}_h$  er en løsning av den homogene ligningen  $A\mathbf{x} = \mathbf{0}$ .*

*Bevis:* Anta først at  $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$  der  $\mathbf{x}_h$  er en løsning av den homogene ligningen. Da er

$$A\mathbf{x} = A(\mathbf{x}_p + \mathbf{x}_h) = A\mathbf{x}_p + A\mathbf{x}_h = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

så  $\mathbf{x}$  er en løsning av ligningen.

Anta så at  $\mathbf{x}$  er en løsning av  $A\mathbf{x} = \mathbf{b}$ , og definer  $\mathbf{x}_h = \mathbf{x} - \mathbf{x}_p$ . Da er  $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$ , og alt vi behøver å vise, er at  $\mathbf{x}_h$  er en løsning av den homogene ligningen. Dette er bare et lite regnestykke:

$$A\mathbf{x}_h = A(\mathbf{x} - \mathbf{x}_p) = A\mathbf{x} - A\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

$\square$

Setningen ovenfor forteller oss at dersom den inhomogene ligningen  $A\mathbf{x} = \mathbf{b}$  har løsninger, så har den like mange løsninger som den homogene løsningen  $A\mathbf{x} = \mathbf{0}$ .

### Simultane løsninger av matriseligninger

Anta at vi ønsker å løse matriseligningen  $A\mathbf{x} = \mathbf{b}$  for flere verdier av  $\mathbf{b}$ , la oss si for  $\mathbf{b} = \mathbf{b}_1, \mathbf{b} = \mathbf{b}_2, \dots, \mathbf{b} = \mathbf{b}_k$ . Vi ønsker med andre ord å løse ligningene

$$A\mathbf{x}_1 = \mathbf{b}_1, A\mathbf{x}_2 = \mathbf{b}_2, \dots, A\mathbf{x}_k = \mathbf{b}_k$$

Som nevnt tidligere er dette en problemstilling som ofte dukker opp i praksis. Den er mest aktuell når ligningssystemene har entydig løsning, så vi antar at  $A$  er en kvadratisk matrise som er radekvivalent med identitetsmatrisen  $I_n$  (husk setning 4.4.2). For å løse ligningene er det naturlige å begynne med å radredusere den utvidede matrisen  $(A, \mathbf{b}_1)$  til den første matriseligningen. Hvis vi gjør dette helt til vi kommer til redusert trappeform, sitter vi igjen med en matrise på formen

$$C_1 = \begin{pmatrix} 1 & 0 & \dots & 0 & \tilde{b}_{11} \\ 0 & 1 & \dots & 0 & \tilde{b}_{21} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \tilde{b}_{n1} \end{pmatrix}$$

der  $\tilde{b}_{11}, \tilde{b}_{21}, \dots, \tilde{b}_{n1}$  er de tallene vi får når vi bruker radoperasjonene på komponentene til  $\mathbf{b}_1$ . Setter vi inn variablene, får vi ligningene

$$\begin{aligned} x_1 &= \tilde{b}_{11} \\ x_2 &= \tilde{b}_{21} \\ &\vdots && \vdots \\ x_n &= \tilde{b}_{n1} \end{aligned}$$

Den siste søylen i  $C_1$  gir oss altså løsningen av den første matriseligningen. Hvis vi gjør det samme med den andre matriseligningen, får vi på tilsvarende måte en matrise

$$C_2 = \begin{pmatrix} 1 & 0 & \dots & 0 & \tilde{b}_{12} \\ 0 & 1 & \dots & 0 & \tilde{b}_{22} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & \tilde{b}_{n2} \end{pmatrix}$$

der  $\tilde{b}_{21}, \tilde{b}_{22}, \dots, \tilde{b}_{n2}$  er løsningen av den andre matriseligningen. Vi ser at vi har gjort nesten nøyaktig de samme operasjonene to ganger; i begge tilfeller har vi radredusert  $A$  til identitetsmatrisen  $I_n$ , den eneste forskjellen er at vi har hatt forskjellige sistesøyler å arbeide med. Vi kan effektivisere arbeidet ved å putte inn alle høyresidene  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$  på en gang. Vi starter altså med matrisen

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_{11} & b_{12} & \dots & b_{1k} \\ a_{21} & a_{22} & \dots & a_{2n} & b_{21} & b_{22} & \dots & b_{2k} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & b_{n1} & b_{n2} & \dots & b_{nk} \end{pmatrix}$$

der

$$\mathbf{b}_1 = \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{n2} \end{pmatrix}, \dots, \mathbf{b}_k = \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{nk} \end{pmatrix}$$

Omformer vi denne matrisen til redusert trappeform, sitter vi igjen med

$$C_1 = \begin{pmatrix} 1 & 0 & \dots & 0 & \tilde{b}_{11} & \tilde{b}_{12} & \dots & \tilde{b}_{1k} \\ 0 & 1 & \dots & 0 & \tilde{b}_{21} & \tilde{b}_{22} & \dots & \tilde{b}_{2k} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 & \tilde{b}_{n1} & \tilde{b}_{n2} & \dots & \tilde{b}_{nk} \end{pmatrix}$$

Løsningene av ligningssystemene er altså søylevektorene som står etter identitetsmatrisen.

La oss se på et enkelt eksempel:

**Eksempel 2:** Vi skal løse ligningene

$$A\mathbf{x}_1 = \mathbf{b}_1, A\mathbf{x}_2 = \mathbf{b}_2, A\mathbf{x}_3 = \mathbf{b}_3$$

der

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

og

$$\mathbf{b}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \mathbf{b}_3 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Vi starter med den utvidede matrisen

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

og omdanner denne til redusert trappeform

$$\begin{aligned} & \begin{pmatrix} 2 & 1 & 1 & 3 & 2 \\ 3 & -1 & 2 & 1 & 1 \end{pmatrix} \xrightarrow{\frac{1}{2}I} \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{3}{2} & 1 \\ 3 & -1 & 2 & 1 & 1 \end{pmatrix} \\ & \xrightarrow{II+(-3)I} \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{3}{2} & 1 \\ 0 & -\frac{5}{2} & \frac{1}{2} & -\frac{7}{2} & -2 \end{pmatrix} \xrightarrow{-\frac{2}{5}II} \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{3}{2} & 1 \\ 0 & 1 & -\frac{1}{5} & \frac{7}{5} & \frac{4}{5} \end{pmatrix} \\ & \xrightarrow{I+(-\frac{1}{2})II} \begin{pmatrix} 1 & 0 & \frac{3}{5} & \frac{4}{5} & \frac{3}{5} \\ 0 & 1 & -\frac{1}{5} & \frac{7}{5} & \frac{4}{5} \end{pmatrix} \end{aligned}$$

Dette betyr at løsningene til de tre matriseligningene er henholdsvis

$$\mathbf{x}_1 = \begin{pmatrix} \frac{3}{5} \\ -\frac{1}{5} \end{pmatrix}, \mathbf{x}_2 = \begin{pmatrix} \frac{4}{5} \\ \frac{7}{5} \end{pmatrix}, \mathbf{x}_3 = \begin{pmatrix} \frac{3}{5} \\ \frac{4}{5} \end{pmatrix}$$

Sjekk svarene ved å sette inn i ligningene!



For store ligningssystemer (som er det man ofte støter på i praksis) lønner det seg å bruke MATLAB til å foreta radreduksjonen.

### Oppgaver til seksjon 4.4

1. Finn alle løsningene av matriseligning  $A\mathbf{x} = \mathbf{b}$  når:

- a)  $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$
- b)  $A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 2 & -1 \\ 2 & -2 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$
- c)  $A = \begin{pmatrix} 2 & 1 & -1 & 2 \\ 1 & 1 & -1 & 2 \\ -1 & 1 & 2 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 3 \\ -1 \\ -2 \end{pmatrix}$
- d)  $A = \begin{pmatrix} 1 & -2 & 3 \\ 2 & -1 & 3 \\ 1 & 0 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}$

2. Løs ligningene  $A\mathbf{x}_1 = \mathbf{b}_1$  og  $A\mathbf{x}_2 = \mathbf{b}_2$  når  $A = \begin{pmatrix} 3 & 1 \\ -1 & 2 \end{pmatrix}, \mathbf{b}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$

3. Løs ligningene  $A\mathbf{x}_1 = \mathbf{b}_1$  og  $A\mathbf{x}_2 = \mathbf{b}_2$  når  $A = \begin{pmatrix} -2 & 1 & 3 \\ 1 & 2 & -1 \\ 0 & 2 & 1 \end{pmatrix}, \mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$

4. (Eksamensoppgave i MAT 1110, 13/6 2005, litt tilpasset)

a) Bring matrisen på trappeform:

$$A = \begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 0 & -1 & 1 \\ 6 & 0 & -6 & 7 \\ 2 & 1 & 0 & 0 \end{pmatrix}$$

b) La

$$\mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ h \\ 0 \end{pmatrix}$$

Avgjør for hvilke verdier av  $h$  ligningen  $A\mathbf{x} = \mathbf{b}$  har løsninger, og finn løsningene når de finnes.

5. (Eksamensoppgave i MAT 1110, 14/6 2006, litt tilpasset) I denne oppgaven er  $C$  matrisen

$$C = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 2 & 1 & a^2 - a & 3 \\ -1 & 1 & -3 & a \end{pmatrix}$$

der  $a$  er et reelt tall.

a) Reduser  $C$  til trappeform.

b) Vi lar  $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & a^2 - a \\ -1 & 1 & -3 \end{pmatrix}$  og  $\mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ a \end{pmatrix}$ . For hvilke verdier av  $a$  har ligningssystemet  $A\mathbf{x} = \mathbf{b}$  henholdsvis én, ingen og uendelig mange løsninger?

**6.** (Prøveeksamen i MAT 1110, våren 2006) I denne oppgaven er  $C$  matrisen

$$C = \begin{pmatrix} 1 & 1 & -1 & 0 \\ 2 & -1 & -5 & 3 \\ -1 & 2 & a^2 + 3a & -3a \end{pmatrix}$$

der  $a$  er et reelt tall.

a) Reduser  $C$  til trappeform.

b) Vi lar  $A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & -1 & -5 \\ -1 & 2 & a^2 + 3a \end{pmatrix}$  og  $\mathbf{b} = \begin{pmatrix} 0 \\ 3 \\ -3a \end{pmatrix}$ . For hvilke verdier av  $a$  har ligningssystemet  $A\mathbf{x} = \mathbf{b}$  henholdsvis én, ingen og uendelig mange løsninger?

## 4.5 Inverse matriser

Husk at hvis  $A$  er en  $n \times n$ -matrise, så er  $A^{-1}$  den inverse matrisen til  $A$  dersom

$$AA^{-1} = I_n \quad \text{og} \quad A^{-1}A = I_n \quad (4.5.1)$$

I seksjon 1.7 viste vi at en kvadratisk matrise kan ha høyst én invers matrise, og at det finnes mange matriser som ikke har en invers. Hittil har vi imidlertid ikke hatt effektive metoder til å finne ut om en matrise er inverterbar, eller til å regne ut den inverse matrisen når den finnes. I denne seksjonen skal vi bruke teorien for matriseligninger til å finne slike metoder.

Det er ikke så rart at det er en nær sammenheng mellom matriseligninger og inverse matriser. Dersom matrisen  $A$  er inverterbar, kan vi nemlig løse matriseligningen  $A\mathbf{x} = \mathbf{b}$  ved å gange med den inverse  $A^{-1}$  på begge sider. Vi får da  $A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{b}$  som kan forenkles til  $\mathbf{x} = A^{-1}\mathbf{b}$  siden  $A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}$ . Dette betyr at dersom  $A$  er inverterbar, har ligningen  $A\mathbf{x} = \mathbf{b}$  en entydig løsning  $\mathbf{x} = A^{-1}\mathbf{b}$ . Vi kan med andre ord ikke regne med å finne en invers matrise med mindre ligningen  $A\mathbf{x} = \mathbf{b}$  har en entydig løsning for alle  $\mathbf{b}$ , dvs. med mindre  $A$  er radekvivalent med identitetsmatrisen  $I_n$  (husk setning 4.4.2). Et av de teoretiske resultatene i denne seksjonen er at vi faktisk har en fullstendig korrespondanse her — en kvadratisk matrise er inverterbar hvis og bare hvis den er radekvivalent med identitetsmatrisen (se setning 4.5.4). Det viser seg at beviset for dette resultatet leder oss til en effektiv måte å finne inverse matriser på, men før vi kommer så langt, er det noen teoretiske spørsmål vi må rydde opp i.

Det første vi skal vise er at en ensidig invers også er en tosidig invers — det vil si at hvis en matrise tilfredsstiller én av betingelsene i (4.5.1), så tilfredsstiller den automatisk den andre. Vi begynner med to hjelpesetninger.

**Lemma 4.5.1** *Anta at  $B$  og  $C$  er to  $m \times n$ -matriser slik at  $B\mathbf{x} = C\mathbf{x}$  for alle  $\mathbf{x} \in \mathbb{R}^n$ . Da er  $B = C$ .*

*Bevis:* Det er nok å vise at dersom  $B \neq C$ , så finnes det en vektor  $\mathbf{x}$  slik at  $B\mathbf{x} \neq C\mathbf{x}$ . Det er ikke så vanskelig: Siden  $B \neq C$ , finnes det minst ett par av indekser  $i, j$  slik at  $b_{ij} \neq c_{ij}$ . Velger vi  $\mathbf{x} = \mathbf{e}_j$ , ser vi at  $B\mathbf{x} \neq C\mathbf{x}$  siden de  $i$ -te komponentene til de to vektorene er henholdsvis  $b_{ij}$  og  $c_{ij}$ .  $\square$

I den neste hjelpesetningen får vi bruk for våre resultater fra forrige seksjon.

**Lemma 4.5.2** *La  $A$  være en  $n \times n$ -matrise og anta at det finnes en  $n \times n$ -matrise  $B$  slik at  $AB = I_n$  ( $B$  er altså en høyreinvers til  $A$ ). Da har matrise-ligningen  $A\mathbf{x} = \mathbf{c}$  en entydig løsning for alle  $\mathbf{c} \in \mathbb{R}^n$ . Søylene  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  til  $B$  er løsningene til ligningene  $A\mathbf{x} = \mathbf{e}_1, A\mathbf{x} = \mathbf{e}_2, \dots, A\mathbf{x} = \mathbf{e}_n$ .*

*Bevis:* La  $\mathbf{b}_j$  være den  $j$ -te soylen i  $B$ . Da er  $A\mathbf{b}_j = \mathbf{e}_j$  (dette skyldes at når du regner ut  $A\mathbf{b}_j$  gjør du akkurat det samme som når du regner ut den  $j$ -te soylen i  $AB$ , og den  $j$ -te soylen i  $AB$  er  $\mathbf{e}_j$  siden  $AB = I_n$ ). Dette betyr at  $\mathbf{b}_j$  er en løsning av ligningen  $A\mathbf{x} = \mathbf{e}_j$ .

For å vise at ligningen  $A\mathbf{x} = \mathbf{c}$  har en løsning for alle  $\mathbf{c} \in \mathbb{R}^n$ , observerer vi først at enhver  $\mathbf{c}$  kan skrives som en lineærkombinasjon av  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ . Vi har:

$$\begin{aligned}\mathbf{c} &= \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + c_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \\ &= c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + \cdots + c_n \mathbf{e}_n\end{aligned}$$

Hvis vi nå lar  $\mathbf{x} = c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \cdots + c_n \mathbf{b}_n$  (der  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  er som ovenfor), får vi

$$\begin{aligned}A\mathbf{x} &= A(c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \cdots + c_n \mathbf{b}_n) = c_1 A\mathbf{b}_1 + c_2 A\mathbf{b}_2 + \cdots + c_n A\mathbf{b}_n = \\ &= c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + \cdots + c_n \mathbf{e}_n = \mathbf{c}\end{aligned}$$

Dette viser at ligningen  $A\mathbf{x} = \mathbf{c}$  har en løsning for alle  $\mathbf{c}$ , og det gjenstår å vise at denne løsningen er entydig.

Her kobler vi inn teorien fra forrige seksjon. Vi tenker oss først at vi bruker radoperasjoner til å redusere  $A$  til en trappematrise  $D$ . Siden ligningen

$A\mathbf{x} = \mathbf{c}$  har en løsning for alle  $\mathbf{c}$ , forteller setning 4.4.2 oss at  $D$  har et pivot-element i hver rad. Siden  $D$  er kvadratisk, må  $D$  da også ha et pivotelement i hver søyle (ellers er det ikke plass til et pivotelement i hver rad), og ifølge setning 4.4.2 er da løsningen av  $A\mathbf{x} = \mathbf{c}$  entydig.  $\square$

Vi kan nå vise at en ensidig invers er en tosidig invers.

**Setning 4.5.3** *Anta at  $A$  og  $B$  er to  $n \times n$ -matriser. Dersom*

$$AB = I_n,$$

*så er  $A$  og  $B$  inverterbare, og  $A^{-1} = B$ ,  $B^{-1} = A$*

*Bevis:* Det nok å vise at  $BA = I_n$  siden vi da har både  $AB = I_n$  og  $BA = I_n$ . Ifølge lemma 4.5.1 holder det å vise at

$$(BA)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}$$

for alle  $\mathbf{x} \in \mathbb{R}^n$ . Lar vi  $\mathbf{y} = (BA)\mathbf{x}$ , er det altså nok å vise at  $\mathbf{y} = \mathbf{x}$ . Bruker vi den assosiativitetsloven for matrisemultiplikasjon (setning 1.6.2(i)) flere ganger, ser vi at

$$A\mathbf{y} = A((BA)\mathbf{x}) = A(B(A\mathbf{x})) = (AB)(A\mathbf{x}) = I_n(A\mathbf{x}) = A\mathbf{x}$$

Setter vi  $\mathbf{b} = A\mathbf{x}$ , har vi dermed både

$$A\mathbf{x} = \mathbf{b} \quad \text{og} \quad A\mathbf{y} = \mathbf{b}$$

Siden ligningen  $A\mathbf{x} = \mathbf{b}$  ifølge lemma 4.5.2 har en entydig løsning, betyr dette at  $\mathbf{x} = \mathbf{y}$ , og dermed er setningen bevist.  $\square$

Det neste resultatet gir vår annonserede beskrivelse av når en matrise er inverterbar.

**Setning 4.5.4** *En  $n \times n$ -matrise  $A$  er inverterbar hvis og bare hvis matrise-ligningen  $A\mathbf{x} = \mathbf{c}$  har en entydig løsning for alle vektorer  $\mathbf{c} \in \mathbb{R}^n$ , det vil si hvis og bare hvis  $A$  er radekvivalent med identitetsmatrisen  $I_n$ .*

*Bevis:* Vi vet fra setning 4.4.2 at ligningen  $A\mathbf{x} = \mathbf{c}$  har en entydig løsning for alle  $\mathbf{c}$  hvis og bare hvis  $A$  er radekvivalent med identitetsmatrisen. Det er derfor nok å vise at  $A$  er inverterbar hvis og bare hvis  $A\mathbf{x} = \mathbf{c}$  har entydig løsning for alle  $\mathbf{c}$ .

Fra setningen 4.5.2 vet vi at dersom  $A$  er inverterbar, så har ligningen  $A\mathbf{x} = \mathbf{c}$  entydig løsning for alle  $\mathbf{c}$ . Anta omvendt at  $A\mathbf{x} = \mathbf{c}$  har en entydig løsning for alle  $\mathbf{c}$ , og la  $\mathbf{b}_j$  være løsningen av ligningen  $A\mathbf{x} = \mathbf{e}_j$ . Hvis  $B$  er matrisen som har  $\mathbf{b}_j$  som  $j$ -te søyle, har vi dermed  $AB = I_n$ , og følgelig er

$A$  inverterbar ifølge setning 4.5.3.  $\square$

Legg merke til at dersom  $A$  er inverterbar, så er løsningen til ligningen  $A\mathbf{x} = \mathbf{c}$  gitt ved

$$\mathbf{x} = A^{-1}\mathbf{c}$$

(gang ligningen  $A\mathbf{x} = \mathbf{c}$  fra venstre med  $A^{-1}$ ).

**Eksempel 1:** Vi skal undersøke om matrisen

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

er inverterbar. Radreduserer vi  $A$ , får vi:

$$A = \begin{pmatrix} 1 & 3 & 2 \\ -1 & 0 & 1 \\ 0 & 3 & 3 \end{pmatrix} \xrightarrow{II+I} \begin{pmatrix} 1 & 3 & 2 \\ 0 & 3 & 3 \\ 0 & 3 & 3 \end{pmatrix} \xrightarrow{III+(-1)II} \begin{pmatrix} 1 & 3 & 2 \\ 0 & 3 & 3 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{\frac{1}{3}II} \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Denne trappematrisen har ikke pivotelementer i siste rad og siste søyle, og ligningen  $A\mathbf{x} = \mathbf{b}$  har derfor *ikke* en entydig løsning for alle  $\mathbf{b}$ . Dermed vet vi at  $A$  ikke er inverterbar.  $\clubsuit$

### En metode for å finne inverse matriser

Vi vet nå at en matrise er inverterbar hvis og bare hvis den er radekvivalent med identitetsmatrisen. Vi vet også at for å finne den inverse matrisen til  $A$ , er det nok å finne en matrise  $B$  slik at

$$AB = I_n$$

I tillegg vet vi at vi kan finne søylene  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n$  i  $B$  ved å løse ligningene

$$A\mathbf{b}_1 = \mathbf{e}_1, A\mathbf{b}_2 = \mathbf{e}_2, \dots, A\mathbf{b}_n = \mathbf{e}_n$$

(husk lemma 4.5.2). Fra forrige seksjon vet vi hvordan vi løser slike ligningssett — vi starter med matrisen

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 & 0 & \dots & 1 \end{pmatrix}$$

der vi har utvidet  $A$  med høyresidene i ligningene  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ , og omformer den til redusert trappeform:

$$C = \begin{pmatrix} 1 & 0 & \dots & 0 & b_{11} & b_{12} & \dots & b_{1n} \\ 0 & 1 & \dots & 0 & b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 & b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

Løsningene av matriseligningene er da søylene til høyre for identitetsmatriksen:

$$\mathbf{b}_1 = \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{pmatrix}, \mathbf{b}_2 = \begin{pmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{n2} \end{pmatrix}, \dots, \mathbf{b}_n = \begin{pmatrix} b_{1n} \\ b_{2n} \\ \vdots \\ b_{nn} \end{pmatrix}$$

og den inverse matrisen blir

$$A^{-1} = B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \dots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix}$$

Vi kan oppsummere metoden slik: Dersom vi omformer matrisen  $(A, I_n)$  til redusert trappeform, får vi matrisen  $(I_n, A^{-1})$ . La oss demonstrere metoden på et eksempel:

**Eksempel 2:** Vi skal bruke metoden til å invertere matrisen

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 2 & -2 & 5 \\ -2 & 1 & 0 \end{pmatrix}$$

Det første vi gjør er “å skjøte på”  $A$  en identitetsmatrise slik at vi får matrisen

$$(A, I_3) = \begin{pmatrix} 1 & -1 & 1 & 1 & 0 & 0 \\ 2 & -2 & 5 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Vi skriver nå denne matrisen på redusert trappeform:

$$(A, I_3) = \begin{pmatrix} 1 & -1 & 1 & 1 & 0 & 0 \\ 2 & -2 & 5 & 0 & 1 & 0 \\ -2 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{II+(-2)I} \begin{pmatrix} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 3 & -2 & 1 & 0 \\ -2 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{III+2I} \begin{pmatrix} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 3 & -2 & 1 & 0 \\ 0 & -1 & 2 & 2 & 0 & 1 \end{pmatrix} \xrightarrow{II \leftrightarrow III} \begin{pmatrix} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & -1 & 2 & 2 & 0 & 1 \\ 0 & 0 & 3 & -2 & 1 & 0 \end{pmatrix}$$

$$\begin{array}{c} \xrightarrow{\frac{(-1)II}{3}III} \\ \left( \begin{array}{cccccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -2 & -2 & 0 & -1 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right) \xrightarrow{II+(-1)III} \\ \left( \begin{array}{cccccc} 1 & -1 & 0 & \frac{5}{3} & -\frac{1}{3} & 0 \\ 0 & 1 & 0 & -\frac{10}{3} & \frac{2}{3} & -1 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right) \\ \xrightarrow{I+II} \\ \left( \begin{array}{cccccc} 1 & 0 & 0 & -\frac{5}{3} & \frac{1}{3} & -1 \\ 0 & 1 & 0 & -\frac{10}{3} & \frac{2}{3} & -1 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right) = (I_3, B) \end{array}$$

Den første halvparten av denne matrisen er identitetsmatrisen, og den andre halvparten er  $B = A^{-1}$ . Vi har altså

$$A^{-1} = \begin{pmatrix} -\frac{5}{3} & \frac{1}{3} & -1 \\ -\frac{10}{3} & \frac{2}{3} & -1 \\ -\frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}$$

Du bør sjekke at dette er riktig ved å utføre multiplikasjonen

$$\begin{pmatrix} 1 & -1 & 1 \\ 2 & -2 & 5 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} -\frac{5}{3} & \frac{1}{3} & -1 \\ -\frac{10}{3} & \frac{2}{3} & -1 \\ -\frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix} \clubsuit$$

Metoden fungerer også i det tilfellet der  $A$  ikke er inverterbar. Da vil den reduserte trappeformen til  $(A, I_n)$  ikke begynne med identitetsmatrisen  $I_n$ , og dette forteller oss at  $A$  ikke er inverterbar.

### Inverse matriser i MATLAB

Dersom du har lastet inn en inverterbar matrise  $A$  i MATLAB, vil kommandoen

```
>> B=inv(A)
```

få MATLAB til å regne ut den inverse matrisen og legge den inn i variabelen  $B$ . Dersom du ønsker å løse matriseligningen

$$Ax = b$$

kan du nå gjøre det ved å skrive

```
>> x=Bb
```

(dette forutsetter selvfølgelig at du allerede har lastet inn  $b$ ). Det er imidlertid mer effektivt å bruke kommandoen

```
>> x=A\b
```

Legg merke til at “brøkstreken” \ går “gal vei” — det skyldes at man her “deler fra venstre” (det vil si at man gjør noe som tilsvarer å gange med  $A^{-1}$  fra venstre). Kommandoene

```
>> x=b/A
```

med “normal” brøkstrek, produserer løsningen til ligningen  $\mathbf{x}A = \mathbf{b}$  (i dette tilfellet må  $\mathbf{x}$  og  $\mathbf{b}$  være radvektorer for at dimensjonene skal passe) fordi vi her kan løse ligningssystemet ved å gange med  $A^{-1}$  fra høyre — dvs. vi deler fra høyre.

Vær oppmerksom på at kommandoene ovenfor bare fungerer etter beskrivelsen når  $A$  er en inverterbar, kvadratisk matrise; vil du løse andre typer ligningssystemer, må du bruke teknikkene vi har sett på tidligere i dette kapitlet. (Advarsel: Det kan hende at du får svar på kommandoene `>> x=A\b` selv om matrisen  $A$  ikke er kvadratisk, men løsningen kan da ha en annen tolkning — prøv `>> help mldivide` for mer informasjon).

### Oppgaver til seksjon 4.5

**1.** Finn den inverse matrisen dersom den finnes:

$$\text{a)} A = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}, \text{ b)} B = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}, \text{ c)} C = \begin{pmatrix} 1 & -2 \\ -3 & 6 \end{pmatrix}$$

**2.** Finn den inverse matrisen dersom den finnes:

$$\text{a)} A = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 1 & 1 \\ 2 & 3 & 3 \end{pmatrix} \quad \text{b)} B = \begin{pmatrix} 1 & -2 & 3 \\ 2 & 4 & 0 \\ 4 & 16 & -6 \end{pmatrix} \quad \text{c)} C = \begin{pmatrix} 0 & 2 & 3 \\ -1 & 1 & 1 \\ 2 & 0 & 3 \end{pmatrix}$$

$$\text{d)} D = \begin{pmatrix} -2 & 3 & 1 \\ 0 & 2 & 1 \\ 1 & -1 & 2 \end{pmatrix}$$

**3.** Bruk MATLAB til å finne den inverse matrisen dersom den finnes:

$$\text{a)} A = \begin{pmatrix} 1 & -2 & 3 & -1 \\ 2 & 3 & -1 & 3 \\ 0 & -1 & 2 & -2 \\ -2 & 2 & -1 & 3 \end{pmatrix}, \text{ b)} B = \begin{pmatrix} 0.1 & 2.5 & 1.3 & 1.1 \\ 0.2 & 3.3 & 1.1 & 0.3 \\ 1.2 & -1.2 & 2.4 & -3.2 \\ -2.2 & 0.2 & -1.1 & 0.2 \end{pmatrix}$$

$$\text{c)} C = \begin{pmatrix} 1.1 & -2.3 & 4.3 & -0.05 & 1 \\ 3.4 & 0.7 & -1 & 3.2 & 4.1 \\ 3 & -1.2 & 4.2 & -3.3 & 0.2 \\ -2 & 2.3 & 3.1 & 1.3 & 2.2 \\ -2.3 & 3 & 2.8 & 1.2 & -1.1 \end{pmatrix}$$

**4.** Bruk MATLAB-kommandoene `>> x=A\b` til å løse matriseligningen  $A\mathbf{x} = \mathbf{b}$  når:

$$\text{a)} A = \begin{pmatrix} 2 & -1 & 3 \\ 0 & -1 & 2 \\ -4 & 3 & 1 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$$

b)  $A = \begin{pmatrix} 3 & 2 & -1 & 3 \\ 1 & -1 & -1 & -1 \\ 2 & -1 & 2 & -3 \\ 2 & -2 & 1 & 0 \end{pmatrix}$ ,  $\mathbf{b} = \begin{pmatrix} 4 \\ -3 \\ 0 \\ 1 \end{pmatrix}$

5. Bruk MATLAB-kommandoen  $\gg \mathbf{x} = \mathbf{b}/\mathbf{A}$  til å løse matriseligningen  $\mathbf{x}\mathbf{A} = \mathbf{b}$  når:

a)  $A = \begin{pmatrix} 2 & -1 & 3 \\ 0 & -1 & 2 \\ -4 & 3 & 1 \end{pmatrix}$ ,  $\mathbf{b} = (-1, 2, 3)$

b)  $A = \begin{pmatrix} 3 & 2 & -1 & 3 \\ 1 & -1 & -1 & -1 \\ 2 & -1 & 2 & -3 \\ 2 & -2 & 1 & 0 \end{pmatrix}$ ,  $\mathbf{b} = (4, -3, 0, 1)$  Sammenlign med svarene på oppgave 4.

6. (Eksamensoppgave fra MAT1110, 14/6 2004, litt tilpasset)

a) Finn den inverse matrisen til

$$B = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix}$$

b) Bruk resultatet i a) til å løse ligningssystemet

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

c) For hvilke verdier av  $a$  og  $b$  har ligningssystemet

$$\begin{aligned} x + 2y &= 5 \\ y + z &= 3 \\ -2y + (a+1)z &= b^2 - 10 \end{aligned}$$

henholdsvis én, ingen og uendelig mange løsninger?

7. Anta at  $A$  er en inverterbar  $n \times n$ -matrise, og at  $\mathbf{b}$  er en radvektor med  $n$  komponenter. Vis at  $\mathbf{x} = \mathbf{b}A^{-1}$  er den entydige løsningen til ligningen  $\mathbf{x}\mathbf{A} = \mathbf{b}$ .

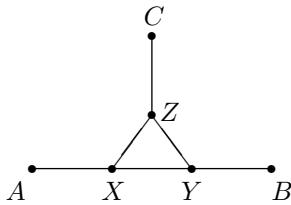
8. Anta at  $A$  er en inverterbar  $n \times n$ -matrise og at  $B$  er en inverterbar  $m \times m$ -matrise. Lag en  $(n+m) \times (n+m)$ -matrise  $C$  ved å sette inn  $A$  i øvre venstre hjørne,  $B$  i nedre høyre hjørne og så fylle ut med nuller. Symbolsk skriver vi:

$$C = \begin{pmatrix} A & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix}$$

Vis at  $C$  er inverterbar og at

$$C^{-1} = \begin{pmatrix} A^{-1} & \mathbf{0} \\ \mathbf{0} & B^{-1} \end{pmatrix}$$

**9.** Figuren nedenfor viser et elektrisk nettverk. Man kan regulere spenningen i de ytre punktene  $A$ ,  $B$  og  $C$ , men spenningen i de indre punktene  $X$ ,  $Y$  og  $Z$  er alltid gjennomsnittet av spenningen i nabopunktene.



- a) La  $a, b, c, x, y, z$  være spenningen i henholdsvis  $A, B, C, X, Y$  og  $Z$ . Vis at dersom

$$\mathbf{b} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{og} \quad \mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

så er  $A\mathbf{x} = \mathbf{b}$  der

$$A = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$

- b) Finn  $A^{-1}$ .  
c) Finn  $x, y$  og  $z$  når  $a = 1, b = 2$  og  $c = 3$ .  
d) Hvordan skal du velge de ytre spenningene  $a, b$  og  $c$  for å få  $x = 1, y = 2, z = 3$ ?

## 4.6 Lineærkombinasjoner og basiser

Anta at vi har  $n$  vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  i  $\mathbb{R}^m$ . En vektor  $\mathbf{b} \in \mathbb{R}^m$  kalles en *lineærkombinasjon* av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  dersom det finnes tall  $x_1, x_2, \dots, x_n \in \mathbb{R}$  slik at

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} \quad (4.6.1)$$

Vi kan tenke på dette som en ligning der  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  og  $\mathbf{b}$  er gitt, og der vi ønsker å finne  $x_1, x_2, \dots, x_n$ .

I lineær algebra er det svært viktig å vite når en vektor  $\mathbf{b}$  kan skrives som en lineærkombinasjon av en utgangsmengde  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ , og på hvor mange måter dette kan gjøres. Det viser seg at dette bare er en ny vri på de spørsmålene vi allerede har studert i dette kapitlet. For å se dette skriver vi vektorene på komponentform på følgende måte:

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

og

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Ligning (4.6.1) kan nå skrives

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

og trekker vi sammen venstresiden, får vi

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Det å finne en lineærkombinasjon av typen (4.6.1) er altså det samme som å løse ligningssystemet

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

Uttrykt ved hjelp av matriseligninger betyr dette at  $\mathbf{b}$  er en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  hvis og bare hvis matriseligningen  $A\mathbf{x} = \mathbf{b}$  (der matrisen  $A$  har vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  som søyler) har en løsning.

**Eksempel 1:** Skriv

$$\mathbf{b} = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$$

som en lineærkombinasjon av

$$\mathbf{a}_1 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} \quad \text{og} \quad \mathbf{a}_2 = \begin{pmatrix} 7 \\ 1 \\ 4 \end{pmatrix}$$

Vi må altså finne tall  $x_1, x_2$  (hvis mulig!) slik at

$$x_1 \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 7 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$$

Det tilsvarende ligningssystemet er

$$\begin{aligned} 3x_1 + 7x_2 &= 1 \\ -x_1 + x_2 &= 3 \\ 2x_1 + 4x_2 &= 0 \end{aligned}$$

med utvidet matrise

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

Vi radreduserer på vanlig måte

$$\begin{aligned} \begin{pmatrix} 3 & 7 & 1 \\ -1 & 1 & 3 \\ 2 & 4 & 0 \end{pmatrix} &\xrightarrow{I \leftrightarrow II} \begin{pmatrix} -1 & 1 & 3 \\ 3 & 7 & 1 \\ 2 & 4 & 0 \end{pmatrix} \xrightarrow{III+3I} \begin{pmatrix} -1 & 1 & 3 \\ 0 & 10 & 10 \\ 0 & 6 & 6 \end{pmatrix} \\ &\xrightarrow[\frac{1}{6}III]{\sim} \begin{pmatrix} -1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{III+(-1)II} \begin{pmatrix} -1 & 1 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{(-1)I} \begin{pmatrix} 1 & -1 & -3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Vi går tilbake til ligningene

$$\begin{aligned} x_1 - x_2 &= -3 \\ x_2 &= 1 \\ 0 &= 0 \end{aligned}$$

som gir  $x_2 = 1$  og  $x_1 = -3 + x_2 = -2$ .  $\square$

Vi legger merke til at vi hadde litt “flaks” i eksemplet ovenfor i og med at vi fikk  $0 = 0$  i den siste ligningen. Det gjenspeiler at en vektor i  $\mathbb{R}^3$  vanligvis ikke kan skrives som en lineærkombinasjon av to gitte vektorer — vi trenger faktisk litt flaks for å få det til! Vi skal nå se nærmere på når en vektor kan skrives som lineærkombinasjoner av andre vektorer — med eller uten “flaks”.

**Setning 4.6.1** *Anta at*

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}, \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

er vektorer i  $\mathbb{R}^m$ . For å undersøke om  $\mathbf{b}$  kan skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ , radreduserer vi matrisen

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}$$

til en trappematrise  $C$ . Da gjelder

- (i) Dersom den siste søylen i  $C$  er en pivotsøyle, er  $\mathbf{b}$  ikke en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ .

Dersom den siste søylen i  $C$  ikke er en pivotsøyle, har vi videre:

- (ii) Dersom alle de andre søylene i  $C$  er pivotsøyler, kan  $\mathbf{b}$  skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  på nøyaktig én måte.
- (iii) Dersom minst én av de andre søylene i  $C$  ikke er en pivotsøyle, kan  $\mathbf{b}$  skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  på uendelig mange måter.

*Bevis:* Dette er bare en omformulering av setning 4.2.4.  $\square$

Et viktig spørsmål er når *alle* vektorer  $\mathbf{b}$  i  $\mathbb{R}^n$  kan skrives som en lineærkombinasjon av en gitt samling vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ . Dette er bare en omformulering av setning 4.2.6.

**Setning 4.6.2** *Anta at*

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

er vektorer i  $\mathbb{R}^m$ , og at matrisen

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

kan radreduseres til trappematrisen  $C$ . Da kan enhver vektor  $\mathbf{b}$  i  $\mathbb{R}^m$  skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  hvis og bare hvis alle radene i  $C$  inneholder et pivotelement.

*Bevis:* Som allerede nevnt er dette bare en omskrivning av setning 4.2.6.  $\square$

Anta at vi har en samling vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  i  $\mathbb{R}^m$ . Med *spennet*

$$\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$$

til  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  mener vi mengden av alle vektorer  $\mathbf{b} \in \mathbb{R}^m$  som kan skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ . Setning 4.6.1 forteller oss hvordan vi kan sjekke om en spesiell vektor  $\mathbf{b}$  hører til  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ ,

mens setning 4.6.2 forteller oss hvordan vi kan sjekke om *alle* vektorer i  $\mathbb{R}^m$  hører til  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ . I det siste tilfellet er  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \mathbb{R}^m$  og vi sier at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  *utspenner hele*  $\mathbb{R}^m$ . Setningen ovenfor har en viktig konsekvens:

**Korollar 4.6.3** *Dersom  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  utspenner hele  $\mathbb{R}^m$ , er  $n \geq m$ .*

*Bevis:* La  $A$  være matrisen med  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  som søyler, og radreduser  $A$  til trappematrisen  $C$ . Dersom vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  utspenner hele  $\mathbb{R}^m$ , må  $C$  ha et pivotelement i hver rad, og det er det bare plass til om  $n \geq m$ .  $\square$

Det er ikke så vanskelig å få en viss geometrisk forståelse av korollaret i  $\mathbb{R}^3$ : Har du to vektorer  $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^3$ , vil alle lineærkombinasjoner av  $\mathbf{a}_1$  og  $\mathbf{a}_2$  ligge i planet som går gjennom punktene  $\mathbf{0}$ ,  $\mathbf{a}_1$  og  $\mathbf{a}_2$ . Du trenger en tredje vektor, som ikke ligger i dette planet, for å kunne skrive enhver vektor i  $\mathbb{R}^3$  som en lineærkombinasjon.

### Lineær uavhengighet

Vi skal være spesielt interessert i situasjoner der elementene i

$$\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$$

kan skrives som lineærkombinasjoner av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  på en *entydig* måte, dvs. at det for hver  $\mathbf{b} \in \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  finnes nøyaktig ett sett av tall  $x_1, x_2, \dots, x_n$  slik at  $\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$ .

**Definisjon 4.6.4** *Vi sier at vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$  er lineært uavhengige dersom hver  $\mathbf{b} \in \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$  kan skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  på en entydig måte. Hvis vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$  ikke er lineært uavhengige, sier vi at de er lineært avhengige.*

Det er ofte nyttig å formulere lineær uavhengighet på en annen måte:

**Setning 4.6.5** *Vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$  er lineært uavhengig hvis og bare hvis følgende betingelse er oppfylt:*

*En lineærkombinasjon  $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$  av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er bare lik  $\mathbf{0}$  dersom alle koeffisientene  $x_1, x_2, \dots, x_n$  er lik 0.*

*Bevis:* Anta først at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er linært uavhengige. Vi kan oppagt skrive  $\mathbf{0}$  som lineærkombinasjonen

$$\mathbf{0} = 0\mathbf{a}_1 + 0\mathbf{a}_2 + \dots + 0\mathbf{a}_n$$

Siden lineærkombinasjoner av lineært uavhengige vektorer er entydige, betyr dette at hvis

$$\mathbf{0} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_n\mathbf{a}_n$$

så er  $x_1 = 0, x_2 = 0, \dots, x_n = 0$ .

Anta så at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  ikke er linært uavhengige. Da må det finnes en vektor  $\mathbf{b}$  som kan skrives som en lineærkombinasjon på to forskjellige måter:

$$\mathbf{b} = y_1 \mathbf{a}_1 + y_2 \mathbf{a}_2 + \cdots + y_n \mathbf{a}_n$$

$$\mathbf{b} = z_1 \mathbf{a}_1 + z_2 \mathbf{a}_2 + \cdots + z_n \mathbf{a}_n$$

Trekker vi disse to ligningene fra hverandre, får vi

$$\mathbf{0} = (y_1 - z_1) \mathbf{a}_1 + (y_2 - z_2) \mathbf{a}_2 + \cdots + (y_n - z_n) \mathbf{a}_n$$

Velger vi  $x_1 = y_1 - z_1, x_2 = y_2 - z_2, \dots, x_n = y_n - z_n$ , må minst én av disse  $x_i$ -ene være forskjellig fra 0 (siden de to lineærkombinasjonene ovenfor er forskjellige), og vi ser dermed at

$$\mathbf{0} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n$$

er oppfylt uten at alle  $x_1, x_2, \dots, x_n$  er lik null .  $\square$

Våre gamle resultater kan nå brukes til å sjekke om vektorer er lineært uavhengige:

**Setning 4.6.6** *Anta at*

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

er vektorer i  $\mathbb{R}^m$ , og at matrisen

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

kan radreduseres til trappematrisen  $C$ . Da er  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  lineært uavhengige hvis og bare hvis alle søylene i  $C$  er pivotsøyler.

Bevis: Fra forrige setning vet vi at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er lineært uavhengige hvis og bare hvis ligningen

$$x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{0} \quad (4.6.2)$$

har en entydig løsning  $x_1 = x_2 = \dots = x_n = 0$ . Den utvidede matrisen til denne ligningen er

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & 0 \\ a_{21} & a_{22} & \dots & a_{2n} & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} & 0 \end{pmatrix}$$

Radreduserer vi denne matrisen til trappeform, vil den siste søylen fortsatt bestå av nuller, mens resten av matrisen vil være lik  $C$ . Vi ser dermed at (4.6.2) har en entydig løsning hvis og bare hvis alle søylene i  $C$  er pivotsøyler.

□

**Eksempel 2:** Vi skal undersøke om vektorene

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 3 \\ -1 \\ 4 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 2 \\ 3 \\ 0 \\ -4 \end{pmatrix} \text{ og } \mathbf{a}_3 = \begin{pmatrix} -3 \\ 3 \\ -2 \\ -1 \end{pmatrix}$$

er lineært uavhengige. Først organiserer vi vektorene som en matrise

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

Putter vi denne inn i MATLAB og kjører `rref`, ser vi at den reduserte trappeformen til  $A$  er

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

Siden denne matrisen bare har pivotsøyler, er vektorene våre lineært uavhengige.

♣

Før vi går videre, tar vi med en viktig konsekvens av setningen ovenfor.

**Korollar 4.6.7** *En lineært uavhengig mengde i  $\mathbb{R}^m$  har  $m$  eller færre elementer.*

*Bevis:* Anta at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er vektorer i  $\mathbb{R}^m$ , og la  $A$  være matrisen med disse vektorene som søyler. Reduser  $A$  til en trappematrise  $C$ . Skal vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  være lineært uavhengige, må  $C$  ha et pivotelement i hver søyle, og det er det bare plass til hvis  $n \leq m$ .

□

Anta nå at vi har en samling vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  som utspenner  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ . Vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  vil vanligvis ikke være lineært uavhengige, og for noen formål er det en stor ulempe. Det neste resultatet viser at det alltid er mulig å plukke ut *noen* av vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  slik at vi får en lineært uavhengig mengde som utspenner hele  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ .

**Setning 4.6.8** *Anta at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er en samling ikke-null vektorer i  $\mathbb{R}^m$ . Da er det mulig å finne en lineært uavhengig delmengde  $\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \dots, \mathbf{a}_{i_k}$  av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  slik at*

$$\text{Sp}(\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \dots, \mathbf{a}_{i_k}) = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$$

*Bevis:* Vi organiserer først de opprinnelige vektorene som en matrise

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

og radreduserer denne til vi får en matrise  $C$  på trappeform. Vi fjerner så de søylene i  $C$  som *ikke* er pivotsøyler, og står da igjen med en matrise  $C'$  der alle søylene er pivotsøyler. Vi går tilbake til  $A$  og fjerner de samme søylene der som vi fjernet i  $C$ . Dette gir oss en matrise  $A'$  som inneholder noen av søylene i  $A$ . Vi ser at vi kan radredusere  $A'$  til  $C'$  ved å bruke de samme radoperasjonene som reduserte  $A$  til  $C$ . Dette betyr at  $A'$  er radekvivalent med en trappematrise med bare pivotsøyler, og følgelig er søylene i  $A'$  lineært uavhengige.

For å fullføre beviset må vi vise at enhver vektor  $\mathbf{b}$  som er en lineær-kombinasjon av søylene i  $A$ , også er en lineærkombinasjon av søylene i  $A'$ . Hvis vi radreduserer den utvidede matrisen  $(A, \mathbf{b})$  ved å bruke de samme operasjonene som ovenfor, vet vi fra setning 4.6.1 at den siste soylen *ikke* er en pivotsøyle. Dersom vi isteden radreduserer  $(A', \mathbf{b})$  ved å bruke de samme operasjonene, vil heller ikke nå den siste soylen være en pivotsøyle (det skyldes at vi ikke “mister” noen pivotelementer når vi bytter ut  $A$  med  $A'$ ). Ifølge setning 4.6.1 er da  $\mathbf{b}$  en lineærkombinasjon av søylene i  $A'$ .  $\square$

Legg merke til at beviset ovenfor inneholder en metode for hvordan man finner de lineært uavhengige elementene  $\mathbf{a}_{i_1}, \mathbf{a}_{i_2}, \dots, \mathbf{a}_{i_k}$ ; man organiserer de opprinnelige vektorene som en matrise  $A$ , radreduserer denne til trappeform, og plukker ut de søylene i  $A$  som korresponderer til pivotsøyler i trappeformen.

**Eksempel 3:** La

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} -2 \\ 0 \\ 2 \\ 1 \\ 3 \end{pmatrix}, \mathbf{a}_3 = \begin{pmatrix} 3 \\ 2 \\ -3 \\ -2 \\ -3 \end{pmatrix}, \mathbf{a}_4 = \begin{pmatrix} 4 \\ -1 \\ 0 \\ 2 \\ 2 \end{pmatrix}$$

Vi skal finne en lineært uavhengig delmengde som utspenner  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4)$ . Bruker vi MATLAB, får vi

```
A=[1 -2 3 4
2 0 2 -1
-1 2 -3 0
-1 1 -2 2
0 3 -3 2];
```

```
>>C=rref(B)
```

C =

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

Vi ser at den første, andre og fjerde søylen er pivotsøyler. Mengden vi er på jakt etter er da de tilsvarende søylene i A, nemlig

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} -2 \\ 0 \\ 2 \\ 1 \\ 3 \end{pmatrix}, \mathbf{a}_4 = \begin{pmatrix} 4 \\ -1 \\ 0 \\ 2 \\ 2 \end{pmatrix}$$

□

## Basiser

*Basis* er kanskje det viktigste begrepet i lineær algebra. Basiser brukes til så mangt, men i dette kapitlet skal vi hovedsakelig benytte dem til å studere egenverdier og egenvektorer. Før vi kommer så langt, trenger vi en kort innføring i noen grunnleggende egenskaper. Vi begynner med definisjonen:

**Definisjon 4.6.9** En basis for  $\mathbb{R}^m$  er en lineært uavhengig mengde vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  som utspenner hele  $\mathbb{R}^m$ , dvs. at  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n) = \mathbb{R}^m$ .

Det er lett å se at enhver basis for  $\mathbb{R}^m$  må ha nøyaktig  $m$  elementer — korollar 4.6.3 forteller oss nemlig at en mengde som utspenner hele  $\mathbb{R}^m$  må ha minst  $m$  elementer, mens korollar 4.6.7 forteller oss at ingen lineært uavhengig mengde i  $\mathbb{R}^m$  kan ha flere enn  $m$  elementer.

Den enkleste basisen er den som består av enhetsvektorene

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \mathbf{e}_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

Dette kalles ofte *standardbasisen* i  $\mathbb{R}^m$ .

La  $A$  være matrisen med vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^m$  som søyler. Da er  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^m$  en basis for  $\mathbb{R}^m$  hvis og bare hvis matriseligningen  $A\mathbf{x} = \mathbf{b}$  har en entydig løsning for alle  $\mathbf{b}$  — vektoren

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

er nemlig en løsning av matriseligningen hvis og bare hvis

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \dots + x_m\mathbf{a}_m = \mathbf{b}$$

Det neste resultatet gjør det lett å sjekke om en samling vektorer er en basis.

**Setning 4.6.10** Anta at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  er vektorer i  $\mathbb{R}^m$ , og la  $A$  være  $m \times m$ -matrisene med disse vektorene som søyler. Da er  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  en basis for  $\mathbb{R}^m$  hvis og bare hvis  $A$  er radekvivalent med identitetsmatrisen  $I_m$ .

*Bevis:* Vi har allerede observert at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  er en basis hvis og bare hvis matriseligningen  $A\mathbf{x} = \mathbf{b}$  har en entydig løsning for alle  $\mathbf{b}$ . Ifølge setning 4.4.2 er dette ekvivalent med at  $A$  er radekvivalent med  $I_m$ .  $\square$

Setningen ovenfor har noen konsekvenser som ofte er nyttige i mer teoretisk arbeid.

**Korollar 4.6.11** Anta at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  er  $m$  vektorer i  $\mathbb{R}^m$ . Dersom vektorene enten er lineært uavhengige eller utspenner hele  $\mathbb{R}^m$ , så danner de en basis.

*Bevis:* La  $A$  være matrisen med  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  som søyler, og la  $D$  være den reduserte trappeformen til  $A$ . Dersom vektorene er lineært uavhengige, er alle søylene i  $D$  pivotsøyler. Siden  $D$  er kvadratisk, er det bare plass til dette om alle pivotelementene står på diagonalen, dvs. hvis  $D = I_m$ . Men da er  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  en basis ifølge setningen ovenfor.

Dersom vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  utspenner hele  $\mathbb{R}^m$  må alle rader i  $D$  inneholde pivotelementer. Siden  $D$  er kvadratisk er det bare plass til dette om alle pivotelementene står på diagonalen. Men dermed er  $D = I_m$  og  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  en basis ifølge setningen ovenfor.  $\square$

Det neste resultatet gjør det enda lettere å finne basiser — det forteller oss at enhver lineært uavhengig mengde kan utvides til en basis:

**Setning 4.6.12** *Anta at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er en lineært uavhengig mengde av vektorer i  $\mathbb{R}^m$ . Da finnes det vektorer  $\mathbf{a}_{n+1}, \mathbf{a}_{n+2}, \dots, \mathbf{a}_m$  slik at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{a}_{n+1}, \dots, \mathbf{a}_m$  er en basis for  $\mathbb{R}^m$ .*

*Bevis:* La  $A$  være matrisen med  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  som søyler, og radreduser  $A$  til en trappematrise  $C$ . Siden vektorene  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  er lineært uavhengige, er alle søyler i  $C$  pivotsøyler. Det betyr at pivotelementene starter øverst i venstre hjørne av matrisen og fortsetter nedover diagonalen inntil de treffer høyre kant av matrisen. Utvid  $C$  til en kvadratisk matrise  $C'$  ved å skjøte på flere søyler med pivotelementer på diagonalen. Reverser de radoperasjonene som forvandlet  $A$  til  $C$ , og bruk dem til å forvandle  $C'$  til en matrise  $A'$ . Da er  $A'$  en utvidelse av  $A$  (det er kommet til nye søyler bakerst), og søylene i  $A'$  er lineært uavhengige siden  $A'$  er radekvivalent med en matrise  $C'$  som bare har pivotsøyler.  $\square$

Beviset ovenfor blir lettere å forstå hvis vi gjennomfører prosedyren på et eksempel:

**Eksempel 4:** Vi skal utvide mengden

$$\mathbf{a}_1 = \begin{pmatrix} -1 \\ -3 \\ 2 \\ 1 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 2 \\ 1 \\ -2 \\ 3 \end{pmatrix}$$

til en basis for  $\mathbb{R}^4$ . Vi begynner med å radredusere matrisen som har  $\mathbf{a}_1$  og  $\mathbf{a}_2$  som søyler:

$$A = \begin{pmatrix} -1 & 2 \\ -3 & 1 \\ 2 & -2 \\ 1 & 3 \end{pmatrix} \xrightarrow{II+(-3)I} \begin{pmatrix} -1 & 2 \\ 0 & -5 \\ 2 & -2 \\ 1 & 3 \end{pmatrix} \xrightarrow{III+2I} \begin{pmatrix} -1 & 2 \\ 0 & -5 \\ 0 & 2 \\ 1 & 3 \end{pmatrix} \xrightarrow{IV+I}$$

$$\xrightarrow{IV+I} \begin{pmatrix} -1 & 2 \\ 0 & -5 \\ 0 & 2 \\ 0 & 5 \end{pmatrix} \xrightarrow{III+\frac{2}{5}II} \begin{pmatrix} -1 & 2 \\ 0 & -5 \\ 0 & 0 \\ 0 & 5 \end{pmatrix} \xrightarrow{IV+II} \begin{pmatrix} -1 & 2 \\ 0 & -5 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = C$$

Vi har nå fått matrisen på så god trappeform som vi trenger (for å få skikklig trappeform bør vi gjøre om “pivotelementene”  $-1$  og  $-5$  til 1’ere, men det fører bare til dobbeltarbeid i dette tilfellet). Nest skritt er å skjøte på  $C$  slik at vi får en diagonalmatrise med pivotelementer på diagonalen:

$$C' = \begin{pmatrix} -1 & 2 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Til slutt bruker vi radoperasjonene ovenfor baklengs:

$$\begin{aligned} C' &= \begin{pmatrix} -1 & 2 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{IV-II} \begin{pmatrix} -1 & 2 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{pmatrix} \\ &\xrightarrow{III-\frac{2}{5}II} \begin{pmatrix} -1 & 2 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{pmatrix} \xrightarrow{IV-I} \begin{pmatrix} -1 & 2 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} \\ &\xrightarrow{III-2I} \begin{pmatrix} -1 & 2 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 2 & -2 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} \xrightarrow{II+3I} \begin{pmatrix} -1 & 2 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 2 & -2 & 1 & 0 \\ 1 & 3 & 0 & 1 \end{pmatrix} = A' \end{aligned}$$

Dette viser at

$$\mathbf{a}_1 = \begin{pmatrix} -1 \\ -3 \\ 2 \\ 1 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 2 \\ 1 \\ -2 \\ 3 \end{pmatrix}, \mathbf{a}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \mathbf{a}_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

er en basis. ♣

**Bemerkning:** Det kan virke på eksemplet ovenfor som om den siste delen av prosessen (nemlig å gå tilbake fra  $C'$  til  $A'$ ) er unødvendig siden de siste to søylene ikke endrer seg. Dette skyldes at vi ikke har noe radombytte blant operasjonene våre — med en gang et slikt ombytte dukker opp, risikerer vi å måtte gjøre endringer i de siste to søylene. La oss legge til at det er mange andre metoder man kan bruke for å utvide en lineært uavhengig mengde til en basis, og at metoden ovenfor slett ikke er den raskeste.

### Basiser og lineæravbildninger

Det er ikke så lett å se på dette stadiet hvorfor basiser er så viktige, men vi skal prøve å antyde det. Den aller viktigste grunnen er at basiser spiller en sentral rolle når man skal studere lineæravbildninger. Husk (fra seksjon 2.8) at en lineæravbildning fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$  er en funksjon  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  som tilfredsstiller kravene:

- (i)  $\mathbf{T}(c\mathbf{x}) = c\mathbf{T}(\mathbf{x})$  for alle  $c \in \mathbb{R}$  og alle  $\mathbf{x} \in \mathbb{R}^n$
- (ii)  $\mathbf{T}(\mathbf{x} + \mathbf{y}) = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$  for alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

Ifølge setning 2.8.2 medfører disse kravene at

$$\mathbf{T}(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k) = c_1\mathbf{T}(\mathbf{x}_1) + c_2\mathbf{T}(\mathbf{x}_2) + \cdots + c_k\mathbf{T}(\mathbf{x}_k)$$

for alle  $c_1, c_2, \dots, c_k \in \mathbb{R}$  og alle  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathbb{R}^n$ .

Anta nå at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er en basis for  $\mathbb{R}^n$ , og at  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  er en lineæravbildning. Anta også at vi kjenner hvordan  $\mathbf{T}$  virker på basisvektorene  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ , la oss si at  $\mathbf{T}(\mathbf{v}_1) = \mathbf{w}_1, \mathbf{T}(\mathbf{v}_2) = \mathbf{w}_2, \dots, \mathbf{T}(\mathbf{v}_n) = \mathbf{w}_n$ . Siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er en basis, kan en hvilket som helst vektor  $\mathbf{x}$  skrives som en lineærkombinasjon

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

av  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ . Bruker vi  $\mathbf{T}$  på dette uttrykket, får vi

$$\mathbf{T}(\mathbf{x}) = \mathbf{T}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) =$$

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

Dette betyr at dersom vi vet hvordan  $\mathbf{T}$  virker på elementene i en basis, så vet vi også hvordan den virker på alle andre vektorer.

Vi kan snu problemstillingen ovenfor på hodet. Anta at vi ikke har en lineæravbildning  $\mathbf{T}$ , men at vi ønsker å definere en lineæravbildning  $\mathbf{T}$  slik at  $\mathbf{T}(\mathbf{v}_1) = \mathbf{w}_1, \mathbf{T}(\mathbf{v}_2) = \mathbf{w}_2, \dots, \mathbf{T}(\mathbf{v}_n) = \mathbf{w}_n$  (dette er en meget vanlig problemstilling i lineær algebra). Den neste setningen forteller oss at dette alltid er mulig:

**Setning 4.6.13** *Anta at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er en basis for  $\mathbb{R}^n$ , og at  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n$  er vektorer i  $\mathbb{R}^m$ . Da finnes det nøyaktig én lineæravbildning  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  slik at  $\mathbf{T}(\mathbf{v}_1) = \mathbf{w}_1, \mathbf{T}(\mathbf{v}_2) = \mathbf{w}_2, \dots, \mathbf{T}(\mathbf{v}_n) = \mathbf{w}_n$ .*

*Bevis:* Argumentet ovenfor viser at det kan være høyst én slik lineæravbildning. For å vise at det virkelig finnes en, bruker vi at siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er en basis, kan enhver vektor  $\mathbf{x} \in \mathbb{R}^n$  skrives som en lineærkombinasjon

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

på nøyaktig én måte. Vi kan derfor *definere* en funksjon  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  ved

$$\mathbf{T}(\mathbf{x}) = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n$$

Siden vi åpenbart har  $\mathbf{T}(\mathbf{v}_1) = \mathbf{w}_1$ ,  $\mathbf{T}(\mathbf{v}_2) = \mathbf{w}_2$ , ...,  $\mathbf{T}(\mathbf{v}_n) = \mathbf{w}_n$ , er det nok å vise at  $\mathbf{T}$  er en lineæravbildning, dvs. at den tilfredsstiller betingelsene (i) og (ii) ovenfor.

For å vise (i), observerer vi at hvis

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

så er

$$c\mathbf{x} = cx_1\mathbf{v}_1 + cx_2\mathbf{v}_2 + \cdots + cx_n\mathbf{v}_n$$

Dermed er

$$\mathbf{T}(c\mathbf{x}) = cx_1\mathbf{w}_1 + cx_2\mathbf{w}_2 + \cdots + cx_n\mathbf{w}_n$$

På den annen side er

$$\begin{aligned} c\mathbf{T}(\mathbf{x}) &= c(x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n) = \\ &= cx_1\mathbf{w}_1 + cx_2\mathbf{w}_2 + \cdots + cx_n\mathbf{w}_n \end{aligned}$$

Dette viser at  $\mathbf{T}(c\mathbf{x}) = c\mathbf{T}(\mathbf{x})$ , så betingelse (i) er oppfylt.

For å vise at betingelse (ii) er oppfylt, begynner vi med to vektorer

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

$$\mathbf{y} = y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n$$

Da er

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n) + (y_1\mathbf{v}_1 + y_2\mathbf{v}_2 + \cdots + y_n\mathbf{v}_n) = \\ &= (x_1 + y_1)\mathbf{v}_1 + (x_2 + y_2)\mathbf{v}_2 + \cdots + (x_n + y_n)\mathbf{v}_n \end{aligned}$$

og følgelig

$$\mathbf{T}(\mathbf{x} + \mathbf{y}) = (x_1 + y_1)\mathbf{w}_1 + (x_2 + y_2)\mathbf{w}_2 + \cdots + (x_n + y_n)\mathbf{w}_n$$

På den annen side er

$$\mathbf{T}(\mathbf{x}) = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n$$

og

$$\mathbf{T}(\mathbf{y}) = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_n\mathbf{w}_n$$

Det betyr at  $\mathbf{T}(\mathbf{x} + \mathbf{y}) = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$ , og følgelig er betingelse (ii) oppfylt.  $\square$

Man kan lure på hvorfor det er bruk for andre basiser i  $\mathbb{R}^n$  enn standardbasisen  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ . Svaret er ganske enkelt at i mange eksempler gir

andre basiser enklere regninger og mer informative svar — spesielt gjelder dette basiser som består av egenvektorer for den lineæravbildningen vi studerer. Husk at en *egenvektor* for en lineæravbildning  $\mathbf{T}$  er en vektor  $\mathbf{v} \neq \mathbf{0}$  slik at  $\mathbf{T}(\mathbf{v}) = \lambda\mathbf{v}$  for et eller annet tall  $\lambda$  ( $\lambda$  kalles *egenverdien* til  $\mathbf{v}$ ). Egenvektorer er spesielt nyttige når vi skal studere gjentatt bruk av avbildningen  $\mathbf{T}$  siden  $\mathbf{T}^n(\mathbf{v}) = \lambda^n\mathbf{v}$ .

Anta nå at vi har en basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  av egenvektorer for  $\mathbf{T}$  med egenverdier  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Enhver vektor  $\mathbf{x}$  kan skrives som en lineærkombinasjon

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_n\mathbf{v}_n$$

Bruker vi  $\mathbf{T}^n$  på dette uttrykket, får vi

$$\mathbf{T}^n(\mathbf{x}) = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2 + \dots + c_n\lambda_n^n\mathbf{v}_n$$

Dersom en av egenverdiene (la oss si  $\lambda_1$ ) er større enn de andre i tallverdi, vil ledet den tilhører dominere over de andre når  $n$  blir stor. Det betyr at størrelsesordenen til  $\mathbf{T}^n(\mathbf{x})$  vokser som  $\lambda_1^n$ , og at fordelingen mellom komponentene i vektoren  $\mathbf{T}^n(\mathbf{x})$  blir mer og mer som fordelingen mellom komponentene i  $\mathbf{v}_1$ . Som vi skal se senere, er informasjon av denne typen uhyre viktige når man skal studere systemer som utvikler seg med tiden.

Det viser seg at de fleste (men ikke alle!) lineæravbildninger har en basis av egenvektorer. Senere i dette kapitlet skal vi lære mer om hvordan vi kan finne egenverdier og egenvektorer.

### Oppgaver til seksjon 4.6

**1.** Skriv  $\mathbf{b}$  som en lineærkombinasjon av  $\mathbf{a}_1$  og  $\mathbf{a}_2$  når  $\mathbf{a}_1 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ ,  $\mathbf{a}_2 = \begin{pmatrix} -4 \\ 2 \end{pmatrix}$  og  $\mathbf{b} = \begin{pmatrix} -3 \\ 3 \end{pmatrix}$ .

**2.** Skriv  $\mathbf{b}$  som en lineærkombinasjon av  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  og  $\mathbf{a}_3$  når  $\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ ,  $\mathbf{a}_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ ,  $\mathbf{a}_3 = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}$  og  $\mathbf{b} = \begin{pmatrix} -2 \\ 5 \\ 1 \end{pmatrix}$ .

**3.** Avgjør om alle vektorer i  $\mathbb{R}^n$  (for relevant  $n$ ) kan skrives som en lineærkombinasjon av vektorene:

a)  $\mathbf{a}_1 = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ ,  $\mathbf{a}_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

b)  $\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ ,  $\mathbf{a}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ ,  $\mathbf{a}_3 = \begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix}$

c)  $\mathbf{a}_1 = \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}$ ,  $\mathbf{a}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ ,  $\mathbf{a}_3 = \begin{pmatrix} -6 \\ 0 \\ 7 \end{pmatrix}$ .

4. Skriv  $\begin{pmatrix} 7 \\ 4 \\ -3 \\ 1 \end{pmatrix}$  som en lineærkombinasjon av  $\begin{pmatrix} 1 \\ 0 \\ -1 \\ 2 \end{pmatrix}$ ,  $\begin{pmatrix} 2 \\ 1 \\ 3 \\ -4 \end{pmatrix}$ ,  $\begin{pmatrix} 5 \\ -7 \\ 6 \\ 3 \end{pmatrix}$ ,  $\begin{pmatrix} 2 \\ -1 \\ 0 \\ 3 \end{pmatrix}$ . Bruk gjerne MATLAB som hjelphemiddel.

5. Bruk MATLAB til å skrive  $\begin{pmatrix} 1 \\ .25 \\ 0 \\ 3 \end{pmatrix}$  som en lineærkombinasjon av  $\begin{pmatrix} 0.75 \\ 0.3 \\ -1 \\ 2 \end{pmatrix}$ ,  $\begin{pmatrix} 4 \\ 0.1 \\ -2 \\ 0.25 \end{pmatrix}$ ,  $\begin{pmatrix} -1 \\ 3 \\ -4 \\ 0.2 \end{pmatrix}$ ,  $\begin{pmatrix} 4 \\ -1 \\ 3 \\ 2 \end{pmatrix}$ .

6. Bruk MATLAB til å sjekke om enhver vektor i  $\mathbb{R}^4$  kan skrives som en lineærkombinasjon av  $\begin{pmatrix} 1 \\ -2 \\ 3 \\ 2 \end{pmatrix}$ ,  $\begin{pmatrix} 0 \\ -3 \\ 4 \\ 1 \end{pmatrix}$ ,  $\begin{pmatrix} 2 \\ -7 \\ 10 \\ 5 \end{pmatrix}$ ,  $\begin{pmatrix} -1 \\ 3 \\ -1 \\ 2 \end{pmatrix}$ ,  $\begin{pmatrix} 0 \\ 2 \\ 1 \\ 0 \end{pmatrix}$ .

7. Avgjør om vektorene er lineært uavhengige:

a)  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

b)  $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

c)  $\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 9 \end{pmatrix}$

d)  $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 5 \end{pmatrix}$

e)  $\begin{pmatrix} 1 \\ -4 \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ -3 \\ 2 \end{pmatrix}$

8. Finn en lineært uavhengig delmengde:

a)  $\begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}$

b)  $\begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 8 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$

c)  $\begin{pmatrix} 4 \\ 1 \\ -3 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 3 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ -5 \\ 2 \end{pmatrix}$

**9.** Avgjør om mengden er en basis for det relevante rommet  $\mathbb{R}^n$  (i noen av tilfellene lønner det seg å tenke før man regner!)

a)  $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

b)  $\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}$

c)  $\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \end{pmatrix}$

d)  $\begin{pmatrix} -1 \\ 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix}$

**10.** Utvid mengden av vektorer til en basis for det relevante rommet  $\mathbb{R}^n$ :

a)  $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 3 \end{pmatrix}$

b)  $\begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$

**11.** I denne oppgaven er  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ .

a) Vis at  $\mathbf{v}_1$  og  $\mathbf{v}_2$  danner en basis for  $\mathbb{R}^2$ .

b) Skriv  $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  og  $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  som lineærkombinasjoner av  $\mathbf{v}_1$  og  $\mathbf{v}_2$ .

c) Forklar hvorfor det finnes nøyaktig én lineæravbildning  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  slik at  $\mathbf{T}(\mathbf{v}_1) = 2\mathbf{v}_1$  og  $\mathbf{T}(\mathbf{v}_2) = -\mathbf{v}_2$ .

d) Finn matrisen  $A$  til lineæravbildningen  $\mathbf{T}$  (dvs.  $2 \times 2$ -matrisen  $A$  slik at  $\mathbf{T}(\mathbf{x}) = A\mathbf{x}$  for alle  $\mathbf{x} \in \mathbb{R}^2$ )

**12.** Anta at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er en mengde av ikke-null vektorer som står normalt på hverandre. Vis at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er lineært uavhengige. (*Hint:* Anta at  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k = \mathbf{0}$  og ta skalarproduktet med  $\mathbf{v}_i$  på begge sider.)

## 4.7 Underrom

I denne seksjonen skal vi utvide, generalisere og systematisere teorien fra forrige seksjon. Resultatene vi skal komme frem til, er svært viktige i videregående lineær algebra, men ved første møte kan de virke vanskelige og uvante. Det kan derfor være en trøst å vite at med unntak av seksjon 4.12 (som hovedsakelig er for spesielt interesserte), har vi ikke bruk for disse resultatene i resten av kapitlet.

Det sentrale begrepet i denne seksjonen er “underrom”. Litt kjapt og upresist kan vi si at underrom er delmengder av  $\mathbb{R}^n$  som oppfører seg som  $\mathbb{R}^m$  for en  $m < n$ . Et enkelt eksempel er mengden av alle vektorer med formen  $(x_1, x_2, \dots, x_m, 0, \dots, 0)$  i  $\mathbb{R}^n$  — denne mengden er åpenbart en kopi av  $\mathbb{R}^m$  inni  $\mathbb{R}^n$ . Vi skal se at det faktisk er mange slike kopier av  $\mathbb{R}^m$  inni  $\mathbb{R}^n$ .

Den grunnleggende definisjonen kan virke ganske abstrakt, men som vi snart skal se, er disse objektene ikke så merkelige som de kan se ut til ved første øyekast.

**Definisjon 4.7.1** En ikke-tom mengde  $H$  av vektorer i  $\mathbb{R}^n$  kalles et underrom av  $\mathbb{R}^n$  dersom følgende betingelser er oppfylt.

- a) Dersom  $\mathbf{u} \in H$ , så er  $c\mathbf{u} \in H$  for alle tall  $c \in \mathbb{R}$ .
- b) Dersom  $\mathbf{u}, \mathbf{v} \in H$ , så er  $\mathbf{u} + \mathbf{v} \in H$ .

**Bemerkning:** Vi sier at  $H$  er *lukket under addisjon* (betingelse b)) og *multiplikasjon med skalar* (betingelse a)).

Det minste underrommet til  $\mathbb{R}^n$  er  $H = \{\mathbf{0}\}$  (mengden som bare består av nullvektoren  $\mathbf{0}$ ) og det største er  $H = \mathbb{R}^n$ . De viktigste eksemplene er de som ligger mellom disse ytterpunktene. Her er en klasse underrom vi kjenner fra før:

**Eksempel 1:** Husk (fra forrige seksjon) at dersom  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  er vektorer i  $\mathbb{R}^n$ , så består *spennet*  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$  av alle lineærkombinasjoner

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m$$

Vi skal vise at  $H = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$  er et underrom av  $\mathbb{R}^n$ .

For å sjekke at betingelse a) er oppfylt, må vi vise at dersom  $\mathbf{u} \in H$  og  $c \in \mathbb{R}$ , så er  $c\mathbf{u} \in H$ . Siden  $\mathbf{u} \in H$ , er  $\mathbf{u}$  en lineærkombinasjon

$$\mathbf{u} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m$$

Dermed er

$$c\mathbf{u} = cc_1\mathbf{a}_1 + cc_2\mathbf{a}_2 + \cdots + cc_m\mathbf{a}_m$$

som er en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ , og følgelig er  $c\mathbf{u} \in H$ .

For å sjekke at betingelse b) er oppfylt, må vi vise at dersom  $\mathbf{u}, \mathbf{v} \in H$ , så er  $\mathbf{u} + \mathbf{v} \in H$ . Siden  $\mathbf{u} \in H$ , er  $\mathbf{u}$  en lineærkombinasjon

$$\mathbf{u} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m$$

og siden  $\mathbf{v} \in H$ , er  $\mathbf{v}$  en lineærkombinasjon

$$\mathbf{v} = d_1\mathbf{a}_1 + d_2\mathbf{a}_2 + \cdots + d_m\mathbf{a}_m$$

Dermed er

$$\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{a}_1 + (c_2 + d_2)\mathbf{a}_2 + \cdots + (c_m + d_m)\mathbf{a}_m$$

en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ , og følgelig er  $\mathbf{u} + \mathbf{v} \in H$ . 

Vi skal snart se at *alle* underrom av  $H$  (unntatt det trivielle underrommet  $\{\mathbf{0}\}$ ) er på formen  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$  for et passende valg av  $\mathbf{a}, \mathbf{a}_2, \dots, \mathbf{a}_m$ . I utgangspunktet kan de imidlertid se ganske annerledes ut:

**Eksempel 2:** La  $\mathbf{a} \in \mathbb{R}^n$ . Vi skal vise at

$$H = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{a} = 0\}$$

er et underrom av  $\mathbb{R}^n$ .  $H$  består altså av alle vektorene som står ortogonalt (normalt) på  $\mathbf{a}$ . Dersom  $\mathbf{a} \neq \mathbf{0}$ , kalles  $H$  *det ortogonale komplementet til  $\mathbf{a}$*  (dersom  $\mathbf{a} = \mathbf{0}$ , er  $H = \mathbb{R}^n$ ).

For å vise at  $H$  er et underrom av  $\mathbb{R}^n$ , må vi sjekke at betingelse a) og b) i definisjon 4.7.1 er oppfylt. For å vise at a) er oppfylt antar vi at  $\mathbf{u} \in H$  og at  $c \in \mathbb{R}$ , og må vise at  $c\mathbf{u} \in H$ . Siden  $\mathbf{u} \in H$ , er  $\mathbf{u} \cdot \mathbf{a} = 0$ . Dermed er  $(c\mathbf{u}) \cdot \mathbf{a} = c(\mathbf{u} \cdot \mathbf{a}) = c \cdot 0 = 0$ , og følgelig er  $c\mathbf{u} \in H$ .

For å vise at b) er oppfylt, antar vi at  $\mathbf{u}, \mathbf{v} \in H$ , og må vise at  $\mathbf{u} + \mathbf{v} \in H$ . Siden  $\mathbf{u}, \mathbf{v} \in H$ , er  $\mathbf{u} \cdot \mathbf{a} = 0$  og  $\mathbf{v} \cdot \mathbf{a} = 0$ . Dermed er  $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{a} = \mathbf{u} \cdot \mathbf{a} + \mathbf{v} \cdot \mathbf{a} = 0 + 0 = 0$ , og følgelig er  $\mathbf{u} + \mathbf{v} \in H$ . 

Vi trenger en liten hjelpesetning før vi går videre:

**Lemma 4.7.2** *Anta at  $H$  er et underrom av  $\mathbb{R}^n$  og at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in H$ . Da er*

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m \in H$$

*for alle*  $c_1, c_2, \dots, c_m \in \mathbb{R}$

*Bevis:* Ifølge punkt a) i definisjon 4.7.1 er  $c_1\mathbf{a}_1, c_2\mathbf{a}_2, \dots, c_m\mathbf{a}_m \in H$ . Ifølge punkt b) er dermed  $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 \in H$ . Bruker vi punkt b) en gang til, ser vi at  $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + c_3\mathbf{a}_3 = (c_1\mathbf{a}_1 + c_2\mathbf{a}_2) + c_3\mathbf{a}_3 \in H$ . Fortsetter vi på denne måten, får vi til slutt at  $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m \in H$  (bruk gjerne induksjon til å føre et formelt bevis!) 

Her kommer resultatet vi annonserte tidligere:

**Setning 4.7.3** Anta at  $H \neq \{\mathbf{0}\}$  er et underrom av  $\mathbb{R}^n$ . Da finnes det en lineært uavhengig mengde vektorer  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in \mathbb{R}^n$  slik at  $H = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$ .

*Bevis:* La  $m$  være det største antall lineært uavhengige vektorer man kan finne i  $H$ , og la  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m \in H$  være en maksimal mengde av lineært uavhengige elementer (vi vet at  $m \leq n$  siden det ikke er mulig å finne mer enn  $n$  lineært uavhengige elementer i hele  $\mathbb{R}^n$ ). Vi skal vise at  $H = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$ .

Fra lemmaet ovenfor vet vi at  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m) \subset H$ , så alt vi behøver å vise er at enhver  $\mathbf{b} \in H$  faktisk ligger i  $\text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$ , dvs. at  $\mathbf{b}$  kan skrives som en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ .

Dette er ikke så vanskelig: Siden  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  er en maksimal lineært uavhengig mengde, vet vi at  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m, \mathbf{b}$  er lineært avhengige, og følgelig finnes det tall  $c_1, c_2, \dots, c_n, d$  som ikke alle er 0, slik at

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_m\mathbf{a}_m + d\mathbf{b} = \mathbf{0}$$

Vi ser at  $d$  ikke kan være 0 for da hadde vi hatt en lineær avhengighet mellom  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ . Men dermed kan vi dele på  $d$  og få

$$\mathbf{b} = (-d^{-1}c_1)\mathbf{a}_1 + (-d^{-1}c_2)\mathbf{a}_2 + \cdots + (-d^{-1}c_m)\mathbf{a}_m$$

Dette viser at  $\mathbf{b}$  er en lineærkombinasjon av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ .  $\square$

Vi kan nå generalisere basisbegrepet til underrom:

**Definisjon 4.7.4** Anta at  $H$  er et underrom av  $\mathbb{R}^n$ . En lineært uavhengig mengde  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  slik at  $H = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$  kalles en basis for  $H$ .

Setningen ovenfor viser at et underrom alltid har en basis. Den neste setningen forteller oss at alle basiser for  $H$  har like mange elementer.

**Setning 4.7.5** Anta  $H$  er et underrom av  $\mathbb{R}^n$ . Da har alle basiser for  $H$  like mange elementer.

*Bevis:* Anta for motsigelse at det finnes to basiser  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$  og  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$  for  $H$  med et ulikt antall elementer, la oss si  $k > m$ . Siden  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$  ligger i  $H = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)$ , kan  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$  skrives som lineærkombinasjoner av  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ . Det finnes altså tall  $c_{ij}$  slik at

$$\begin{aligned} \mathbf{b}_1 &= c_{11}\mathbf{a}_1 + c_{12}\mathbf{a}_2 + \cdots + c_{1m}\mathbf{a}_m \\ \mathbf{b}_2 &= c_{21}\mathbf{a}_1 + c_{22}\mathbf{a}_2 + \cdots + c_{2m}\mathbf{a}_m \\ &\vdots && \vdots && \vdots \\ \mathbf{b}_k &= c_{k1}\mathbf{a}_1 + c_{k2}\mathbf{a}_2 + \cdots + c_{km}\mathbf{a}_m \end{aligned}$$

Vektorene

$$\mathbf{c}_1 = \begin{pmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1m} \end{pmatrix}, \mathbf{c}_2 = \begin{pmatrix} c_{21} \\ c_{22} \\ \vdots \\ c_{2m} \end{pmatrix}, \dots, \mathbf{c}_k = \begin{pmatrix} c_{k1} \\ c_{k2} \\ \vdots \\ c_{km} \end{pmatrix}$$

må være lineært avhengige siden  $k > m$ . Det finnes altså tall  $x_1, x_2, \dots, x_k$  som ikke alle er 0, slik at

$$x_1\mathbf{c}_1 + x_2\mathbf{c}_2 + \dots + x_k\mathbf{c}_k = \mathbf{0}$$

Skriver vi ut dette uttrykket, ser vi at

$$\begin{pmatrix} x_1c_{11} + x_2c_{21} + \dots + x_kc_{k1} \\ x_1c_{12} + x_2c_{22} + \dots + x_kc_{k2} \\ \vdots \\ x_1c_{1m} + x_2c_{2m} + \dots + x_kc_{km} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Men dermed er

$$\begin{aligned} & x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \dots + x_k\mathbf{b}_k = \\ &= x_1(c_{11}\mathbf{a}_1 + c_{12}\mathbf{a}_2 + \dots + c_{1m}\mathbf{a}_m) \\ &\quad + x_2(c_{21}\mathbf{a}_1 + c_{22}\mathbf{a}_2 + \dots + c_{2m}\mathbf{a}_m) \\ &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ &\quad + x_k(c_{k1}\mathbf{a}_1 + c_{k2}\mathbf{a}_2 + \dots + c_{km}\mathbf{a}_m) = \\ &= (x_1c_{11} + x_2c_{21} + \dots + x_kc_{k1})\mathbf{a}_1 + \\ &\quad + (x_1c_{12} + x_2c_{22} + \dots + x_kc_{k2})\mathbf{a}_2 + \\ &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ &\quad + (x_1c_{1m} + x_2c_{2m} + \dots + x_kc_{km})\mathbf{a}_m = \\ &= 0\mathbf{a}_1 + 0\mathbf{a}_2 + \dots + 0\mathbf{a}_m = \mathbf{0} \end{aligned}$$

Dette er en selvmotsigelse siden  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$  er lineært uavhengige. Eneste mulige konklusjon er at alle basiser for  $H$  har like mange elementer.  $\square$

Vi vet fra før at alle basiser for det  $n$ -dimensjonale rommet  $\mathbb{R}^n$  har nøyaktig  $n$  elementer. Det er derfor naturlig å utvide dimensjonsbegrepet til underrom på denne måten:

**Definisjon 4.7.6** La  $H \neq \{0\}$  være et underrom av  $\mathbb{R}^n$ . Med dimensjonen til  $H$  mener vi antall elementer i en basis for  $H$ . Dimensjonen til  $H$  betegnes med  $\dim(H)$ .

### Rangen til en matrise

Det er to typer underrom som dukker opp naturlig når vi studerer en  $m \times n$ -matrise  $A$ . *Søylerommet* til  $A$  er underrommet av  $\mathbb{R}^m$  utspent av søylene i  $A$ , men *radrommet* til  $A$  er underrommet av  $\mathbb{R}^n$  utspent av radene i  $A$ . Vi skal nå studere dimensjonen til disse rommene. Det viser seg at vi allerede har full kontroll over søylerommet.

**Setning 4.7.7** *Anta at  $A$  er en  $m \times n$ -matrise. Da er dimensjonen til søylerommet til  $A$  lik antall pivotsøyler i trappeformen til  $A$ .*

*Bevis:* Dette følger fra beviset for setning 4.6.8 (se også teksten rett etter slutten av beviset).  $\square$

For å bestemme dimensjonen til radrommet trenger vi en liten observasjon.

**Lemma 4.7.8** *Anta at  $A'$  fremkommer når vi gjør en radoperasjon på  $A$ . Da har  $A$  og  $A'$  samme radrom.*

*Bevis:* Det er lett å se at dersom vi bytter om to rader eller ganger en rad med et tall  $s \neq 0$ , så endrer ikke radrommet seg. Vi ser derfor på tilfellet der  $A'$  fremkommer fra  $A$  ved at vi bytter ut den  $i$ -te raden  $\mathbf{a}_i$  med  $\mathbf{a}_i + s\mathbf{a}_j$  der  $\mathbf{a}_j$  er den  $j$ -te raden. Anta først at  $\mathbf{b}$  er i radrommet til  $A$ . Da finnes det tall  $c_1, \dots, c_i, \dots, c_j, \dots, c_m$  slik at

$$\mathbf{b} = c_1\mathbf{a}_1 + \dots + c_i\mathbf{a}_i + \dots + c_j\mathbf{a}_j + \dots + c_m\mathbf{a}_m$$

Men da er

$$\mathbf{b} = c_1\mathbf{a}_1 + \dots + c_i(\mathbf{a}_i + s\mathbf{a}_j) + \dots + (c_j - c_is)\mathbf{a}_j + \dots + c_m\mathbf{a}_m$$

som viser at  $\mathbf{b}$  er i radrommet til  $A'$ . Anta så omvendt at  $\mathbf{b}$  ligger i radrommet til  $A'$ , dvs. at

$$\mathbf{b} = d_1\mathbf{a}_1 + \dots + d_i(\mathbf{a}_i + s\mathbf{a}_j) + \dots + d_j\mathbf{a}_j + \dots + d_m\mathbf{a}_m$$

for passende tall  $d_1, \dots, d_m$ . Da er

$$\mathbf{b} = d_1\mathbf{a}_1 + \dots + d_i\mathbf{a}_i + \dots + (d_j + d_is)\mathbf{a}_j + \dots + d_m\mathbf{a}_m$$

som viser at  $\mathbf{b}$  er i radrommet til  $A$ .  $\square$

Vi kan nå bevise et nyttig teorem — det sier at i en matrise kan vi alltid finne like mange lineært uavhengige rader som vi kan finne lineært uavhengige søyler. Vi skal få bruk for dette resultatet når vi studerer Lagrange-multiplikatorer i neste kapittel.

**Teorem 4.7.9 (Rangteoremet)** *Dimensjonen til søylerommet til en matrise er lik dimensjonen til radrommet.*

*Bevis:* Vi har allerede sett at dimensjonen til søylerommet er lik antall pivotsøyler (la oss si dette er  $k$ ), så vi må vise at radrommet også har dimensjon  $k$ . La  $A'$  være den reduserte trappeformen til  $A$ . Siden  $A'$  fremkommer fra  $A$  ved en sekvens av radoperasjoner, vet vi fra lemmaet ovenfor at  $A$  og  $A'$  har samme radrom, så alt vi trenger å vise er at radrommet til  $A'$  har dimensjon  $k$ . Siden det et et  $k$  pivotelementer, har  $A'$  bare  $k$  ikke-null rader, så det holder å vise at disse er lineært uavhengige. Dette er enkelt.: Siden  $A'$  er på redusert trappeform, er pivotelementene de eneste ikke-null elementene i sin søyle. Dette betyr at hvis en lineærkombinasjon av “pivotradene” skal være null, må hver koeffisient være null.  $\square$

Med *rangen* til en matrise mener vi dimensjonen til søylerommet (eller, om man vil, dimensjonen til radrommet). For å finne rangen, skriver vi matrisen på trappeform og teller opp antall pivotelementer. Er matrisen stor, er dette mye arbeid, og det kan være lurt å få MATLAB til å gjøre jobben for oss. Har du lastet inn en matrise  $A$ , vil kommandoen

```
>> rank(A)
```

gi deg rangen til matrisen.

### Ortonormale basiser

Standardbasisen  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  i  $\mathbb{R}^n$  er ikke bare en basis, den er også en *ortonormal mengde*, dvs. at elementene  $\mathbf{e}_i$  har lengde én og står normalt på hverandre. Ortonormale basiser har mange fordeler, og vi skal nå vise at alle underrom har slike basiser. Metoden vi skal bruke til å skaffe oss en ortonormal basis, kalles *Gram-Schmidt-ortogonalisering* og er nyttig i mange sammenhenger.

**Setning 4.7.10** *Ethvert underrom  $H \neq \{\mathbf{0}\}$  av  $\mathbb{R}^n$  har en ortonormal basis, dvs. en basis  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$  slik at*

$$\mathbf{w}_i \cdot \mathbf{w}_j = \begin{cases} 0 & \text{hvis } i \neq j \\ 1 & \text{hvis } i = j \end{cases}$$

*Bevis:* La  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  være en basis for  $H$ . Vi skal omdanne denne basisen til en ortonormal basis  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ . Vi gjør dette trinnvis, og sørger hele tiden for at  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$  er en ortonormal mengde som utspenner det samme underrommet som  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ .

Prosedyren starter med at vi setter  $\mathbf{w}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}$ . Vi normaliserer altså  $\mathbf{v}_1$  slik at den får lengde 1. Det er klart at  $\mathbf{v}_1$  og  $\mathbf{w}_1$  utspenner det samme underrommet.

Nest skritt i prosedyren er å finne  $\mathbf{w}_2$ . Vi definerer først en vektor  $\mathbf{w}'_2$  ved

$$\mathbf{w}'_2 = \mathbf{v}_2 - (\mathbf{w}_1 \cdot \mathbf{v}_2)\mathbf{w}_1$$

Tar vi skalarproduktet med  $\mathbf{w}_1$  på begge sider, ser vi at

$$\mathbf{w}_1 \cdot \mathbf{w}'_2 = \mathbf{w}_1 \cdot \mathbf{v}_2 - (\mathbf{w}_1 \cdot \mathbf{v}_2)(\mathbf{w}_1 \cdot \mathbf{w}_1) = 0$$

siden  $\mathbf{w}_1 \cdot \mathbf{w}_1 = |\mathbf{w}_1|^2 = 1$ . Dette viser at  $\mathbf{w}'_2$  og  $\mathbf{w}_1$  er ortogonale. Siden  $\mathbf{w}'_2$  kan uttrykkes ved hjelp av  $\mathbf{v}_2$  og  $\mathbf{w}_1$  (og  $\mathbf{v}_2$  kan uttrykkes ved hjelp av  $\mathbf{w}'_2$  og  $\mathbf{w}_1$ ), er det lett å se at  $\mathbf{w}_1$  og  $\mathbf{w}'_2$  utspenner samme underrom som  $\mathbf{v}_1$  og  $\mathbf{v}_2$ . Vi avslutter dette trinnet med å sette  $\mathbf{w}_2 = \frac{\mathbf{w}'_2}{|\mathbf{w}'_2|}$  for å få en vektor med lengde én (dette ødelegger ikke de egenskapene vi allerede har sjekket).

Vi beskriver nå det generelle skrittet i prosedyren. Anta at vi har greid å finne en ortonormal mengde  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$  som utspenner samme underrom som  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ . Vi definerer først vektoren  $\mathbf{w}'_{k+1}$  ved:

$$\mathbf{w}'_{k+1} = \mathbf{v}_{k+1} - (\mathbf{w}_1 \cdot \mathbf{v}_{k+1})\mathbf{w}_1 - (\mathbf{w}_2 \cdot \mathbf{v}_{k+1})\mathbf{w}_2 - \cdots - (\mathbf{w}_k \cdot \mathbf{v}_{k+1})\mathbf{w}_k \quad (4.7.1)$$

Tar vi skalarproduktet med  $\mathbf{w}_i$  (der  $i \leq k$ ) på begge sider, får vi

$$\mathbf{w}_i \cdot \mathbf{w}'_{k+1} = \mathbf{w}_i \cdot \mathbf{v}_{k+1} - (\mathbf{w}_i \cdot \mathbf{v}_{k+1}) \cdot (\mathbf{w}_i \cdot \mathbf{w}_i) = 0$$

der vi har brukt at  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$  er ortonormale. Dette betyr at  $\mathbf{w}_{k+1}$  står normalt på  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$  som allerede står normalt på hverandre. Ved hjelp av ligning (4.7.1) er det lett å se at  $\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{w}'_{k+1}$  utspenner samme mengde som  $\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}_{k+1}$ , som igjen utspenner samme mengde som  $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{v}_{k+1}$ . Til slutt setter vi  $\mathbf{w}_{k+1} = \frac{\mathbf{w}'_{k+1}}{|\mathbf{w}'_{k+1}|}$  for å få en vektor med lengde 1 (hvordan vet du at  $|\mathbf{w}'_{k+1}|$  ikke er lik 0?). Dermed har vi fått en ortonormal mengde  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{k+1}$  som utspenner samme underrom som  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}$ .

Fortsetter vi denne prosedyren helt til det ikke er flere  $\mathbf{v}$ 'er igjen, får vi en ortonormal basis  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$  for  $H$ .  $\square$

Legg merke til at beviset ovenfor inneholder en oppskrift vi kan følge for å finne en ortonormal basis. Det er imidlertid ikke morsomt å bruke denne oppskriften for hånd, og heldigvis har MATLAB en kommando som gjør arbeidet for oss. Har du lastet inn en matrise  $A$ , så vil kommandoen

```
>> orth(A)
```

returnere en ortonormal basis for soylerommet til  $A$ .

En av grunnene til at ortonormale basiser er så populære, er at det er svært enkelt å finne ut hvordan man skriver en vektor som en lineær kombinasjon av en slik basis.

**Setning 4.7.11** Anta at  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$  er en ortonormal basis for et underrom  $H$  av  $\mathbb{R}^n$ . Hvis  $\mathbf{v} \in H$ , så er

$$\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_m \mathbf{v}_m$$

der  $c_i = \mathbf{v} \cdot \mathbf{v}_i$

*Bevis:* Siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  er en basis for  $H$ , vet vi at det finnes tall  $c_1, c_2, \dots, c_m$  slik at

$$\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_m \mathbf{v}_m$$

Tar vi skalarproduktet med  $\mathbf{v}_i$  på begge sider og bruker at

$$\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{hvis } i \neq j \\ 1 & \text{hvis } i = j \end{cases},$$

får vi  $\mathbf{v} \cdot \mathbf{v}_i = c_i$

□

### Oppgaver til seksjon 4.7

1. Finn en basis for underrommet  $H$  når

- a)  $H = \text{Sp}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$  der  $\mathbf{a}_1 = (1, -2, 1)$ ,  $\mathbf{a}_2 = (2, -1, 3)$ ,  $\mathbf{a}_3 = (-1, -1, -2)$ .
- b)  $H = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} \cdot \mathbf{j} = 0\}$  (husk at  $\mathbf{j} = (0, 0, 1)$ ).
- c)  $H = \{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{x} \cdot \mathbf{a} = 0\}$  der  $\mathbf{a} = (1, -2, 1)$ .
- d)  $H$  er søylerommet til matrisen  $A = \begin{pmatrix} 3 & -1 \\ 2 & 0 \\ 1 & 2 \end{pmatrix}$ .
- e)  $H$  er radrommet til matrisen  $A = \begin{pmatrix} 3 & -1 \\ 2 & 0 \\ 1 & 2 \end{pmatrix}$ .

2. Finn rangen til  $A$  når

- a)  $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$
- b)  $A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 3 & 4 \end{pmatrix}$
- c)  $A = \begin{pmatrix} 1 & 2 & 3 \\ -1 & 4 & 1 \\ 0 & 6 & 4 \end{pmatrix}$

3. Bruk MATLAB til å finne rangen til matrisene

$$A = \begin{pmatrix} 1 & 1 & -2 & 3 & -2 \\ 3 & 1 & -2 & 3 & 1 \\ 2 & 3 & 4 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & -2 & 4 & 3 \\ 3 & -1 & -2 & 3 & 1 \\ 6 & 2 & -5 & 11 & 6 \\ 2 & 3 & -1 & 4 & 2 \end{pmatrix}$$

**4.** Bruk MATLAB til å finne en ortonormal basis for rommet  $H$  utspent av

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}, \quad \mathbf{a}_2 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad \mathbf{a}_3 = \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix}$$

**5.** Vis at  $\mathbf{v}_1 = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ ,  $\mathbf{v}_2 = \left(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$  er en ortonormal basis for  $\mathbb{R}^2$ . Skriv  $\mathbf{v} = (3, -1)$  som en lineærkombinasjon av  $\mathbf{v}_1$  og  $\mathbf{v}_2$ .

**6.** I denne oppgaven er vi i  $\mathbb{R}^3$ .

- a) Anta at  $H$  er en rett linje gjennom origo. Forklar at  $H$  er et underrom av  $\mathbb{R}^3$ . Hva er dimensjonen til  $H$ ?
- b) Anta at  $H$  er et plan gjennom origo. Forklar at  $H$  er et underrom av  $\mathbb{R}^3$ . Hva er dimensjonen til  $H$ ?

**7.** Anta at  $A$  er en ikke-tom mengde vektorer i  $\mathbb{R}^n$ . Vis at

$$H = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \cdot \mathbf{a} = 0 \text{ for alle } \mathbf{a} \in A\}$$

er et underrom av  $\mathbb{R}^n$ .  $H$  kalles det *ortogonale komplementet* til  $A$ .

**8.** Anta at  $H_1$  og  $H_2$  er underrom av  $\mathbb{R}^n$ .

- a) Vis at  $H = H_1 \cap H_2$  også er et underrom av  $\mathbb{R}^n$ . (Husk at *snittet*  $H_1 \cap H_2$  består av de vektorene som er med i både  $H_1$  og  $H_2$ .)
- b) Vis ved et eksempel at  $H = H_1 \cup H_2$  ikke alltid er et underrom av  $\mathbb{R}^n$ . (Husk at *unionen*  $H_1 \cup H_2$  består av de vektorene som er med i minst én av mengdene  $H_1$  og  $H_2$ .)

**9.** Anta at  $H$  er et underrom av  $\mathbb{R}^n$ .

- a) Vis at dersom  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  er lineært uavhengige vektorer i  $H$ , så kan  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  utvides til en basis for  $H$  (det kan tenkes at denne utvidelsen er lik  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ ).
- b) Anta at  $K$  er et annet underrom av  $\mathbb{R}^m$  slik at  $K \subset H$ . Vis  $\dim(K) < \dim(H)$ .

**10.** I denne oppgaven er  $A$  en  $m \times n$ -matrise.

- a) Vis at

$$H = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

er et underrom av  $\mathbb{R}^n$ .  $H$  kalles *nullrommet* eller *kjernen* til  $A$ .

- b) Vis at dimensjonen til  $H$  er  $n - k$  der  $k$  er antall pivotsøyler i den reduserte trappeformen til  $A$ . (*Hint:* Tenk på hva som skjer når du løser ligningssystemet  $A\mathbf{x} = \mathbf{0}$  ved trappereduksjon. Du får også behov for å vise at noen vektorer er lineært uavhengige).
- c) Bevis *dimensjonsteoremet* som sier at i en  $m \times n$ -matrise er summen av dimensjonen til soylerommet og dimensjonen til nullrommet alltid lik  $n$ .

**11.** Anta at  $H_1$  og  $H_2$  er underrom av  $\mathbb{R}^n$ . Vi lar

$$H_1 + H_2 = \{\mathbf{u}_1 + \mathbf{u}_2 \mid \mathbf{u}_1 \in H_1, \mathbf{u}_2 \in H_2\}$$

dvs. at  $H_1 + H_2$  består av alle elementene i  $\mathbb{R}^n$  som kan skrives som en sum av et element i  $H_1$  og et element i  $H_2$ .

- a) Vis at  $H_1 + H_2$  er et underrom av  $\mathbb{R}^n$
- b) Vis at  $\dim(H_1 + H_2) = \dim(H_1) + \dim(H_2) - \dim(H_1 \cap H_2)$

**12.** I denne oppgaven skal vi se på projeksjoner av vektorer ned på underrom. Dersom  $H$  er et underrom av  $\mathbb{R}^n$  og  $\mathbf{x} \in \mathbb{R}^n$ , sier vi at  $\mathbf{y}$  er en *projeksjon* av  $\mathbf{x}$  ned på  $H$  dersom  $\mathbf{y} \in H$  og  $\mathbf{x} - \mathbf{y}$  står normalt på alle vektorer  $\mathbf{v} \in H$ .

- a) Tenk deg (bare i dette punktet) at vi er i  $\mathbb{R}^3$  og at  $H$  er et plan gjennom origo. Lag en tegning av en projeksjon av en vektor  $\mathbf{x}$  ned på  $H$ .
- b) Vis at dersom  $y$  og  $z$  er to projeksjoner av  $\mathbf{x}$  ned på  $H$ , så er  $\mathbf{y} = \mathbf{x}$ , med andre ord at det finnes høyst én projeksjon av  $\mathbf{x}$  ned på  $H$ . (*Hint:* Siden både  $\mathbf{y}$  og  $\mathbf{z}$  ligger i  $H$ , gjør  $\mathbf{v} = \mathbf{y} - \mathbf{z}$  det også. Bruk betingelsene  $\mathbf{x} - \mathbf{y} \perp \mathbf{v}$  og  $\mathbf{x} - \mathbf{z} \perp \mathbf{v}$  til å utlede at  $|\mathbf{y} - \mathbf{z}| = 0$ ).
- c) La  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  være en ortonormal basis for  $H$ . Forklar at denne basisen kan utvides til en ortonormal basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m, \mathbf{v}_{m+1}, \dots, \mathbf{v}_n$  for hele  $\mathbb{R}^n$ . Dersom

$$\mathbf{x} = c_1 \mathbf{v}_1 + \cdots + c_m \mathbf{v}_m + c_{m+1} \mathbf{v}_{m+1} + \cdots + \cdots + c_n \mathbf{v}_n.$$

definerer vi

$$\mathbf{P}(\mathbf{x}) = c_1 \mathbf{v}_1 + \cdots + c_m \mathbf{v}_m$$

Vis at  $\mathbf{P}(\mathbf{x})$  er en projeksjon av  $\mathbf{x}$  ned på  $H$ .

- d) Vis at  $\mathbf{P}(\mathbf{x}) = \mathbf{x}$  for alle  $\mathbf{x} \in H$  og at  $\mathbf{P}(\mathbf{P}(\mathbf{x})) = \mathbf{x}$  for alle  $\mathbf{x} \in \mathbb{R}^n$ .
- e) Vis at  $\mathbf{P} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  er en lineæravbildning.

## 4.8 Elementære matriser

Alt vi hittil har gjort i dette kapitlet er basert på radoperasjoner. I denne seksjonen skal vi se at det å utføre en radoperasjon, kan tolkes som multiplikasjon med en spesiell type matrise — en såkalt *elementær matrise*. I utgangspunktet kan dette virke som en ganske pussig og unyttig observasjon, men som vi skal se senere, har den mer sprekraft enn man skulle tro. La oss begynne med definisjonen:

**Definisjon 4.8.1** En  $m \times m$  elementær matrise er en matrise som fremkommer når vi gjør én radoperasjon på identitetsmatrisen  $I_m$ . Enhver elementær matrise korresponderer altså til en radoperasjon.

Siden det finnes tre typer radoperasjoner, finnes det også tre typer elementære matriser.

- (i) Elementære matriser som fremkommer ved å bytte om to rader i identitetsmatrisen: Et typisk eksempel er

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

der vi har byttet om rad 2 og 4.

- (ii) Elementære matriser der vi har multiplisert en av radene i identitetsmatrisen med et tall forskjellig fra 0: Et typisk eksempel er

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

der tredje rad er multiplisert med  $-3$ .

- (iii) Elementære matriser der vi har lagt et multiplum av en rad til en av de andre radene: Et typisk eksempel er

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

der vi har lagt  $-2$  ganger rad 4 til rad 2.

Den grunnleggende observasjonen om elementære matriser er denne.

**Setning 4.8.2** *Anta at  $E$  er en elementær  $m \times m$ -matrise og la  $A$  være en vilkårlig  $m \times n$ -matrise. La  $A'$  være den matrisen vi får når vi bruker radoperasjonen som korresponderer til  $E$  på  $A$ . Da er  $A' = EA$ .*

*Bevis:* Det er lettest å sjekke dette selv ved å se hva som skjer når vi ganger en matrise med de forskjellige typene elementære matriser.  $\square$

Den neste setningen er helt essensiell for å utnytte elementære matriser.

**Setning 4.8.3** *Enhver elementær matrise er inverterbar, og den inverse er også en elementær matrise.*

*Bevis:* Vi har allerede observert (se seksjon 4.2) at enhver radoperasjon kan “reverseres”, det vil si at det finnes en annen radoperasjon som fører matrisen tilbake til utgangspunktet. Det er lett å sjekke at de korresponderende elementærmatrisene er inverse. Vi sjekker dette for det vanskeligste tilfellet som er elementærmatriser av type (iii) ovenfor:

Anta at  $E$  er elementærmatrisen som tilsvarer å addere  $s$  ganger rad  $j$  til rad  $i$ . Den inverse matrisen  $E'$  er da den som tilsvarer å addere  $-s$  ganger rad  $j$  til rad  $i$ . Det er lett å sjekke at  $E'E = I_n$ : Sammenlignet med identitetsmatrisen har  $E$  en komponent  $s$  “for mye” i posisjon  $(i, j)$ , og  $E'$  fjerner denne ved å addere  $-s$  i komponent  $(i, j)$ .  $\square$

Vi har nå kommet frem til det første hovedresultatet vårt.

**Setning 4.8.4** *Enhver  $m \times n$ -matrise  $A$  kan skrives som et produkt*

$$A = E_1 E_2 \dots E_k B$$

*der  $E_1, E_2, \dots, E_k$  er elementære matriser og  $B$  er den reduserte trappeformen til  $A$ . Dersom  $A$  er en inverterbar, kvadratisk matrise, kan  $A$  altså skrives som et produkt  $A = E_1 E_2 \dots E_k$  av elementære matriser.*

*Bevis:* Vi vet at  $A$  kan omdannes til  $B$  ved hjelp av en sekvens av radoperasjoner. Hvis  $F_1, F_2, \dots, F_k$  er de korresponderende elementærmatrisene, vet vi fra setning 4.8.2 at

$$B = F_k \dots F_2 F_1 A$$

Ganger vi denne ligningen fra venstre med  $F_1^{-1} F_2^{-1} \dots F_k^{-1}$ , får vi

$$F_1^{-1} F_2^{-1} \dots F_k^{-1} B = A$$

Setter vi  $E_1 = F_1^{-1}, E_2 = F_2^{-1}, \dots, E_k = F_k^{-1}$ , vet vi fra setningen ovenfor at  $E_1, E_2, \dots, E_k$  er elementære matriser. Dermed er

$$E_1 E_2 \dots E_k B = A$$

som er formelen i setningen. Dersom  $A$  er en inverterbar, kvadratisk matrise, er  $B = I_n$  og vi får

$$A = E_1 E_2 \dots E_k I_n = E_1 E_2 \dots E_k$$

$\square$

**Bemerkning:** Legg merke til at de elementære matrisene i setningen ovenfor er de inverse til dem som omdanner  $A$  til redusert trappeform, og at de kommer i omvendt rekkefølge.

Vi skal også se raskt på sammenhengen mellom elementære matriser og transponering. Husk at den transponerte matrisen  $A^T$  til  $A$  er den matrisen vi får når vi bytter om rader og søyler

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

**Setning 4.8.5** Den transponerte  $E^T$  til en elementær matrise  $E$  er selv en elementær matrise. Dersom  $E$  korresponderer til å bytte om to rader eller til å gange en rad med et tall  $s$ , så er  $E = E^T$ . Dersom  $E$  korresponderer til å addere  $s$  ganger linje  $j$  til linje  $i$ , så korresponderer  $E^T$  til å addere  $s$  ganger linje  $i$  til linje  $j$ .

*Bevis:* Dette er greit å sjekke selv.  $\square$

Kombinert med neste resultat vil setningen ovenfor komme til nytte når vi i neste seksjon skal regne ut determinanten til den transponerte matrisen.

**Setning 4.8.6** Anta at  $A = E_1 E_2 \dots E_k B$  der  $E_1, E_2, \dots, E_k$  er elementære matriser og  $B$  er på redusert trappeform. Da er

$$A^T = B^T E_k^T \dots E_2^T E_1^T$$

*Bevis:* Det følger fra setning 1.5.2(v) at den transponerte til et produkt er produktet av de transponerte i omvendt rekkefølge, dvs.  $(A_1 A_2 \dots A_k)^T = A_k^T \dots A_2^T A_1^T$ . Setningen følger umiddelbart fra dette.  $\square$

### Opgaver til seksjon 4.8

1. Vis at matrisene er elementære:

a)  $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$  b)  $\begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$  c)  $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$  d)  $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix}$  e)  $\begin{pmatrix} 4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

2. Skriv  $\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$  som et produkt av elementære matriser.

3. Skriv  $\begin{pmatrix} 1 & 2 & 1 \\ -1 & -2 & 3 \\ 0 & 1 & 0 \end{pmatrix}$  som et produkt av elementære matriser.

4. Gjennomfør beviset for setning 4.8.2.

5. Gjennomfør resten av beviset for setning 4.8.3.

6. Gjennomfør beviset for setning 4.8.5.

## 4.9 Determinanter

I seksjon 1.8 studerte vi  $2 \times 2$ - og  $3 \times 3$ -determinanter, og vi kikket så vidt på hvordan man kan definere determinanten til  $4 \times 4$ - og  $5 \times 5$ -matriser. I denne seksjonen skal vi se på teorien for generelle  $n \times n$ -determinanter. Definisjonen følger det samme mønsteret som tidligere — hvis vi allerede vet hvordan vi regner ut  $(n-1) \times (n-1)$ -determinanter, definerer vi  $n \times n$ -determinanter ved hjelp av denne formelen:

$$\begin{aligned} & \left| \begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & \\ a_{21} & a_{22} & \cdots & a_{2n} & \\ \vdots & \vdots & \cdots & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{array} \right| = a_{11} \left| \begin{array}{cccc|c} a_{22} & a_{23} & \cdots & a_{2n} & \\ a_{32} & a_{33} & \cdots & a_{3n} & \\ \vdots & \vdots & \cdots & \vdots & \\ a_{n2} & a_{n3} & \cdots & a_{nn} & \end{array} \right| - \\ & -a_{12} \left| \begin{array}{cccc|c} a_{21} & a_{23} & \cdots & a_{2n} & \\ a_{31} & a_{33} & \cdots & a_{3n} & \\ \vdots & \vdots & \cdots & \vdots & \\ a_{n1} & a_{n3} & \cdots & a_{nn} & \end{array} \right| + \dots + (-1)^{n+1} a_{1n} \left| \begin{array}{cccc|c} a_{21} & a_{22} & \cdots & a_{2,n-1} & \\ a_{31} & a_{32} & \cdots & a_{3,n-1} & \\ \vdots & \vdots & \cdots & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} & \end{array} \right| \end{aligned}$$

Vi ser hvordan mønsteret er: Hvert element i den første raden ganges med den  $(n-1) \times (n-1)$ -determinanten vi får når vi stryker raden og søylen gjennom elementet. Alle leddene legges så sammen med alternerende (skiftende) fortegn.

**Eksempel 1:** Vi skal starte utregningen av  $5 \times 5$ -determinanten

$$\left| \begin{array}{ccccc} 2 & 3 & -4 & 5 & 1 \\ 1 & 0 & 3 & 4 & -1 \\ -2 & 1 & 1 & 0 & -2 \\ 2 & -2 & 1 & 3 & 1 \\ 0 & 0 & 2 & 1 & 3 \end{array} \right|$$

(det blir for mye kjedelig arbeid å gjennomføre hele beregningen!) Ifølge definisjonen ovenfor er

$$\begin{aligned} & \left| \begin{array}{ccccc} 2 & 3 & -4 & 5 & 1 \\ 1 & 0 & 3 & 4 & -1 \\ -2 & 1 & 1 & 0 & -2 \\ 2 & -2 & 1 & 3 & 1 \\ 0 & 0 & 2 & 1 & 3 \end{array} \right| = 2 \left| \begin{array}{ccccc} 0 & 3 & 4 & -1 & \\ 1 & 1 & 0 & -2 & \\ -2 & 1 & 3 & 1 & \\ 0 & 2 & 1 & 3 & \end{array} \right| - 3 \left| \begin{array}{ccccc} 1 & 3 & 4 & -1 & \\ -2 & 1 & 0 & -2 & \\ 2 & 1 & 3 & 1 & \\ 0 & 2 & 1 & 3 & \end{array} \right| + \\ & + (-4) \left| \begin{array}{ccccc} 1 & 0 & 4 & -1 & \\ -2 & 1 & 0 & -2 & \\ 2 & -2 & 3 & 1 & \\ 0 & 0 & 1 & 3 & \end{array} \right| - 5 \left| \begin{array}{ccccc} 1 & 0 & 3 & -1 & \\ -2 & 1 & 1 & -2 & \\ 2 & -2 & 1 & 1 & \\ 0 & 0 & 2 & 3 & \end{array} \right| + 1 \left| \begin{array}{ccccc} 1 & 0 & 3 & 4 & \\ -2 & 1 & 1 & 0 & \\ 2 & -2 & 1 & 3 & \\ 0 & 0 & 2 & 1 & \end{array} \right| \end{aligned}$$

Nå fortsetter du på samme måte med hver av  $4 \times 4$ -determinantene osv. ♣

Eksemplet ovenfor viser at definisjonen er ubrukelig til å regne ut verdien til store determinanter. Det er heller ikke så lett å se de teoretiske egenskapene til determinanter direkte fra definisjonen. Vi skal derfor bruke litt tid på å beskrive determinanter på andre måter.

**Bemerkning:** Determinanter er definert *induktivt* ved at vi definerer determinanter av en viss størrelse ved hjelp av determinanter én størrelse mindre. Dette medfører at induksjonsbevis er en naturlig bevismetode når man skal vise egenskaper ved determinanter. Er du uvant med induksjonsbevis (eller synes bevisene nedenfor er vanskelige å forstå), kan det være lurt å ta en kikk på seksjon 1.2 i *Kalkulus*.

Vi skal først se på noen tilfeller der determinanten er spesielt enkel å regne ut.

**Lemma 4.9.1** *Anta at  $A$  er en kvadratisk matrise der enten en rad eller en søyle bare består av nuller. Da er  $\det(A) = 0$ .*

*Bevis:* Vi ser på tilfellet der en søyle er null — beviset for det andre tilfellet ligner, men er litt lettere.

Vi skal vise resultatet for  $n \times n$ -matriser ved induksjon på  $n$ . Det er lett å sjekke at en  $2 \times 2$ -determinant er null dersom en av søylene er null. Anta at resultatet holder for  $n \times n$ -matriser, og at  $A$  er en  $(n+1) \times (n+1)$ -matrise der  $j$ -te søyle er 0. Ifølge definisjonen er

$$\begin{aligned} \det(A) &= \left| \begin{array}{cccc|c} a_{11} & \dots & a_{1j} & \dots & a_{1,n+1} \\ a_{21} & \dots & a_{2j} & \dots & a_{2,n+1} \\ \vdots & \dots & \vdots & \dots & \vdots \\ a_{n+1,1} & \dots & a_{n+1,j} & \dots & a_{n+1,n+1} \end{array} \right| = \\ &= a_{11} \left| \begin{array}{cccc|c} a_{22} & \dots & a_{2j} & \dots & a_{2,n+1} \\ a_{32} & \dots & a_{3j} & \dots & a_{3,n+1} \\ \vdots & \dots & \vdots & \dots & \vdots \\ a_{n+1,2} & \dots & a_{n+1,j} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\ &\quad \dots + (-1)^{j+1} a_{1j} \left| \begin{array}{cccc|c} a_{21} & a_{22} & \dots & a_{2,n+1} \\ a_{31} & a_{32} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\ &\quad \dots + (-1)^{n+2} a_{1,n+1} \left| \begin{array}{cccc|c} a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n} \end{array} \right| \end{aligned}$$

I dette uttrykket er ledet

$$(-1)^{j+1} a_{1j} \begin{vmatrix} a_{21} & a_{22} & \dots & a_{2,n+1} \\ a_{31} & a_{32} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{vmatrix}$$

null fordi  $a_{1j} = 0$ . De andre leddene er null fordi determinanten inneholder en rad som bare består av nuller, og induksjonshypotesen forteller oss at en slik  $n \times n$ -determinant er null.  $\square$

De neste matrisene vi skal beregne determinanten til, er triangulære matriser. Vi kaller en kvadratisk matrise *øvre triangulær* dersom alle elementene under diagonalen er null, og vi kaller den *nedre triangulær* dersom alle elementer over diagonalen er null. Matrisen  $A$  nedenfor er øvre triangulær, mens  $B$  er nedre triangulær:

$$A = \begin{pmatrix} 2 & 1 & 0 & 6 \\ 0 & 3 & -1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -3 & 4 & 6 & 0 \\ 0 & 1 & 3 & 4 \end{pmatrix}$$

Legg merke til at alle trappematriser er øvre triangulære.

**Lemma 4.9.2** *Dersom matrisen  $A$  er øvre eller nedre triangulær, er determinanten til  $A$  lik produktet av elementene på diagonalen til  $A$ .*

*Bevis:* Vi skal bare vise resultatet for øvre triangulære matriser — beviset for nedre triangulære matriser går på samme måte, men er lettere. Også i dette tilfellet går beviset ved induksjon på størrelsen til  $A$ . Er  $A$  en øvre triangulær  $2 \times 2$ -matrise  $A = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ , ser vi at  $\det(A) = ad - 0b = ad$ , så resultatet stemmer for  $2 \times 2$ -matriser.

Anta nå at resultatet stemmer for øvre triangulære  $n \times n$ -matriser, og la

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1,n+1} \\ 0 & a_{22} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{n+1,n+1} \end{pmatrix}$$

være en øvre triangulær  $(n+1) \times (n+1)$ -matrise. Vi har

$$\det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2,n+1} \\ 0 & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{n+1,n+1} \end{vmatrix} - a_{12} \begin{vmatrix} 0 & a_{23} & \dots & a_{2,n+1} \\ 0 & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{n+1,n+1} \end{vmatrix} + \dots$$

$$\cdots + (-1)^{n+2} a_{1,n+1} \begin{vmatrix} 0 & a_{22} & \dots & a_{2n} \\ 0 & 0 & \dots & a_{3n} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2,n+1} \\ 0 & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{n+1,n+1} \end{vmatrix}$$

der vi har brukt det forrige lemmaet til å kvitte oss med de determinantene som har en søyle med bare nuller. Ifølge induksjonshypotesen er

$$\begin{vmatrix} a_{22} & a_{23} & \dots & a_{2,n+1} \\ 0 & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & a_{n+1,n+1} \end{vmatrix} = a_{22}a_{33} \cdots a_{n+1,n+1}$$

og dermed får vi  $\det(A) = a_{11}a_{22} \cdots a_{n+1,n+1}$ .  $\square$

### Determinanter og radoperasjoner

Vi er nå klare til å starte et litt større prosjekt — vi skal undersøke hva som skjer med determinanten til en matrise når vi utfører en radoperasjon. Dette vil gi oss en mer effektiv måte å regne ut determinanter på, og det vil også gi oss en bedre forståelse av hva determinanter er. Inspirasjon til arbeidet finner vi i setningene 1.8.3 og 1.8.5 som forteller oss hva som skjer med  $2 \times 2$ - og  $3 \times 3$ -determinanter når vi bruker radoperasjoner. Målet er å vise tilsvarende resultater for generelle determinanter. Dette krever en del arbeid (og tålmodighet!)

Vi starter med den enkleste radoperasjonen.

**Lemma 4.9.3** *Anta at  $B$  er den matrisen vi får når vi ganger den  $i$ -te raden i  $A$  med tallet  $s$ . Da er  $\det(B) = s \det(A)$ .*

*Bevis:* Igjen bruker vi induksjon på størrelsen til matrisen  $A$ . Hvis  $A$  er en  $2 \times 2$ -determinant, er påstanden lett å sjekke (vi kjener den dessuten fra setning 1.8.3). Anta nå at setningen holder for  $n \times n$ -matriser, og at  $A$  er en  $(n+1) \times (n+1)$ -matrise. Anta først at  $B$  er den matrisen vi får når vi ganger den første raden i  $A$  med  $s$ . Da er

$$\begin{aligned} \det(B) &= \begin{vmatrix} sa_{11} & sa_{12} & \dots & sa_{1,n+1} \\ a_{21} & a_{21} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{vmatrix} = \\ &= sa_{11} \begin{vmatrix} a_{22} & a_{23} & \dots & a_{2,n+1} \\ a_{32} & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,2} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{vmatrix} - \end{aligned}$$

$$\begin{aligned}
 & -sa_{12} \left| \begin{array}{cccc} a_{21} & a_{23} & \dots & a_{2,n+1} \\ a_{31} & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\
 & \dots + (-1)^{n+2} sa_{1,n+1} \left| \begin{array}{cccc} a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n} \end{array} \right| = s \det(A)
 \end{aligned}$$

der vi har brukt at hvert ledd i uttrykket for  $\det(B)$  har en ekstra  $s$  sammenlignet med uttrykket for  $\det(A)$ .

La oss nå anta at  $\det$  er en rad  $i > 1$  som ganges med  $s$ . Da er

$$\begin{aligned}
 \det(B) &= \left| \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1,n+1} \\ \vdots & \vdots & \dots & \vdots \\ sa_{i1} & sa_{i2} & \dots & sa_{i,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{array} \right| = \\
 &= a_{11} \left| \begin{array}{cccc} a_{22} & a_{23} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ sa_{i2} & sa_{i3} & \dots & sa_{i,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,2} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| - \\
 &\quad -a_{12} \left| \begin{array}{cccc} a_{21} & a_{23} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ sa_{i1} & sa_{i3} & \dots & sa_{i,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\
 &\dots + (-1)^{n+2} a_{1,n+1} \left| \begin{array}{cccc} a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ sa_{i1} & sa_{i2} & \dots & sa_{in} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n} \end{array} \right| = \\
 &= sa_{11} \left| \begin{array}{cccc} a_{22} & a_{23} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{i2} & a_{i3} & \dots & a_{i,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,2} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| -
 \end{aligned}$$

$$\begin{array}{c|cccc}
 & a_{21} & a_{23} & \dots & a_{2,n+1} \\
 \vdots & \vdots & \dots & & \vdots \\
 -sa_{12} & a_{i1} & a_{i3} & \dots & a_{i,n+1} \\
 \vdots & \vdots & \dots & & \vdots \\
 & a_{n+1,1} & a_{n+1,3} & \dots & a_{n+1,n+1} \\
 \hline
 & a_{21} & a_{22} & \dots & a_{2n} \\
 \vdots & \vdots & \dots & & \vdots \\
 \cdots + (-1)^{n+2}sa_{1,n+1} & a_{i1} & a_{i2} & \dots & a_{in} \\
 \vdots & \vdots & \dots & & \vdots \\
 & a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n}
 \end{array} = s \det(A)$$

der vi har brukt induksjonshypotesen til å trekke  $s$  utenfor  $n \times n$ -matrisene.  $\square$

Vi skal nå se på hva som skjer når vi bytter om to rader. Dette er litt komplisert, men nøkkelen ligger i først å vise hva som skjer når vi bytter om de to øverste radene.

**Lemma 4.9.4** *Anta at  $B$  er den matrisen vi får når vi bytter om de to øverste radene i  $A$ . Da er  $\det(A) = -\det(B)$ .*

*Bevis:* Vi tenker oss  $A$  er en  $n \times n$ -matrise og at vi gjennomfører de to første skrittene i utregningen av  $\det(A)$  — først uttrykker vi  $\det(A)$  ved hjelp av  $(n-1) \times (n-1)$ -matriser, og så uttrykker vi hver av disse igjen ved hjelp av  $(n-2) \times (n-2)$ -matriser. De  $(n-2) \times (n-2)$ -matrisene vi nå har, er fremkommet ved at vi har fjernet de to øverste radene i  $A$  samt to av søylene. Dersom  $i < j$ , lar vi  $A_{ij}$  være den  $(n-2) \times (n-2)$ -determinanten vi får når vi fjerner den  $i$ -te og den  $j$ -te søylen i tillegg til de to første radene. I utregningene våre oppstår denne determinanten på to måter: Vi kan enten fjerne den  $i$ -te søylen i første omgang og den  $j$ -te søylen i andre, eller omvendt. I det første tilfellet får vi en faktor  $(-1)^{i+1}a_{1i}$  i første omgang og en faktor  $(-1)^j a_{2j}$  i andre omgang (observer at siden  $i < j$ , har  $a_{2j}$  nå rykket frem til  $(j-1)$ -te søylen, og fortegnsfaktoren blir derfor  $(-1)^j$  og ikke  $(-1)^{j+1}$  som man kanskje ville vente). Totalt gir dette en faktor  $(-1)^{i+j+1}a_{1i}a_{2j}$ . I det andre tilfellet får vi en faktor  $(-1)^{j+1}a_{1j}$  i første omgang og en faktor  $(-1)^{i+1}a_{2i}$  i andre omgang. Totalt gir dette en faktor  $(-1)^{i+j+2}a_{1j}a_{2i}$ . Alt i alt har vi dermed et bidrag til  $\det(A)$  på

$$(-1)^{i+j+1} (a_{1i}a_{2j} - a_{1j}a_{2i}) A_{ij} \quad (4.9.1)$$

Dersom vi gjør tilsvarende beregninger for  $B$ , får vi samme svar bortsett fra at elementene i første og annen rad har byttet plass. Dette gir et bidrag på

$$(-1)^{i+j+1} (a_{2i}a_{1j} - a_{2j}a_{1i}) A_{ij} \quad (4.9.2)$$

Disse uttrykkene er motsatt like store, og dermed ser vi at  $\det(B) = -\det(A)$ .  $\square$

La oss utvide resultatet til alle naborader:

**Lemma 4.9.5** *Anta at  $B$  er en matrise som fremkommer ved at vi bytter om to naborader i  $A$ . Da er  $\det(B) = -\det(A)$ .*

*Bevis:* Det er lett å se at resultatet holder for  $2 \times 2$ -matriser, og vi bruker induksjon til å vise det generelt. Vi antar at  $A$  er en  $(n+1) \times (n+1)$ -matrise, og at resultatet holder for  $n \times n$ -matriser. Dersom vi bytter om de to første radene, følger resultatet fra foregående lemma. Dersom vi bytter om to andre rader, følger resultatet fra induksjonshypotesen. Vi har nemlig

$$\begin{aligned} \det(A) &= \left| \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1,n+1} \\ a_{21} & a_{22} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{array} \right| = \\ &= a_{11} \left| \begin{array}{cccc} a_{22} & a_{23} & \dots & a_{2,n+1} \\ a_{32} & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,2} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\ &\quad \dots + (-1)^{j+1} a_{1j} \left| \begin{array}{cccc} a_{21} & a_{22} & \dots & a_{2,n+1} \\ a_{31} & a_{32} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\ &\quad \dots + (-1)^{n+2} a_{1,n+1} \left| \begin{array}{cccc} a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n} \end{array} \right| \end{aligned}$$

Når vi bytter om de to radene i  $\det(A)$ , bytter vi også om to rader i hver av de mindre determinantene. Ifølge induksjonshypotesen bytter de da fortegn, og dermed bytter  $\det(A)$  fortegn.  $\square$

Til slutt utvider vi resultatet til alle rader.

**Lemma 4.9.6** *Dersom  $B$  fremkommer ved at vi bytter om to rader i  $A$ , så er  $\det(B) = -\det(A)$ .*

*Bevis:* Vi kan bytte om to vilkårlige rader ved systematisk ombytte av naborader: Anta at vi skal bytte om rad  $i$  og rad  $i + k$ . Vi bytter først om rad  $i + k$  med rad  $i + k - 1$ , deretter bytter vi den med rad  $i + k - 2$  osv. Etter  $k$  slike bytter, har raden flyttet seg til  $i$ -te posisjon, og den opprinnelige rad  $i$  er skjøvet ned til  $(i + 1)$ -te posisjon. Vi bytter nå denne raden nedover til den er i  $(i + k)$ -te posisjon — dette tar  $k - 1$  nabobytter. Nå er vi fremme ved matrisen  $B$  og har foretatt  $2k - 1$  ombytter av naborader. Det betyr at determinanten har byttet fortegn et oddet antall ganger, og følgelig er  $\det(B) = -\det(A)$ .  $\square$

Før vi går videre, tar vi med oss en nyttig konsekvens av resultatet ovenfor.

**Korollar 4.9.7** *Dersom to av radene i  $A$  er like, er  $\det(A) = 0$ .*

*Bevis:* Dersom vi bytter om de to like radene, vil determinanten ifølge lemmaet ovenfor bytte fortegn. Men determinanten har jo ikke endret seg siden radene vi byttet om var like. Den eneste løsningen på dette tilsynelatende paradokset, er at  $\det(A) = 0$ .  $\square$

Vi er nå fremme ved den siste typen radoperasjon — de hvor vi adderer et multiplum av en rad til en annen.

**Lemma 4.9.8** *Anta at  $B$  fremkommer fra  $A$  ved at vi adderer et multiplum av en av radene i  $A$  til en av de andre radene. Da er  $\det(B) = \det(A)$ .*

*Bevis:* Igjen går beviset ved induksjon på størrelsen til matrisen. For  $2 \times 2$ -matriser vet vi fra setning 1.8.3 at resultatet holder. Vi antar at resultatet holder for  $n \times n$ -matriser, og at  $A$  er en  $(n+1) \times (n+1)$ -matrise. Anta først at vi adderer  $s$  ganger rad  $j$  til rad  $i$  der  $i > 1$ . Da er

$$\begin{aligned} \det(B) &= \left| \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{i1} + sa_{j1} & a_{i2} + sa_{j2} & \dots & a_{i,n+1} + sa_{j,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{array} \right| = \\ &= a_{11} \left| \begin{array}{cccc} a_{22} & a_{23} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{i2} + sa_{j2} & a_{i3} + sa_{j3} & \dots & a_{i,n+1} + sa_{j,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,2} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| - \end{aligned}$$

$$\begin{aligned}
 & -a_{12} \left| \begin{array}{cccc} a_{21} & a_{23} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{i1} + sa_{j1} & a_{i3} + sa_{j3} & \dots & a_{i,n+1} + sa_{j,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\
 & \dots + (-1)^{n+2} a_{1,n+1} \left| \begin{array}{cccc} a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{i1} + sa_{j1} & a_{i2} + sa_{j2} & \dots & a_{3n} + sa_{jn} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n} \end{array} \right|
 \end{aligned}$$

Siden resultatet gjelder for  $n \times n$ -matriser, vet vi at disse determinantene ikke endrer seg om vi fjerner ” $s$ ”-leddene. Følgelig er  $\det(B) = \det(A)$ .

La oss nå se på det gjenstående tilfellet der vi adderer  $s$  ganger rad  $j$  til den *første* raden. Da har vi:

$$\begin{aligned}
 \det(B) &= \left| \begin{array}{cccc} a_{11} + sa_{j1} & a_{12} + sa_{j2} & \dots & a_{1,n+1} + sa_{j,n+1} \\ a_{21} & a_{22} & \dots & a_{2,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n+1} \end{array} \right| = \\
 &= (a_{11} + sa_{j1}) \left| \begin{array}{cccc} a_{22} & a_{23} & \dots & a_{2,n+1} \\ a_{32} & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,2} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| - \\
 &\quad -(a_{12} + sa_{j2}) \left| \begin{array}{cccc} a_{21} & a_{23} & \dots & a_{2,n+1} \\ a_{31} & a_{33} & \dots & a_{3,n+1} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,3} & \dots & a_{n+1,n+1} \end{array} \right| + \dots \\
 &\dots + (-1)^{n+2} (a_{1,n+1} + sa_{j,n+1}) \left| \begin{array}{cccc} a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \dots & a_{n+1,n} \end{array} \right|
 \end{aligned}$$

Ganger vi ut parentesene og samler ledd med og uten  $s$  hver for seg, ser vi at  $\det(B) = \det(A) + s \det(\tilde{A})$  der  $\tilde{A}$  er den matrisen vi får når vi erstatter den første raden med den  $j$ -te (hvis du ikke ser dette direkte, så skriv opp det første trinnet i utregningen av  $\det(A)$  og  $\det(\tilde{A})$ ). Ifølge korollaret ovenfor er  $\det(\tilde{A}) = 0$  siden to av radene er like. Følgelig er  $\det(B) = \det(A)$  også i

dette tilfellet.  $\square$

Vi har nå nådd vårt mål og kan oppsummere hva som skjer med determinanten til en matrise når vi bruker en radoperasjon.

**Teorem 4.9.9** *Anta at  $A$  er en kvadratisk matrise. Da gjelder:*

- (i) *Hvis  $A$  er øvre eller nedre triangulær, er determinanten lik produktet av diagonalelementene.*
- (ii) *Bytter vi om to rader, bytter determinanten fortegn (men beholder sin tallverdi).*
- (iii) *Ganger vi en rad med et tall  $s$ , endres determinanten med en faktor  $s$ .*
- (iv) *Adderer vi et multiplum av en rad til en annen rad, endres ikke determinanten*

*Bevis:* Dette er bare en oppsummering av lemmaene 4.9.2, 4.9.3, 4.9.6 og 4.9.8.  $\square$

Resultatet ovenfor gir oss en effektiv metode for å regne ut determinanter. Selv om effektiviteten først viser seg for alvor på større matriser, illustrerer vi metoden i  $3 \times 3$ -tilfellet.

**Eksempel 2:** Vi skal regne ut determinanten til matrisen

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

Vi radreduserer  $A$  mens vi holder styr på hvilke operasjoner vi bruker:

$$\begin{aligned} A &\xrightarrow{I \leftrightarrow II} \begin{pmatrix} 1 & 0 & -1 \\ 3 & 3 & -1 \\ 2 & 2 & 2 \end{pmatrix} \xrightarrow{II + (-3)I} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 3 & 2 \\ 2 & 2 & 2 \end{pmatrix} \xrightarrow{III + (-2)I} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 3 & 2 \\ 0 & 2 & 4 \end{pmatrix} \\ &\xrightarrow[\frac{1}{3}II]{\sim} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & \frac{2}{3} \\ 0 & 2 & 4 \end{pmatrix} \xrightarrow{III + (-2)II} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & \frac{2}{3} \\ 0 & 0 & \frac{8}{3} \end{pmatrix} = B \end{aligned}$$

Matrisen  $B$  er øvre triangulær, og determinanten er produktet av diagonalelementene:  $\det(B) = 1 \cdot 1 \cdot \frac{8}{3} = \frac{8}{3}$ . Vi kan nå regne oss bakover til determinanten til  $A$  ved å bruke teoremet ovenfor — hver ganger vi bytter om to rader, bytter vi fortegn på determinanten, og hver gang vi ganger en rad med en faktor  $s$ , endrer determinanten seg med en faktor  $s$ . Den vanligste operasjonen (å addere et multiplum av en rad til en annen) endrer

ikke determinanten i det hele tatt. I prosessen ovenfor er det to operasjoner som endrer determinanten — ett radombytte og en multiplikasjon med  $\frac{1}{3}$ . Det betyr at  $\det(B) = (-1) \cdot \frac{1}{3} \cdot \det(A)$  (husk på at det er  $\det(A)$  som endres til  $\det(B)$ ). Dermed har vi

$$\det(A) = (-1) \cdot 3 \cdot \det(B) = (-1) \cdot 3 \cdot \frac{8}{3} = -8$$



Prosedyren i eksemplet ovenfor kan brukes på alle kvadratiske matriser — vil vi regne ut determinanten til  $A$ , reduserer vi  $A$  til en trappematrise  $B$ , og regner ut determinanten til  $B$  ved å gange sammen diagonalelementene. Deretter finner vi  $\det(A)$  ved formelen

$$\det(A) = s_1^{-1} s_2^{-1} \cdots s_k^{-1} \det(B)$$

der  $s_1, s_2, \dots, s_k$  er *faktorene* til de radoperasjonene vi brukte når vi reduserte  $A$  til  $B$ . Det burde være klart hva vi mener med *faktoren* til en radoperasjon — faktoren til et radombytte er  $-1$ , faktoren til det å gange en rad med  $s \neq 0$  er  $s$ , mens faktoren til det å addere et multiplum av én rad til en annen, er  $1$ .

Prosedyren ovenfor har også viktige teoretiske konsekvenser slik det neste teoremet viser. Vi benytter anledningen til også å oppsummere noen tidligere resultater:

**Teorem 4.9.10** *For  $n \times n$ -matriser  $A$  er følgende ekvivalent:*

- (i)  $\det(A) \neq 0$
- (ii)  $A$  er inverterbar
- (iii) Matriseligningen  $A\mathbf{x} = \mathbf{b}$  har en entydig løsning for alle  $\mathbf{b}$
- (iv) Matriseligningen  $A\mathbf{x} = \mathbf{0}$  har bare løsningen  $\mathbf{x} = \mathbf{0}$
- (v) Søylene i  $A$  danner en basis for  $\mathbb{R}^n$
- (vi)  $A$  er radekvivalent med  $I_n$

*Bevis:* Vi vet allerede at (ii), (iii) (iv) og (v) er ekvivalente med (vi) (det følger fra hhv. setning 4.5.4, setning 4.3.3, setning 4.4.3 og setning 4.6.10), så det gjenstår bare å vise at (i) er ekvivalent med (vi). Anta at  $B$  er den reduserte trappeformen til  $A$  og la  $s_1, s_2, \dots, s_k$  være faktorene til de radoperasjonene vi bruker når vi reduserer  $A$  til  $B$ . Da er

$$\det(A) = s_1^{-1} s_2^{-1} \cdots s_k^{-1} \det(B)$$

og følgelig er  $\det(A) \neq 0$  hvis og bare hvis  $\det(B) \neq 0$ . En matrise  $B$  på redusert trappeform er øvre triangulær, og determinanten er derfor lik produktet av diagonalelementene. Det betyr at determinanten er forskjellig fra null hvis og bare hvis alle pivotelementene står på diagonalen, dvs. når  $B = I_n$ .  $\square$

### Determinanten til et produkt

Vi skal nå se hvordan vi kan bruke elementære matriser til å bevise noen viktige setninger om determinanter. Siden en elementær matrise fremkommer ved å gjøre en enkelt radoperasjon på identitetsmatrisen, er det lett å regne ut determinanten.

**Lemma 4.9.11** *Anta at  $E$  er en elementær  $n \times n$ -matrise. Da er determinanten til  $E$  lik faktoren til den tilhørende radoperasjonen. Svarer  $E$  til å bytte om to rader, er altså  $\det(E) = -1$ , svarer  $E$  til å gange en rad med  $s$ , er  $\det(E) = s$ , og svarer  $E$  til å addere et multiplum av en rad til en annen, er  $\det(E) = 1$ .*

*Bevis:* Dette følger direkte fra Teorem 4.9.9 siden en elementær matrise fremkommer fra identitetsmatrisen når vi bruker den korresponderende radoperasjonen.  $\square$

Det neste lemmaet er også en omformulering av tidligere resultater.

**Lemma 4.9.12** *Anta  $C = EB$  der  $E$  er en elementær matrise. Da er  $\det(C) = \det(E) \det(B)$ .*

*Bevis:* Vi vet fra Setning 4.8.2 at  $C$  er den matrisen som fremkommer når vi bruker radoperasjonen til  $E$  på  $B$ . Fra Teorem 4.9.9 vet vi at da er  $\det(C) = s \det(B)$ , der  $s$  er faktoren til denne radoperasjonen. Ifølge foregående lemma er  $s = \det(E)$ , og dermed får vi  $\det(C) = \det(E) \det(B)$ .  $\square$

Konklusjonen i lemmaet ovenfor kan skrives slik:  $\det(EB) = \det(E) \det(B)$ . Målet vårt er å utvide denne formelen til alle matriser  $A$  slik at vi generelt får  $\det(AB) = \det(A) \det(B)$ . Før vi går løs på den generelle formelen, trenger vi litt informasjon om hva som skjer når en av matrisene  $A, B$  har determinant lik 0.

**Lemma 4.9.13** *Anta at  $A$  og  $B$  er to  $n \times n$ -matriser. Da er produktmatrisen  $C = AB$  inverterbar hvis og bare hvis både  $A$  og  $B$  er inverterbare.*

*Bevis:* Dersom både  $A$  og  $B$  er inverterbare, vet vi fra setning 1.7.4 at  $C$  er inverterbar med  $C^{-1} = B^{-1}A^{-1}$ . Det gjenstår dermed å vise at dersom enten  $A$  eller  $B$  ikke er inverterbar, så er heller ikke  $C$  inverterbar.

Anta først at  $B$  ikke er inverterbar. Den reduserte trappeformen til  $B$  mangler da pivotelement i (minst) en søyle, og ligningen  $B\mathbf{x} = \mathbf{0}$  har følgelig en løsning  $\mathbf{x} \neq \mathbf{0}$ . Dermed er  $C\mathbf{x} = (AB)\mathbf{x} = A(B\mathbf{x}) = A\mathbf{0} = \mathbf{0}$ . Dersom  $C$  var inverterbar, ville ligningen  $C\mathbf{x} = \mathbf{0}$  bare hatt én løsning  $\mathbf{x} = \mathbf{0}$ , så dette betyr at  $C$  ikke er inverterbar.

Anta så at  $A$  ikke er inverterbar. Da må den reduserte trappeformen mangle pivotelement i en rad, så det finnes en vektor  $\mathbf{b}$  slik at ligningen  $A\mathbf{x} = \mathbf{b}$  ikke har løsning. Dermed kan heller ikke ligningen  $C\mathbf{y} = \mathbf{b}$  ha en løsning, for hvis  $\mathbf{y}$  var en løsning av denne ligningen, ville  $\mathbf{x} = B\mathbf{y}$  være en løsning av  $A\mathbf{x} = \mathbf{b}$ . Vi har nemlig

$$A\mathbf{x} = A(B\mathbf{y}) = (AB)\mathbf{y} = C\mathbf{y} = \mathbf{b}$$

Dette betyr at ligningen  $C\mathbf{y} = \mathbf{b}$  ikke har løsning, og følgelig er  $C$  ikke inverterbar.  $\square$

Resultatet ovenfor forteller oss at formelen  $\det(AB) = \det(A)\det(B)$  holder dersom en av matrisene  $A, B$  er singulær (dvs. ikke inverterbar) — i så fall er en av faktorene på høyresiden lik null, og det samme er  $\det(AB)$  ifølge lemmaet. Vi kan derfor konsentrere oss om å vise formelen  $\det(AB) = \det(A)\det(B)$  når  $A$  og  $B$  er inverterbare.

**Setning 4.9.14** *For alle  $n \times n$ -matriser  $A, B$  er*

$$\det(AB) = \det(A)\det(B)$$

*Bevis:* Vi vet allerede at formelen stemmer når  $A$  eller  $B$  er singulær, så vi kan anta at både  $A$  og  $B$  er inverterbare. Ifølge setning 4.8.4 er da både  $A$  og  $B$  produkter av elementære matriser

$$A = E_1 E_2 \cdots E_m \quad \text{og} \quad B = F_1 F_2 \cdots F_k$$

Ved gjentatt bruk av lemma 4.9.12, ser vi at

$$\det(A) = \det(E_1) \det(E_2) \cdots \det(E_m)$$

og

$$\det(B) = \det(F_1) \det(F_2) \cdots \det(F_k)$$

Siden  $AB = E_1 E_2 \cdots E_m F_1 F_2 \cdots F_k$  får vi på tilsvarende måte

$$\det(AB) = \det(E_1) \det(E_2) \cdots \det(E_m) \det(F_1) \det(F_2) \cdots \det(F_k)$$

og følgelig er  $\det(AB) = \det(A)\det(B)$ .  $\square$

**Bemerkning:** Dersom vi tenker på determinanten som en forstørrelsesfaktor (se seksjon 2.9), er resultatet ovenfor som forventet — forstørrer

vi først med en faktor  $\det(B)$  og så med en faktor  $\det(A)$ , bør den samlede forstørrelsesfaktoren  $\det(AB)$  være lik produktet  $\det(A)\det(B)$ . Vær forøvrig oppmerksom på at det ikke finnes noen tilsvarende enkel metode for å regne ut determinanten til en sum.

Setningen ovenfor har mange nyttige konsekvenser:

**Korollar 4.9.15** *For alle inverterbare matriser  $A$  er*

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

*Bevis:* Ifølge setningen ovenfor er

$$\det(A)\det(A^{-1}) = \det(I_n) = 1$$

□

**Korollar 4.9.16** *For alle  $n \times n$ -matriser er*

$$\det(A^T) = \det(A)$$

*Bevis:* Observer først at dersom  $E$  er en elementær matrise, så forteller setning 4.8.5 oss at  $E^T$  er en elementær matrise med samme faktor. Dermed er  $\det(E^T) = \det(E)$  ifølge lemma 4.9.11.

Ifølge setning 4.8.4 kan enhver matrise  $A$  skrives som et produkt  $A = E_1 E_2 \cdots E_m B$  der  $E_1, E_2, \dots, E_m$  er elementære matriser, og  $B$  er på redusert trappeform. Ifølge setningen ovenfor er

$$\det(A) = \det(E_1) \det(E_2) \cdots \det(E_m) \det(B)$$

Siden  $A^T = B^T E_m^T \cdots E_2^T E_1^T$  (se setning 4.8.6 om du ikke ser dette direkte), får vi tilsvarende

$$\det(A^T) = \det(B^T) \det(E_m^T) \cdots \det(E_2^T) \det(E_1^T)$$

Vi har allerede observert at  $\det(E_1^T) = \det(E_1)$ ,  $\det(E_2^T) = \det(E_2), \dots$ ,  $\det(E_m^T) = \det(E_m)$ , så det er nok å vise at  $\det(B^T) = \det(B)$ . Siden  $B$  er en øvre triangulær matrise, er  $B^T$  nedre triangulær, og begge determinantene er da lik produktet av diagonalelementene ifølge lemma 4.9.2. Diagonalelementene endrer seg ikke når vi transponerer, og dermed er  $\det(B^T) = \det(B)$ . □

Resultatet ovenfor er nyttig når vi skal regne ut visse determinanter:

**Eksempel 3:** Tenk deg at vi skal regne ut determinanten til

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

Ifølge korollaret ovenfor kan vi like godt regne ut determinanten til  $\det(A^T)$ , og da får vi

$$\det(A) = \det(A^T) = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & -2 & -3 \\ -3 & -2 & -1 & 1 \\ 4 & 3 & 4 & 1 \end{vmatrix} = \begin{vmatrix} 1 & -2 & -3 \\ -2 & -1 & 1 \\ 3 & 4 & 1 \end{vmatrix}$$

der vi har utnyttet alle nullene til å forenkle regnearbeidet (fullfør utregningene selv).  $\square$

### Utvikling langs rader og søyler

Når vi bruker definisjonen til å uttrykke en  $n \times n$ -determinant ved hjelp av  $(n-1) \times (n-1)$ -determinanter, sier vi at vi *utvikler* eller *ekspanderer* determinanten langs første rad. Det viser seg at vi kan utvikle en determinant også langs andre rader — og faktisk også langs søyler (her bruker vi trikset i eksempel 3; vi transponerer matrisen for å gjøre om søyler til rader).

Når vi utvikler determinanten til en  $n \times n$ -matrise  $A$  langs  $i$ -te rad, tar vi det første elementet  $a_{i1}$  i raden og ganger med den  $(n-1) \times (n-1)$ -matrisen vi får når vi fjerner raden og søylen gjennom  $a_{i1}$ . Deretter gjør vi det samme med det andre elementet i raden; vi ganger  $a_{i2}$  med den  $(n-1) \times (n-1)$ -determinanten vi får når vi fjerner raden og søylen gjennom  $a_{i2}$ . Vi fortsetter på denne måten bortover hele raden, og til slutt legger vi sammen alle leddene med vekslende fortegn. Her må vi være litt forsiktige — når vi utvikler langs første rad, begynner vi alltid med positivt fortegn, men det gjelder ikke generelt. Regelen er at dersom nummeret  $i$  på raden er et oddetall, begynner vi med positivt fortegn, dersom  $i$  er et partall, begynner vi med negativt fortegn. Figuren nedenfor viser hvilket fortegn de forskjellige leddene får når vi utvikler en determinant.

$$\begin{pmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ + & - & + & - & \dots \\ - & + & - & + & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix}$$

Du kan sjekke at fortegnet til det  $(i, j)$ -te elementet er gitt ved  $(-1)^{i+j}$ . Når vi har utviklet langs  $i$ -te rad, sitter vi dermed igjen med

$$\det(A) = (-1)^{i+1}a_{i1}A_{i1} + (-1)^{i+2}a_{i2}A_{i2} + \cdots + (-1)^{i+n}a_{in}A_{in}$$

der  $A_{ij}$  er determinanten vi får når vi fjerner raden og søylen gjennom  $a_{ij}$ .

Det er ofte lurt å utvikle en determinant langs en annen rad enn den første fordi vi da kan utnytte spesielle egenskaper ved determinanten. Spesielt er det lurt å utvikle langs en rad med mange nuller. La oss se på et

eksempel:

**Eksempel 4:** Vi skal regne ut determinanten til

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

Vi velger å utvikle langs fjerde rad siden den inneholder to nuller. Siden dette er en “partallsrad”, må vi begynne med negativt fortegn:

$$\begin{aligned} \det(A) &= -0 \cdot \begin{vmatrix} 1 & 3 & -1 \\ -2 & 2 & 3 \\ -1 & 2 & 1 \end{vmatrix} + 1 \cdot \begin{vmatrix} 2 & 3 & -1 \\ 0 & 2 & 3 \\ 3 & 2 & 1 \end{vmatrix} - \\ &\quad -2 \cdot \begin{vmatrix} 2 & 1 & -1 \\ 0 & -2 & 3 \\ 3 & -1 & 1 \end{vmatrix} + 0 \cdot \begin{vmatrix} 2 & 1 & 3 \\ 0 & -2 & 2 \\ 3 & -1 & 2 \end{vmatrix} = \\ &= \begin{vmatrix} 2 & 3 & -1 \\ 0 & 2 & 3 \\ 3 & 2 & 1 \end{vmatrix} - 2 \cdot \begin{vmatrix} 2 & 1 & -1 \\ 0 & -2 & 3 \\ 3 & -1 & 1 \end{vmatrix} \end{aligned}$$

Vi fortsetter på samme måte med  $3 \times 3$ -matrisene. På grunn av nullen, lønner det seg nå å utvikle langs annen rad. For oversiktens skyld tar vi determinantene hver for seg. Vi får

$$\begin{aligned} \begin{vmatrix} 2 & 3 & -1 \\ 0 & 2 & 3 \\ 3 & 2 & 1 \end{vmatrix} &= -0 \cdot \begin{vmatrix} 3 & -1 \\ 2 & 1 \end{vmatrix} + 2 \cdot \begin{vmatrix} 2 & -1 \\ 3 & 1 \end{vmatrix} - 3 \cdot \begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix} = \\ &= 2 \cdot (2 \cdot 1 - 3 \cdot (-1)) - 3 \cdot (2 \cdot 2 - 3 \cdot 3) = 25 \end{aligned}$$

og

$$\begin{aligned} \begin{vmatrix} 2 & 1 & -1 \\ 0 & -2 & 3 \\ 3 & -1 & 1 \end{vmatrix} &= -0 \cdot \begin{vmatrix} 1 & -1 \\ -1 & 1 \end{vmatrix} + (-2) \cdot \begin{vmatrix} 2 & -1 \\ 3 & 1 \end{vmatrix} - 3 \cdot \begin{vmatrix} 2 & 1 \\ 3 & -1 \end{vmatrix} = 5 \\ &= (-2) \cdot (2 \cdot 1 - 3 \cdot (-1)) - 3 \cdot (2 \cdot (-1) - 3 \cdot 1) = 5 \end{aligned}$$

Dermed har vi  $\det(A) = 25 - 2 \cdot 5 = 15$ .  $\square$

Hvorfor får vi riktig svar når vi utvikler en determinant langs en annen rad enn den første? Det er ikke så vanskelig å forklare. Anta at vi ønsker å utvikle determinanten til  $A$  langs  $i$ -te rad. Vi kan flytte denne raden til

toppen av matrisen gjennom  $i - 1$  radombytter — først bytter vi raden med den  $(i - 1)$ -te, så med rad nummer  $(i - 2)$  osv. Til slutt får vi en matrise  $B$  der den  $i$ -te raden i  $A$  står øverst, mens alle de andre kommer i opprinnelig rekkefølge. Siden determinanten bytter fortegn for hvert radombytte, ser vi at  $\det(B) = (-1)^{i-1} \det(A)$ , dvs. at  $\det(B) = \det(A)$  hvis  $i$  er et oddetall, og  $\det(B) = -\det(A)$  hvis  $i$  er et partall. Hvis vi nå regner ut determinanten til  $B$  ved å utvikle langs første rad (altså i henhold til definisjonen), får vi nøyaktig de samme leddene som når vi utvikler  $\det(A)$  langs  $i$ -te rad, bortsett fra at fortegnene til leddene er motsatt dersom  $i$  er et partall. Dette fortegnsbyttet kompenserer for at  $\det(B) = -\det(A)$  når  $i$  er et partall, og dermed ser vi at prosedyren gir riktig svar i alle tilfeller.

Vi kan også utvikle determinanter langs søyler istedenfor rader. Ønsker vi å utvikle determinanten til  $n \times n$ -matrisen  $A$  langs  $j$ -te søylen, starter vi med det første elementet  $a_{1j}$  i søylen, og ganger det med den  $(n - 1) \times (n - 1)$ -determinanten vi får når vi fjerner raden og søylen gjennom  $a_{1j}$ . Vi gjør tilsvarende med de andre elementene i søylen, og til slutt legger vi sammen alle leddene med vekslende fortegn. Fortegnet følger det samme sjakkbrømmönsteret som tidligere. Det er lett å se hvorfor denne prosedyren gir riktig svar — å utvikle determinanten til  $A$  etter  $j$ -te søylen, er det samme som å utvikle determinanten til  $A^T$  etter  $j$ -te rad, og vi vet at  $\det(A^T) = \det(A)$ . La oss se på et enkelt eksempel.

**Eksempel 5:** Vi skal regne ut determinanten til

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 4 & 2 & 3 \\ 3 & 1 & 0 \end{pmatrix}$$

Det lønner seg å utvikle langs siste søylen siden den inneholder to nuller:

$$\begin{aligned} \det(A) &= 0 \cdot \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} - 3 \cdot \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} + 0 \cdot \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix} = \\ &= -3 \cdot (2 \cdot 1 - 3 \cdot 1) = 3 \end{aligned}$$

□

Vi har nå den kunnskapen vi trenger om determinanter. I neste seksjon skal vi se hvordan vi kan bruke denne kunnskapen til å finne egenverdier og egenvektorer.

### Oppgaver til seksjon 4.9

1. Bruk definisjonen av determinant til å regne ut determinanten til matrisene:

a)  $\begin{pmatrix} 1 & -2 & -1 \\ 3 & 1 & -2 \\ 1 & 0 & 1 \end{pmatrix}$  b)  $\begin{pmatrix} 2 & 0 & 3 \\ 1 & -1 & 2 \\ 0 & 1 & 2 \end{pmatrix}$  c)  $\begin{pmatrix} 0 & 2 & 3 \\ -1 & 2 & 0 \\ 2 & -1 & 1 \end{pmatrix}$

**2.** Bruk radoperasjoner til å regne ut determinanten til matrisene:

a)  $\begin{pmatrix} 1 & -3 & 0 \\ 2 & -1 & -2 \\ 1 & -1 & 1 \end{pmatrix}$  b)  $\begin{pmatrix} 2 & 1 & -1 \\ 1 & -1 & 1 \\ 2 & 1 & 2 \end{pmatrix}$  c)  $\begin{pmatrix} 0 & -1 & 1 \\ -1 & 3 & 2 \\ 2 & -2 & 1 \end{pmatrix}$

**3.** Regn ut determinanten ved å ekspandere langs velvalgte søyler og rader:

a)  $\begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & -1 \\ 3 & 7 & 2 \end{pmatrix}$  b)  $\begin{pmatrix} 1 & 0 & 3 \\ 2 & -3 & 1 \\ 4 & 0 & 2 \end{pmatrix}$  c)  $\begin{pmatrix} 3 & 1 & 0 & 4 \\ 2 & 1 & 2 & 0 \\ 0 & 0 & 1 & -2 \\ 1 & 2 & 1 & 2 \end{pmatrix}$

**4.** Bruk MATLAB til å regne ut determinanten til matrisen:

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

**5.** Bevis radtilfellet av lemma 4.9.1.

**6.** Bevis lemma 4.9.2 for nedre triangulære matriser.

**7.** Vis at dersom  $A$  er en  $n \times n$ -matrise og  $r$  er et tall, så er  $\det(rA) = r^n \det(A)$ .

**8.** Vis at  $\det(A^n) = \det(A)^n$  for alle hele tall  $n$  (på høyre side er det  $\det(A)$  som opphøyes i  $n$ -te).

**9.** Bruk teorem 4.9.9 til å vise at dersom radene til  $A$  er lineært avhengige, så er  $\det(A) = 0$ . (*Hint:* Bruk radoperasjoner til å skaffe deg en rad som bare består av nuller). Vis at dette også følger fra teorem 4.9.10 og korollar 4.9.16.

**10.** En  $n \times n$ -matrise kalles *ortogonal* dersom  $U^{-1} = U^T$ . Vis at  $\det(U)$  er enten 1 eller -1.

**11.** I denne oppgaven skal vi bruke følgende notasjon: Dersom  $A$  er en  $n \times n$ -matrise og  $\mathbf{b} \in \mathbb{R}^n$  er en søylevektor, så er  $A_i(\mathbf{b})$  matrisen vi får når vi erstatter den  $i$ -te søylen til  $A$  med  $\mathbf{b}$ . Vi skal vise at dersom  $A$  er inverterbar, så er løsningen

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

til ligningen  $A\mathbf{x} = \mathbf{b}$  gitt ved

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det(A)}$$

Dette kalles *Cramers regel*.

a) Vis at dersom  $I$  er  $n \times n$ -identitetsmatrisen, så er  $\det(I_i(\mathbf{x})) = x_i$ .

- b) Vis at  $AI_i(\mathbf{x}) = A_i(\mathbf{b})$ .
- c) Bevis Cramers regel.
- d) Bruk Cramers regel til å løse ligningssystemet

$$\begin{aligned} 2x - 3y &= 4 \\ x - 4y &= -2 \end{aligned}$$

## 4.10 Egenvektorer og egenverdier

Dersom  $A$  er en  $n \times n$ -matrise, kalles en vektor  $\mathbf{v} \neq \mathbf{0}$  en *egenvektor* for  $A$  dersom det finnes et tall  $\lambda$  slik at  $A\mathbf{v} = \lambda\mathbf{v}$ . Tallet  $\lambda$  kalles en *egenverdi* for  $A$ . Vi har vært borti egenvektorer og egenverdier tidligere, men siden vi ikke har hatt noen metode for å finne dem, har det vært vanskelig å utnytte dem skikkelig. Med det vi nå har lært om determinanter, har vi de redskapene vi trenger.

Det er et par ting vi bør avklare før vi begynner. For det første legger vi merke til at dersom  $\mathbf{v}$  er en egenvektor med egenverdi  $\lambda$ , så er også enhver parallel vektor  $a\mathbf{v}$  (der  $a$  er et tall forskjellig fra 0) en egenvektor med egenverdi  $\lambda$  siden

$$A(a\mathbf{v}) = aA(\mathbf{v}) = a(\lambda\mathbf{v}) = \lambda(a\mathbf{v})$$

Når vi er på jakt etter egenvektorer, er det nok å finne én av disse parallele vektorene, og vi velger da som regel en som har “pene” (f.eks. heltallige) komponenter.

Den andre tingen vi bør være klar over, er at selv om en matrise  $A$  er reell, kan det hende at den har egenverdier og egenvektorer som er komplekse (se eksempel 5 i seksjon 2.8). Skal man løse et praktisk problem, er det ofte helt nødvendig å studere disse komplekse egenvektorene og egenverdiene, og det medfører at vi er nødt til å bruke teorien fra de tidligere seksjonene på komplekse tall og komplekse vektorer. I noen tilfeller går vi da utover det vi strengt tatt har bevist, men det er ikke vanskelig å sjekke at de tidlige resultatene i dette kapitlet også holder for komplekse tall og komplekse vektorer. Det eneste vi må passe litt ekstra på når vi regner med komplekse vektorer, er skalarproduktet. Som du kanskje husker fra seksjon 1.3, inneholder skalarproduktet av to komplekse vektorer  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ,  $\mathbf{y} = (y_1, y_2, \dots, y_n)$  en komplekskonjugasjon — vi har

$$\mathbf{x} \cdot \mathbf{y} = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}$$

Når vi skal finne egenvektorer og egenverdier, lønner det seg som regel å finne egenverdiene først. Ifølge definisjonen er  $\lambda$  en egenverdi for  $A$  dersom det finnes en ikke-null vektor  $\mathbf{v}$  slik at

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

Sagt på en annen måte:  $\lambda$  er en egenverdi for  $A$  dersom ligningen  $A\mathbf{v} = \lambda\mathbf{v}$  har en ikke-null løsning  $\mathbf{v}$ . Skriver vi høyresiden av denne ligningen som  $\lambda I_n \mathbf{v}$  og flytter leddet  $A\mathbf{v}$  over på den andre siden, får vi

$$(\lambda I_n - A)\mathbf{v} = \mathbf{0} \quad (4.10.1)$$

Dette betyr at  $\lambda$  er en egenverdi for  $A$  hvis og bare hvis ligning (4.10.1) har en løsning  $\mathbf{v} \neq \mathbf{0}$ , dvs. hvis og bare hvis ligningen har mer enn én løsning ( $\mathbf{v} = \mathbf{0}$  er alltid en løsning). Ifølge teorem 4.9.10 skjer dette hvis og bare hvis  $\det(\lambda I_n - A) = 0$ . Vi har dermed vist:

**Lemma 4.10.1**  $\lambda$  er en egenverdi for  $n \times n$ -matrisen  $A$  hvis og bare hvis  $\det(\lambda I_n - A) = 0$ .  $\square$

Observer at dersom

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

så er

$$\lambda I_n - A = \begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{pmatrix}$$

La oss se på et eksempel.

**Eksempel 1:** Vi skal finne egenverdiene til matrisen

$$A = \begin{pmatrix} 4 & -1 \\ 5 & -2 \end{pmatrix}$$

Vi ser at

$$\det(\lambda I_2 - A) = \begin{vmatrix} \lambda - 4 & 1 \\ -5 & \lambda + 2 \end{vmatrix} =$$

$$= (\lambda - 4)(\lambda + 2) - (-5) \cdot 1 = \lambda^2 - 2\lambda - 8 + 5 = \lambda^2 - 2\lambda - 3$$

Egenverdiene er altså løsningene til annengradslikningen

$$\lambda^2 - 2\lambda - 3 = 0$$

Løser vi denne, får vi egenverdiene  $\lambda_1 = 3$ ,  $\lambda_2 = -1$ .

Når vi har funnet egenverdiene, er det ingen sak å finne egenvektorene. En egenvektor  $\mathbf{v}_1$  med egenverdi  $\lambda_1 = 3$ , er en løsning av ligningen

$$A\mathbf{v}_1 = 3\mathbf{v}_1$$

Setter vi inn  $\mathbf{v}_1 = \begin{pmatrix} x \\ y \end{pmatrix}$ , får vi

$$\begin{aligned} 4x - y &= 3x \\ 5x - 2y &= 3y \end{aligned}$$

som kan omformes til

$$\begin{aligned} x - y &= 0 \\ 5x - 5y &= 0 \end{aligned}$$

Vi kan selvfølgelig løse dette systemet ved Gauss-eliminasjon, men det er egentlig unødvendig — vi ser at ligningene har de samme løsningene, og at vi derfor kan nøye oss med å løse den første. Den har åpenbart løsningene  $x = y$ . Dette betyr at vi kan velge  $y$  fritt og så regne ut  $x$ . Alle vektorer vi får på denne måten er parallele, og vi plukker derfor bare ut én av dem. Velger vi  $y = 1$ , får vi  $x = 1$ , og vi ser dermed at  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  er en egenvektor.

La oss også finne en egenvektor  $\mathbf{v}_2$  for den andre egenverdien  $\lambda_2 = -1$ . Denne vektoren må være en løsning av ligningen

$$A\mathbf{v}_1 = (-1)\mathbf{v}_1 = -\mathbf{v}_1$$

Setter vi inn  $\mathbf{v}_1 = \begin{pmatrix} x \\ y \end{pmatrix}$ , får vi

$$\begin{aligned} 4x - y &= -x \\ 5x - 2y &= -y \end{aligned}$$

som kan omformes til

$$\begin{aligned} 5x - y &= 0 \\ 5x - y &= 0 \end{aligned}$$

Igjen ser vi at vi kan regne ut  $x$  når vi har valgt  $y$ . Velger vi  $y = 5$ , får vi  $x = 1$ , og dermed har vi  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$ . Vi har altså to egenvektorer

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ og } \mathbf{v}_2 = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

Det er lett å se at disse er lineært uavhengige og dermed danner en basis for  $\mathbb{R}^2$ .



I eksemplet ovenfor måtte vi løse en annengradsligning for å finne egenverdiene til en  $2 \times 2$ -matrise. Dette er typisk — skal vi finne egenverdiene til en  $n \times n$ -matrise, må vi løse en  $n$ -tegradsligning. Grunnen er at når vi regner ut  $\det(\lambda I_n - A)$ , får vi et  $n$ -tegradspolynom i  $\lambda$ . For å overbevise deg om dette behøver du bare å tenke på hva som skjer når du utvikler en matrise etter første rad (hvis du vil, kan du føre et induksjonsbevis).

**Definisjon 4.10.2** Dersom  $A$  er  $n \times n$ -matrise, kalles  $n$ -tegradspolynomet

$$P_A(\lambda) = \det(\lambda I_n - A)$$

det karakteristiske polynomet<sup>1</sup> til  $A$ .

Fra algebraens fundamentalteorem (se *Kalkulus*, teorem 3.5.1) vet vi at et  $n$ -tegradspolynom alltid har  $n$  røtter når vi teller med multiplisitet og tillater komplekse løsninger. Det vanligste er at alle røttene er forskjellige. Det neste resultatet viser at hvis det karakteristiske polynomet  $P_A$  har  $n$  forskjellige røtter, så har vi en basis av egenvektorer for  $A$ .

**Setning 4.10.3** La  $A$  være en  $n \times n$ -matrise, og anta at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er egenvektorer med forskjellige egenverdier. Da er  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  lineært uavhengige. Dersom  $A$  har  $n$  forskjellige egenverdier, finnes det altså en basis som består av egenvektorer for  $A$ .

*Bevis:* Anta (for motsigelse) at det finnes lineært avhengige egenvektorer med forskjellige egenverdier, og la  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  være en slik mengde med færrest mulige elementer. Vi ser at  $k$  er minst 2 siden en mengde som bare består av ett ikke-null element umulig kan være lineært avhengig. Siden vektorene er lineært avhengige, finnes det tall  $c_1, c_2, \dots, c_k$  som ikke alle er null, slik at

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k = \mathbf{0} \quad (4.10.2)$$

Faktisk må alle  $c$ 'ene være forskjellige fra null, for hvis ikke finnes det en mindre, lineært avhengig mengde av egenvektorer. Hvis vi ganger ligning (4.10.2) fra venstre med  $A$ , får vi

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \dots + c_k\lambda_k\mathbf{v}_k = \mathbf{0} \quad (4.10.3)$$

der  $\lambda_1, \lambda_2, \dots, \lambda_k$  er egenverdiene. Ganger vi ligning (4.10.2) med  $-\lambda_1$  og legger resultatet til (4.10.3), får vi

$$c_2(\lambda_2 - \lambda_1)\mathbf{v}_2 + c_3(\lambda_3 - \lambda_1)\mathbf{v}_3 + \dots + c_k(\lambda_k - \lambda_1)\mathbf{v}_k = \mathbf{0}$$

Siden alle  $c$ 'ene er forskjellige fra null og egenverdiene er forskjellige, viser dette at  $\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k$  er lineært avhengige, men det er umulig siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er en *minimal*, lineært avhengig mengde. Dette betyr at antagelsen om at det finnes lineært avhengige egenvektorer med forskjellige egenverdier, er gal, og dermed er den første delen av setningen bevist. Den andre delen følger siden enhver lineært uavhengig mengde med  $n$  elementer i  $\mathbb{R}^n$  er en basis (se korollar 4.6.11).  $\square$

---

<sup>1</sup>De fleste bøker bruker definisjonen  $P(\lambda) = \det(A - \lambda I_n)$  isteden. De to definisjonene skiller seg med en faktor  $(-1)^n$ , og er derfor så godt som ekvivalente. Fordelen med vårt valg er at vi slipper en del unødvendige minuser når vi skal finne egenverdier i praksis.

### Multiple egenverdier

Vi har nå sett at dersom alle egenverdiene til en matrise er forskjellige, så finnes det en basis av egenvektorer. Det er naturlig å tenke seg at dette også gjelder dersom noen av egenverdiene er sammenfallende — det naturlige tipset er at dersom  $\lambda$  er en egenverdi med multiplisitet  $k$ , så finnes det  $k$  lineært uavhengige egenvektorer med egenverdi  $\lambda$ . Dette er imidlertid ikke tilfellet, og dermed er det heller ikke slik at enhver matrise har en basis av egenvektorer. La oss se på et eksempel.

**Eksempel 2:** Vi skal finne egenverdiene og egenvektorene til matrisen

$$A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$$

Det karakteristiske polynomet er

$$P_A(\lambda) = \begin{pmatrix} \lambda - 2 & 1 \\ -1 & \lambda \end{pmatrix} = (\lambda - 2)\lambda - (-1) \cdot 1 = \lambda^2 - 2\lambda + 1$$

som bare har én rot  $\lambda = 1$ . Denne roten har multiplisitet 2 siden

$$P_A(\lambda) = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2$$

En egenvektor  $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$  må tilfredsstille  $A\mathbf{v} = \mathbf{v}$ , dvs.

$$\begin{aligned} 2x - y &= x \\ x &= y \end{aligned}$$

Dette ligningssystemet er oppfylt hvis  $x = y$ , dvs. hvis  $\mathbf{v} = \begin{pmatrix} x \\ x \end{pmatrix}$ . Alle disse vektorene er lineært avhengige, så selv om  $\lambda$  har multiplisitet 2, finnes det ikke to lineært uavhengige vektorer med egenverdi 1. Det finnes heller ikke noen basis bestående av egenvektorer til  $A$ . ♣

Vi skal se på et eksempel til. Dette eksemplet viser at vi ikke må bli *for* pessimistiske; vi kan godt ha en basis av egenvektorer selv om ikke alle egenverdiene er forskjellige. Eksemplet demonstrerer også noen av de regnetekniske utfordringene vi får, når vi skal finne egenverdiene til litt større systemer.

**Eksempel 3:** Vi skal finne egenverdiene og egenvektorene til matrisen

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

Det karakteristiske polynomet er gitt ved:

$$P_A(\lambda) = \begin{vmatrix} \lambda - \frac{5}{3} & -\frac{1}{3} & \frac{1}{3} \\ 0 & \lambda - 1 & 0 \\ \frac{2}{3} & \frac{1}{3} & \lambda - \frac{4}{3} \end{vmatrix}$$

For å regne ut polynomet, ekspanderer vi langs andre rad og får

$$\begin{aligned} P_A(\lambda) &= (\lambda - 1) \begin{vmatrix} \lambda - \frac{5}{3} & \frac{1}{3} \\ \frac{2}{3} & \lambda - \frac{4}{3} \end{vmatrix} = (\lambda - 1) \left( (\lambda - \frac{5}{3})(\lambda - \frac{4}{3}) - \frac{2}{3} \cdot \frac{1}{3} \right) = \\ &= (\lambda - 1)(\lambda^2 - 3\lambda + 2) \end{aligned}$$

Annengradspolynomet  $\lambda^2 - 3\lambda + 2$  har røttene 2 og 1 (sjekk!), og dermed har  $P_A$  røttene 2 og 1 — den siste med multiplisitet 2. Egenverdiene til  $A$  er dermed også 2 og 1.

La oss først finne egenvektorene med egenverdi 1. En slik egenvektor

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

må tilfredsstille ligningen  $A\mathbf{v} = \mathbf{v}$ , dvs.

$$\frac{5}{3}x + \frac{1}{3}y - \frac{1}{3}z = x$$

$$y = y$$

$$-\frac{2}{3}x - \frac{1}{3}y + \frac{4}{3}z = z$$

Dette systemet kan også skrives

$$\frac{2}{3}x + \frac{1}{3}y - \frac{1}{3}z = 0$$

$$0 = 0$$

$$-\frac{2}{3}x - \frac{1}{3}y + \frac{1}{3}z = 0$$

Det er mange måter å løse dette systemet på, men siden det er viktig ikke å miste noen løsninger, kobler vi inn den “offisielle” metoden vår. Den utvidede matrisen til ligningssystemet er

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

Vi radreduserer denne matrisen ved å legge den første raden til den tredje, og deretter gange den første raden med  $\frac{3}{2}$ . Vi sitter da igjen med matrisen

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

Vi ser at  $y$  og  $z$  er frie variable, mens  $x$  er en basisvariabel. Velger vi verdier for  $y$  og  $z$ , får vi

$$x = -\frac{1}{2}y + \frac{1}{2}z$$

En egenvektor må derfor være på formen

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}y + \frac{1}{2}z \\ y \\ z \end{pmatrix} = y \begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$$

Dette viser at

$$\mathbf{v}_1 = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_2 = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$$

genererer alle egenvektorer med egenverdi 1, og det er lett å sjekke at  $\mathbf{v}_1$  og  $\mathbf{v}_2$  er lineært uavhengige.

Vi ser nå på den andre egenverdien 2. En egenvektor

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

med denne egenverdien må tilfredsstille ligningen  $A\mathbf{v} = 2\mathbf{v}$ , dvs.

$$\frac{5}{3}x + \frac{1}{3}y - \frac{1}{3}z = 2x$$

$$y = 2y$$

$$-\frac{2}{3}x - \frac{1}{3}y + \frac{4}{3}z = 2z$$

Dette systemet kan også skrives

$$-\frac{1}{3}x + \frac{1}{3}y - \frac{1}{3}z = 0$$

$$-y = 0$$

$$-\frac{2}{3}x - \frac{1}{3}y - \frac{2}{3}z = 0$$

Den utvidede matrisen til dette ligningssystemet er

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

Vi radreduserer denne matrisen:

$$\begin{aligned} \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -1 & 0 & 0 \\ -\frac{2}{3} & -\frac{1}{3} & -\frac{2}{3} & 0 \end{pmatrix} &\xrightarrow{III+(-2)I} \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \\ &\xrightarrow{III+(-1)II} \begin{pmatrix} -\frac{1}{3} & \frac{1}{3} & -\frac{1}{3} & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{(-3)I} \begin{pmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Vi ser at  $z$  er en fri variabel, mens  $x$  og  $y$  er basisvariable. Velger vi en verdi for  $z$ , får vi

$$y = 0$$

og

$$x = y - z = -z$$

En egenvektor må derfor være på formen

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -z \\ 0 \\ z \end{pmatrix} = z \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

Alle disse vektorene er parallelle, og vi velger

$$\mathbf{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

som vår representant. Vi har dermed tre egenvektorer

$$\mathbf{v}_1 = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

Det er lett å sjekke at disse er lineært uavhengige og dermed danner en basis for  $\mathbb{R}^3$ . ♣

Eksemplene ovenfor viser at det ikke er så lett å vite hva som skjer med egenvektorene når vi har sammenfallende egenverdier — i noen tilfeller vil vi ha en basis av egenvektorer, i andre tilfeller ikke. Vi kan ikke gjøre så mye annet enn å undersøke hvert enkelt tilfelle. Det finnes mer avanserte verktøy man kan bruke, men de får vente til en annen anledning.

### Komplekse egenverdier

Vi tar også med et eksempel der egenverdiene er komplekse. Fremgangsmåten er akkurat den samme som i det reelle tilfellet, men regningene kan bli litt styggere siden vi nå får ligningssystemer med komplekse koeffisienter.

**Eksempel 4:** Vi skal finne egenverdiene og egenvektorene til matrisen

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

Det karakteristiske polynomet er

$$P_A(\lambda) = \begin{vmatrix} \lambda - 1 & 2 \\ -2 & \lambda - 1 \end{vmatrix} = \lambda^2 - 2\lambda + 5$$

Setter vi dette uttrykket lik null og løser annengradslikningen, får vi

$$\lambda = \frac{-(-2) \pm \sqrt{(-2)^2 - 4 \cdot 1 \cdot 5}}{2} = 1 \pm 2i$$

Egenverdiene er altså  $\lambda_1 = 1 + 2i$  og  $\lambda_2 = 1 - 2i$ .

Vi finner først en egenvektor  $\mathbf{v}_1 = \begin{pmatrix} x \\ y \end{pmatrix}$  med egenverdi  $\lambda_1$ . En slik vektor må oppfylle ligningen

$$\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (1 + 2i) \begin{pmatrix} x \\ y \end{pmatrix}$$

det vil si

$$\begin{aligned} x - 2y &= (1 + 2i)x \\ 2x + y &= (1 + 2i)y \end{aligned}$$

Flytter vi leddene på høyre side over på den andre siden, og forkorter med 2, får vi

$$\begin{aligned} -ix - y &= 0 \\ x - iy &= 0 \end{aligned}$$

Ganger vi den øverste ligningen med  $i$ , får vi den nederste ligningen, og det er derfor nok å finne en løsning til den ene av ligningene. Velger vi  $y$  lik 1 i den nederste ligningen, får vi  $x = i$ , og dermed har vi funnet egenvektoren

$$\mathbf{v}_1 = \begin{pmatrix} i \\ 1 \end{pmatrix}$$

Vi kan finne en egenvektor  $\mathbf{v}_2 = \begin{pmatrix} x \\ y \end{pmatrix}$  med egenverdi  $\lambda_2$  på akkurat samme måte. En slik vektor må oppfylle ligningen

$$\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = (1 - 2i) \begin{pmatrix} x \\ y \end{pmatrix}$$

det vil si

$$\begin{aligned} x - 2y &= (1 - 2i)x \\ 2x + y &= (1 - 2i)y \end{aligned}$$

Flytter vi leddene på høyre side over på den andre siden, og forkorter med 2, får vi

$$\begin{aligned} ix - y &= 0 \\ x + iy &= 0 \end{aligned}$$

Ganger vi den første ligningen med  $-i$ , får vi den andre, så det er nok å løse én av ligningene. Velger vi  $y = 1$  i den nederste, får vi  $x = -i$ . Dermed har vi egenvektoren

$$\mathbf{v}_2 = \begin{pmatrix} -i \\ 1 \end{pmatrix}$$



I eksemplet ovenfor ser vi at de to egenverdiene og de to egenvektorene er komplekskonjugerte av hverandre. At egenverdiene er komplekskonjugerte, er ikke noe mysterium — det følger av at de komplekse røttene til et reelt polynom alltid kommer i komplekskonjugerte par (se *Kalkulus*, lemma 3.5.3). For å sjekke at det samme gjelder egenvektorene, trenger vi et lite resonnement.

**Setning 4.10.4** *Anta at  $A$  er en reell  $n \times n$ -matrise, og at  $\mathbf{v}$  er en kompleks egenvektor med egenverdi  $\lambda$ . Da er  $\bar{\mathbf{v}}$  en egenvektor med egenverdi  $\bar{\lambda}$  (her er  $\bar{\mathbf{v}}$  den vektoren vi får når vi konjugerer alle komponentene til  $\mathbf{v}$ ).*

*Bevis:* (I dette beviset bruker vi regnereglene for konjugasjon på vektorer og matriser, og ikke bare på tall. Du bør sjekke at dette er tillatt.) Siden  $A$  er reell, har vi

$$\overline{A\mathbf{v}} = \overline{A}\bar{\mathbf{v}} = A\bar{\mathbf{v}}$$

På den annen side er

$$\overline{A\mathbf{v}} = \overline{\lambda}\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$$

Kombinerer vi disse to uttrykkene får vi

$$A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$$

som viser at  $\bar{\mathbf{v}}$  er en egenvektor med egenverdi  $\bar{\lambda}$ .  $\square$

Legg merke til at vi kan bruke denne setningen til å forenkle arbeidet med å finne komplekse egenvektorer. I eksempel 4 kunne vi ha brukt den til å skrive opp  $\mathbf{v}_2$  med en gang vi hadde funnet  $\mathbf{v}_1$ .

### Eigenverdier til symmetriske matriser

Vi må innrømme at teorien vår har sine ubehagelige sider — det er ikke alle matriser som har en basis av egenvektorer, og det kan godt tenkes at eigenverdiene er komplekse selv om matrisen er reell. Dette må vi bare leve med — det er nå slik verden engang er. Det finnes imidlertid noen matriser som oppfører seg slik vi kunne ønske oss, nemlig de symmetriske.

**Definisjon 4.10.5** En  $n \times n$ -matrise  $A$  er *symmetrisk* dersom  $A = A^T$ .

Navnet *symmetrisk* kommer av at en symmetrisk matrise ikke endrer seg når vi speiler den om diagonalen. Disse matrisene er symmetriske:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 2 & 3 \\ 2 & 0 & 5 \\ 3 & 5 & 7 \end{pmatrix}, \quad C = \begin{pmatrix} -1 & 3 & -2 & 1 \\ 3 & -4 & 5 & 0 \\ -2 & 5 & -1 & 4 \\ 1 & 0 & 4 & \pi \end{pmatrix}$$

De symmetriske matrisene kan virke spesielle, men de dukker opp i forbausende mange sammenhenger. I neste kapittel skal vi utnytte at matriser bestående av de annenderiverte til en funksjon (såkalte Hesse-matriser) er symmetriske.

Vi trenger en definisjon til før vi kan skrive opp hovedresultatet for symmetriske matriser (har du lest seksjon 4.7 vil du ha sett både denne definisjonen og setning 4.10.7 før): En basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er *ortonormal* dersom alle vektorene i basisen har lengde 1 og står ortogonalt (normalt) på hverandre — med andre ord dersom

$$\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{hvis } i \neq j \\ 1 & \text{hvis } i = j \end{cases}$$

**Teorem 4.10.6 (Spektralteoremet for symmetriske matriser)** Anta at  $A$  er en symmetrisk  $n \times n$ -matrise. Da er alle eigenverdiene til  $A$  reelle, og det finnes en ortonormal basis for  $\mathbb{R}^n$  som består av egenvektorer til  $A$ .

Vi utsetter beviset for dette teoremet til seksjon 4.12 — det er ikke spesielt vanskelig, men det krever en del forberedelser som vi ikke har gjort ennå.

Ortonormale basiser har mange fordeler — blant annet er det raskt å finne ut hvordan man kan skrive en vilkårlig vektor som en lineærkombinasjon av basisvektorene:

**Setning 4.10.7** *Anta at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er en ortonormal basis for  $\mathbb{R}^n$ . For ethvert element  $\mathbf{v} \in \mathbb{R}^n$  er da*

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

der  $c_i = \mathbf{v} \cdot \mathbf{v}_i$  for  $i = 1, 2, \dots, n$ .

*Bevis:* Siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  er en basis, vet vi at  $\mathbf{v}$  kan skrives som en lineær-kombinasjon

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

Tar vi skalarproduktet med  $\mathbf{v}_i$  på begge sider, får vi

$$\mathbf{v} \cdot \mathbf{v}_i = c_1\mathbf{v}_1 \cdot \mathbf{v}_i + c_2\mathbf{v}_2 \cdot \mathbf{v}_i + \cdots + c_i\mathbf{v}_i \cdot \mathbf{v}_i + \cdots + c_n\mathbf{v}_n \cdot \mathbf{v}_i = c_i$$

der vi har brukt at

$$\mathbf{v}_j \cdot \mathbf{v}_i = \begin{cases} 0 & \text{hvis } i \neq j \\ 1 & \text{hvis } i = j \end{cases}$$

□

### Diagonalisering av matriser

Dersom en matrise har en basis av egenvektorer, kan den *diagonaliseres* på en måte som ofte er nyttig.

**Setning 4.10.8** *Anta at  $A$  er en  $n \times n$ -matrise med en basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  av egenvektorer, og la  $\lambda_1, \lambda_2, \dots, \lambda_n$  være de tilhørende egenverdiene. La  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  være lineæravbildningen som for alle  $i$  avbilder  $\mathbf{e}_i$  på  $\mathbf{v}_i$ , og la  $M$  være matrisen til  $\mathbf{T}$ . Da er  $M$  inverterbar, og*

$$M^{-1}AM = D$$

der  $D$  er diagonalmatrisen

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

*Bevis:* Ifølge setning 4.6.13 finnes det virkelig en lineæravbildning  $\mathbf{T}$  slik at  $\mathbf{T}(\mathbf{e}_i) = \mathbf{v}_i$  for alle  $i$ . Dersom  $M$  er matrisen til  $T$ , er da  $M\mathbf{e}_i = \mathbf{v}_i$  for alle  $i$ . Ved setning 4.6.13 finnes det også en lineæravbildning  $\hat{\mathbf{T}}$  slik at  $\hat{\mathbf{T}}(\mathbf{v}_i) = \mathbf{e}_i$  for alle  $i$ . Hvis  $\hat{M}$  er matrisen til denne lineæravbildningen, er  $\hat{M}\mathbf{v}_i = \mathbf{e}_i$  for alle  $i$ . Dermed er  $\hat{M}M\mathbf{e}_i = \hat{M}\mathbf{v}_i = \mathbf{e}_i$ . Dette betyr at  $\hat{M}M\mathbf{e}_i = I_n\mathbf{e}_i$  for alle basislementer  $\mathbf{e}_i$ , og ifølge setning 4.6.12 er da  $\hat{M}M = I_n$ . Dette betyr at  $M$  er inverterbar med  $M^{-1} = \hat{M}$

For å vise at  $D = M^{-1}AM$ , observerer vi først at  $D\mathbf{e}_i = \lambda_i\mathbf{e}_i$ . Ifølge setning 4.6.13 er da nok å vise at  $M^{-1}AM\mathbf{e}_i = \lambda_i\mathbf{e}_i$  for alle  $i$ . Men det er lett:

$$M^{-1}AM\mathbf{e}_i = M^{-1}A\mathbf{v}_i = M^{-1}\lambda_i\mathbf{e}_i = \lambda_iM^{-1}\mathbf{v}_i = \lambda_i\mathbf{e}_i$$

Dermed er setningen bevist  $\square$

I praksis er det lett å finne matrisen  $M$  i setningen ovenfor — fra setning 2.8.4 vet vi nemlig at den  $i$ -te søylen til  $M$  rett og slett er søylevektoren  $\mathbf{v}_i$ . Det er litt mer arbeid å finne  $M^{-1}$  siden vi da må invertere matrisen  $M$ . Dersom matrisen  $A$  er symmetrisk, har vi ifølge spektralteoremet 4.10.6 en ortonormal basis av egenvektorer. I dette tilfellet er det enkelt å finne  $M^{-1}$ ; det viser seg nemlig at  $M^{-1}$  er lik den transponerte matrisen  $M^T$ .

**Korollar 4.10.9** *Anta at  $A$  er en symmetrisk  $n \times n$ -matrise, la  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  være en ortonormal basis av egenvektorer, og la  $\lambda_1, \lambda_2, \dots, \lambda_n$  være de tilhørende egenverdiene. La  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  være lineæravbildningen som for alle  $i$  avbilder  $\mathbf{e}_i$  på  $\mathbf{v}_i$ , og la  $M$  være matrisen til  $\mathbf{T}$ . Da er*

$$M^T AM = D$$

der  $D$  er diagonalmatrisen

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

$\square$

*Bevis:* Det holder å vise at  $M^T = M^{-1}$ , dvs. at  $M^T M = I_n$ . Per definisjon av matrisemultiplikasjon er det  $ij$ -te elementet i  $M^T M$  lik prikkproduktet av den  $i$ -te linjen i  $M^T$  med den  $j$ -te søylen i  $M$ . Den  $i$ -te linjen i  $M^T$  er lik den  $i$ -te søylen i  $M$  som er lik  $\mathbf{v}_i$ , og den  $j$ -te søylen i  $M$  er lik  $\mathbf{v}_j$ . Altså er det  $ij$ -te elementet i  $M^T M$  lik  $\mathbf{v}_i \cdot \mathbf{v}_j$  som er 1 hvis  $i = j$  og 0 ellers (her bruker vi at basisen er ortonormal). Følgelig er  $M^T M = I_n$ , og beviset er fullført.  $\square$

Setningen ovenfor gir oss en viktig forbindelse mellom egenverdier og determinanter.

**Korollar 4.10.10** *Anta at  $A$  er en  $n \times n$ -matrise med en basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  av egenvektorer, og la  $\lambda_1, \lambda_2, \dots, \lambda_n$  være de tilhørende egenverdiene. Da er*

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$$

Determinanten er altså lik produktet av egenverdiene.

*Bevis:* La  $D$  være diagonalmatrisen i setningen ovenfor. Da er (ifølge lemma 4.9.2)

$$\det(D) = \lambda_1 \lambda_2 \cdots \lambda_n$$

Bruker vi isteden setning 4.9.14, får vi

$$\det(D) = \det(M^{-1}AM) = \det(M^{-1}) \det(A) \det(M) = \det(A)$$

der vi også har brukt at ifølge korollar 4.9.15 er  $\det(M^{-1}) = \frac{1}{\det(M)}$ .  $\square$

### Egenverdier med MATLAB

Det er lett å finne egenverdier og egenvektorer med MATLAB. Det er flere kommandoer du kan bruke, men den nyttigste er som regel

```
>> [u,v]=eig(A)
```

Denne kommandoen definerer to matriser  $u$  og  $v$ . Søylene i matrisen  $u$  er egenvektorene til  $A$ , mens  $v$  er en diagonalmatrise der elementene på diagonalen er egenverdiene til  $A$ . Egenvektorene og egenverdiene kommer i samme rekkefølge slik at den første egenverdien tilhører den første egenvektoren osv. Her er et eksempel på en kjøring:

```
>> B=[2 1 3
     4 0 3
     1 1 -2];

>> [u,v]=eig(B)

u =

```

-0.2864	-0.0000	0.3833
-0.9143	0.9487	-0.8404
-0.2864	0.3162	0.3833

```
v =

```

2.1926	0	0
0	1.0000	0
0	0	-3.1926

Vær oppmerksom på at MATLAB alltid velger egenvektorer med lengde 1. Dette er praktisk for noen formål, men fører ofte til at egenvektorene

blir mer uoversiktlig enn nødvendig. De fleste av oss ville f.eks. ha oppgitt den andre egenvektoren ovenfor som  $\begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$ , men MATLAB velger altså en normalisert variant. Mer ubegripelig er MATLABs forkjærighet for å velge negative komponenter i egenvektorene; for de fleste formål ville det være mer naturlig å velge den første egenvektoren til å være

$$\begin{pmatrix} 0.2864 \\ 0.9143 \\ 0.2864 \end{pmatrix} \text{ istedenfor } \begin{pmatrix} -0.2864 \\ -0.9143 \\ -0.2864 \end{pmatrix}$$

Når man regner videre med egenvektorer man har fått av MATLAB, kan det derfor være lurt å se om man kan forenkle dem ved å velge en annen skalering eller et annet fortegn.

Det er en ting til man bør være klar over. MATLAB vil av og til operere med en liten imaginærdel i en egenverdi/egenvektor som egentlig er reell. Det skyldes at MATLAB er et numerisk beregningsverktøy som regner med avrundede tall. Får du egenverdier/egenvektorer med en ørliten imaginærdel (eller en ørliten realdel), kan det være lurt å sjekke om dette er en avrundingsfeil før du går videre.

### Oppgaver til seksjon 4.10

**1.** Finn egenverdiene og egenvektorene til matrisen:

a)  $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$  b)  $\begin{pmatrix} 1 & 4 \\ 1 & 1 \end{pmatrix}$  c)  $\begin{pmatrix} 4 & 3 \\ 1 & 2 \end{pmatrix}$ ,

d)  $\begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$  e)  $\begin{pmatrix} 1 & 1 \\ 3 & 1 \end{pmatrix}$  f)  $\begin{pmatrix} 5 & 2 \\ -1 & 3 \end{pmatrix}$

**2.** Finn egenverdier og egenvektorene til matrisen:

a)  $\begin{pmatrix} 2 & 1 & -3 \\ 4 & 2 & 3 \\ 0 & 0 & 1 \end{pmatrix}$  b)  $\begin{pmatrix} 1 & 3 & -1 \\ 2 & 0 & 1 \\ -1 & -1 & 2 \end{pmatrix}$  (Hint: Tipp en rot i polynomet) c)  $\begin{pmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \\ 0 & 1 & 1 \end{pmatrix}$

**3.** Bruk MATLAB til å finne egenvektorene og egenverdiene til matrisen :

a)  $\begin{pmatrix} 2 & -1 & 0.5 \\ 3 & -2 & 1 \\ 3 & -1 & 2 \end{pmatrix}$  b)  $\begin{pmatrix} 2 & 0.4 & 10 \\ -2.4 & 7.3 & 0.05 \\ 4.2 & 1 & -3.2 \end{pmatrix}$  c)  $\begin{pmatrix} 3 & -2 & -2 & 4 \\ -5 & 2 & -3 & 2 \\ -2 & 2 & -8 & 3 \\ -4 & 1 & 6 & 4 \end{pmatrix}$

**4.** Finn egenverdiene og egenvektorene til matrisen  $A$  og skriv vektoren  $\mathbf{x}$  som en lineærkombinasjon av egenvektorer:

a)  $A = \begin{pmatrix} 2 & 2 \\ 2 & -1 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$

b)  $A = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} -2 \\ -6 \end{pmatrix}$

$$\text{c)} \quad A = \begin{pmatrix} -\frac{1}{2} & -\frac{4}{3} & \frac{3}{2} \\ 0 & -1 & 0 \\ \frac{3}{2} & -\frac{2}{3} & -\frac{1}{2} \end{pmatrix}, \mathbf{x} = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$$

**5.** Bruk MATLAB til å finne egenverdien og egenvektorene til matrisen  $A$ . Bruk også MATLAB til å skrive vektoren  $\mathbf{x}$  som en lineærkombinasjon av egenvektorene:

$$\text{a)} \quad A = \begin{pmatrix} 1 & -2 & 4 \\ 0 & 5 & 4 \\ 2.5 & -3 & 4 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} 0.3 \\ 2.4 \\ -3.4 \end{pmatrix}.$$

$$\text{b)} \quad A = \begin{pmatrix} 2.3 & -0.3 & 1.2 & 3 \\ 1.2 & 3 & 2.4 & -1.2 \\ 3.3 & -1.2 & 0.5 & 7 \\ -2 & 3.1 & -2.1 & 1.3 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} -1.3 \\ 2.4 \\ 0.04 \\ 4.1 \end{pmatrix}$$

**6.** La  $A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}$ . Finn en diagonalmatrise  $D$  og en inverterbar matrise  $M$  slik at  $D = M^T D M$ .

**7.** Vis at  $A$  og  $A^T$  har de samme egenverdiene. Har de også de samme egenvektorene?

**8.** Anta at  $\mathbf{v}$  er en egenvektor for både  $A$  og  $B$ . Vis at  $\mathbf{v}$  er en egenvektor for  $A + B$ .

**9.** Anta at  $\mathbf{v}$  er en egenvektor for både  $A$  og  $B$ . Vis at  $\mathbf{v}$  er en egenvektor for  $AB$ .

**10.** To  $n \times n$ -matriser  $A$  og  $B$  kalles *similære* dersom det finnes en inverterbar matrise  $P$  slik at  $B = P^{-1}AP$ . Vis at  $A$  og  $B$  da har de samme egenverdiene. Finn egenvektorene til  $B$  uttrykt ved hjelp av  $P$  og egenvektorene til  $A$ .

**11.** Anta at  $A$  er en inverterbare matrise og at  $\mathbf{v}$  er en egenvektor for  $A$  med egenverdi  $\lambda \neq 0$ . Vis at  $\mathbf{v}$  er en egenvektor for  $A^{-1}$  med egenverdi  $\lambda^{-1}$ .

**12.** Vis at dersom alle søylene i en matrise har samme sum, så er dette tallet en egenverdi for matrisen (*Hint*: Gjør noen radoperasjoner før du regner ut determinanten til  $\lambda I_n - A$ ). Bruk dette til å finne egenverdiene og egenvektorene til matrisen

$$A = \begin{pmatrix} -2 & 3 & 1 \\ 2 & 1 & 2 \\ 5 & 1 & 2 \end{pmatrix}$$

**13.** Vis at egenverdien til en  $2 \times 2$ -matrise

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

er

$$\lambda = \frac{a+d \pm \sqrt{(a-d)^2 + 4bc}}{2}$$

Bruk denne formelen til å forklare at egenverdiene til en symmetrisk (reell)  $2 \times 2$ -matrise alltid er reelle.

**14.** En symmetrisk  $n \times n$ -matrise  $A$  kalles *positiv definit* dersom  $(A\mathbf{x}) \cdot \mathbf{x} > 0$  for alle  $\mathbf{x} \neq \mathbf{0}$ .

- a) Vis at  $A$  er positiv definit hvis og bare hvis alle egenverdiene til  $A$  er strengt positive.
- b) Anta at  $A$  og  $B$  er to symmetriske matriser med strengt positive egenverdier. Vis at alle egenverdiene til  $A + B$  er strengt positive.

**15.** Anta at  $P(x) = a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$  er et polynom og at  $A$  er en kvadratisk matrise. Da er  $P(A)$  matrisen

$$P(A) = a_nA^n + a_{n-1}A^{n-1} + \dots + a_1A + a_0$$

- a) Vis at dersom  $\mathbf{v}$  er en egenvektor for  $A$  med egenverdi  $\lambda$ , så er  $\mathbf{v}$  en egenvektor for  $P(A)$  med egenverdi  $P(\lambda)$ .
- b) Vi lar nå  $P_A$  være det karakteristiske polynomet til  $A$ . Vis at  $P_A(A)\mathbf{v} = \mathbf{0}$  for alle egenvektorer  $\mathbf{v}$  til  $A$ .
- c) Vis at dersom  $A$  har en basis av egenvektorer, så er  $P_A(A) = 0$ .

(Kommentar: Det viser seg at  $P_A(A) = 0$  også når  $A$  ikke har en basis av egenvektorer. Dette kalles *Cayley-Hamiltons teorem*.)

## 4.11 Egenvektorer i praksis

I denne seksjonen skal vi se på tre eksempler som illustrerer hvordan egenvektorer og egenverdier kan brukes i praksis. Disse eksemplene er lange og ganske kompliserte, men de viser på en realistisk måte hva vi må gjøre for å analysere problemer fra den virkelige verden. Det siste eksemplet viser også hvor nyttig det er å ha et verktøy som MATLAB når matrisene blir store og uttrykkene stygge.

Før vi begynner, minner vi om følgende viktige observasjon fra kapittel 2:

**Setning 4.11.1** *Anta at  $\mathbf{v}$  er en egenvektor for  $A$  med egenverdi  $\lambda$ . Da er  $\mathbf{v}$  en egenvektor for  $A^n$  med egenverdi  $\lambda^n$ , dvs.*

$$A^n\mathbf{v} = \lambda^n\mathbf{v}$$

*Bevis:* Vi har

$$A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda A\mathbf{v} = \lambda^2\mathbf{v}$$

$$A^3\mathbf{v} = A(A^2\mathbf{v}) = A(\lambda^2\mathbf{v}) = \lambda^2 A\mathbf{v} = \lambda^3\mathbf{v}$$

osv. Før gjerne et induksjonsbevis om du vil! □

**Eksempel 1:** Vi går tilbake til handlevogneksemplet i seksjon 1.5: Et kjøpesenter har tre stativ  $X$ ,  $Y$  og  $Z$  hvor du kan hente og avlevere handlevogner. Av de vognene som starter dagen i stativ  $X$ , vil 70% avslutte den på samme sted,

10% vil ha endt opp i  $Y$ , og 20% i  $Z$ . Av de vognene som startet dagen i stativ  $Y$ , vil 30% avslutte dagen i stativ  $X$ , mens henholdsvis 50% og 20% vil havne i stativene  $Y$  og  $Z$ . De tilsvarende tallene for vogner som starter i  $Z$ , er at 40% ender dagen i  $X$ , 20% i  $Y$  og 40% i  $Z$ . Vi ordner disse tallene i en matrise  $A$  der første søyle gir fordelingen av de vognene som startet i  $X$ , andre søyle gir fordelingen av de vognene som startet i  $Y$  og tredje søyle gir fordelingen av vognene som startet i  $Z$ :

$$A = \begin{pmatrix} 0.7 & 0.3 & 0.4 \\ 0.1 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.4 \end{pmatrix}$$

Vi ser at hvis vi starter dagen med  $x_0$  handlevogner i stativ  $X$ ,  $y_0$  handlevogner i stativ  $Y$  og  $z_0$  handlevogner i stativ  $Z$ , og lar

$$\mathbf{r}_0 = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$$

så vil vektoren

$$\mathbf{r}_1 = A\mathbf{r}_0$$

gi oss fordelingen av handlevogner på slutten av dagen. Hvis handlesenteret aldri rydder opp i handlevognene, men lar dem bli stående der kundene setter dem, vil fordelingen etter  $n$  dager være gitt ved vektoren

$$\mathbf{r}_n = A^n \mathbf{r}_0$$

Vi skal se hvordan vi kan bruke egenverdiene og egenvektorene til  $A$  til å finne et uttrykk for  $\mathbf{r}_n$ .

Vi regner først ut egenverdiene til  $A$ . Etter en del regning finner vi at

$$\begin{aligned} \det(\lambda I_3 - A) &= \begin{vmatrix} \lambda - 0.7 & -0.3 & -0.4 \\ -0.1 & \lambda - 0.5 & -0.2 \\ -0.2 & -0.2 & \lambda - 0.4 \end{vmatrix} = \\ &= \lambda^3 - 1.6\lambda^2 + 0.68\lambda - 0.08 \end{aligned}$$

For å finne egenverdiene må vi altså løse tredjegradslikningen

$$\lambda^3 - 1.6\lambda^2 + 0.68\lambda - 0.08 = 0$$

Dette kan høres vanskelig ut, men ved innsetting ser vi at  $\lambda = 1$  er en løsning. Vi kan derfor polynomdividere med  $\lambda - 1$  og få

$$\lambda^3 - 1.6x\lambda^2 + 0.68\lambda - 0.08 = (\lambda - 1)(\lambda^2 - 0.6\lambda + 0.08)$$

Løsningene til annengradslikningen  $\lambda^2 - 0.6\lambda + 0.08 = 0$  er

$$\lambda = \frac{-(-0.6) \pm \sqrt{(-0.6)^2 - 4 \cdot 1 \cdot 0.08}}{2 \cdot 1} = \frac{0.6 \pm 0.2}{2} = \begin{cases} 0.4 \\ 0.2 \end{cases}$$

Egenverdiene til  $A$  er dermed  $\lambda_1 = 1$ ,  $\lambda_2 = 0.4$  og  $\lambda_3 = 0.2$ .

Neste punkt på programmet er å finne egenvektorene. Dersom  $\mathbf{v}_1 = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$  er en egenvektor med egenverdi 1, må vi ha

$$\begin{pmatrix} 0.7 & 0.3 & 0.4 \\ 0.1 & 0.5 & 0.2 \\ 0.2 & 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

Dette gir ligningssystemet

$$\begin{aligned} -0.3x + 0.3y + 0.4z &= 0 \\ 0.1x - 0.5y + 0.2z &= 0 \\ 0.2x + 0.2y - 0.6z &= 0 \end{aligned}$$

Vi ganger ligningssystemet med 10 for å slippe desimaltall og skriver deretter opp den utvidede matrisen:

$$B = \begin{pmatrix} -3 & 3 & 4 & 0 \\ 1 & -5 & 2 & 0 \\ 2 & 2 & -6 & 0 \end{pmatrix}$$

Radreduserer vi matrisen, får vi:

$$\begin{aligned} B &\stackrel{I \leftrightarrow II}{\sim} \begin{pmatrix} 1 & -5 & 2 & 0 \\ -3 & 3 & 4 & 0 \\ 2 & 2 & -6 & 0 \end{pmatrix} \stackrel{III+3I}{\sim} \begin{pmatrix} 1 & -5 & 2 & 0 \\ 0 & -12 & 10 & 0 \\ 0 & 12 & -10 & 0 \end{pmatrix} \\ &\stackrel{III+II}{\sim} \begin{pmatrix} 1 & -5 & 2 & 0 \\ 0 & -12 & 10 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \stackrel{-\frac{1}{12}II}{\sim} \begin{pmatrix} 1 & -5 & 2 & 0 \\ 0 & 1 & -\frac{5}{6} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

Vi ser at  $z$  er en fri variabel. Gitt  $z$ , kan vi regne ut  $y = \frac{5}{6}z$ ,  $x = 5y - 2z = 5 \cdot \frac{5}{6}z - 2z = \frac{13}{6}z$ . Velger vi derfor  $z = 6$ , får vi  $y = 5$  og  $x = 13$ . Dette gir egenvektoren

$$\mathbf{v}_1 = \begin{pmatrix} 13 \\ 5 \\ 6 \end{pmatrix}$$

Vi kan finne egenvektorene knyttet til de andre egenverdiene på tilsvarende måte. Vi får

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{og} \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

Siden egenverdiene er forskjellige, vet vi at  $\mathbf{v}_1$ ,  $\mathbf{v}_2$  og  $\mathbf{v}_3$  danner en basis. Begynnelsestilstanden  $\mathbf{r}_0$  kan derfor skrives som en lineærkombinasjon av egenvektorene

$$\mathbf{r}_0 = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3 \quad (4.11.1)$$

Vi skal finne konstantene  $c_1$ ,  $c_2$ ,  $c_3$  senere, men la oss foreløpig arbeide videre med uttrykket ovenfor. Ganger vi med  $A^n$  på begge sider, får vi

$$\begin{aligned} \mathbf{r}_n &= A^n \mathbf{r}_0 = c_1 A^n \mathbf{v}_1 + c_2 A^n \mathbf{v}_2 + c_3 A^n \mathbf{v}_3 = \\ &= c_1 \lambda_1^n \mathbf{v}_1 + c_2 \lambda_2^n \mathbf{v}_2 + c_3 \lambda_3^n \mathbf{v}_3 = c_1 \mathbf{v}_1 + c_2 \cdot (0.4)^n \mathbf{v}_2 + c_3 \cdot (0.2)^n \mathbf{v}_3 \end{aligned}$$

Tar vi grensen når  $n \rightarrow \infty$ , blir de to siste leddene borte, og vi sitter igjen med

$$\mathbf{r}_n \rightarrow c_1 \mathbf{v}_1 \quad \text{når } n \rightarrow \infty$$

Dette betyr at fordelingen av handlevogner nærmer seg en likevektstilstand når  $n$  går mot uendelig, og denne fordelingen er bestemt av egenvektoren til den største egenverdien.

La oss til slutt se hvordan vi kan finne konstantene  $c_1$ ,  $c_2$ ,  $c_3$ . Vi må da spesifisere begynnelsestilstanden  $\mathbf{r}_0$ , og la oss anta at handlesenteret har 144 handlevogner som alle blir plassert i stativ  $X$  i utgangspunktet. Det betyr at

$$\mathbf{r}_0 = \begin{pmatrix} 144 \\ 0 \\ 0 \end{pmatrix}$$

og at ligning (4.11.1) ovenfor kan skrives:

$$c_1 \begin{pmatrix} 13 \\ 5 \\ 6 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 144 \\ 0 \\ 0 \end{pmatrix}$$

Dette er ekvivalent med ligningssystemet

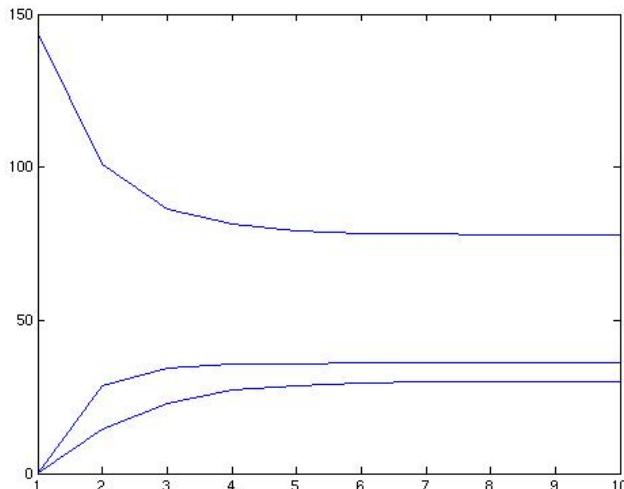
$$\begin{aligned} 13c_1 + c_2 + c_3 &= 144 \\ 5c_1 - c_2 + c_3 &= 0 \\ 6c_1 - 2c_2 &= 0 \end{aligned}$$

som har løsningene  $c_1 = 6$ ,  $c_2 = 48$ ,  $c_3 = 18$ . Setter vi dette inn i uttrykket for  $\mathbf{r}_n$  ovenfor, får vi

$$\mathbf{r}_n = \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = 6 \begin{pmatrix} 13 \\ 5 \\ 6 \end{pmatrix} + 48 \cdot (0.4)^n \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + 18 \cdot (0.2)^n \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

Likevektstilstanden i dette tilfellet er gitt ved  $6 \begin{pmatrix} 13 \\ 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 78 \\ 30 \\ 36 \end{pmatrix}$ , dvs. 78 handlevogner i stativ  $X$ , 30 i stativ  $Y$  og 36 i stativ  $Z$ .

Figuren nedenfor viser hvordan fordelingen nærmer seg likevektstilstanden. Den øverste kurven viser antall vogner i stativ  $X$ , den nest øverste antall vogner i stativ  $Z$  og den nederste antall vogner i stativ  $Y$ . ♣



**Bemerkning:** Oppførselen i eksemplet ovenfor er typisk for systemer der vi har en konstant mengde (i eksemplet: antall handlevogner) som omfordoles mellom tilstander (i eksemplet: stativene  $X$ ,  $Y$ ,  $Z$ ). I slike systemer er 1 alltid en egenverdi, og den tilhørende egenvektoren beskriver en likevekts-tilstand for systemet.

I det neste eksemplet skal vi se på et system av differensialligninger.

**Eksempel 2:** Dyreslagene I og II lever i det samme området. Dyreslag II er avhengig av dyreslag I som føde for å kunne overleve i området. Store mengder av dyreslag II vil derfor bremse veksten til dyreslag I, mens store mengder av dyreslag I fremmer veksten til dyreslag II. Dersom  $x(t)$  og  $y(t)$  er mengden av hhv. dyreslag I og dyreslag II ved tiden  $t$ , antar vi at ligningene

som styrer veksten til de to dyreslagene, er

$$\begin{aligned}x'(t) &= \frac{1}{5}x(t) - \frac{1}{20}y(t) \\y'(t) &= \frac{1}{4}x(t) - \frac{1}{10}y(t)\end{aligned}$$

Vår oppgave er å løse ligningssystemet og finne uttrykk for  $x(t)$  og  $y(t)$ . Siden ligningssystemet kobler de to ukjente funksjonene til hverandre, kan vi ikke bruke våre vanlige differensialligningsteknikker til å finne  $x(t)$  og  $y(t)$  hver for seg. Vi skal se hvordan vi kan bruke egenverdier og egenvektorer til å “dekoble” ligningssystemet slik at vi får to ligninger som kan løses hver for seg.

Vi observerer først at dersom vi innfører vektorfunksjonen

$$\mathbf{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix},$$

kan ligningssystemet skrives

$$\mathbf{r}'(t) = A\mathbf{r}(t)$$

der

$$A = \begin{pmatrix} \frac{1}{5} & -\frac{1}{20} \\ \frac{1}{4} & -\frac{1}{10} \end{pmatrix}$$

Vi finner først egenverdiene og egenvektorene til matrisen  $A$ . Det karakteristiske polynomet

$$P_A(\lambda) = \begin{vmatrix} \lambda - \frac{1}{5} & \frac{1}{20} \\ -\frac{1}{4} & \lambda + \frac{1}{10} \end{vmatrix} = \lambda^2 - \frac{1}{10}\lambda - \frac{3}{400}$$

har røttene

$$\lambda = \frac{-(-\frac{1}{10}) \pm \sqrt{(-\frac{1}{10})^2 - 4 \cdot 1 \cdot (-\frac{3}{400})}}{2 \cdot 1} = \frac{\frac{1}{10} \pm \frac{2}{10}}{2} = \begin{cases} \frac{3}{20} \\ -\frac{1}{20} \end{cases}$$

Egenverdiene er altså  $\lambda_1 = \frac{3}{20}$  og  $\lambda_2 = -\frac{1}{20}$ .

En egenvektor  $\mathbf{v}_1 = \begin{pmatrix} x \\ y \end{pmatrix}$  med egenverdi  $\lambda_1 = \frac{3}{20}$  må oppfylle ligningen

$$\begin{pmatrix} \frac{1}{5} & -\frac{1}{20} \\ \frac{1}{4} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{3}{20} \begin{pmatrix} x \\ y \end{pmatrix}$$

Multipliserer vi ut, får vi ligningene

$$\begin{aligned}\frac{1}{5}x - \frac{1}{20}y &= \frac{3}{20}x \\ \frac{1}{4}x - \frac{1}{10}y &= \frac{3}{20}y\end{aligned}$$

Flytter vi over og rydder opp litt, ser vi at begge disse ligningene er ekvivalente med

$$x - y = 0$$

Det betyr at vi kan velge  $y$  fritt, men at  $x$  da er gitt ved  $x = y$ . Velger vi  $y = 1$ , får vi  $x = 1$ , og den første egenvektoren vår er dermed

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

På tilsvarende måte må en egenvektor  $\mathbf{v}_2 = \begin{pmatrix} x \\ y \end{pmatrix}$  med egenverdi  $\lambda_2 = -\frac{1}{20}$  oppfylle ligningen

$$\begin{pmatrix} \frac{1}{5} & -\frac{1}{20} \\ \frac{1}{4} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{20} \begin{pmatrix} x \\ y \end{pmatrix}$$

Multipliserer vi ut, får vi ligningene

$$\begin{aligned} \frac{1}{5}x - \frac{1}{20}y &= -\frac{1}{20}x \\ \frac{1}{4}x - \frac{1}{10}y &= -\frac{1}{20}y \end{aligned}$$

Flytter vi over og rydder opp litt, ser vi at begge disse ligningene er ekvivalente med

$$5x - y = 0$$

Det betyr at vi kan velge  $y$  fritt, men at  $x$  da er gitt ved  $5x = y$ . Velger vi  $y = 5$ , får vi  $x = 1$ , og den andre egenvektoren vår er dermed

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

Vi går nå tilbake til differensialligningene våre. Siden  $\mathbf{v}_1, \mathbf{v}_2$  er en basis for  $\mathbb{R}^2$ , kan enhver vektor skrives som en lineærkombinasjon av  $\mathbf{v}_1$  og  $\mathbf{v}_2$ . Det betyr at det for hver  $t$  finnes tall  $c_1(t)$  og  $c_2(t)$  slik at

$$\mathbf{r}(t) = c_1(t)\mathbf{v}_1 + c_2(t)\mathbf{v}_2$$

Deriverer vi, får vi

$$\mathbf{r}'(t) = c'_1(t)\mathbf{v}_1 + c'_2(t)\mathbf{v}_2$$

Sette vi dette inn i ligningen  $\mathbf{r}'(t) = A\mathbf{r}(t)$ , ser vi at

$$\begin{aligned} c'_1(t)\mathbf{v}_1 + c'_2(t)\mathbf{v}_2 &= A(c_1(t)\mathbf{v}_1 + c_2(t)\mathbf{v}_2) = \\ &= c_1(t)A\mathbf{v}_1 + c_2(t)A\mathbf{v}_2 = c_1(t)\lambda_1\mathbf{v}_1 + c_2(t)\lambda_2\mathbf{v}_2 \end{aligned}$$

Vi har altså

$$c'_1(t)\mathbf{v}_1 + c'_2(t)\mathbf{v}_2 = c_1(t)\lambda_1\mathbf{v}_1 + c_2(t)\lambda_2\mathbf{v}_2$$

og siden  $\mathbf{v}_1, \mathbf{v}_2$  er lineært uavhengige, betyr dette at

$$c'_1(t) = \lambda_1 c_1(t) \quad \text{og} \quad c'_2(t) = \lambda_2 c_2(t)$$

Legg merke til at vi nå har “dekoblet” ligningssystemet og fått to differensielligner som kan løses hver for seg. Gjør vi det, får vi

$$c_1(t) = C_1 e^{\lambda_1 t} = C_1 e^{\frac{3}{20}t} \quad \text{og} \quad c_2(t) = C_2 e^{\lambda_2 t} = C_2 e^{-\frac{1}{20}t}$$

der  $C_1$  og  $C_2$  er konstanter. Dermed har vi

$$\mathbf{r}(t) = C_1 e^{\frac{3}{20}t}\mathbf{v}_1 + C_2 e^{-\frac{1}{20}t}\mathbf{v}_2$$

For å bestemme konstantene  $C_1$  og  $C_2$  trenger vi flere opplysninger om dyrestammene. La oss anta at det ved tiden  $t = 0$  er 3 000 dyr av slag I og 11 000 av slag II. Det betyr at

$$\begin{pmatrix} 3000 \\ 11000 \end{pmatrix} = \mathbf{r}(0) = C_1\mathbf{v}_1 + C_2\mathbf{v}_2 = C_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2 \begin{pmatrix} 1 \\ 5 \end{pmatrix}$$

Løser vi dette ligningssystemet, får vi  $C_1 = 1000$  og  $C_2 = 2000$ . Dermed har vi

$$\begin{aligned} \mathbf{r}(t) &= 1000e^{\frac{3}{20}t}\mathbf{v}_1 + 2000e^{-\frac{1}{20}t}\mathbf{v}_2 = 1000e^{\frac{3}{20}t} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2000e^{-\frac{1}{20}t} \begin{pmatrix} 1 \\ 5 \end{pmatrix} = \\ &= \begin{pmatrix} 1000e^{\frac{3}{20}t} + 2000e^{-\frac{1}{20}t} \\ 1000e^{\frac{3}{20}t} + 10000e^{-\frac{1}{20}t} \end{pmatrix} \end{aligned}$$

Antall dyr av slag I ved tiden  $t$  er dermed  $x(t) = 1000e^{\frac{3}{20}t} + 2000e^{-\frac{1}{20}t}$ , mens antall dyr av slag II er  $y(t) = 1000e^{\frac{3}{20}t} + 10000e^{-\frac{1}{20}t}$ .

Til slutt legger vi nok en gang merke til hvordan vi i dette eksemplet brukte egenvektorer til å “dekoble” systemet — de opprinnelige funksjonene  $x(t)$  og  $y(t)$  er koblet sammen gjennom ligningssystemet, mens de nye funksjonene  $c_1$  og  $c_2$  er “frakoblet” hverandre og oppfyller hver sin ligning. Slike “dekoblinger” står sentralt i mange anvendelser av egenvektorer. ♣

La oss til slutt se på et litt mer komplisert eksempel der vi får god bruk for MATLAB til å holde styr på egenverdier og egenvektorer. Eksemplet minner en del om eksempel 1, men vi ser nå på et system som vokser, og der mye av poenget er å finne hvor stor veksten er. Eksemplet viser også hva som skjer dersom vi har komplekse egenverdier.

**Eksempel 3:** Et dyreslag har en levealder på fire år. Det første året er dyrene *unger*, det andre året er de *ungdommer*, det tredje året er de *voksne* og det fjerde året er de *eldre*. Av ungene overlever 50% til året etter, av ungdommene overlever 80% til året etter og av de voksne overlever 20% til året etter. En ungdom gir i gjennomsnitt opphav til 0.5 unger som blir født året etter, en voksen gir i gjennomsnitt opphav til 2 unger som blir født året etter, og et eldre dyr gir i gjennomsnitt opphav til 0.1 unge som blir født året etter. Vi antar at vi starter med 200 dyr i hver aldersklasse, og ønsker å finne ut hvordan stammen utvikler seg.

La  $x_n$ ,  $y_n$ ,  $z_n$  og  $u_n$  være henholdsvis antall unger, ungdommer, voksne og eldre i år  $n$ . Da er

$$\begin{aligned}x_{n+1} &= 0.5y_n + 2z_n + 0.1u_n \\y_{n+1} &= 0.5x_n \\z_{n+1} &= 0.8y_n \\u_{n+1} &= 0.2z_n\end{aligned}$$

I tillegg vet vi at  $x_1 = y_1 = z_1 = u_1 = 200$ .

Det er flere måter å angripe dette problemet på. La oss først se hva som skjer når vi bruker MATLAB til å regne ut utviklingen de 50 første årene. Vi lager m-filen

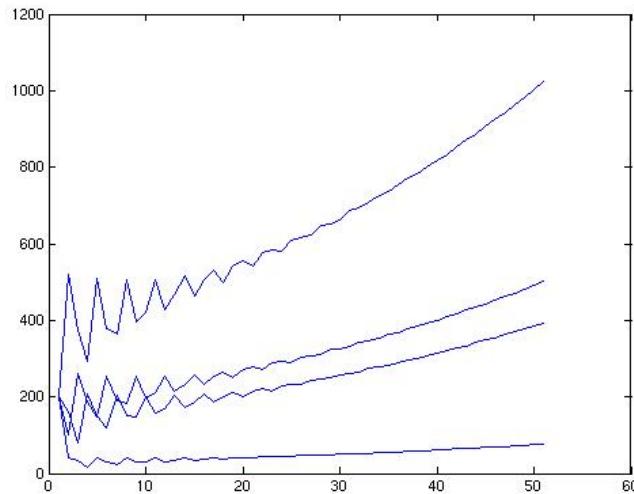
```
function [x,y,z,u]=dyrestamme(a,b,c,d,N)
x(1)=a;
y(1)=b;
z(1)=c;
u(1)=d;
for n=1:N
    x(n+1)=.5*y(n)+2*z(n)+.1*u(n);
    y(n+1)=.5*x(n);
    z(n+1)=.8*y(n);
    u(n+1)=.2*z(n);
end
```

Den neste kommandosekvensen får MATLAB til å plotte ut følgene i samme figur:

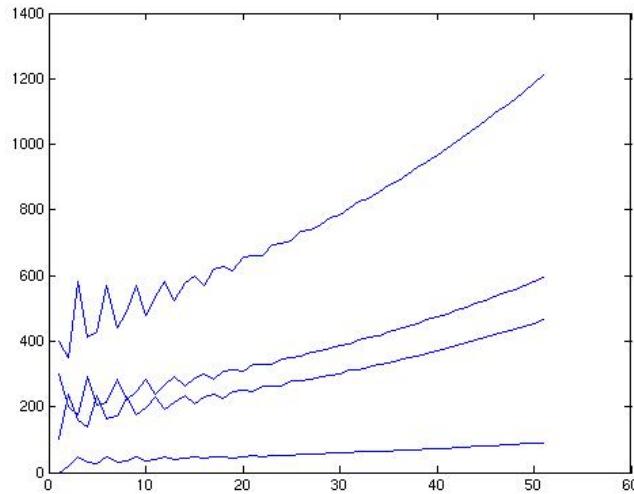
```
>> [x,y,z,u]=dyrestamme(200,200,200,200,49);
>> plot(x)
>> hold on
>> plot(y)
>> plot(z)
>> plot(u)
```

Resultatet er figuren nedenfor der den øverste kurven gir antall unger, den nest øverste antall ungdommer, den tredje øverste antall voksne og den nederste antall eldre.

Disse kurvene er ikke så lette å tolke. Det ser ut som de etter noen innledende svingninger går over i jevn vekst, og at fordelingen mellom de forskjellige aldersgruppene nærmer seg en likevekt. Men hvor kommer svingningene fra, hvor rask er veksten, og hvordan finner vi likevektsfordelingen mellom aldersgruppene?



La oss kjøre programmet en gang til med startverdier  $x_1 = 400$ ,  $y_1 = 300$ ,  $z_1 = 100$ ,  $u_1 = 0$ . Resultatet ser du på figuren nedenfor, og i hovedtrekk ligner det forbløffende på det vi fikk i stad; etter noen innledende svingninger går kurvene over i jevn vekst, og forholdet mellom aldersgruppene ligner på det vi fikk ovenfor.



Vi skal nå se hvordan vi kan bruke egenverdier og egenvektorer til å forklare disse resultatene. Det første vi observerer, er at dersom vi innfører

vektorene

$$\mathbf{r}_n = \begin{pmatrix} x_n \\ y_n \\ z_n \\ u_n \end{pmatrix}$$

så kan ligningssystemet ovenfor skrives

$$\mathbf{r}_{n+1} = A\mathbf{r}_n$$

der  $A$  er matrisen

$$A = \begin{pmatrix} 0 & .5 & 2 & .1 \\ .5 & 0 & 0 & 0 \\ 0 & .8 & 0 & 0 \\ 0 & 0 & .2 & 0 \end{pmatrix}$$

Bruker vi denne formelen gjentatte ganger, får vi

$$\mathbf{r}_n = A^{n-1}\mathbf{r}_1$$

Legg merke til at siden vi kaller begynnelsesbestanden  $\mathbf{r}_1$  og ikke  $\mathbf{r}_0$ , må  $A$  oppphøyes i  $n - 1$  og ikke  $n$ . Matematisk sett hadde det vært greiere å begynne med  $\mathbf{r}_0$  slik vi gjorde i forrige eksempel, men MATLAB begynner alltid nummereringer på 1, og vi har derfor valgt å holde oss til det siden vi MATLAB bruker såpass mye i dette eksemplet.

La oss benytte MATLAB til å finne egenverdiene og egenvektorene til  $A$ :

```
>> A=[0 .5 2 .1
.5 0 0 0
0 .8 0 0
0 0 .2 0];

>> [u,v]=eig(A)

u =
Columns 1 through 3

-0.8472      0.7917      0.7917
-0.4151    -0.2560 - 0.3675i   -0.2560 + 0.3675i
-0.3254    -0.1405 + 0.3802i   -0.1405 - 0.3802i
-0.0638     0.0887 - 0.0231i    0.0887 + 0.0231i

Column 4

-0.0000
```

```

0.0006
-0.0501
0.9987

```

v =

Columns 1 through 3

1.0206	0	0
0	-0.5053 + 0.7253i	0
0	0	-0.5053 - 0.7253i
0	0	0

Column 4

0
0
0
-0.0100

Vi har altså egenverdiene  $\lambda_1 = 1.0206$ ,  $\lambda_2 = -0.5053 + 0.7253i$ ,  $\lambda_3 = -0.5053 - 0.7253i$ ,  $\lambda_4 = -0.01$  med tilhørende egenvektorer (vi bytter fortegn på den første av dem for å slippe minuser):

$$\mathbf{v}_1 = \begin{pmatrix} 0.8472 \\ 0.4151 \\ 0.3254 \\ 0.0638 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0.7917 \\ -0.2560 - 0.3675i \\ -0.1405 + 0.3802i \\ 0.0887 - 0.0231i \end{pmatrix}$$

$$\mathbf{v}_3 = \begin{pmatrix} 0.7917 \\ -0.2560 + 0.3675i \\ -0.1405 - 0.3802i \\ 0.0887 + 0.0231i \end{pmatrix}, \quad \mathbf{v}_4 = \begin{pmatrix} 0 \\ 0.0006 \\ -0.0501 \\ 0.9987 \end{pmatrix}$$

Vi ser at de komplekse egenverdiene og egenvektorene er konjugerte av hverandre slik setning 4.10.4 sier. Vi ser også at egenverdiene er ordnet i avtagende rekkefølge:  $|\lambda_1| > |\lambda_2| = |\lambda_3| > |\lambda_4|$ .

Siden egenverdiene er forskjellige, vet vi at  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$  danner en basis. Vi kan derfor skrive starttilstanden

$$\mathbf{r}_1 = \begin{pmatrix} 200 \\ 200 \\ 200 \\ 200 \end{pmatrix}$$

som en lineærkombinasjon

$$\mathbf{r}_1 = \begin{pmatrix} 200 \\ 200 \\ 200 \\ 200 \end{pmatrix} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3 + c_4 \mathbf{v}_4$$

Vi skal bruke MATLAB til å finne koeffisientene  $c_1, c_2, c_3, c_4$ , men la oss først se hva som skjer når vi bruker  $A^{n-1}$  på ligningen ovenfor. Vi får

$$\begin{aligned} \mathbf{r}_n &= A^{n-1} \mathbf{r}_1 = c_1 A^{n-1} \mathbf{v}_1 + c_2 A^{n-1} \mathbf{v}_2 + c_3 A^{n-1} \mathbf{v}_3 + c_4 A^{n-1} \mathbf{v}_4 = \\ &c_1 \lambda_1^{n-1} \mathbf{v}_1 + c_2 \lambda_2^{n-1} \mathbf{v}_2 + c_3 \lambda_3^{n-1} \mathbf{v}_3 + c_4 \lambda_4^{n-1} \mathbf{v}_4 \end{aligned}$$

Vi setter den største egenverdien  $\lambda_1^{n-1}$  utenfor en parentes

$$\mathbf{r}_n = \lambda_1^{n-1} \left( c_1 \mathbf{v}_1 + c_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{n-1} \mathbf{v}_2 + c_3 \left( \frac{\lambda_3}{\lambda_1} \right)^{n-1} \mathbf{v}_3 + c_4 \left( \frac{\lambda_4}{\lambda_1} \right)^{n-1} \mathbf{v}_4 \right)$$

Siden  $\lambda_1$  har størst tallverdi av egenverdiene, vil alle faktorene  $\left(\frac{\lambda_2}{\lambda_1}\right)^{n-1}$ ,  $\left(\frac{\lambda_3}{\lambda_1}\right)^{n-1}$ ,  $\left(\frac{\lambda_4}{\lambda_1}\right)^{n-1}$  gå mot null når  $n$  går mot uendelig. Det betyr at

$$\lim_{n \rightarrow \infty} \left( c_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{n-1} \mathbf{v}_2 + c_3 \left( \frac{\lambda_3}{\lambda_1} \right)^{n-1} \mathbf{v}_3 + c_4 \left( \frac{\lambda_4}{\lambda_1} \right)^{n-1} \mathbf{v}_4 \right) = 0$$

Definerer vi

$$\boldsymbol{\sigma}(n) = c_2 \left( \frac{\lambda_2}{\lambda_1} \right)^{n-1} \mathbf{v}_2 + c_3 \left( \frac{\lambda_3}{\lambda_1} \right)^{n-1} \mathbf{v}_3 + c_4 \left( \frac{\lambda_4}{\lambda_1} \right)^{n-1} \mathbf{v}_4,$$

kan vi derfor skrive

$$\mathbf{r}_n = \lambda_1^{n-1} (c_1 \mathbf{v}_1 + \boldsymbol{\sigma}(n))$$

der  $\boldsymbol{\sigma}(n) \rightarrow 0$  når  $n \rightarrow \infty$ . Skriver vi ut komponentene og setter in  $\lambda_1 = 1.0206$ , får vi

$$\begin{pmatrix} x_n \\ y_n \\ z_n \\ u_n \end{pmatrix} = 1.0206^{n-1} \left( c_1 \begin{pmatrix} 0.8472 \\ 0.4151 \\ 0.3254 \\ 0.0638 \end{pmatrix} + \boldsymbol{\sigma}(n) \right)$$

Dette betyr at når  $n$  blir stor, er veksten bestemt av den største egenverdien  $\lambda_1 = 1.0206$ , og fordelingen mellom komponentene er bestemt av den tilhørende egenvektoren  $\mathbf{v}_1$ . Som du ser, minner disse resultatene om det vi fikk i Eksempel 1, men vi har fått med en vekstfaktor i tillegg.

La oss nå finne konstantene  $c_1, c_2, c_3$  og  $c_4$ . Dersom vi velger den opprinnelige begynnelsestilstanden

$$\mathbf{r}_1 = \begin{pmatrix} 200 \\ 200 \\ 200 \\ 200 \end{pmatrix}$$

får vi ligningen

$$\begin{pmatrix} 200 \\ 200 \\ 200 \\ 200 \end{pmatrix} = c_1 \begin{pmatrix} 0.8472 \\ 0.4151 \\ 0.3254 \\ 0.0638 \end{pmatrix} + c_2 \begin{pmatrix} 0.7917 \\ -0.2560 - 0.3675i \\ -0.1405 + 0.3802i \\ 0.0887 - 0.0231i \end{pmatrix} + \\ + c_3 \begin{pmatrix} 0.7917 \\ -0.2560 + 0.3675i \\ -0.1405 - 0.3802i \\ 0.0887 + 0.0231i \end{pmatrix} + c_4 \begin{pmatrix} 0 \\ 0.0006 \\ -0.0501 \\ 0.9987 \end{pmatrix}$$

Innfører vi matrisen

$$D = \begin{pmatrix} 0.8472 & 0.7917 & 0.7917 & 0 \\ 0.4151 & -0.2560 - 0.3675i & -0.2560 + 0.3675i & 0.0006 \\ 0.3254 & -0.1405 + 0.3802i & -0.1405 - 0.3802i & -0.0501 \\ 0.0638 & 0.0887 - 0.0231i & 0.0887 + 0.0231i & 0.9987 \end{pmatrix}$$

kan vi bruke MATLAB til å finne vektoren

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}$$

ved å taste

`>> c=D\r1`

Vi får  $c_1 = 436.59, c_2 = -107.29 - 49.34i, c_3 = -107.29 + 49.34i, c_4 = 193.72$ . Legg merke til at koeffisientene  $c_2$  og  $c_3$  til de komplekse egenverdiene er konjugerte.

Vi har ennå ikke forklart hvor svingningene i figuren kommer fra. Det viser seg at de kommer fra de komplekse egenverdiene. Skriver vi den komplekse egenverdien  $\lambda_2$  på polarform  $\lambda_2 = re^{i\theta}$ , ser vi at

$$\lambda_2^{n-1} = r^{n-1}e^{i(n-1)\theta} = r^{n-1} \left( \cos((n-1)\theta) + i \sin((n-1)\theta) \right)$$

Cosinus- og sinus-leddene får uttrykket til å svinge, men i dette tilfellet vil svingningene dø ut etter hvert fordi  $r < 1$  og  $r^{n-1} \rightarrow 0$  når  $n \rightarrow \infty$ .



### Oppgaver til seksjon 4.11

**1.** Finn to følger  $\{x_n\}$ ,  $\{y_n\}$  slik at

$$\begin{aligned}x_{n+1} &= x_n + 3y_n \\y_{n+1} &= 2x_n + 2y_n\end{aligned}$$

når  $x_0 = 5$ ,  $y_0 = -5$ .

**2.** Finn funksjonene  $x(t)$ ,  $y(t)$  slik at

$$\begin{aligned}x'(t) &= x(t) + 8y(t) \\y'(t) &= 2x(t) + y(t)\end{aligned}$$

og  $x(0) = 1$ ,  $y(0) = 6$ .

**3.** (Eksamens i MAT1100, 13/, 2008)

a) Finn egenverdiene og egenvektorene til matrisen  $A = \begin{pmatrix} 1.1 & -0.2 \\ 0.1 & 0.8 \end{pmatrix}$

b) To dyreslag bor i det samme området. Dersom det er  $x_n$  og  $y_n$  dyr av hvert slag ett år, vil det året etter være

$$\begin{aligned}x_{n+1} &= 1.1x_n - 0.2y_n \\y_{n+1} &= 0.1x_n + 0.8y_n\end{aligned}$$

dyr av hvert slag. Finn uttrykk for  $x_n$  og  $y_n$  dersom  $x_0 = 3000$ ,  $y_0 = 1000$ . Hva skjer med bestandene når  $n$  går mot uendelig?

**4.** (Eksamens i MA 001, 31/5 1994) I barnehagen har Viktoria og Emil fått hvert sitt glass saft med nøyaktig like mye saft til hver. Viktoria er imidlertid ikke helt fornøyd siden saften til Emil inneholder dobbelt så mye sukker som hennes. Glassene er ikke fullere enn at det går an å helle litt fra det ene over i det andre, og smart som hun er, får Viktoria med Emil på følgende lek: Hun heller  $\frac{1}{9}$  av sin saft over i glasset til Emil, ber ham røre godt rundt og så helle den samme mengden saft tilbake i hennes glass slik at de igjen har like mye saft.

Blandeprosedyren ovenfor gjentas flere ganger. La  $x_n$  og  $y_n$  være sukkermengden i glassene til henholdsvis Viktoria og Emil etter at prosedyren er utført  $n$  ganger.

a) Vis at

$$\begin{aligned}x_{n+1} &= 0.9x_n + 0.1y_n \\y_{n+1} &= 0.1x_n + 0.9y_n\end{aligned}$$

b) La  $M$  være matrisen slik at  $\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = M \begin{pmatrix} x_n \\ y_n \end{pmatrix}$ . Finn egenverdiene og egenvektorene til  $M$ .

c) Skriv  $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$  som en lineærkombinasjon av egenvektorer for  $M$ , og finn  $M^n \begin{pmatrix} 2 \\ 4 \end{pmatrix}$ .

- d) Hvor mange ganger må blandeprosedyren utføres for at forholdet mellom sukkerinnholdet i Viktorias saft og Emils saft er minst 0.95?

**5.** (Eksamensoppgave fra MA 001, 2/6 1997) I denne oppgaven er

$$M = \begin{pmatrix} \frac{1}{2} & \frac{5}{9} \\ \frac{5}{9} & \frac{1}{2} \end{pmatrix}$$

- a) Finn egenverdiene og egenvektorene til  $M$ .

I resten av oppgaven skal vi studere en modell for hvordan en ufarlig infeksjons-sykdom sprer seg i en befolkning. Vi deler befolkningen i to grupper — de som er immune for sykdommen og de som er mottagelige for smitte. De fleste som nylig har hatt sykdommen vil være immune, men mange vil miste immuniteten etter som tiden går.

I modellen ønsker vi å studere hvor mange som er immune, og hvor mange som er mottagelige for smitte etter 0, 10, 20, 30, ... år. Vi lar  $y_n$  være antall immune etter  $10n$  år og  $x_n$  antall mottagelige ved samme tidspunkt. Vi har følgende observasjoner:

Av dem som er immune et år, vil  $\frac{4}{9}$  fortsatt være immune 10 år senere,  $\frac{1}{9}$  vil være døde og resten vil være mottagelige for smitte.

Av dem som er mottagelige for smitte et år, vil halvparten være immune 10 år senere,  $\frac{1}{9}$  vil være døde og resten vil være mottagelige for smitte.

I løpet av en 10-årsperiode vil befolkningen få et tilskudd pga. fødsel og innvandring. Dette tilskuddet er  $\frac{1}{6}$  av befolkningstallet ved begynnelsen av perioden, og ved slutten av perioden vil  $\frac{1}{3}$  av de nye individene være immune og resten mottagelige for smitte.

- b) Vis at

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = M \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

- c) Anta at  $x_0 = 8$  millioner og at  $y_0 = 2$  millioner. Finn  $x_n$  og  $y_n$ .

- d) Etter som tiden går vil prosentdelen av immune nærme seg en grense. Hva er denne grensen?

**6.** (Eksamensoppgave fra MA 001, 3/6 1996) En oljemilliardær bestemmer seg for å satse på turisme. Hun kjøper 1 000 hytter på fjellet. Hyttene leies ut for ett år av gangen. Hytteeieren finner ut at 80% av hyttene som er leid ut ett år, også er leid ut året etter, mens 70% av hyttene som er tomme ett år, også er tomme neste år.

La  $x_n$  være antall uteleide og  $y_n$  antall tomme hytter i år  $n$ .

- a) Finn en matrise  $M$  slik at

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = M \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

- b) Ett år er 550 hytter uteleid. Hvor mange var uteleid året før?

- c) Finn egenverdiene og egenvektorene til  $M$ .

Det første året (år 0) er halvparten av hyttene leid ut mens resten står tomme.

- d) Finn  $x_n$  og  $y_n$ .

- e) Det første året er nettofortjenesten pr. uteid hytte 10 000 kroner, mens utgiftene forbundet med en tom hytte er 4 000 kroner. På grunn av elde og slitasje øker utgiftene for tomme hytter med 10% per år, mens nettofortjenesten for uteide hytter ligger stabilt på 10 000 kroner i året. La  $P_n$  være nettofortjeneste i år  $n$ , dvs. inntekter minus utgifter. Finn  $P_n$  uttrykt ved  $n$  og begrunn at hytteierene etter hvert taper penger.

**7.** (Eksamens i MA 001, 29/11 1996, litt tilpasset) Det var en gang en bestand av biller som levde i en gammel verneverdig trebygning. Vi deler billebestanden inn i tre aldersgrupper: *nyfødte* (0 uker gamle), *voksne* (1 uke gammel) og *gamle* (2 uker gamle). La  $x_n, y_n, z_n$  være henholdsvis antall nyfødte, voksne og gamle biller ved tiden  $t = n$ , der tiden regnes i uker. Vi antar at alle billene som er nyfødte en uke, overlever til uken etter, men at bare halvparten av de voksne billene overlever til neste uke, og at ingen gamle biller lever en uke til. En voksen bille gir i gjennomsnitt opphav til 3 nye biller som blir født uken etter, mens en gammel bille i gjennomsnitt gir opphav til 4 nye biller som blir født uken etter.

- a) Finn en matrise  $M$  slik at

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} = M \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}$$

- b) Finn egenverdiene og egenvektorene til  $M$ .  
c) Anta at det er 24 nyfødte, ingen voksne og 6 gamle biller i trebygningen ved  $t = 0$ . Hvor mange biller er det i hver aldersgruppe  $n$  uker senere?

**8.** (Eksamens i MA 001, 2/6 1993, litt tilpasset) I en by finnes det tre avisar, en skandaleavis  $A$ , en rimelig seriøs avis  $B$  og en svært seriøs avis  $C$ . I løpet av fem år skjer det følgende forandringer:

Avisene  $A$  og  $C$  får et antall nye kjøpere (som ikke har kjøpt noen avis tidligere) tilsvarende 10% av det antall kjøpere de hadde ved starten av perioden, mens avis  $C$  får en tilvekst av nye kjøpere på 20%. 10% av leserne av avis  $A$  slutter med  $A$  og går over til  $B$ . 10% av leserne av  $B$  slutter med  $B$  og går over til  $A$  og en annen gruppe på 10% går over til  $C$ . 10% av kjøperne av  $C$  slutter med  $C$  og begynner å kjøpe  $B$ . Ellers beholder alle kjøperne sin gamle avis.

La  $x_n, y_n, z_n$  være salgstallene for henholdsvis avis  $A$ ,  $B$  og  $C$  i året  $5n$ , for  $n = 1, 2, 3, \dots$

- a) Finn en matrise  $M$  slik at

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \\ z_{n+1} \end{pmatrix} = M \begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix}$$

- b) Finn egenverdiene og egenvektorene til  $M$ .  
c) Finn  $x_n, y_n, z_n$  uttrykt ved  $x_0, y_0, z_0$  og  $n$ . Vis at forholdet mellom salgstallene nærmer seg grenser som er uavhengig av starttilstanden. Finn  $k_1 = \lim_{n \rightarrow \infty} \frac{x_n}{y_n}$  og  $k_2 = \lim_{n \rightarrow \infty} \frac{y_n}{z_n}$ .

**9.** To fiskeslag lever i samme innsjø. Fiskeslag II er avhengig av fiskeslag I for å opprettholde bestanden. Dersom  $x(t)$  er antall fisk av slag I ved tiden  $t$ , og  $y(t)$  er antall fisk av slag II ved tiden  $t$ , regner vi at

$$x'(t) = 0.02x(t) - 0.03y(t)$$

$$y'(t) = 0.01x(t) - 0.02y(t)$$

- a) Vis at egenverdiene og egenvektorene til matrisen  $A = \begin{pmatrix} 0.02 & -0.03 \\ 0.01 & -0.02 \end{pmatrix}$  er  $\lambda_1 = 0.01$ ,  $\lambda_2 = -0.01$ ,  $\mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

- b) La  $\mathbf{r}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}$  og skriv  $\mathbf{r}(t) = c_1(t)\mathbf{v}_1 + c_2(t)\mathbf{v}_2$ . Vis at  $c_1$  og  $c_2$  tilfredsstiller differensialligningene

$$c'_1(t) = 0.01c_1(t)$$

$$c'_2(t) = -0.01c_2(t)$$

- c) Anta at  $x(0) = 5\,000$  og  $y(0) = 1\,000$ . Finn  $x(t)$  og  $y(t)$ . Hvordan går det med forholdet  $\frac{x(t)}{y(t)}$  mellom antall fisk av slag I og antall fisk av slag II når  $t$  blir stor?
- d) Bruk MATLAB til å plotte  $x(t)$  og  $y(t)$  i samme koordinatsystem.

**10.** (Eksamens i MA 001, 3/12 1993) To arter, et rovdyr og et byttedyr, lever i samme område. La  $x(t)$  være antall rovdyr og  $y(t)$  antall byttedyr ved tiden  $t$  (t måles i år). Anta at  $x(0) = 500$ ,  $y(0) = 1000$ . Vi skal betrakte to enkle modeller for  $x(t)$  og  $y(t)$ .

- a) I den første modellen antar vi at  $x$  og  $y$  tilfredsstiller

$$\begin{aligned} x'(t) &= x(t) + y(t) \\ y'(t) &= -x(t) + y(t) \end{aligned}$$

Hva blir  $x(t)$  og  $y(t)$  i dette tilfellet (husk at  $e^{a+ib} = e^a(\cos b + i \sin b)$ )?

- b) I den andre modellen antar vi at  $x$  og  $y$  tilfredsstiller

$$\begin{aligned} x'(t) &= -2x(t) + 4y(t) \\ y'(t) &= x(t) - 2y(t) \end{aligned}$$

Hva blir  $x(t)$  og  $y(t)$  da?

- c) Observasjonene våre tyder på at det på et visst tidspunkt ikke er flere byttedyr igjen. Hvilken modell passer med disse observasjonene? Hvor mange måneder tar det ifølge modellen før dette skjer? I den andre modellen vil antall rovdyr og antall byttedyr etter hvert stabilisere seg seg på visse verdier. Bestem disse verdiene.

**11.** En by har nettopp innført et system med “bysykler” der man kan låne en sykkel fra et sykkelstativ og levere den fra seg ved et annet (eller ved det samme om man bare skal en tur i nærområdet). Foreløpig har byen 4 stativer som vi kaller  $X, Y, Z, U$ . Myndighetene er interessert i å undersøke lånemønsteret for syklene,

Utgangspunkt	Prosentfordeling neste måned				
	i $X$	i $Y$	i $Z$	i $U$	ute av drift
$X$	40%	20%	20%	10%	10%
$Y$	10%	40%	20%	20%	10%
$Z$	10%	20%	25%	25%	20%
$U$	30%	20%	20%	20%	10%

og har derfor en månedlig undersøkelse av hvor de forskjellige syklene befinner seg. Denne undersøkelsen viser at av de syklene som befant seg i stativ  $X$  en måned, befinner 40% seg i  $X$  måneden etter, 20% befinner seg i  $Y$ , 20% i  $Z$  og 10% i  $U$ , mens 10% er ute av drift fordi de enten er forsvunnet eller inne til vedlikehold. Tilsvarende tall for syklene som opprinnelig var i  $Y, Z$  og  $U$ , fremgår av tabellen ovenfor. I forbindelse med undersøkelsen får hvert stativ påfyll med nye/reparerte sykler. Påfyllet tilsvarer 15% av antall sykler som står i stativet.

a) La  $x_n, y_n, z_n, u_n$  være antall sykler i henholdsvis  $X, Y, Z, U$  rett etter den  $n$ -te undersøkelsen (og rett etter påfyllet av nye/reparerte sykler). Forklar at

$$\begin{aligned}x_{n+1} &= 0.46x_n + 0.115y_n + 0.115z_n + 0.345u_n \\y_{n+1} &= 0.23x_n + 0.46y_n + 0.23z_n + 0.23u_n \\z_{n+1} &= 0.23x_n + 0.23y_n + 0.2875z_n + 0.23u_n \\u_{n+1} &= 0.115x_n + 0.23y_n + 0.2875z_n + 0.23u_n\end{aligned}$$

b) Skriv en m-fil som gitt  $x_1, y_1, z_1, u_1$  returnerer  $x_n, y_n, z_n, u_n$  for  $n$  fra 1 til 50.

c) Velg  $x_1 = y_1 = z_1 = u_1 = 100$ , og bruk MATLAB til å tegne følgene  $\{x_n\}, \{y_n\}, \{z_n\}, \{u_n\}$  i samme koordinatsystem.

d) Lag en ny MATLAB-figur der du plotter følgene  $\{\frac{x_n}{1.0071^{n-1}}\}, \{\frac{y_n}{1.0071^{n-1}}\}, \{\frac{z_n}{1.0071^{n-1}}\}, \{\frac{u_n}{1.0071^{n-1}}\}$  i samme koordinatsystem.

e) Gjenta plottingen i c) og d), men bruk starttilstanden  $x_1 = 200, y_1 = 0, z_1 = 0, u_1 = 200$ . Ser du et mønster? Eksperimenter gjerne med andre startverdier.

f) La

$$\mathbf{w}_n = \begin{bmatrix} x_n \\ y_n \\ z_n \\ u_n \end{bmatrix}$$

være fordelingen av sykler den  $n$ -te måneden. Forklar at  $\mathbf{w}_{n+1} = A\mathbf{w}_n$  der  $A$  er matrisen

$$A = \begin{bmatrix} 0.46 & 0.115 & 0.115 & 0.345 \\ 0.23 & 0.46 & 0.23 & 0.23 \\ 0.23 & 0.23 & 0.2875 & 0.23 \\ 0.115 & 0.23 & 0.2875 & 0.23 \end{bmatrix}$$

Forklar også hvorfor  $\mathbf{w}_n = A^{n-1}\mathbf{w}_1$ .

g) Bruk MATLAB til å vise at matrisen  $A$  ovenfor har fire egenverdier  $\lambda_1, \lambda_2, \lambda_3, \lambda_4$  med tilhørende egenvektorer  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ . Sørg for å ordne rekkefølgen slik at  $\lambda_1$  er egenverdiens med størst tallverdi. (NB: MATLAB gir av og til egenvektorer der alle komponentene er negative. For å få en egenvektor som er greiere å arbeide med, kan du da bare fjerne alle minustegnene i denne vektoren.)

h) Vis at egenvektorene  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$  til  $A$  danner en basis for  $\mathbf{R}^4$ , dvs. at de er lineært uavhengige og utspenner hele  $\mathbf{R}^4$  (bruk gjerne MATLAB).

i) Velg

$$\mathbf{w}_1 = \begin{bmatrix} 100 \\ 100 \\ 100 \\ 100 \end{bmatrix}$$

og skriv denne vektoren som en lineærkombinasjon av  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$  (bruk gjerne MATLAB). Finn

$$\lim_{n \rightarrow \infty} \frac{A^{n-1}\mathbf{w}_1}{\lambda_1^{n-1}}$$

Sammenlign med resultatene i oppgave 1c) og d). Hva regner du med å få dersom du velger en annen starttilstand  $\mathbf{w}_1$ ?

## 4.12 Spektralteoremet

I denne seksjonen skal vi bevise spektralteoremet for symmetriske matriser. Dette er et av de viktigste resultatene i lineær algebra og det er også modell for mer generelle resultater i andre deler av matematikken. Beviset er i seg selv ikke vanskeligere enn andre vi allerede har vært borti, men det er ganske langt fordi vi ennå ikke har gjort alle forberedelsene vi trenger. Vær oppmerksom på at beviset bygger på seksjon 4.7, og har du ikke lest den seksjonen før, bør du gjøre det nå.

La oss først minne om hva spektralteoremet sier.

**Teorem 4.12.1 (Spektralteoremet for symmetriske matriser)** *Anta at  $A$  er en symmetrisk  $n \times n$ -matrise. Da er alle egenverdiene til  $A$  reelle, og det finnes en ortonormal basis for  $\mathbb{R}^n$  som består av egenvektorer til  $A$ .*

Vi skal først vise at alle egenverdiene og egenvektorene til en symmetrisk matrise er reelle. Nøkkelen er dette enkle, men viktige resultatet:

**Setning 4.12.2** Anta at  $A$  er en reell  $n \times n$ -matrise og at  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$  er to komplekse søylevektorer. Da er

$$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^T \mathbf{y})$$

*Bevis:* For alle søylevektorer  $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$  er

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \bar{\mathbf{v}}$$

der produktet på venstresiden er et skalarprodukt mens produktet på høyresiden er et matriseprodukt (ser du ikke dette med det samme, så skriv ut begge sidene av formelen og husk at komplekse skalarprodukter inneholder en komplekskonjugering i annen faktor). Bruker vi denne formelen samt regnereglene for transponering og matriseprodukter, får vi

$$(A\mathbf{x}) \cdot \mathbf{y} = (A\mathbf{x})^T \bar{\mathbf{y}} = (\mathbf{x}^T A^T) \bar{\mathbf{y}} = \mathbf{x}^T (A^T \bar{\mathbf{y}})$$

Siden  $A$  er reell, er  $A^T = \overline{A^T}$ , og dermed får vi videre

$$\mathbf{x}^T (A^T \bar{\mathbf{y}}) = \mathbf{x}^T (\overline{A^T \mathbf{y}}) = \mathbf{x} \cdot (A^T \mathbf{y})$$

□

**Bemerkning:** Argumentet ovenfor kan være litt vanskelig å lese fordi notasjonene for matriseprodukt og skalarprodukt er så like. I mer avanserte bøker er det derfor vanlig å bruke andre måter å angi skalarprodukt på, f.eks.  $\langle \mathbf{x}, \mathbf{y} \rangle$  istedenfor  $\mathbf{x} \cdot \mathbf{y}$ . Med denne notasjonen kan resultatet ovenfor skrives  $\langle A\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, A^T \mathbf{y} \rangle$  og nå blir det kanskje enklere å se hva setningen sier — vi kan føre matrisen  $A$  over på den andre siden av skalarproduktet forutsatt at vi transponerer den.

Vi tar med et viktig korollar av setningen ovenfor:

**Korollar 4.12.3** En reell  $n \times n$ -matrise er symmetrisk hvis og bare hvis

$$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})$$

for alle  $x, y \in \mathbb{R}^n$ .

*Bevis:* Dersom  $A$  er symmetrisk, er

$$(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^T \mathbf{y}) = \mathbf{x} \cdot (A\mathbf{y})$$

ifølge setningen ovenfor.

Dersom  $A$  ikke er symmetrisk, finnes det minst ett par av indekser  $(i, j)$  slik at den  $(i, j)$ -te komponenten  $a_{ij}$  til  $A$  er forskjellig fra den  $(i, j)$ -te komponenten  $a_{ji}$  til  $A^T$ . Velger vi  $\mathbf{x} = \mathbf{e}_i$  og  $\mathbf{y} = \mathbf{e}_j$ , får vi

$$(A\mathbf{x}) \cdot \mathbf{y} = a_{ji} \quad \text{og} \quad \mathbf{x} \cdot (A\mathbf{y}) = a_{ij}$$

Altså er  $(A\mathbf{x}) \cdot \mathbf{y} \neq \mathbf{x} \cdot (A\mathbf{y})$ .  $\square$

Vi kan nå vise det første resultatet vi er på jakt etter.

**Setning 4.12.4** *Dersom  $A$  er en symmetrisk (reell)  $n \times n$ -matrise, så er alle egenverdiene til  $A$  reelle. Til enhver reell egenverdi hører det en reell egenvektor.*

*Bevis:* Anta at  $\lambda$  er en (muligens kompleks) egenverdi for  $A$  med en (muligens kompleks) egenvektor  $\mathbf{v}$ . Da er

$$A\mathbf{v} \cdot \mathbf{v} = (\lambda\mathbf{v}) \cdot \mathbf{v} = \lambda(\mathbf{v} \cdot \mathbf{v}) = \lambda|\mathbf{v}|^2$$

På den annen side er (husk at vi må bruke komplekse skalarprodukter siden  $\lambda$  og  $\mathbf{v}$  kan være komplekse)

$$A\mathbf{v} \cdot \mathbf{v} = \mathbf{v} \cdot (A\mathbf{v}) = \mathbf{v} \cdot (\lambda\mathbf{v}) = \bar{\lambda}(\mathbf{v} \cdot \mathbf{v}) = \bar{\lambda}|\mathbf{v}|^2$$

(se punkt f) i setning 1.3.1 dersom du ikke skjønner hvor komplekskonjugeringen kommer fra). Dermed har vi  $\lambda|\mathbf{v}|^2 = \bar{\lambda}|\mathbf{v}|^2$ , og forkorter vi med  $|\mathbf{v}|^2$ , får vi  $\lambda = \bar{\lambda}$  som viser at  $\lambda$  er reell. Det gjenstår å vise at det finnes en *reell* egenvektor med  $\lambda$ . Dette kan gjøres på flere måter — det enkleste er kanskje å observere at siden  $\det(\lambda I_n - A) = 0$ , så har ligningen  $(\lambda I_n - A)\mathbf{x} = \mathbf{0}$  ifølge teorien vår en ikke-null, reell løsning.  $\square$

Den neste setningen er det første spede skrittet på veien mot eksistensdelen av spekralteoremet — den viser at en symmetrisk matrise i hvert fall har én (reell) egenvektor.

**Setning 4.12.5** *Enhver symmetrisk matrise  $A$  har minst én (reell) egenvektor.*

*Bevis:* Ifølge algebraens fundamentalteorem (*Kalkulus*, teorem 3.5.1) har det karakteristiske polynomet  $P_A(\lambda)$  minst én rot. Det betyr at  $A$  har minst én egenverdi, og ifølge setningen ovenfor er denne egenverdien reell og har en reell egenvektor.  $\square$

Denne setningen er brekkstangen vi trenger for å komme igang med et induksjonsbevis for spekralteoremet. Beviset går ved induksjon på dimensjonen til rommet som matrisen virker på. Disse rommene er generelle underrom av  $\mathbb{R}^n$ , og for å få induksjonsbeviset til å fungere, må vi først generalisere symmetribegrepet til underrom.

### Symmetriske lineæravbildninger

Husk (fra seksjon 2.8) at en *lineæravbildning* fra  $\mathbb{R}^n$  til  $\mathbb{R}^m$  er en funksjon  $\mathbf{T} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  slik at

- (i)  $\mathbf{T}(c\mathbf{x}) = c\mathbf{T}(\mathbf{x})$  for alle  $c \in \mathbb{R}$  og alle  $\mathbf{x} \in \mathbb{R}^n$
- (ii)  $\mathbf{T}(\mathbf{x} + \mathbf{y}) = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$  for alle  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

Til enhver slik lineæravbildning finnes det en matrise  $A$  slik at  $A\mathbf{x} = \mathbf{T}(\mathbf{x})$  for alle  $\mathbf{x} \in \mathbb{R}^n$  (se setning 2.8.4), og vi har derfor stort sett arbeidet med matriser istedenfor lineæravbildninger. I denne seksjonen trenger vi imidlertid å arbeide med lineæravbildninger definert på underrom av  $\mathbb{R}^n$ , og da er det tungvint å bruke matriser istedenfor lineæravbildninger.

Vi begynner med å generalisere begrepet lineæravbildning til underrom.

**Definisjon 4.12.6** *Anta at  $H$  er et underrom av  $\mathbb{R}^n$ . En funksjon  $\mathbf{T} : H \rightarrow H$  kalles en lineæravbildning dersom*

- (i)  $\mathbf{T}(c\mathbf{x}) = c\mathbf{T}(\mathbf{x})$  for alle  $c \in \mathbb{R}$  og alle  $\mathbf{x} \in H$
- (ii)  $\mathbf{T}(\mathbf{x} + \mathbf{y}) = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$  for alle  $\mathbf{x}, \mathbf{y} \in H$

Vi skal ofte arbeide med lineæravbildninger som i utgangspunktet er definert på en større mengde enn  $H$  (gjerne på hele  $\mathbb{R}^n$ ). Vi kan oppfatte disse som lineæravbildninger fra  $H$  til  $H$  dersom  $\mathbf{T}(\mathbf{x}) \in H$  for alle  $\mathbf{x} \in H$ . I så fall sier vi at  $\mathbf{T}$  *avbilder  $H$  inn i  $H$* .

Med korollar 4.12.3 som inspirasjon er det lett å definere *symmetriske* lineæravbildninger.

**Definisjon 4.12.7** *En lineæravbildning  $\mathbf{T} : H \rightarrow H$  kalles symmetrisk dersom*

$$\mathbf{T}(\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot \mathbf{T}(\mathbf{y}) \quad \text{for alle } \mathbf{x}, \mathbf{y} \in H$$

Det er også lett å generalisere begrepene egenverdi og egenvektor til underrom:

**Definisjon 4.12.8** *En ikke-null vektor  $\mathbf{v} \in H$  kalles en egenvektor for lineæravbildningen  $\mathbf{T} : H \rightarrow H$  dersom det finnes et tall  $\lambda \in \mathbb{R}$  slik at  $\mathbf{T}(\mathbf{v}) = \lambda\mathbf{v}$ . Tallet  $\lambda$  kalles en egenverdi for  $\mathbf{T}$ .*

**Bemerkning:** Legg merke til at vi nå bare tillater *reelle* egenverdier og egenvektorer. Det er for å slippe å snakke om den komplekse versjonen av underrommet  $H$ . Siden vi for øyeblikket kun er interessert i symmetriske lineæravbildninger (som bare har reelle egenvektorer og egenverdier), taper vi ikke noe på å holde oss til det reelle tilfellet.

Det neste resultatet generaliserer setning 4.12.5 til symmetriske lineæravbildninger på underrom. Beviset kan se ut som et skittent triks, men det er faktisk eksempel på en generell teknikk (men hva er en matematisk teknikk annet enn et skittent triks som brukes mer enn én gang?) Synes du beviset er mysteriøst, kan det være lurt å ta en kikk på bemerkningen som følger etter det (men bemerkningen forutsetter nok at du har prøvd å lese beviset først!)

**Setning 4.12.9** *Anta at  $H \neq \{\mathbf{0}\}$  er et underrom av  $\mathbb{R}^n$ . Da har enhver symmetrisk lineæravbildning  $\mathbf{T} : H \rightarrow H$  minst én egenvektor.*

*Bevis:* Velg en ortonormal basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  for  $H$ . Vi kan skrive bildene  $\mathbf{T}(\mathbf{v}_1), \mathbf{T}(\mathbf{v}_2), \dots, \mathbf{T}(\mathbf{v}_m)$  som lineærkombinasjoner av basisvektorene:

$$\begin{aligned}\mathbf{T}(\mathbf{v}_1) &= c_{11}\mathbf{v}_1 + c_{21}\mathbf{v}_2 + \cdots + c_{m1}\mathbf{v}_m \\ \mathbf{T}(\mathbf{v}_2) &= c_{12}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \cdots + c_{m2}\mathbf{v}_m \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \\ \mathbf{T}(\mathbf{v}_m) &= c_{1m}\mathbf{v}_1 + c_{2m}\mathbf{v}_2 + \cdots + c_{mm}\mathbf{v}_m\end{aligned}$$

Siden basisen er ortonormal, vet vi fra setning 4.10.7 at

$$c_{ij} = \mathbf{T}(\mathbf{v}_j) \cdot \mathbf{v}_i$$

Tilsvarende er

$$c_{ji} = \mathbf{T}(\mathbf{v}_i) \cdot \mathbf{v}_j$$

Siden  $\mathbf{T}$  er symmetrisk, er  $\mathbf{T}(\mathbf{v}_j) \cdot \mathbf{v}_i = \mathbf{v}_j \cdot \mathbf{T}(\mathbf{v}_i) = \mathbf{T}(\mathbf{v}_i) \cdot \mathbf{v}_j$ . Det betyr at  $c_{ji} = c_{ij}$ , og følgelig er matrisen

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{pmatrix}$$

symmetrisk. Vi vet fra setning 4.12.5 at  $C$  har minst én reell egenverdi  $\lambda$  med tilhørende (reell) egenvektor

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

Dette betyr at  $C\mathbf{x} = \lambda\mathbf{x}$ , og skriver vi ut denne ligningen komponentvis, får vi

$$\begin{aligned} x_1c_{11} + x_2c_{12} + \dots + x_mc_{1m} &= \lambda x_1 \\ x_1c_{21} + x_2c_{22} + \dots + x_mc_{2m} &= \lambda x_2 \\ &\vdots && \vdots && \vdots \\ x_1c_{m1} + x_2c_{m2} + \dots + x_mc_{mm} &= \lambda x_m \end{aligned}$$

Vi er nå kommet til poenget som er å vise at  $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \dots + x_m\mathbf{v}_m$  er en egenvektor for  $\mathbf{T}$  med egenverdi  $\lambda$ . Dette er bare et regnestykke (legg merke til hvordan vi bruker ligningene ovenfor):

$$\begin{aligned} \mathbf{T}(\mathbf{v}) &= x_1\mathbf{T}(\mathbf{v}_1) + x_2\mathbf{T}(\mathbf{v}_2) + \dots + x_m\mathbf{T}(\mathbf{v}_m) \\ &= x_1(c_{11}\mathbf{v}_1 + c_{21}\mathbf{v}_2 + \dots + c_{m1}\mathbf{v}_m) \\ &\quad + x_2(c_{12}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \dots + c_{m2}\mathbf{v}_m) \\ &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ &\quad + x_m(c_{1m}\mathbf{v}_1 + c_{2m}\mathbf{v}_2 + \dots + c_{mm}\mathbf{v}_m) \\ &= (x_1c_{11} + x_2c_{12} + \dots + x_mc_{1m})\mathbf{v}_1 \\ &\quad + (x_1c_{21} + x_2c_{22} + \dots + x_mc_{2m})\mathbf{v}_2 \\ &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ &\quad + (x_1c_{m1} + x_2c_{m2} + \dots + x_mc_{mm})\mathbf{v}_m \\ &= \lambda x_1\mathbf{v}_1 + \lambda x_2\mathbf{v}_2 + \dots + \lambda x_m\mathbf{v}_m = \lambda\mathbf{v} \end{aligned}$$

Dette viser at  $\mathbf{v}$  er en egenvektor med egenverdi  $\lambda$ , og dermed er beviset fullført.  $\square$

**Bemerkning:** Ved første øyekast kan beviset ovenfor se ut som et umotivert regnestykke som på mystisk vis ender opp med det svaret vi ønsker oss. Beviset har imidlertid en klar idé: Vi konstruerer en matrise  $C$  slik at den tilhørende lineæravbildningen (la oss kalle den  $\hat{\mathbf{T}}$ ) oppfører seg på akkurat samme måte overfor basisen  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$  i  $\mathbb{R}^m$  som  $\mathbf{T}$  oppfører seg overfor basisen  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  i  $H$ . På grunn av setning 4.12.5 vet vi at  $\hat{\mathbf{T}}$  har en egenvektor  $\mathbf{x}$ . Vi konstruerer så en vektor  $\mathbf{v}$  som forholder seg på akkurat samme måte til basisen  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  som  $\mathbf{x}$  forholder seg til basisen  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ . Da er det ikke så merkelig at  $\mathbf{v}$  er en egenvektor for  $\mathbf{T}$  med samme egenverdi som  $\mathbf{x}$ .

Vi er nå kommet frem til de to siste resultatene vi trenger før vi kan gå løs på selve beviset for spektralteoremet. Det første skal hjelpe oss med dimensjonsregnskapet.

**Setning 4.12.10** Anta at  $H \neq \{\mathbf{0}\}$  er et underrom av  $\mathbb{R}^n$ , og la  $\mathbf{a}$  være en ikke-null vektor i  $H$ . Da er

$$H_{\mathbf{a}^\perp} = \{\mathbf{x} \in H \mid \mathbf{x} \cdot \mathbf{a} = 0\}$$

et underrom av  $\mathbb{R}^n$ , og  $\dim(H_{\mathbf{a}^\perp}) = \dim(H) - 1$ . Dersom  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er en basis for  $H_{\mathbf{a}^\perp}$ , så er  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$  en basis for  $H$  (dersom  $H_{\mathbf{a}^\perp} = \{\mathbf{0}\}$  betyr dette at  $\mathbf{a}$  alene utgjør en basis for  $H$ ).

*Bevis:* At  $H_{\mathbf{a}^\perp}$  er et underrom, sjekkes på akkurat samme måte som i eksempel 4.7.2. Vi overlater tilfellet der  $H_{\mathbf{a}^\perp} = \{\mathbf{0}\}$  til leserne, og antar at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er en basis for  $H_{\mathbf{a}^\perp}$ . Vår jobb er å bevise at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$  er en basis for  $H$ . Vi må da sjekke at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$  er lineært uavhengige og utspenner hele  $H$ .

Vi begynner med uavhengigheten. Vi må vise at dersom

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k + d\mathbf{a} = \mathbf{0}$$

så er  $c_1 = c_2 = \cdots = c_k = d = 0$ . Tar vi skalarproduktet med  $\mathbf{a}$  på begge sider av formelen ovenfor og bruker at  $\mathbf{a} \cdot \mathbf{v}_i = 0$  (siden  $\mathbf{v}_i \in H_{\mathbf{a}^\perp}$ ), får vi

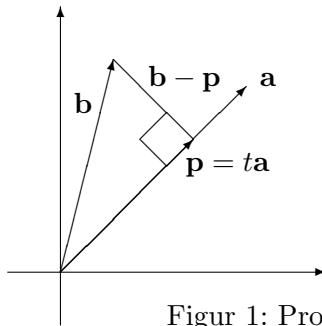
$$d|\mathbf{a}|^2 = 0$$

Siden  $\mathbf{a} \neq \mathbf{0}$ , betyr dette at  $d = 0$ . Dermed står vi igjen med

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

og siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er lineært uavhengig, betyr dette at  $c_1 = c_2 = \cdots = c_k = 0$ . Vi har altså vist at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$  er lineært uavhengige.

Det gjenstår å vise at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$  utspenner hele  $H$ , dvs. at enhver  $\mathbf{b} \in H$  kan skrives som en lineærkombinasjon av  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$ . Gitt en  $\mathbf{b} \in H$ , lar vi  $\mathbf{p}$  være projeksjonen av  $\mathbf{b}$  ned på  $\mathbf{a}$  (se figur 1 og husk setning 1.2.2).



Figur 1: Projeksjonen  $\mathbf{p}$  av  $\mathbf{b}$  ned på  $\mathbf{a}$

Legg merke til at siden  $\mathbf{p} = t\mathbf{a}$  for et tall  $t$ , så er  $\mathbf{b} - \mathbf{p} = \mathbf{b} + (-t)\mathbf{a}$  med i  $H$ . I tillegg står  $\mathbf{b} - \mathbf{p}$  per konstruksjon normalt på  $\mathbf{a}$ , og er følgelig et

element i  $H_{\mathbf{a}^\perp}$ . Siden  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  er en basis for  $H_{\mathbf{a}^\perp}$ , finnes det dermed tall  $c_1, c_2, \dots, c_k$  slik at

$$\mathbf{b} - \mathbf{p} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_k \mathbf{v}_k$$

Siden  $\mathbf{p}$  er parallel med  $\mathbf{a}$ , finnes det som allerede nevnt et tall  $t$  slik at  $\mathbf{p} = t\mathbf{a}$ . Dermed er

$$\mathbf{b} = t\mathbf{a} + c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \cdots + c_k \mathbf{v}_k$$

Dette viser at  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k, \mathbf{a}$  utspenner hele  $H$ , og beviset er fullført.  $\square$

Det neste resultatet peker på en viktig egenskap som skiller symmetriske lineæravbildninger fra generelle lineæravbildninger, og som er helt nødvendig for å få beviset for spektralteoremet til å fungere.

**Setning 4.12.11** La  $H$  være et underrom av  $\mathbb{R}^n$  og anta at  $\mathbf{T} : H \rightarrow H$  er en symmetrisk lineæravbildning. Anta videre at  $\mathbf{a}$  er en egenvektor for  $T$ , og la

$$H_{\mathbf{a}^\perp} = \{\mathbf{x} \in H \mid \mathbf{x} \cdot \mathbf{a} = 0\}$$

være det ortogonale komplementet til  $\mathbf{a}$  i  $H$ . Da avbilder  $T$  underrommet  $H_{\mathbf{a}^\perp}$  inn i seg selv, dvs. at dersom  $x \in H_{\mathbf{a}^\perp}$ , så er også  $T(\mathbf{x}) \in H_{\mathbf{a}^\perp}$ .

*Bevis:* Anta at  $x \in H_{\mathbf{a}^\perp}$ . Da er  $\mathbf{x} \cdot \mathbf{a} = 0$ , og siden  $\mathbf{T}$  er symmetrisk, har vi

$$\mathbf{T}(\mathbf{x}) \cdot \mathbf{a} = \mathbf{x} \cdot \mathbf{T}(\mathbf{a}) = \mathbf{x} \cdot (\lambda \mathbf{a}) = \lambda(\mathbf{x} \cdot \mathbf{a}) = 0$$

der  $\lambda$  er egenverdien til  $\mathbf{a}$ . Dette viser at  $T(\mathbf{x}) \in H_{\mathbf{a}^\perp}$ .  $\square$

### Bevis for spektralteoremet

Vi er nå klare for å bevise spektralteoremet. For å få induksjonsargumentet vårt til å fungere, må vi bevise et litt mer generelt resultat:

**Teorem 4.12.12 (Spektralteoremet for underrom)** Anta at  $H \neq \{\mathbf{0}\}$  er et underrom av  $\mathbb{R}^n$ , og at  $\mathbf{T} : H \rightarrow H$  er en symmetrisk lineæravbildning. Da har  $H$  en ortonormal basis som består av egenvektorer for  $\mathbf{T}$ .

**Bemerkning:** Siden  $H = \mathbb{R}^n$  er et underrom av  $\mathbb{R}^n$  og alle symmetriske  $n \times n$ -matriser  $A$  definerer en symmetrisk lineæravbildning  $\mathbf{T}$  ved  $\mathbf{T}(\mathbf{x}) = A\mathbf{x}$ , ser vi at teoremet ovenfor medfører den opprinnelige versjonen av spektralteoremet i 4.12.1.

*Bevis for teorem 4.12.12:* Vi beviser teoremet ved induksjon på dimensjonen til underrommet  $H$ . Anta først at  $H$  har dimensjon 1. Da har  $H$  en basis som

bare består av én vektor  $\mathbf{v}$ . Vi kan anta at  $|\mathbf{v}| = 1$  (hvis ikke bytter vi bare ut  $\mathbf{v}$  med  $\frac{\mathbf{v}}{|\mathbf{v}|}$ ). Siden  $\mathbf{T}(\mathbf{v}) \in H$ , må det finnes et tall  $c$  slik at  $\mathbf{T}(\mathbf{v}) = c\mathbf{v}$ . Men da er  $\mathbf{v}$  en egenvektor med egenverdi  $c$ , og basisen som bare består av  $\mathbf{v}$ , er derfor en ortonormal basis av egenvektorer (ortogonaliteten er trivelt oppfylt siden  $\mathbf{v}$  ikke har noen basiskompanjoner å stå ortogonalt på).

Anta så at teoremet gjelder for alle underrom med dimensjon  $m$  og at  $H$  har dimensjon  $m + 1$ . Ifølge setning 4.12.9 har en symmetrisk lineæravbildning  $\mathbf{T} : H \rightarrow H$  i hvert fall én egenvektor  $\mathbf{a}$  som vi kan velge til å ha lengde 1. Ifølge setning 4.12.11 avbilder  $T$  det ortogonale komplementet  $H_{\mathbf{a}^\perp}$  inn i seg selv. Setning 4.12.10 forteller oss at  $H_{\mathbf{a}^\perp}$  har dimensjon  $m$ , og ifølge induksjonsantagelsen har dermed  $H_{\mathbf{a}^\perp}$  en ortonormal basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  av egenvektorer for  $\mathbf{T}$ . Ifølge setning 4.12.10 er da  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m, \mathbf{a}$  en basis for  $H$ , og den består åpenbart bare av egenvektorer. Siden  $\mathbf{a}$  står ortogonalt på alle vektorer i  $H_{\mathbf{a}^\perp}$  (og dermed spesielt på  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ ), er basisen ortonormal.

□

### Oppgaver til seksjon 4.12

1. En matrise kalles *skjevtsymmetrisk* dersom  $A^T = -A$ . Vis at alle egenverdiene til en skjevtsymmetrisk matrise er imaginære (dvs. at de er komplekse tall på formen  $z = ib$ ).
2. En kvadratisk matrise  $U$  kalles *orthogonal* dersom  $U^{-1} = U^T$ . Vis at dersom  $\lambda$  er en (reell eller kompleks) egenverdi for  $U$ , så er  $|\lambda| = 1$ . (*Hint:* Anta at  $\mathbf{v}$  er en egenvektor med egenverdi  $\lambda$ , og regn ut  $U\mathbf{v} \cdot U\mathbf{v}$  på to forskjellige måter.)
3. Anta at  $A$  er en  $m \times n$ -matrise.
  - a) Vis at  $B = A^T A$  er en symmetrisk  $n \times n$ -matrise.
  - b) Vis at alle egenverdiene  $\lambda_1, \lambda_2, \dots, \lambda_n$  til  $B$  er ikke-negative.
  - c) Sett  $\mu_i = \sqrt{\lambda_i}$  og vis at det finnes en ortonormal basis  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  slik at

$$(A\mathbf{v}_i) \cdot (A\mathbf{v}_j) = \begin{cases} \mu_i^2 & \text{hvis } i = j \\ 0 & \text{ellers} \end{cases}$$

Tallene  $\mu_1, \mu_2, \dots, \mu_n$  kalles *singulærverdiene* til  $A$ .

## Kapittel 5

# Iterasjon og optimering

Forrige kapittel handlet om *lineære* ligningssystemer og hvordan de kan brukes til å løse problemer knyttet til matriser og vektorer. Selv om det ikke er like lett å få øye på, handler dette kapitlet i stor grad om *ikke-lineære* ligningssystemer og hvordan de kan brukes til å løse problemer knyttet til følger og funksjoner (et ikke-lineært ligningssystem er rett og slett et ligningssystem som ikke er lineært). Vi skal blant annet se hvordan vi kan bruke slike ligningssystemer i *optimering*, dvs. i studiet av forskjellige typer maksimums- og minimumsproblemer.

Alle lineære ligningssystemer kan i prinsippet løses eksakt ved hjelp av Gauss-eliminasjon. For ikke-lineære ligningssystemer er situasjonen en helt annen — for disse ligningssystemene er det som oftes umulig å finne eksakte løsninger, og man må derfor nøye seg med metoder som enten viser at løsninger eksisterer (uten å finne dem), eller som finner tilnærmede løsninger ved hjelp av numeriske teknikker. Ofte går disse to angrepsmåtene hånd i hånd; det har som regel liten nytteverdi å vite at en løsning finnes uten å ane hvordan den ser ut, og man kan kaste bort mye tid på numerisk leting etter løsninger som ikke finnes. Dette kapitlet er derfor både blant de mest praktiske og de mest teoretiske i kompendiet — vi skal beskrive metoder og teknikker som brukes for å løse praktiske problemer, men vi skal samtidig se på teorien som forklarer når og hvorfor disse metodene virker.

Numeriske løsningsmetoder er ofte basert på *iterasjon*, dvs. at man starter med en tilnærmet løsning, og så kjører den samme prosessen om og om igjen for å få stadig bedre løsningskandidater. Newtons metode (se *Kalkulus* seksjon 7.3) er et typisk eksempel på en slik iterativ metode, og en av problemstillingene vi skal se på, er hvordan denne metoden kan generaliseres til en effektiv teknikk for å finne tilnærmede løsninger til ikke-lineære ligningssystemer.

Iterasjon er nært knyttet til følger; vi får en følge av tilnærmede løsninger som helst bør konvergere mot en virkelig løsning. For å forstå hvordan iterasjonsteknikker fungerer, trenger vi å forstå hva det vil si at en følge kon-

vergerer, og kapitlet starter derfor med et teoretisk studium av følger i  $\mathbb{R}^m$ . Det viktigste begrepet vi skal se på, er *kompletthet* — en omformulering og videreføring av kompletthetsprinsippet i *Kalkulus*. Kompletthet er en forutsetning for alt vi gjør i dette kapitlet, og det kommer til å bli et enda viktigere verktøy når du går videre for å studere mer avansert matematikk — differensialligningsteori handler f.eks. i stor grad om å utnytte kompletthet av rom som består av funksjoner (og ikke av vektorer slik som våre rom hittil har gjort).

Selv om dette kapitlet handler om ikke-lineære ligningssystemer, betyr ikke det at vi kan glemme alt vi har lært om lineære systemer — vi kommer tvert i mot til å få bruk for det meste av det vi vet om vektorer, matriser og lineære ligningssystemer i løpet av kapitlet. Grunnen er at vi i mange situasjoner skal tilnærme ikke-lineære ligningssystemer med lineære systemer som har omtrent samme oppførsel i det området vi er interessert i. Et typisk eksempel er Newtons metode der vi i hvert skritt i utregningen erstatter det opprinnelige, ikke-lineære ligningssystemet med et lineært system som er enklere å løse. Nøkkelen til sammenhengen mellom ikke-lineære og lineære systemer er begrepet *linearisering* som vi studerte i seksjon 2.9 (husk at lineariseringen til en funksjon  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  i punktet  $\mathbf{a}$  er den affinavbildningen som ligger tettest opptil  $\mathbf{F}$  i området rundt  $\mathbf{a}$ ).

## 5.1 Litt topologi i $\mathbb{R}^m$

Når vi arbeider med funksjoner av én variabel, vil definisjonsområdet som regel være et intervall eller en enkel sammensetning av intervaller. For funksjoner av flere variable finnes det mange flere muligheter for hvordan definisjonsområdet kan være, og vi må derfor være litt mer formelle i vår omgang med mengder. I denne seksjonen skal vi innføre *åpne* og *lukkede* mengder. Disse mengdene spiller på mange måter den samme rollen i flervariabelteori som åpne og lukkede intervaller spiller i teorien for funksjoner av én variabel. En liten bemerkning om notasjon før vi starter: I dette kapitlet vil vi stort sett referere til det underliggende rommet som  $\mathbb{R}^m$  og ikke  $\mathbb{R}^n$  slik vi vanligvis gjør i de andre kapitlene. Det er rett og slett fordi vi kommer til å arbeide mye med følger og ønsker å ha bokstaven  $n$  ledig for å referere til det  $n$ -te ledet  $x_n$  i en følge.

Vi begynner med å repete litt fra kapittel 2. Dersom  $\mathbf{a}$  er et punkt i  $\mathbb{R}^m$ , så er (*den åpne*) *kulen om  $\mathbf{a} \in \mathbb{R}^m$*  med radius  $r$  definert som

$$B(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^m : |\mathbf{x} - \mathbf{a}| < r\}$$

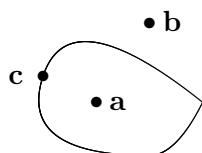
Den består altså av de punktene i  $\mathbb{R}^m$  som har avstand mindre enn  $r$  til  $\mathbf{a}$ . I  $\mathbb{R}^3$  er dette virkelig en (åpen) kule i tradisjonell forstand, mens det i  $\mathbb{R}^2$  er en (åpen) sirkelskive og i  $\mathbb{R}$  et åpent intervall. Vi velger å bruke “kule” som et fellesord i alle dimensjoner selv om det til å begynne med kan virke

litt uvant når vi arbeider i planet eller på tallinjen. De fleste illustrasjonene våre vil være i planet, og der vil kuler fremstå som sirkler. Noen ganger skal vi også ha bruk for de *lukkede kulene*

$$\overline{B}(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^m : |\mathbf{x} - \mathbf{a}| \leq r\},$$

men det er de åpne som vil spille hovedrollen til å begynne med.

Figur 1 viser en mengde i planet (området innenfor kurven) og tre punkter **a**, **b** og **c**.



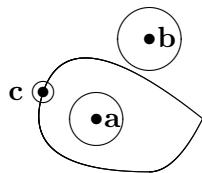
Figur 1

De tre punktene ligger på forskjellig måte i forhold til mengden — **a** ligger klart på innsiden av mengden, **b** ligger klart på utsiden, mens **c** ligger på grensen mellom mengden og omgivelsene. Vi kaller **a** et *indre punkt*, **b** et *ytre punkt* og **c** et *randpunkt* for mengden. I høyere dimensjoner kan vi ikke støtte oss på figurer, og vi trenger derfor en mer formell definisjon av indre punkter, ytre punkter og randpunkter:

**Definisjon 5.1.1** La  $A$  være en delmengde av  $\mathbb{R}^m$ .

- (i) Et punkt  $\mathbf{a} \in \mathbb{R}^m$  kalles et *indre punkt* for  $A$  dersom det finnes en kule  $B(\mathbf{a}, r)$  om  $\mathbf{a}$  som bare inneholder punkter som er med i  $A$ .
- (ii) Et punkt  $\mathbf{b} \in \mathbb{R}^m$  kalles et *ytre punkt* for  $A$  dersom det finnes en kule  $B(\mathbf{b}, r)$  om  $\mathbf{b}$  som ikke inneholder noen punkter som er med i  $A$ .
- (iii) Et punkt  $\mathbf{c} \in \mathbb{R}^m$  kalles et *randpunkt* for  $A$  dersom enhver kule  $B(\mathbf{c}, r)$  om  $\mathbf{c}$  inneholder både punkter som er med i  $A$  og punkter som ikke er med i  $A$ .

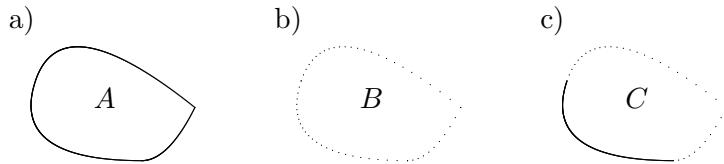
Legg merke til at de tre delene av definisjonen uttømmer alle muligheter, så et punkt i  $\mathbb{R}^m$  må enten være et indre punkt, et ytre punkt eller et randpunkt for  $A$ . I figur 2 har vi illustrert definisjonen ved å legge “kuler” (dvs. sirkler) rundt punktene **a**, **b** og **c**.



Figur 2

Et indre punkt hører alltid med til mengden  $A$ , mens et ytre punkt aldri hører med til mengden. For randpunkter er det ingen generell regel; de vil

noen ganger høre med til mengden og andre ganger ikke. Figur 3 viser dette for tre mengder i planet. I punkt a) er randen tegnet med en hel strek — det markerer at alle punktene på randen hører med til mengden  $A$ . En slik mengde kalles *lukket*. I punkt b) er randen stiplet — det markerer at ingen av punktene på randen hører med til mengden  $B$ . En slik mengde kalles *åpen*. I punkt c) er noe av randen heltrukken og resten stiplet — det markerer at noen av punktene på randen hører med til mengden  $C$ , mens andre ikke gjør det. En slik mengde er hverken åpen eller lukket (på figuren har vi kalt den “halvåpen”, men det er ingen offisiell betegnelse).



Figur 3: Lukket, åpen og “halvåpen” mengde

La oss skrive opp definisjonen av åpne og lukkede mengder litt mer formelt:

**Definisjon 5.1.2** *En mengde  $A \subset \mathbb{R}^m$  er lukket dersom den inneholder alle sine randpunkter, og åpen dersom den ikke inneholder noen randpunkter.*

Åpne og lukkede mengder spiller omrent samme rolle i teorien for funksjoner av flere variable som åpne og lukkede intervaller gjør i teorien for funksjoner av én variabel. Det er imidlertid en viktig forskjell — åpne og lukkede mengder kan ha mange forskjellige former, og det gjør at flervariabelteorien har en del geometriske komplikasjoner som envariabelteorien ikke har.

**Eksempel 1:** Vi skal undersøke om mengden

$$A = \left\{ (x, y) \in \mathbb{R}^2 : \frac{x^2}{9} + \frac{y^2}{4} \leq 1 \right\}$$

er lukket. Mengden består av alle punkter som ligger på eller innenfor en ellipse (lag en skisse!) Randen består av alle punktene som ligger på selve ellipsen, og mengden er lukket siden alle disse randpunktene hører med til  $A$ .

### Følger i $\mathbb{R}^m$

I resten av denne seksjonen skal vi se litt på konvergens av følger i  $\mathbb{R}^m$ . Dette er et begrep som kommer til å stå sentralt i de neste seksjonene.

En *følge* i  $\mathbb{R}^m$  er en uendelig sekvens

$$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n, \dots$$

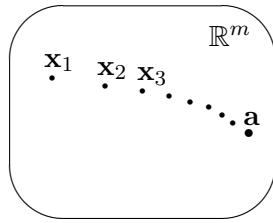
av punkter  $\mathbf{x}_n \in \mathbb{R}^m$ . Akkurat som for tallfølger bruker vi  $\{\mathbf{x}_n\}$  som en kortfattet skrivemåte for disse følgene. Vi tillater at følgen starter med andre indeks enn 1, f.eks.

$$\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \dots, \mathbf{x}_n, \dots$$

eller

$$\mathbf{x}_{-3}, \mathbf{x}_{-2}, \mathbf{x}_{-1}, \dots, \mathbf{x}_n, \dots$$

Dersom det er viktig å vite hvor følgen starter, kan vi markere det i den kortfattede skrivemåten ved å skrive  $\{\mathbf{x}_n\}_{n=3}^{\infty}$  og  $\{\mathbf{x}_n\}_{n=-3}^{\infty}$  for de to følgene ovenfor.



Figur 4: En følge som konvergerer til  $\mathbf{a}$

Intuitivt sier vi at en følge  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^m$  nærmer seg punktet  $\mathbf{a}$  som grenseverdi dersom vi kan få avstanden mellom  $\mathbf{x}_n$  og  $\mathbf{a}$  så liten vi måtte ønske ved å velge  $n$  tilstrekkelig stor. Vi kan formulere dette på akkurat samme måte som i det en-dimensjonale tilfellet (sammenlign med definisjon 4.3.1 i *Kalkulus*):

**Definisjon 5.1.3** *Følgen  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^m$  konvergerer mot punktet  $\mathbf{a} \in \mathbb{R}^m$  dersom det til enhver  $\epsilon > 0$  finnes en  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{a}| < \epsilon$  for alle  $n \geq N$ . Vi skriver*

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{a}$$

Vi har akkurat de samme regnereglene som for tallfølger:

**Setning 5.1.4** *Anta at  $\{\mathbf{x}_n\}$  og  $\{\mathbf{y}_n\}$  er to følger i  $\mathbb{R}^m$  som konvergerer mot henholdsvis  $\mathbf{x}$  og  $\mathbf{y}$ . Da har vi:*

- (i) *Følgen  $\{c\mathbf{x}_n\}$  konvergerer for ethvert tall  $c$ , og  $\lim_{n \rightarrow \infty} (c\mathbf{x}_n) = c\mathbf{x}$*
- (ii) *Følgen  $\{\mathbf{x}_n + \mathbf{y}_n\}$  konvergerer, og  $\lim_{n \rightarrow \infty} (\mathbf{x}_n + \mathbf{y}_n) = \mathbf{x} + \mathbf{y}$*
- (iii) *Følgen  $\{\mathbf{x}_n - \mathbf{y}_n\}$  konvergerer, og  $\lim_{n \rightarrow \infty} (\mathbf{x}_n - \mathbf{y}_n) = \mathbf{x} - \mathbf{y}$*
- (iv) *Følgen  $\{\mathbf{x}_n \cdot \mathbf{y}_n\}$  konvergerer, og  $\lim_{n \rightarrow \infty} (\mathbf{x}_n \cdot \mathbf{y}_n) = \mathbf{x} \cdot \mathbf{y}$  (legg merke til at dette er en tallfølge og ikke en følge av vektorer).*

*Bevis:* Bevisene er nesten identiske med de tilsvarende bevisene for tallfølger (se 4.3.3 i *Kalkulus*), den eneste forskjellen er at vi nå må bruke Schwarz' ulikhet (setning 1.2.3) og trekantulikheten for vektorer (setning 1.2.4) istedenfor de tilsvarende ulikhettene for tall. For å illustrere bruken av disse ulikhettene, tar vi med bevisene for (ii) og (iv).

(ii) Vi må vise at gitt en  $\epsilon > 0$ , kan vi alltid finne en  $N \in \mathbb{N}$  slik at  $|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{x} + \mathbf{y})| < \epsilon$  for alle  $n \geq N$ . Det første vi gjør er å omgruppere leddene slik at vi kan behandle  $\{\mathbf{x}_n\}$  og  $\{\mathbf{y}_n\}$  hver for seg:

$$|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{x} + \mathbf{y})| = |(\mathbf{x}_n - \mathbf{x}) + (\mathbf{y}_n - \mathbf{y})| \leq |\mathbf{x}_n - \mathbf{x}| + |\mathbf{y}_n - \mathbf{y}|$$

der vi i det siste skrittet har brukt trekantulikheten (setning 1.2.4). Siden  $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$ , må det finnes en  $N_1 \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{x}| < \frac{\epsilon}{2}$  for alle  $n \geq N_1$ , og siden  $\lim_{n \rightarrow \infty} \mathbf{y}_n = \mathbf{y}$ , må det finnes en  $N_2 \in \mathbb{N}$  slik at  $|\mathbf{y}_n - \mathbf{y}| < \frac{\epsilon}{2}$  for alle  $n \geq N_2$ . Velger vi  $N$  lik det største av tallene  $N_1, N_2$ , ser vi at når  $n \geq N$ , er

$$|(\mathbf{x}_n + \mathbf{y}_n) - (\mathbf{x} + \mathbf{y})| \leq |\mathbf{x}_n - \mathbf{x}| + |\mathbf{y}_n - \mathbf{y}| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Dermed er (ii) bevist.

(iv) Vi må vise at gitt en  $\epsilon > 0$ , kan vi alltid finne en  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}| < \epsilon$  for alle  $n \geq N$ . Vi bruker først trikset med å legge til og trekke fra leddet  $\mathbf{x} \cdot \mathbf{y}_n$ , og benytter deretter trekantulikheten (setning 1.2.4) og Schwarz' ulikhet (setning 1.2.3):

$$\begin{aligned} |\mathbf{x}_n \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}| &= |\mathbf{x}_n \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}_n + \mathbf{x} \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}| \leq \\ &\leq |\mathbf{x}_n \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}_n| + |\mathbf{x} \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}_n - \mathbf{x}| |\mathbf{y}_n| + |\mathbf{x}| |\mathbf{y}_n - \mathbf{y}| \end{aligned}$$

Vi skal nå vise at vi kan få hvert av de to leddene  $|\mathbf{x}_n - \mathbf{x}| |\mathbf{y}_n|$  og  $|\mathbf{x}| |\mathbf{y}_n - \mathbf{y}|$  mindre enn  $\frac{\epsilon}{2}$  ved å velge  $n$  stor nok. Det siste leddet er enklest, og vi starter med det. Siden dette leddet opplagt er mindre enn  $\frac{\epsilon}{2}$  hvis  $|\mathbf{x}| = 0$ , kan vi konsentrere oss om tilfellet  $|\mathbf{x}| \neq 0$ . Siden  $\lim_{n \rightarrow \infty} \mathbf{y}_n = \mathbf{y}$ , må det finnes en  $N_1 \in \mathbb{N}$  slik at  $|\mathbf{y}_n - \mathbf{y}| < \frac{\epsilon}{2|\mathbf{x}|}$  for alle  $n \geq N_1$ . Dermed er  $|\mathbf{x}| |\mathbf{y}_n - \mathbf{y}| < |\mathbf{x}| \cdot \frac{\epsilon}{2|\mathbf{x}|} = \frac{\epsilon}{2}$  når  $N \geq N_1$ .

Det første leddet  $|\mathbf{x} - \mathbf{x}_n| |\mathbf{y}_n|$  er litt verre. Vi observerer først at siden  $\lim_{n \rightarrow \infty} \mathbf{y}_n = \mathbf{y}$ , finnes det et tall  $N_2$  slik at  $|\mathbf{y}_n - \mathbf{y}| \leq 1$  når  $n \geq N_2$ . Dermed er

$$|\mathbf{y}_n| = |\mathbf{y} + (\mathbf{y}_n - \mathbf{y})| \leq |\mathbf{y}| + |\mathbf{y}_n - \mathbf{y}| \leq |\mathbf{y}| + 1$$

Siden  $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$ , må det finnes en  $N_3 \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{x}| < \frac{\epsilon}{2(|\mathbf{y}| + 1)}$  for alle  $n \geq N_3$ . Hvis  $n$  er større enn eller lik både  $N_2$  og  $N_3$ , er dermed

$$|\mathbf{x}_n - \mathbf{x}| |\mathbf{y}_n| < \frac{\epsilon}{2(|\mathbf{y}| + 1)} \cdot (|\mathbf{y}| + 1) = \frac{\epsilon}{2}$$

Velger vi nå  $N$  til å være det største av tallene  $N_1, N_2, N_3$ , ser vi at for  $n \geq N$  er

$$|\mathbf{x}_n \cdot \mathbf{y}_n - \mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}_n - \mathbf{x}| |\mathbf{y}_n| + |\mathbf{x}| |\mathbf{y}_n - \mathbf{y}| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Dermed er (iv) bevist.  $\square$

Vi tok med bevisene ovenfor for å demonstrere hvordan trekantulikheten og Schwarz' ulikhet ofte brukes. Det viser seg at vi kan bevise disse resultatene vel så enkelt ved å føre dem tilbake til tilsvarende resultater for tallfølger. Det neste resultatet vil gi oss det redskapet vi trenger. Litt notasjon før vi begynner: Dersom  $\{\mathbf{x}_n\}$  er en følge i  $\mathbb{R}^m$ , skriver vi koordinatene til  $\mathbf{x}_n$  slik:

$$\mathbf{x}_n = (x_1^{(n)}, x_2^{(n)}, \dots, x_m^{(n)})$$

Vi skriver indeksen  $n$  opp for ikke å blande den sammen med koordinatene til  $\mathbf{x}_n$ , og vi putter den inn i en parentes for å gjøre det klart at den ikke er en eksponent.

**Setning 5.1.5** *Anta at  $\{\mathbf{x}_n\}$  er en følge i  $\mathbb{R}^m$  med komponenter*

$$\mathbf{x}_n = (x_1^{(n)}, x_2^{(n)}, \dots, x_m^{(n)})$$

*og at  $\mathbf{x} \in \mathbb{R}^m$  har komponenter  $\mathbf{x} = (x_1, x_2, \dots, x_m)$ . Da er*

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$$

*hvis og bare hvis*

$$\lim_{n \rightarrow \infty} x_i^{(n)} = x_i \text{ for alle } i = 1, 2, \dots, m$$

*Med andre ord:  $\{\mathbf{x}_n\}$  konvergerer mot  $\mathbf{x}$  hvis og bare hvis hver komponent i  $\mathbf{x}_n$  konvergerer mot tilsvarende komponent i  $\mathbf{x}$ .*

Dersom du synes det er vanskelig å forstå hva setningen sier, er det lurt å ta en kikk på eksemplet nedenfor før du går videre.

*Bevis:* Anta først at  $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$ . Vi skal vise at  $\lim_{n \rightarrow \infty} x_i^{(n)} = x_i$ . Det betyr at gitt en  $\epsilon > 0$ , må vi vise at det alltid finnes en  $N \in \mathbb{N}$  slik at  $|x_i^{(n)} - x_i| < \epsilon$  for alle  $n \geq N$ . Siden  $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$ , finnes det en  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{x}| < \epsilon$  når  $n \geq N$ . Siden

$$\begin{aligned} |x_i^{(n)} - x_i| &= \sqrt{(x_i^{(n)} - x_i)^2} \leq \\ &\leq \sqrt{(x_1^{(n)} - x_1)^2 + \dots + (x_i^{(n)} - x_i)^2 + \dots + (x_m^{(n)} - x_m)^2} = |\mathbf{x}_n - \mathbf{x}| \end{aligned}$$

medfører dette at  $|x_i^{(n)} - x_i| < \epsilon$  for alle  $n \geq N$ , og det er akkurat det vi skulle vise.

Anta så at  $\lim_{n \rightarrow \infty} x_i^{(n)} = x_i$  for  $i = 1, 2, \dots, m$ . Vi skal vise at  $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}$ . Det betyr at gitt en  $\epsilon > 0$ , må vi produsere en  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{x}| < \epsilon$  når  $n \geq N$ . Siden  $\lim_{n \rightarrow \infty} x_i^{(n)} = x_i$ , finnes det for hver  $i$  en  $N_i \in \mathbb{N}$  slik at  $|x_i^{(n)} - x_i| < \frac{\epsilon}{\sqrt{m}}$  når  $n \geq N_i$ . La  $N$  være den største av  $N_1, N_2, \dots, N_m$ . For  $n \geq N$  er da

$$\begin{aligned} |\mathbf{x}_n - \mathbf{x}| &= \sqrt{(x_1^{(n)} - x_1)^2 + \dots + (x_i^{(n)} - x_i)^2 + \dots + (x_m^{(n)} - x_m)^2} \leq \\ &\leq \sqrt{\left(\frac{\epsilon}{\sqrt{m}}\right)^2 + \dots + \left(\frac{\epsilon}{\sqrt{m}}\right)^2 + \dots + \left(\frac{\epsilon}{\sqrt{m}}\right)^2} = \sqrt{m \frac{\epsilon^2}{m}} = \sqrt{\epsilon^2} = \epsilon \end{aligned}$$

□

**Eksempel 2:** Finn grenseverdien

$$\lim_{n \rightarrow \infty} \begin{pmatrix} \frac{n^2}{n^2+1} \\ n \sin\left(\frac{1}{n}\right) \\ \left(1 + \frac{2}{n}\right)^n \end{pmatrix}$$

Ifølge setningen ovenfor behøver vi bare å regne ut grensen til hver komponent. Den første er enkel:

$$\lim_{n \rightarrow \infty} \frac{n^2}{n^2 + 1} \stackrel{\text{L'H}}{=} \lim_{n \rightarrow \infty} \frac{2n}{2n} = 1$$

Den andre går også greit:

$$\lim_{n \rightarrow \infty} n \sin\left(\frac{1}{n}\right) = \lim_{n \rightarrow \infty} \frac{\sin\left(\frac{1}{n}\right)}{\frac{1}{n}} \stackrel{\text{L'H}}{=} \lim_{n \rightarrow \infty} \frac{\cos\left(\frac{1}{n}\right)\left(-\frac{1}{n^2}\right)}{-\frac{1}{n^2}} = \lim_{n \rightarrow \infty} \cos\left(\frac{1}{n}\right) = 1$$

Den tredje skriver vi først om:

$$\left(1 + \frac{2}{n}\right)^n = e^{n \log\left(1 + \frac{2}{n}\right)}$$

og bruker deretter L'Hôpitals regel på eksponenten:

$$\lim_{n \rightarrow \infty} n \log\left(1 + \frac{2}{n}\right) = \lim_{n \rightarrow \infty} \frac{\log\left(1 + \frac{2}{n}\right)}{\frac{1}{n}} \stackrel{\text{L'H}}{=} \lim_{n \rightarrow \infty} \frac{\frac{1}{(1+\frac{2}{n})} \cdot (-\frac{2}{n^2})}{-\frac{1}{n^2}} = 2$$

Altså er

$$\lim_{n \rightarrow \infty} \left(1 + \frac{2}{n}\right)^n = e^2$$

Kombinerer vi alt dette, får vi

$$\lim_{n \rightarrow \infty} \begin{pmatrix} \frac{n^2}{n^2+1} \\ n \sin\left(\frac{1}{n}\right) \\ \left(1 + \frac{2}{n}\right)^n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ e^2 \end{pmatrix}$$

♣

Det neste resultatet gir et nytt innblikk i hvorfor lukkede mengder kalles ”lukkede”:

**Setning 5.1.6** *Anta at  $A \subset \mathbb{R}^m$  er lukket, og at  $\{\mathbf{x}_n\}$  er en følge fra  $A$  som konvergerer mot et punkt  $\mathbf{x}$ . Da er  $\mathbf{x} \in A$ .*

*Bevis:* Anta for motsigelse at  $\mathbf{x} \notin A$ . Siden  $A$  er lukket, må  $\mathbf{x}$  da være et ytre punkt (alle randpunktene hører jo med til  $A$  når  $A$  er lukket). Det betyr at det finnes en kule  $B(\mathbf{x}, r)$  om  $\mathbf{x}$  som ikke inneholder noe punkt fra  $A$ . Spesielt kan ingen av leddene i følgen  $\{\mathbf{x}_n\}$  ligge i  $B(\mathbf{x}, r)$ . Det betyr at  $|\mathbf{x}_n - \mathbf{x}| \geq r$  for alle  $n$ , og følgelig kan ikke  $\{\mathbf{x}_n\}$  konvergere mot  $\mathbf{x}$ . Dette gir oss den selvmotsigelsen vi er på jakt etter.  $\square$

Det siste resultatet i denne seksjonen binder sammen konvergens og kontinuitet (se *Kalkulus*, setning 5.1.10 for den endimensjonale versjonen).

**Setning 5.1.7** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av flere variable, og at  $\mathbf{a}$  er et punkt i definisjonsområdet  $A$  til  $\mathbf{F}$ . Da er  $\mathbf{F}$  kontinuerlig i  $\mathbf{a}$  hvis og bare hvis  $\mathbf{F}(\mathbf{x}_n) \rightarrow \mathbf{F}(\mathbf{a})$  for alle følger  $\{\mathbf{x}_n\}$  fra  $A$  slik at  $\mathbf{x}_n \rightarrow \mathbf{a}$ .*

*Bevis:* Anta først at  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ . Gitt en følge  $\{\mathbf{x}_n\}$  av punkter i  $A$  slik at  $\mathbf{x}_n \rightarrow \mathbf{a}$ , må vi vise at det for enhver  $\epsilon > 0$ , finnes en  $N \in \mathbb{N}$  slik at  $|\mathbf{F}(\mathbf{x}_n) - \mathbf{F}(\mathbf{a})| < \epsilon$  for alle  $n \geq N$ . Siden  $\mathbf{F}$  er kontinuerlig i  $\mathbf{a}$ , finnes det en  $\delta > 0$  slik at  $|\mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{a})| < \epsilon$  for alle  $\mathbf{y} \in A$  slik at  $|\mathbf{y} - \mathbf{a}| < \delta$ . Siden  $\mathbf{x}_n \rightarrow \mathbf{a}$ , finnes det en  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{a}| < \delta$  når  $n \geq N$ . Men dette betyr at når  $n \geq N$ , så er  $|\mathbf{x}_n - \mathbf{a}| < \delta$  og følgelig  $|\mathbf{F}(\mathbf{x}_n) - \mathbf{F}(\mathbf{a})| < \epsilon$ .

Anta så at  $\mathbf{F}$  ikke er kontinuerlig i  $\mathbf{a}$ . Vi må vise at det finnes i hvert fall én følge  $\{\mathbf{x}_n\}$  fra  $A$  slik at  $\mathbf{x}_n \rightarrow \mathbf{a}$ , men  $\mathbf{F}(\mathbf{x}_n) \not\rightarrow \mathbf{F}(\mathbf{a})$ . Siden  $\mathbf{F}$  ikke er kontinuerlig i  $\mathbf{a}$ , må det finnes en  $\epsilon > 0$  slik at uansett hvor liten vi velger  $\delta > 0$ , så eksisterer det en  $\mathbf{x} \in A$  slik at  $|\mathbf{x} - \mathbf{a}| < \delta$ , men  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{a})| \geq \epsilon$ . Velger vi  $\delta = \frac{1}{n}$ , finner vi på denne måten et punkt  $\mathbf{x}_n \in A$  slik at  $|\mathbf{x}_n - \mathbf{a}| < \frac{1}{n}$ , men  $|\mathbf{F}(\mathbf{x}_n) - \mathbf{F}(\mathbf{a})| \geq \epsilon$ . Følgen  $\{\mathbf{x}_n\}$  konvergerer mot  $\mathbf{a}$ , men  $\{\mathbf{F}(\mathbf{x}_n)\}$  kan ikke konvergere mot  $\mathbf{F}(\mathbf{a})$  siden  $|\mathbf{F}(\mathbf{x}_n) - \mathbf{F}(\mathbf{a})| \geq \epsilon$  for alle  $n$ .  $\square$

### Oppgaver til seksjon 5.1

**1.** Avgjør om mengden er åpen eller lukket eller ingen av delene. Det holder med en figur og et uformelt argument.

- a)  $\{(x, y) \in \mathbb{R}^2 : |x| \leq 1 \text{ og } |y| \leq 1\}$
- b)  $\{(x, y) \in \mathbb{R}^2 : |x| < 1 \text{ og } |y| < 1\}$
- c)  $\{(x, y) \in \mathbb{R}^2 : |x| \leq 1 \text{ og } |y| < 1\}$
- d)  $\{(x, y) \in \mathbb{R}^2 : |x| \leq 1\}$
- e)  $\{(x, y) \in \mathbb{R}^2 : x + 2y < 1\}$
- f)  $\{(x, y) \in \mathbb{R}^2 : x + 2y = 1\}$
- g)  $\{(x, y) \in \mathbb{R}^2 : x \text{ og } y \text{ er rasjonale}\}$
- h)  $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 < 1\}$
- i)  $\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \geq 1\}$

**2.** Finn grensen til følgene  $\{\mathbf{x}_n\}$  når:

- a)  $\mathbf{x}_n = (\frac{2n^2+1}{n^2+3n}, \frac{3n}{1-2n})$ .
- b)  $\mathbf{x}_n = (n \sin \frac{1}{n}, n(1 - e^{2/n}))$ .
- c)  $\mathbf{x}_n = (\sqrt{n^2 + 2n} - n, \cos \frac{1}{n}, (\cos \frac{1}{n})^{n^2})$

**3.** Bruk setning 5.1.5 og grensesetningene for “vanlige” følger (dvs. følger med verdier i  $\mathbb{R}$ ) til å bevise setning 5.1.4.

**4.** Anta at  $\mathbf{a}$  og  $\mathbf{b}$  er to punkter i  $\mathbb{R}^m$ , og at  $\{\mathbf{x}_n\}$  er en følge som konvergere mot  $\mathbf{b}$ . Vis at  $\lim_{n \rightarrow \infty} |\mathbf{x}_n - \mathbf{a}| = |\mathbf{b} - \mathbf{a}|$ . (*Hint:* Vis først at  $||\mathbf{x}_n - \mathbf{a}| - |\mathbf{b} - \mathbf{a}|| \leq |\mathbf{x}_n - \mathbf{b}|$ ).

**5.** Vis at den åpnekulen  $B(\mathbf{a}, r)$  er en åpen mengde og at den lukkedekulen  $\overline{B}(\mathbf{a}, r)$  er en lukket mengde.

**6.** Vis at dersom  $\mathbf{c}$  er et randpunkt for  $A \subset \mathbb{R}^m$ , så finnes det to følger  $\{\mathbf{x}_n\}$  og  $\{\mathbf{y}_n\}$  som konvergerer mot  $\mathbf{c}$  slik at den ene følgen bare består av punkter som er med i  $A$ , mens den andre følgen bare består av punkter som ikke er med i  $A$ .

**7.** I denne oppgaven skal vi se litt på egenskapene til de åpne mengdene.

- a) Anta at  $A, B \subset \mathbb{R}^m$  er åpne mengder. Vis at  $A \cup B$  og  $A \cap B$  er åpne mengder.  
(Husk at  $A \cup B = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{x} \text{ er med i minst én av mengdene } A \text{ og } B\}$  og  $A \cap B = \{\mathbf{x} \in \mathbb{R}^m : \mathbf{x} \text{ er med i begge mengdene } A \text{ og } B\}$ .)

- b) Hvis  $A \subset \mathbb{R}^m$ , så kalles mengden

$$A^c = \{x \in \mathbb{R}^m : x \notin A\}$$

komplementet til  $A$ . Vis at  $A$  og  $A^c$  har samme rand.

- c) Vis at en mengde  $A \subset \mathbb{R}^m$  er åpen hvis og bare hvis komplementet  $A^c$  er lukket.

**8.** I denne oppgaven skal vi se litt på egenskapene til de lukkede mengdene. Det er lurt å gjøre oppgaven ovenfor før du begynner på denne.

- a) Anta at  $A, B \subset \mathbb{R}^m$  er lukkede mengder. Vis at  $A \cup B$  og  $A \cap B$  er lukkede mengder.

- b) Anta at  $A \subset \mathbb{R}^m$  er lukket og at  $B \subset \mathbb{R}^m$  er åpen. Vis at

$$A \setminus B = \{x \in A : x \notin B\}$$

er lukket.

- c) Vis at  $A \subset \mathbb{R}^m$  er lukket hvis og bare hvis  $A = \overline{A}$ , der *tillukningen*  $\overline{A}$  består av alle punkter som er med i enten  $A$  eller randen til  $A$  (dvs.  $\overline{A} = A \cup \partial A$  der  $\partial A$  er randen til  $A$ ).

**9.** (Eksamenspapir MAT 1110 13/6, 2008)

- a) Vis at dersom  $A$  er en  $m \times m$ -matrise og  $\{\mathbf{x}_n\}$  er en følge i  $\mathbb{R}^m$  som konvergerer mot  $\mathbf{0}$ , så konvergerer også  $\{A\mathbf{x}_n\}$  mot  $\mathbf{0}$ .
- b) Vis at dersom  $B$  er en inverterbar  $m \times m$ -matrise og følgen  $\{B\mathbf{x}_n\}$  konvergerer mot  $\mathbf{0}$ , så konvergerer også  $\{\mathbf{x}_n\}$  mot  $\mathbf{0}$ .
- c) Vis til slutt at dersom  $C$  er en ikke-inverterbar  $m \times m$ -matrise, så finnes det alltid en følge  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^m$  slik at  $\{C\mathbf{x}_n\}$  konvergerer mot  $\mathbf{0}$ , mens  $\{\mathbf{x}_n\}$  ikke konvergerer mot  $\mathbf{0}$ .

## 5.2 Kompletthet av $\mathbb{R}^m$

I denne seksjonen skal vi se på konvergens av følger fra en teoretisk synsvinkel, men resultatene vi kommer frem til, har stor praktisk nytte — blant annet når vi skal bruke datamaskiner til å finne numeriske løsninger på matematiske problemer. Du vil få se flere eksempler på dette i de neste seksjonene.

Vi starter med en følge av punkter i  $\mathbb{R}^m$

$$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}, \mathbf{x}_{11}, \dots, \mathbf{x}_n, \dots$$

og tenker oss at vi plukker ut uendelig mange av punktene i følgen (men sannsynligvis ikke alle). Vi kan for eksempel begynne med å plukke ut de elementene som vi har satt en strek under her:

$$\mathbf{x}_1, \mathbf{x}_2, \underline{\mathbf{x}_3}, \mathbf{x}_4, \underline{\mathbf{x}_5}, \underline{\mathbf{x}_6}, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \underline{\mathbf{x}_{10}}, \mathbf{x}_{11}, \dots, \mathbf{x}_n, \dots$$

På denne måten får vi en ny følge som begynner

$$\mathbf{x}_3, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_{10}, \dots$$

Denne nye følgen kalles en *delfølge* av den opprinnelige følgen.

La oss se litt mer formelt på dette. Dersom de leddene som vi plukker ut i den opprinnelige følgen, har nummer  $n_1, n_2, n_3, \dots, n_k, \dots$ , vil den nye følgen (delfølgen) ha elementene  $x_{n_1}, x_{n_2}, x_{n_3}, \dots, x_{n_k}, \dots$ . Kaller vi delfølgen  $\{\mathbf{y}_k\}$ , har vi altså  $\{\mathbf{y}_k\} = \{\mathbf{x}_{n_k}\}$ . Vi kan nå gi den presise definisjonen av en delfølge:

**Definisjon 5.2.1** Anta at  $\{\mathbf{x}_n\}$  er en følge av punkter i  $\mathbb{R}^m$  og at

$$n_1 < n_2 < n_3 < n_4 < \dots < n_k < \dots$$

er en strengt voksende følge av naturlige tall. Da kalles følgen  $\{\mathbf{y}_k\}$  der  $\mathbf{y}_k = \mathbf{x}_{n_k}$  en delfølge av  $\{\mathbf{x}_n\}$ .

Vi skal være interessert i samspillet mellom konvergens av følger og konvergens av delfølger. Det første resultatet er enkelt.

**Setning 5.2.2** Anta at en følge  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^m$  konvergerer mot et punkt  $\mathbf{x}$ . Da konvergerer også alle delfølger av  $\{\mathbf{x}_n\}$  mot  $\mathbf{x}$ .

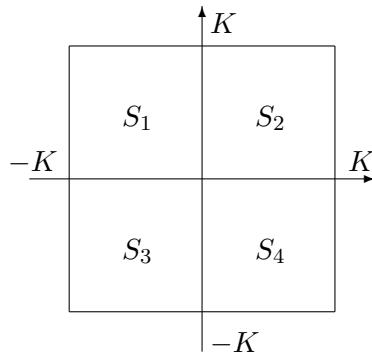
*Bevis:* Beviset overlates til leserne.  $\square$

Vi sier at en følge  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^m$  er *begrenset* dersom det finnes et tall  $K$  slik at  $|\mathbf{x}_n| \leq K$  for alle  $n$ . Det neste resultatet er nøkkelen til resten av denne seksjonen.

**Teorem 5.2.3 (Bolzano-Weierstrass' teorem)** Alle begrensede følger i  $\mathbb{R}^m$  har en konvergent delfølge.

*Bevis:* Beviset er enklest å oppfatte for følger i  $\mathbb{R}^2$ , men er lett å generalisere til høyere dimensjoner når man først har forstått det. Vi skal derfor bevise resultatet for en følge  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^2$  og overlate resten til leserne.

Siden følgen er begrenset, vet vi at det finnes et tall  $K$  slik at  $|\mathbf{x}_n| < K$  for alle  $n$ . Det betyr spesielt at alle elementene i følgen ligger innenfor det store kvadratet i figur 1.

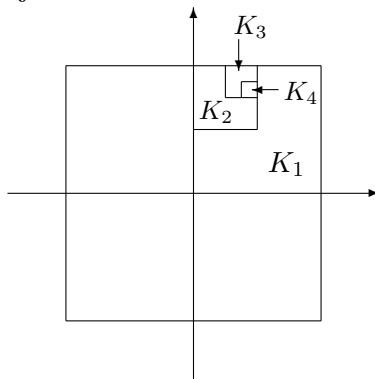


Figur 1: Kvadratene  $S_1, S_2, S_3$  og  $S_4$

Vi deler kvadratet opp i fire mindre kvadrater  $S_1, S_2, S_3, S_4$  som vist på figuren (disse kvadratene er lukkede og utgjør “hvert sitt hjørne” av det opprinnelige kvadratet). Siden følgen har uendelig mange ledd  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots$ , må minst ett av kvadratene  $S_1, S_2, S_3, S_4$  også inneholde uendelig mange ledd. Kall dette kvadratet  $K_1$  (hvis flere av kvadratene  $S_1, S_2, S_3, S_4$  inneholder

uendelig mange ledd, velger vi bare ett av dem). Vi deler nå dette kvadratet  $K_1$  i fire nye kvadrater på samme måte som før, og observerer at minst én av delene må inneholde uendelig mange ledd fra følgen. Dette kvadratet kaller vi  $K_2$ . Fortsetter vi på denne måten, får vi en følge av kvadrater  $K_1, K_2, K_3, \dots$  som ligger inni hverandre og som alle inneholder uendelig mange ledd fra følgen. Figur 2 viser hvordan en slik følge kan se ut.

Ideen er nå å plukke en delfølge  $\{\mathbf{x}_{n_k}\}$  av  $\{\mathbf{x}_n\}$  slik at  $\mathbf{x}_{n_k} \in K_k$  for alle  $k$ . Dette er mulig siden hvert kvadrat inneholder uendelig mange av leddene i den opprinnelige følgen, og det er intuitivt rimelig at en slik delfølge må konvergere siden leddene “er fangst i” mindre og mindre kvadrater. La oss se på detaljene.



Figur 2: Kvadratene  $K_1, K_2, K_3$  og  $K_4\dots$

For å konstruere delfølgen  $\{\mathbf{x}_{n_k}\}$  lar vi først  $n_1$  være det minste tallet slik at  $\mathbf{x}_{n_1} \in K_1$ . Deretter velger vi  $n_2$  til å være det første tallet etter  $n_1$  slik at  $\mathbf{x}_{n_2}$  er med i  $K_2$  — et slikt element må finnes siden  $K_2$  inneholder uendelig mange ledd fra følgen. På tilsvarende måte velger vi  $n_3$  til å være det første tallet etter  $n_2$  slik at  $\mathbf{x}_{n_3} \in K_3$ , osv. Vi har nå funnet en delfølge  $\{\mathbf{x}_{n_k}\}$  av  $\{\mathbf{x}_n\}$  slik at  $\mathbf{x}_{n_k} \in K_k$ , og det gjenstår å vise at den konvergerer.

La  $(a_k, b_k)$  være koordinatene til det nedre, venstre hjørne til kvadratet  $K_k$ . Det følger fra konstruksjonen at  $\{a_k\}$  og  $\{b_k\}$  er voksende (dvs. ikke avtagende), begrensede følger. Ifølge teorem 4.3.9 i *Kalkulus* må de da konvergere mot hver sin grenseverdi  $a$  og  $b$ , og følgelig vil punktene  $\mathbf{z}_k = (a_k, b_k)$  konvergere mot  $\mathbf{z} = (a, b)$  (husk setning 5.1.5). Siden både  $\mathbf{z}_k$  og  $\mathbf{x}_{n_k}$  ligger i kvadratet  $K_k$ , og størrelsen av dette kvadratet går mot null, må da også følgen  $\{\mathbf{x}_{n_k}\}$  konvergere mot  $\mathbf{z}$ . Dermed er teoremet bevist.  $\square$

Som du snart vil få anledning til å se, er teoremet ovenfor et usedvanlig nyttig redskap når man arbeider med følger. Et enda nyttigere redskap er *Cauchy-følger*.

**Definisjon 5.2.4** En følge  $\{\mathbf{x}_n\}$  i  $\mathbb{R}^m$  er en Cauchy-følge dersom det for enhver  $\epsilon > 0$ , finnes en  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{x}_k| < \epsilon$  for alle  $n, k \geq N$

En følge er altså en Cauchy-følge dersom vi kan få avstanden mellom to ledd til å bli vilkårlig liten ved å gå tilstrekkelig langt ut i følgen. Det er ikke vanskelig å se at alle konvergente følger er Cauchy-følger.

**Lemma 5.2.5** *Enhver konvergent følge i  $\mathbb{R}^m$  er en Cauchy-følge*

*Bevis:* Anta at  $\mathbf{x}_n$  konvergerer mot  $\mathbf{x}$ . Gitt et tall  $\epsilon > 0$ , vet vi da at det finnes et tall  $N \in \mathbb{N}$  slik at  $|\mathbf{x}_n - \mathbf{x}| < \frac{\epsilon}{2}$  når  $n \geq N$ . Hvis både  $n$  og  $k$  er større enn eller lik  $N$ , har vi dermed ved trekantulikheten

$$|\mathbf{x}_n - \mathbf{x}_k| = |(\mathbf{x}_n - \mathbf{x}) + (\mathbf{x} - \mathbf{x}_k)| \leq |\mathbf{x}_n - \mathbf{x}| + |\mathbf{x} - \mathbf{x}_k| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Dermed har vi vist at  $\mathbf{x}_n$  er en Cauchy-følge.  $\square$

Det neste resultatet er mer overraskende — og atskillig mer nyttig.

**Teorem 5.2.6** *Alle Cauchy-følger i  $\mathbb{R}^m$  konvergerer.*

*Bevis:* Gangen i beviset er som følger: Først viser vi at enhver Cauchy-følge er begrenset, deretter bruker vi teorem 5.2.3 til å plukke ut en konvergent delfølge, og til slutt viser vi at Cauchy-følgen konvergerer mot det samme punktet som delfølgen.

Anta at  $\{\mathbf{x}_n\}$  er en Cauchy-følge. For å vise at  $\{\mathbf{x}_n\}$  er begrenset, velger vi en  $\epsilon > 0$  (f.eks.  $\epsilon = 1$ ). Da finnes det en  $N$  slik at  $|\mathbf{x}_n - \mathbf{x}_k| < \epsilon$  når  $n, k \geq N$ . Spesielt må  $|\mathbf{x}_n - \mathbf{x}_N| < \epsilon$  for alle  $n \geq N$ , og ved trekantulikheten betyr det at

$$|\mathbf{x}_n| = |\mathbf{x}_N + (\mathbf{x}_n - \mathbf{x}_N)| \leq |\mathbf{x}_N| + |\mathbf{x}_n - \mathbf{x}_N| < |\mathbf{x}_N| + \epsilon$$

når  $n \geq N$ . Dette betyr at følgen er begrenset av det største av tallene

$$|\mathbf{x}_1|, |\mathbf{x}_2|, |\mathbf{x}_3|, \dots, |\mathbf{x}_{N-1}|, |\mathbf{x}_N| + \epsilon$$

Siden følgen  $\{\mathbf{x}_n\}$  er begrenset, har den ifølge teorem 5.2.3 en konvergent delfølge  $\{\mathbf{x}_{n_k}\}$  med en grenseverdi  $\mathbf{x}$ . Vi skal vise at også den opprinnelige følgen  $\{\mathbf{x}_n\}$  konvergerer mot  $\mathbf{x}$ . Gitt en  $\epsilon > 0$ , må vi da vise at det finnes en  $N$  slik at  $|\mathbf{x}_n - \mathbf{x}| < \epsilon$  når  $n \geq N$ . Det er ikke så vanskelig: Siden  $\{\mathbf{x}_n\}$  er en Cauchy-følge, finnes det en  $N$  slik at  $|\mathbf{x}_n - \mathbf{x}_k| < \frac{\epsilon}{2}$  når  $n, k \geq N$ . Siden delfølgen  $\{\mathbf{x}_{n_k}\}$  konvergerer mot  $\mathbf{x}$ , finnes det et element  $\mathbf{x}_{n_K}$  i denne følgen slik at  $n_K \geq N$  og  $|\mathbf{x}_{n_K} - \mathbf{x}| < \frac{\epsilon}{2}$ . Hvis  $n \geq N$ , er dermed

$$|\mathbf{x}_n - \mathbf{x}| = |(\mathbf{x}_n - \mathbf{x}_{n_K}) + (\mathbf{x}_{n_K} - \mathbf{x})| \leq |\mathbf{x}_n - \mathbf{x}_{n_K}| + |\mathbf{x}_{n_K} - \mathbf{x}| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Dermed er teoremet bevist.  $\square$

Kombinerer vi teoremet ovenfor med det foregående lemmaet, får vi denne eksakte sammenhengen:

**Korollar 5.2.7** En følge i  $\mathbb{R}^m$  konvergerer hvis og bare hvis den er en Cauchy-følge.

**Bemerkning:** Noen synes kanskje at resultatene ovenfor er så opplagte at det ikke er noen grunn til å bevise dem — det er da en selvfølge at Cauchy-følger og konvergente følger er det samme! Det er imidlertid lett å overbevise seg om at dette ikke nødvendigvis er tilfellet. Anta at vi foretrakk å gjøre vår matematikk i mengden  $\mathbb{Q}$  av rasjonale tall istedenfor i mengden  $\mathbb{R}$  av reelle tall. Følgen

$$x_0 = 1, x_2 = 1.4, x_3 = 1.41, x_4 = 1.4142, x_5 = 1.41421, \dots$$

(bestående av lengre og lengre desimaltallstilnærmingar til  $\sqrt{2}$ ) er en Cauchy-følge i  $\mathbb{Q}$ , men den konvergerer ikke i  $\mathbb{Q}$  siden grensen er det irrasjonale tallet  $\sqrt{2}$ .

Det er mulig å definere konvergens og Cauchy-følger i andre mengder (matematikere kaller dem gjerne *metriske rom*) enn  $\mathbb{R}^m$  — det viser seg at alt vi trenger, er et mål for avstanden mellom to punkter (en *metrikk*). Et slikt metrisk rom kalles *komplett* dersom alle Cauchy-følger konvergerer. Teoremet ovenfor forteller oss altså at  $\mathbb{R}^m$  er komplett, mens eksemplet i begynnelsen av denne bemerkningen viser at  $\mathbb{Q}$  ikke er komplett. Du har støtt på ordet *komplett* før i forbindelse med reelle tall — i seksjon 2.3 i *Kalkulus* studerte vi kompletthetsprinsippet for  $\mathbb{R}$  (det sier at alle begrensete, ikke-tomme delmengder av  $\mathbb{R}$  har en minste øvre skranke). Det viser seg å være en nær sammenheng mellom disse to formene for kompletthet, og vi kunne ha basert vår diskusjon av  $\mathbb{R}$  på konvergens av Cauchy-følger istedenfor eksistens av minste øvre skranker.

Selv om det kankje ikke er så lett å se ved første øyekast, er teoremet ovenfor et meget viktig redskap når man skal studere konvergens av følger — det er ofte mye lettere å vise at en følge  $\{\mathbf{x}_n\}$  konvergerer ved å vise at den er en Cauchy-følge enn ved å bruke definisjonen av konvergens. Grunnen er at for å bruke definisjonen, må vi først finne grensepunktet  $\mathbf{x}$  og så vise at vi kan få  $|\mathbf{x}_n - \mathbf{x}| < \epsilon$  mindre enn  $\epsilon$  ved å velge  $N$  stor nok. For å vise at  $\{\mathbf{x}_n\}$  er en Cauchy-følge, trenger vi ikke kjenne grensepunktet — alt vi skal sjekke, er at vi kan få  $|\mathbf{x}_n - \mathbf{x}_k| < \epsilon$  mindre enn  $\epsilon$  ved å velge  $n$  og  $k$  store nok. Vi arbeider altså med de gitte verdiene  $\mathbf{x}_n$  og  $\mathbf{x}_k$  og trenger ikke å vite noe om grensepunktet på forhånd. Vi skal se slående eksempler på denne teknikken i de neste tre seksjonene.

## Oppgaver til seksjon 5.2

1. Bevis setning 5.2.2.
2. Bevis setning 5.2.3 for følger i et vilkårlig rom  $\mathbb{R}^m$ .

**3.** Anta  $I_n$  er en følge av intervaller i  $\mathbb{R}$  slik at  $I_{n+1} \subset I_n$  og lengden til  $I_n$  går mot null. Vis at det finnes nøyaktig ett tall som er med i alle intervallene  $I_n$ .

**4.** Anta at  $\{\mathbf{x}_n\}$  er en følge i  $\mathbb{R}^m$ . Et punkt  $\mathbf{x}$  i  $\mathbb{R}^m$  kalles et *oppfhopningspunkt* for  $\{\mathbf{x}_n\}$  hvis enhver kule  $B(\mathbf{x}, r)$  inneholder uendelig mange elementer i følgen.

- a) Vis at  $\mathbf{x}$  et et oppfhopningspunkt for  $\{\mathbf{x}_n\}$  hvis og bare hvis  $\{\mathbf{x}_n\}$  har en delfølge som konvergerer mot  $\mathbf{x}$
- b) Anta at  $A$  er en lukket, begrenset delmengde av  $\mathbb{R}^m$ . Vis at enhver følge i  $A$  har et oppfhopningspunkt i  $A$ .
- c) Anta at  $A \subset \mathbb{R}^m$  ikke er lukket. Vis at det finnes en følge i  $A$  som ikke har et oppfhopningspunkt i  $A$ .
- d) Anta at  $A \subset \mathbb{R}^m$  ikke er begrenset. Vis at det finnes en følge i  $A$  som ikke har et oppfhopningspunkt i  $A$ .

**5.** Anta at  $\mathbf{a}$  er et punkt i  $\mathbb{R}^m$  og at  $B$  er en delmengde i  $\mathbb{R}^m$ . Vi definerer *avstanden fra  $\mathbf{a}$  til  $B$*  til å være

$$d(\mathbf{a}, B) = \inf\{|\mathbf{a} - \mathbf{b}| : \mathbf{b} \in B\}$$

- a) La  $B = \{(x, y) \in \mathbb{R}^2 : x > 1\}$ . Vis at  $d(\mathbf{0}, B) = 1$ , men at det ikke finnes noe punkt  $\mathbf{b} \in B$  slik at  $|\mathbf{0} - \mathbf{b}| = 1$ .
- b) Vis at dersom  $\mathbf{a} \in \mathbb{R}^m$  og  $B$  er en lukket, begrenset delmengde av  $\mathbb{R}^m$ , så finnes det et punkt  $\mathbf{b} \in B$  slik at  $|\mathbf{a} - \mathbf{b}| = d(\mathbf{a}, B)$ . Vis ved et eksempel at det godt kan være flere slike punkter  $\mathbf{b}$ .
- c) Vis at dersom  $\mathbf{a} \in \mathbb{R}^m$  og  $B$  er en lukket (men ikke nødvendigvis begrenset) delmengde av  $\mathbb{R}^m$ , så finnes det et punkt  $\mathbf{b} \in B$  slik at  $|\mathbf{a} - \mathbf{b}| = d(\mathbf{a}, B)$ .
- d) Mengden  $B$  kalles *konveks* dersom den har følgende egenskap: Dersom  $\mathbf{c}$  og  $\mathbf{d}$  er to punkter i  $B$ , så ligger også hele linjestykket som forbinder  $\mathbf{c}$  og  $\mathbf{d}$  i  $B$  (se figur 5.5.1 hvis du føler behov for en illustrasjon). Vis at dersom  $B$  er en lukket, konveks delmengde av  $\mathbb{R}^m$ , så finnes det et *entydig* punkt  $\mathbf{b} \in B$  slik at  $|\mathbf{a} - \mathbf{b}| = d(\mathbf{a}, B)$ .

### 5.3 Noen konsekvenser av kompletthet

I denne seksjonen skal vi se på en del anvendelser av resultatene i forrige seksjon. Seksjonen kan virke som en liten avstikker fra hovedtemaet i dette kapitlet, og det er mulig å hoppe over den og gå direkte løs på seksjon 5.4. Man bør da være klar over at teorem 5.3.2 brukes for å bevise integrerbarhet av kontinuerlige funksjoner i seksjon 6.1 og at teorem 5.3.11 brukes i det (\*-merkede) beviset for Kantorovichs teorem i seksjon 5.6.

#### Uniform kontinuitet

Vi skal først se på et begrep som blir svært viktig når vi skal studere integrasjon av funksjoner av flere variable i kapittel 6. I denne seksjonen skal vi

bruke det til å studere derivasjon under integraltegnet og uavhengigheten av veien for linjeintegraler.

Vi begynner med en definisjon:

**Definisjon 5.3.1** *Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable. Vi sier at  $f$  er uniformt kontinuerlig på en mengde  $B \subset A$  dersom det til enhver  $\epsilon > 0$  finnes en  $\delta > 0$ , slik at hvis  $\mathbf{u}, \mathbf{v} \in B$  og  $|\mathbf{u} - \mathbf{v}| < \delta$ , så er  $|f(\mathbf{u}) - f(\mathbf{v})| < \epsilon$ .*

Ved første øyekast er det ikke lett å se forskjell på denne definisjonen og definisjonen av vanlig kontinuitet. Nøkkelen ligger i ordet “uniform” som antyder at funksjonen skal være “like kontinuerlig” i alle punkter, dvs. at gitt en  $\epsilon > 0$ , kan vi finne en  $\delta > 0$  som fungerer uansett hvilket punkt  $\mathbf{u} \in B$  vi befinner oss i. Med vanlig kontinuitet kan det tenkes at vi til samme  $\epsilon$  må bruke forskjellig  $\delta$  avhengig av hvilket punkt  $\mathbf{u}$  vi sjekker kontinuiteten i, og at det ikke er noen  $\delta$  som fungerer i alle punkter. Et eksempel på en kontinuerlig, men ikke uniformt kontinuerlig funksjon av én variabel, er  $f(x) = x^2$  med  $B = \mathbb{R}$  — denne funksjonen blir brattere og brattere dess større  $x$  blir i tallverdi, og vi trenger derfor mindre og mindre  $\delta$  til å takle samme  $\epsilon$ .

Er du fortsatt forvirret over sammenhengen mellom uniform kontinuitet og vanlig kontinuitet, er det ikke sikkert det neste resultatet hjelper — det sier at på lukkede, begrensete mengder er det ikke noen forskjell på de to begrepene! Beviset bygger på Bolzano-Weierstrass' teorem fra forrige seksjon.

**Teorem 5.3.2** *Anta at  $K$  er en lukket, begrenset delmengde av  $\mathbb{R}^n$ . Enhver funksjon  $f$  som er kontinuerlig på  $K$ , er også uniformt kontinuerlig på  $K$ .*

*Bevis:* Anta at  $f$  ikke er uniformt kontinuerlig på  $K$ . Vi må vise at da kan  $f$  heller ikke være (vanlig) kontinuerlig på hele  $K$ . Siden  $f$  ikke er uniformt kontinuerlig, finnes det en  $\epsilon > 0$  slik at uansett hvor liten vi velger  $\delta > 0$ , så finnes det punkter  $\mathbf{u}, \mathbf{v} \in K$  slik at  $|\mathbf{u} - \mathbf{v}| < \delta$ , men  $|f(\mathbf{u}) - f(\mathbf{v})| \geq \epsilon$ . Spesielt kan vi for hver  $n \in \mathbb{N}$ , finne punkter  $\mathbf{u}_n, \mathbf{v}_n \in K$  slik at  $|\mathbf{u}_n - \mathbf{v}_n| < \frac{1}{n}$ , men  $|f(\mathbf{u}_n) - f(\mathbf{v}_n)| \geq \epsilon$ . Siden  $K$  er lukket og begrenset, vet vi fra Bolzano-Weierstrass' teorem at  $\{\mathbf{u}_n\}$  har en delfølge  $\{\mathbf{u}_{n_k}\}$  som konvergerer mot et punkt  $\mathbf{u} \in K$ . Siden avstanden mellom  $\mathbf{u}_{n_k}$  og  $\mathbf{v}_{n_k}$  går mot null, må også  $\mathbf{v}_{n_k}$  gå mot  $\mathbf{u}$ . Det er nå lett å se at  $f$  ikke kan være kontinuerlig i  $\mathbf{u}$ . Hadde den vært det, ville (ifølge setning 5.1.7) både  $f(\mathbf{u}_{n_k})$  og  $f(\mathbf{v}_{n_k})$  gått mot  $f(\mathbf{u})$ , men det er umulig siden  $|f(\mathbf{u}_{n_k}) - f(\mathbf{v}_{n_k})| \geq \epsilon$ . Dette viser at vi har funnet et punkt  $\mathbf{u} \in K$  der  $f$  ikke er kontinuerlig, og beviset er fullført.  $\square$

Vi skal nå se hvordan uniform kontinuitet kan brukes til å begrunne en nyttig teknikk som kalles “derivasjon under integraltegnet”. Sannsynligvis vil du før eller senere støte på denne teknikken i andre fag, men siden vi

ikke har bruk for den andre steder i dette kompendiet, kan du godt hoppe over resten av seksjonen uten å gå glipp av noe i fortsettelsen.

Tenk deg at du har en funksjon av flere variable  $\mathbf{x} = (x_1, x_2, \dots, x_m)$  definert som et integral

$$F(\mathbf{x}) = \int_a^b G(\mathbf{x}, t) dt$$

og at du har behov for å derivere  $F$  med hensyn på en av variablene  $x_1, x_2, \dots, x_m$ . Under rimelige betingelser kan du da “derivere under integraltegnet” og få

$$\frac{\partial F}{\partial x_i}(\mathbf{x}) = \int_a^b \frac{\partial G}{\partial x_i}(\mathbf{x}, t) dt$$

Det er ikke så vanskelig å se at denne formelen “burde være” riktig; bruker vi definisjonen av partiellderivert, får vi

$$\begin{aligned} \frac{\partial F}{\partial x_i}(\mathbf{x}) &= \lim_{h \rightarrow 0} \frac{F(x_1, \dots, x_i + h, \dots, x_m) - F(x_1, \dots, x_i, \dots, x_m)}{h} = \\ &= \lim_{h \rightarrow 0} \int_a^b \frac{G(x_1, \dots, x_i + h, \dots, x_m, t) - G(x_1, \dots, x_i, \dots, x_m, t)}{h} dt \end{aligned}$$

Hvis vi tillater oss å flytte grenseverdien forbi integraltegnet, får vi

$$\begin{aligned} \frac{\partial F}{\partial x_i}(\mathbf{x}) &= \lim_{h \rightarrow 0} \int_a^b \frac{G(x_1, \dots, x_i + h, \dots, x_m, t) - G(x_1, \dots, x_i, \dots, x_m, t)}{h} dt = \\ &= \int_a^b \lim_{h \rightarrow 0} \frac{G(x_1, \dots, x_i + h, \dots, x_m, t) - G(x_1, \dots, x_i, \dots, x_m, t)}{h} dt = \int_a^b \frac{\partial G}{\partial x_i}(\mathbf{x}, t) dt \end{aligned}$$

som er resultatet vi ønsker oss. Det viser seg imidlertid at vi ikke uhemmet kan flytte grenseverdier forbi integraltegn — i noen tilfeller vil det faktisk gi galt svar! Vi skal nå bruke uniform kontinuitet til å vise at svaret blir riktig i dette tilfellet.

**Teorem 5.3.3 (Derivasjon under integraltegnet)** *Anta  $A \subset \mathbb{R}^m$  er en åpen mengde i  $\mathbb{R}^m$  og at  $a, b$  er to reelle tall,  $a < b$ . Anta videre at  $G(\mathbf{x}, t)$  er en funksjon av  $m + 1$ -variable som er kontinuerlig og har kontinuerlige partiellderiverte  $\frac{\partial G}{\partial x_i}(\mathbf{x}, t)$  i alle punkter  $(\mathbf{x}, t)$  der  $\mathbf{x} \in A$  og  $a \leq t \leq b$ . Da har funksjonen  $F(\mathbf{x}) = \int_a^b G(\mathbf{x}, t) dt$  kontinuerlige partiellderiverte i alle punkter  $\mathbf{x} \in A$  og*

$$\frac{\partial F}{\partial x_i}(\mathbf{x}) = \int_a^b \frac{\partial G}{\partial x_i}(\mathbf{x}, t) dt$$

*Bevis:* Velg et punkt  $\mathbf{x} = (x_1, x_2, \dots, x_m) \in A$ . Siden  $A$  er åpen, finnes det en lukket kule  $\bar{B}(\mathbf{x}, r)$  som ligger inni  $A$ . Da er mengden  $C = \{(\mathbf{x}, t) \in$

$\mathbb{R}^{m+1} \mid \mathbf{x} \in \overline{B}(\mathbf{x}), a \leq t \leq b\}$  lukket og begrenset, og funksjonen  $\frac{\partial G}{\partial x_i}$  er ifølge teorem 5.3.2 uniformt kontinuerlig på  $C$ .

For hver  $h < r$  finnes det ifølge middelverdisetningen (se seksjon 6.2 i *Kalkulus*) et tall  $c_h$  mellom  $x_i$  og  $x_i + h$  slik at

$$\frac{G(x_1, \dots, x_i + h, \dots, x_m) - G(x_1, \dots, x_i, \dots, x_m)}{h} = \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m)$$

Dette betyr at

$$\begin{aligned} \frac{F(x_1, \dots, x_i + h, \dots, x_m) - F(x_1, \dots, x_i, \dots, x_m)}{h} &= \\ &= \int_a^b \frac{G(x_1, \dots, x_i + h, \dots, x_m, t) - G(x_1, \dots, x_i, \dots, x_m, t)}{h} dt \\ &= \int_a^b \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m) dt \end{aligned}$$

Hvis vi legger til og trekker fra  $\int_a^b \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m) dt$  i det siste uttrykket, får vi

$$\begin{aligned} \frac{F(x_1, \dots, x_i + h, \dots, x_m) - F(x_1, \dots, x_i, \dots, x_m)}{h} &= \\ &= \int_a^b \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m, t) dt + \\ &\quad + \int_a^b \left( \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m, t) - \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m, t) \right) dt \end{aligned}$$

For å vise at  $\frac{\partial F}{\partial x_i}(\mathbf{x}) = \int_a^b \frac{\partial G}{\partial x_i}(\mathbf{x}, t) dt$  er det dermed nok å vise at

$$\lim_{h \rightarrow 0} \int_a^b \left( \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m, t) - \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m, t) \right) dt = 0$$

Vi bruker definisjonen av grenseverdi; gitt en  $\epsilon > 0$ , må vi finne en  $\delta > 0$  slik at

$$\left| \int_a^b \left( \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m, t) - \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m, t) \right) dt \right| < \epsilon$$

når  $h < \delta$ . Siden  $\frac{\partial G}{\partial x_i}$  er uniformt kontinuerlig på  $C$ , vet vi at det finnes en  $\delta > 0$  slik at hvis  $\mathbf{u}, \mathbf{v} \in C$  og  $|\mathbf{u} - \mathbf{v}| < \delta$ , så er  $|G(\mathbf{u}) - G(\mathbf{v})| < \frac{\epsilon}{b-a}$ . Spesielt betyr dette at hvis  $h < \delta$ , så er

$$\left| \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m, t) - \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m, t) \right| < \frac{\epsilon}{b-a}$$

for alle  $t \in [a, b]$ , og dermed er

$$\begin{aligned} & \left| \int_a^b \left( \frac{\partial G}{\partial x_i}(x_1, \dots, c_h, \dots, x_m, t) - \frac{\partial G}{\partial x_i}(x_1, \dots, x_i, \dots, x_m, t) \right) dt \right| < \\ & < \int_a^b \frac{\epsilon}{b-a} dt = \frac{\epsilon}{b-a} \cdot (b-a) = \epsilon \end{aligned}$$

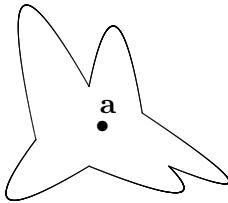
Det gjenstår å vise at  $\frac{\partial F}{\partial x_i}(\mathbf{x})$  er kontinuerlig. Siden

$$\left| \frac{\partial F}{\partial x_i}(\mathbf{x}) - \frac{\partial F}{\partial x_i}(\mathbf{y}) \right| = \left| \int_a^b \left( \frac{\partial G}{\partial x_i}(\mathbf{x}, t) - \frac{\partial G}{\partial x_i}(\mathbf{y}, t) \right) dt \right|$$

kan dette gjøres ved å bruke uniform kontinuitet på akkurat samme måte som ovenfor. Detaljene overlates til leserne.  $\square$

For å vise en typisk anvendelse av derivasjon under integraltegnet, tar vi turen tilbake til seksjon 3.5. Der viste vi i setning 3.5.3 at dersom et vektorfelt  $\mathbf{F}$  er konservativt i et område  $A$ , så er  $\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$  for alle  $\mathbf{x} \in A$ . Vi viste også ved et eksempel at det er mulig å ha betingelsene  $\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$  oppfylt for alle  $\mathbf{x}$  i et område  $A$  uten at  $\mathbf{F}$  er konservativt i  $A$ . Problemet oppstår hvis området vi ser på er gjennomhullet, og vi nevnte at dersom området  $A$  er såkalt *enkeltsammenhengende*, så medfører betingelsene  $\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$  for alle  $\mathbf{x} \in A$  (og alle indeks  $i, j$ ) at  $\mathbf{F}$  er konservativt i  $A$ . Dette resultatet er fortsatt litt for vanskelig for oss, men ved hjelp av derivasjon under integraltegnet er det lett å vise at konklusjonen holder for en enklere klasse områder, de som kalles stjerneformede.

Et område  $A \subset \mathbb{R}^m$  er *stjerneformet* dersom det finnes et punkt  $\mathbf{a} \in A$  slik at for alle andre punkter  $\mathbf{x}$  i  $A$  ligger hele linjestykket fra  $\mathbf{a}$  til  $\mathbf{x}$  i  $A$ . Legg merke til at en kule er et stjerneformet i henhold til denne definisjonen, så et stjerneformet område behøve slett ikke å ligne så mye på en stjerne — poenget er heller at stjerner er typiske eksempler på stjerneformede områder.



Figur 1: Et stjerneformet område

**Setning 5.3.4** *Anta at  $A \subset \mathbb{R}^m$  er et åpent og stjerneformet område, og at vektorfeltet  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  har kontinuerlige partiellderiverte slik at*

$$\frac{\partial F_i}{\partial x_j}(\mathbf{x}) = \frac{\partial F_j}{\partial x_i}(\mathbf{x})$$

*for alle  $\mathbf{x} \in A$  og alle  $i, j = 1, 2, \dots, m$ . Da er  $\mathbf{F}$  konservativt i  $A$ .*

*Bevis:* For å slippe at notasjonen blir altfor uoversiktelig, antar vi at “sentret”  $\mathbf{a}$  i  $A$  er origo (beviset for det generelle tilfellet går på akkurat samme måte). Linjestykket fra  $\mathbf{a} = \mathbf{0}$  er da gitt ved parametriseringen

$$\mathbf{r}(t) = t\mathbf{x} = (tx_1, tx_2, \dots, tx_m) \quad t \in [0, 1]$$

som gir  $\mathbf{r}'(t) = \mathbf{x}$ . Vi definerer funksjonen  $\phi : A \rightarrow \mathbb{R}$  ved

$$\begin{aligned} \phi(\mathbf{x}) &= \int_0^1 \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt = \int_0^1 \mathbf{F}(t\mathbf{x}) \cdot \mathbf{x} dt = \\ &= \int_0^1 \sum_{j=1}^m F_j(tx_1, \dots, tx_i, \dots, tx_m) x_j dt \end{aligned}$$

Målet er å vise at  $\phi$  er en potensialfunksjon til  $\mathbf{F}$ , dvs. at  $\frac{\partial \phi}{\partial x_i}(\mathbf{x}) = \mathbf{F}_i(\mathbf{x})$ . Deriverer vi under integraltegnet i uttrykket ovenfor, får vi

$$\begin{aligned} \frac{\partial \phi}{\partial x_i}(\mathbf{x}) &= \int_0^1 \frac{\partial}{\partial x_i} \left( F_j(tx_1, \dots, tx_i, \dots, tx_m) x_j \right) dt \\ &= \int_0^1 \left( \sum_{j=1}^m t x_j \frac{\partial F_j}{\partial x_i}(tx_1, \dots, tx_i, \dots, tx_m) + F_i(tx_1, \dots, tx_i, \dots, x_m) \right) dt \end{aligned}$$

der vi har brukt både kjerneregelen og produktregelen under integraltegnet. Benytter vi betingelsen  $\frac{\partial F_j}{\partial x_i} = \frac{\partial F_i}{\partial x_j}$ , får vi

$$\frac{\partial \phi}{\partial x_i}(\mathbf{x}) = \int_0^1 \left( \sum_{j=1}^m t x_j \frac{\partial F_i}{\partial x_j}(tx_1, \dots, tx_i, \dots, tx_m) + F_i(tx_1, \dots, tx_i, \dots, x_m) \right) dt$$

Observer nå at ved produkt- og kjerneregelen er

$$\begin{aligned} \frac{d}{dt} (t F_i(tx_1, \dots, tx_i, \dots, tx_m)) &= F_i(tx_1, \dots, tx_i, \dots, x_m) \\ &\quad + \sum_{j=1}^m t x_j \frac{\partial F_i}{\partial x_j}(tx_1, \dots, tx_i, \dots, tx_m) \end{aligned}$$

Ved analysens fundamentalteorem er dermed

$$\begin{aligned} \frac{\partial \phi}{\partial x_i}(\mathbf{x}) &= \int_0^1 \frac{d}{dt} (t F_i(tx_1, \dots, tx_i, \dots, tx_m)) dt = \\ &= \left[ t F_i(tx_1, \dots, tx_i, \dots, tx_m) \right]_0^1 = F_i(\mathbf{x}) \end{aligned}$$

og beviset er ferdig.  $\square$

Beviset ovenfor er ganske typisk — mange funksjoner i matematikk og andre fag er definert ved hjelp av bestemte integraler, og skal vi derivere slike funksjoner, må vi vanligvis derivere under integraltegnet.

### \*Operatornorm og inverterbarhet

I dette avsnittet skal vi bruke kompletthet til å utlede et nyttig kriterium for inverterbarhet av matriser. Vi skal få bruk for dette kriteriet når vi beviser Kantorovitsjs teorem om konvergens av Newtons metode i seksjon 5.6. Det beviset er imidlertid ment for de spesielt interesserte, så du kan hoppe over dette avsnittet uten å miste noe i fortsettelsen.

Inspirasjonskilden til kriteriet er summeformelen for en geometrisk rekke

$$\frac{1}{1-x} = 1 + x + x^2 + \dots + x^n + \dots \quad \text{for } |x| < 1$$

Vi skal vise en tilsvarende formel for  $m \times m$ -matriser  $A$ , nemlig at hvis  $|A| < 1$  (vi skal forklare hva dette betyr om et øyeblikk), så er  $I_m - A$  inverterbar og

$$(I_m - A)^{-1} = I_m + A + A^2 + \dots + A^n + \dots$$

En slik geometrisk rekke av matriser kalles ofte en *Neumann-rekke* etter den tyske matematikeren Carl Neumann (1832-1925). Formelen ovenfor er nyttig for å vise inverterbarhet av matriser som ikke avviker for mye fra identitetsmatrisen.

La oss begynne med å forklare hva vi mener med  $|A|$ . I slutten av seksjon 1.6 innførte vi *normen*  $\|A\|$  til en  $m \times m$ -matrise

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

ved

$$\|A\| = \sqrt{\sum_{1 \leq i,j \leq m} a_{ij}^2}$$

og beviste (setning 1.6.3) at

$$|Ax| \leq \|A\| |x| \quad \text{for alle } \mathbf{x} \in \mathbb{R}^m$$

Deler vi med  $|\mathbf{x}|$  i den siste ligningen, får vi

$$\frac{|Ax|}{|\mathbf{x}|} \leq \|A\| \quad \text{for alle } \mathbf{x} \neq \mathbf{0} \tag{5.3.1}$$

Vi skal nå innføre et annet mål på størrelsen til en matrise som ofte er mer effektivt enn normen.

**Definisjon 5.3.5** *Anta at  $A$  er en  $m \times m$ -matrise. Operatornormen  $|A|$  til  $A$  er definert ved*

$$|A| = \sup \left\{ \frac{|Ax|}{|\mathbf{x}|} : \mathbf{x} \in \mathbb{R}^m, \mathbf{x} \neq \mathbf{0} \right\}$$

Legg merke til at ifølge formel (5.3.1) ovenfor er mengden i definisjonen begrenset av  $\|A\|$ , så den minste øvre skranken  $|A|$  finnes. Fra formel (5.3.1) følger det også at

$$|A| \leq \|A\|$$

Vi understreker igjen at normen  $\|A\|$  og operatornormen  $|A|$  er to forskjellige måter å måle størrelsen til en matrise på. Normen  $\|A\|$  er som regel lettest å beregne, mens operatornormen  $|A|$  ofte er lettere å bruke i teoretiske argumenter. På grunn av ulikheten  $|A| \leq \|A\|$  er det ofte mulig å erstatte operatornormen med normen i beregningsproblemer og dermed få enklere beregninger.

Det følger fra definisjonen av operatornorm at

$$|Ax| \leq |A||\mathbf{x}| \quad \text{for alle } \mathbf{x} \in \mathbb{R}^m$$

Ved hjelp av denne ulikheten kan vi vise våre første resultater.

**Lemma 5.3.6** *For enhver kvadratisk matrise  $A$  er*

$$|a_{ij}| \leq |A|$$

*for alle elementer  $a_{ij}$  i matrisen.*

*Bevis:* Ganger vi  $A$  med enhetsvektoren  $\mathbf{e}_j$ , får vi

$$A\mathbf{e}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

Tar vi normen på begge sider og bruker formelen ovenfor, får vi (husk at  $|\mathbf{e}_j| = 1$ ):

$$|A| = |A||\mathbf{e}_j| \geq |A\mathbf{e}_j| = \left| \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \right| = \sqrt{a_{1j}^2 + a_{2j}^2 + \cdots + a_{mj}^2} \geq |a_{ij}| \quad \square$$

**Lemma 5.3.7** *Hvis  $A, B$  er  $m \times m$ -matriser, så er*

$$(i) |AB| \leq |A||B|$$

(ii)  $|A + B| \leq |A| + |B|$

*Bevis:* (i) For enhver  $\mathbf{x} \neq \mathbf{0}$  i  $\mathbb{R}^m$  har vi

$$|(AB)\mathbf{x}| = |A(B\mathbf{x})| \leq |A||B\mathbf{x}| \leq |A||B||\mathbf{x}|$$

der vi har brukt ulikheten  $|C\mathbf{y}| \leq |C||\mathbf{y}|$  to ganger, først med  $C = A$  og  $\mathbf{y} = B\mathbf{x}$ , og så med  $C = B$  og  $\mathbf{y} = \mathbf{x}$ . Deler vi på  $|\mathbf{x}|$  i ulikheten ovenfor, får vi

$$\frac{|(AB)\mathbf{x}|}{|\mathbf{x}|} \leq |A||B|$$

for alle  $\mathbf{x} \neq \mathbf{0}$ . Fra definisjonen av operatornorm får vi dermed at  $|AB| \leq |A||B|$

(ii) For enhver  $\mathbf{x} \neq \mathbf{0}$  i  $\mathbb{R}^m$  har vi

$$|(A + B)\mathbf{x}| = |A\mathbf{x} + B\mathbf{x}| \leq |A\mathbf{x}| + |B\mathbf{x}| \leq |A||\mathbf{x}| + |B||\mathbf{x}| = (|A| + |B|)|\mathbf{x}|$$

Deler vi på  $|\mathbf{x}|$ , får vi

$$\frac{|(A + B)\mathbf{x}|}{|\mathbf{x}|} \leq |A| + |B|$$

for alle  $\mathbf{x} \neq \mathbf{0}$ . Fra definisjonen av operatornorm får vi dermed at  $|A + B| \leq |A| + |B|$ .  $\square$

Resultatet ovenfor gjelder selvfølgelig også for flere matriser enn to; vi har

$$|A_1 A_2 \dots A_n| \leq |A_1| |A_2| \dots |A_n|$$

og

$$|A_1 + A_2 + \dots + A_n| \leq |A_1| + |A_2| + \dots + |A_n|$$

Spesielt er

$$|A^n| \leq |A|^n$$

som vi snart skal få bruk for.

Vi er nå nesten ferdig med forberedelsene, alt som gjenstår er å definere grenseverdien til en følge  $\{A_n\}$  av matriser. Det er lett; vi sier rett og slett at  $\{A_n\}$  konvergerer mot  $B$  dersom hvert element i  $A_n$  konvergerer mot tilsvarende element i  $B$ .

**Lemma 5.3.8** *Dersom  $A$  er en  $m \times m$ -matrise med  $|A| < 1$ , så konvergerer den geometriske rekken*

$$I_m + A + A^2 + \dots + A^n + \dots$$

*mot en matrise  $B$  i den forstand at  $B = \lim_{n \rightarrow \infty} (I_m + A + A^2 + \dots + A^n)$ .*

*Bevis:* La  $a_{ij}^{(k)}$  være det  $ij$ -te elementet til  $A^k$  (vi skriver  $k$ -en i  $a_{ij}^{(k)}$  i parentes for å understreke at det ikke er snakk om en potens). Vi må vise at følgen av delsummer  $\{s_n\}$ , der  $s_n = \sum_{k=0}^n a_{ij}^{(k)}$ , konvergerer. Ifølge teorem 5.2.6 er det nok å vise at denne følgen er en Cauchy-følge. Dersom  $N > n$ , har vi ifølge de to lemmaene ovenfor

$$\begin{aligned}|s_N - s_n| &= \left| \sum_{k=n+1}^N a_{ij}^{(k)} \right| \leq \sum_{k=n+1}^N |a_{ij}^{(k)}| \leq \sum_{k=n+1}^N |A^k| \leq \sum_{k=n+1}^N |A|^k = \\ &= \frac{|A|^{n+1}(1 - |A|^{N-n})}{1 - |A|} \leq \frac{|A|^{n+1}}{1 - |A|}\end{aligned}$$

der vi har summert en geometrisk rekke (husk at  $|A|$  er et tall slik at det er en vanlig geometrisk rekke av tall vi summerer her). Siden  $|A| < 1$ , kan vi få uttrykket på høyre side så lite vi vil ved å velge  $n$  tilstrekkelig stor. Følgelig er  $\{s_n\}$  en Cauchy-følge og må konvergere.  $\square$

**Bemerkning:** Argumentet ovenfor er ganske typisk for hvordan man bruker Cauchy-følger i praksis. I dette tilfellet har vi ikke full kontroll over hvordan elementene  $s_n = \sum_{k=0}^n a_{ij}^{(k)}$  i følgen ser ut (du kan prøve å regne dem ut, men uttrykkene blir utrolig kompliserte), men vi har en ulikhet som forteller oss noe om størrelsen. Ved hjelp av denne ulikheten kan vi vise at følgen er en Cauchy-følge, og dermed er konvergensen etablert til tross for at vi fortsatt ikke vet noe særlig hverken om leddene i følgen eller grenseverdien!

Vi kan nå bevise hovedresultatet vårt:

**Teorem 5.3.9** *Anta at  $A$  er en  $m \times m$ -matrise og at  $|A| < 1$ . Da er  $I_m - A$  inverterbar og*

$$(I_m - A)^{-1} = I_m + A + A^2 + \dots + A^n + \dots$$

*Bevis:* La  $B_n = (I_m + A + A^2 + \dots + A^n)$  og  $B = \lim_{n \rightarrow \infty} B_n = I_m + A + A^2 + \dots + A^n + \dots$ . Multipliserer vi ut og forkorter, ser vi at

$$(I_m - A)B_n = (I_m - A)((I_m + A + A^2 + \dots + A^n)) = I_m - A^{n+1}$$

Lar vi  $n$  gå mot uendelig i denne ligningen, får vi

$$(I_m - A)B = \lim_{n \rightarrow \infty} (I_m - A)B_n = \lim_{n \rightarrow \infty} (I_m - A^{n+1}) = I_m$$

der vi har brukt at siden  $|A| < 1$ , går  $|A|^{n+1}$  mot null (tenk gjennom hva du egentlig bruker i denne overgangen!). Ifølge teorem 4.5.3 er da  $I_m - A$  inverterbar med invers  $B$ .  $\square$

Vi tar med en reformulering av dette resultatet som ofte er mer naturlig å bruke:

**Korollar 5.3.10** *Anta at  $C$  er en  $m \times m$ -matrise slik at  $|I_m - C| < 1$ . Da er  $C$  inverterbar og*

$$C^{-1} = I_m + (I_m - C) + (I_m - C)^2 + \dots + (I_m - C)^n + \dots$$

*Bevis:* Sett  $A = I_m - C$  og bruk teoremet ovenfor.  $\square$

Til slutt skal vi utvide resultatet vårt slik at det ikke bare gjelder for matriser som ligger nær identitetsmatrisen, men for matriser som ligger nær en hvilken som helst inverterbar matrise.

**Teorem 5.3.11 (Banachs lemma)** *Anta at  $B$  er en inverterbar  $m \times m$ -matrise og at  $A$  er en annen  $m \times m$ -matrise. Dersom*

$$|B - A| < |B^{-1}|^{-1}$$

*så er også  $A$  inverterbar og*

$$|A^{-1}| \leq \frac{|B^{-1}|}{1 - |B^{-1}| |B - A|}$$

*Bevis:* Siden  $B$  er inverterbar, kan vi skrive

$$A = B - (B - A) = B (I_m - B^{-1}(B - A))$$

Ifølge teorem 5.3.9 er matrisen  $I_m - B^{-1}(B - A)$  inverterbar dersom

$$|B^{-1}(B - A)| < 1$$

Siden  $|B^{-1}(B - A)| \leq |B^{-1}| |B - A|$ , følger det at  $I_m - B^{-1}(B - A)$  er inverterbar dersom  $|B^{-1}| |B - A| < 1$ , dvs. dersom

$$|B - A| < |B^{-1}|^{-1}$$

(legg merke til at dette er betingelsen i teoremet). I så fall er også

$$A = B (I_m - B^{-1}(B - A))$$

inverterbar med

$$A^{-1} = (I_m - B^{-1}(B - A))^{-1} B^{-1}$$

(husk setning 1.7.4(ii)). Det gjenstår å estimere normen til  $A^{-1}$ . Siden

$$|A^{-1}| \leq |(I_m - B^{-1}(B - A))^{-1}| |B^{-1}|$$

og (ifølge teorem 5.3.9)

$$|(I_m - B^{-1}(B - A))^{-1}| = |\sum_{k=0}^{\infty} (B^{-1}(B - A))^k| \leq$$

$$\leq \sum_{k=0}^{\infty} (|B^{-1}| |(B - A)|)^k \leq \frac{1}{1 - |B^{-1}| |B - A|}$$

før vi

$$|A^{-1}| \leq \frac{|B^{-1}|}{1 - |B^{-1}| |B - A|}$$

□

### Oppgaver til seksjon 5.3

**1.** Vis at funksjonen  $f : \mathbb{R} \rightarrow \mathbb{R}$  gitt ved  $f(x) = x^2$  er kontinuerlig, men ikke uniformt kontinuerlig.

**2.** Vis at funksjonen  $f : (0, 1) \rightarrow \mathbb{R}$  gitt ved  $f(x) = \frac{1}{x}$  er kontinuerlig, men ikke uniformt kontinuerlig.

**3.** Anta at  $A \subset \mathbb{R}^m$  og at  $f$  er en funksjon  $f : A \rightarrow \mathbb{R}$ . Vis at dersom det finnes en konstant  $K$  slik at  $|f(\mathbf{u}) - f(\mathbf{v})| \leq K|\mathbf{u} - \mathbf{v}|$  for alle  $\mathbf{u}, \mathbf{v} \in A$ , så er  $f$  uniformt kontinuerlig i  $A$ .

**4.** Anta at  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  er en kontinuerlig funksjon slik at  $\lim_{|\mathbf{x}| \rightarrow \infty} f(\mathbf{x}) = 0$ . Vis at  $f$  er uniformt kontinuerlig på  $\mathbb{R}^m$ .

**5.** Fullfør beviset for at de partiellderiverte  $\frac{\partial F}{\partial x_i}$  i teorem 5.3.3 er kontinuerlige.

**6.** La  $f(s) = \int_0^1 e^{sx^2} dx$ . Vis at  $f'(s) = \int_0^1 x^2 e^{sx^2} dx$ .

**7.** Funksjonen  $f : \mathbb{R} \rightarrow \mathbb{R}$  er definert ved  $f(t) = \int_{\frac{1}{2}}^1 \frac{\sin(tx)}{x} dx$ . Vis at for  $t \neq 0$  er  $f'(t) = \frac{\sin t - \sin(\frac{t}{2})}{t}$ . Vis også at  $f'(0) = \frac{1}{2}$ .

**8.** Funksjonen  $h : \mathbb{R}^2 \rightarrow \mathbb{R}$  er kontinuerlig og har begrensede partiellderiverte. Funksjonen  $f : \mathbb{R} \rightarrow \mathbb{R}$  er gitt ved

$$f(x) = \int_0^{g(x)} h(x, t) dt$$

der  $g : \mathbb{R} \rightarrow \mathbb{R}$  er en deriverbar funksjon. Vis at

$$f'(x) = \int_0^{g(x)} \frac{\partial h(x, t)}{\partial x} dt + h(x, g(x))g'(x)$$

**9.** Vis at for alle inverterbare  $m \times m$ -matriser  $A$  er  $|A^{-1}|^{-1} \leq |A|$ .

- 10.** Anta at  $A$  er en inverterbar  $m \times m$ -matrise og at  $\{A_n\}$  er en følge av  $m \times m$ -matriser som konvergerer mot  $A$ . Vis at  $A_n$  er inverterbar for alle tilstrekkelig store  $n$  og at  $\{A_n^{-1}\}$  konvergerer mot  $A^{-1}$ .

## 5.4 Iterasjon av funksjoner

Vi er nå klare til å ta fatt på det første hovedtemaet i dette kapitlet — iterasjon. Anta at vi har en funksjon  $\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ . Dersom vi velger et startpunkt  $\mathbf{x}_0 \in \mathbb{R}^m$ , kan vi bruke  $\mathbf{F}$  til å skaffe oss en følge  $\{\mathbf{x}_n\}$  på denne måten:

$$\mathbf{x}_1 = \mathbf{F}(\mathbf{x}_0), \mathbf{x}_2 = \mathbf{F}(\mathbf{x}_1), \mathbf{x}_3 = \mathbf{F}(\mathbf{x}_2), \dots, \mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n), \dots$$

Hvert ledd i følgen fremkommer altså ved at vi bruker  $\mathbf{F}$  på det foregående leddet. Vi sier at følgen  $\{\mathbf{x}_n\}$  oppstår ved *iterasjon* av  $\mathbf{F}$ .

I denne seksjonen skal vi se først og fremst bli kjent med iterasjon som fenomen. Vi skal se på en del praktiske situasjoner som kan beskrives ved hjelp av iterasjoner, og vi skal bruke MATLAB til å illustrere de forskjellige typene oppførsel som finnes hos itererte systemer. I neste seksjon skal vi se nærmere på teorien bak den aller viktigste formen for oppførsel — konvergens mot et fikspunkt. Før vi begynner, kan det være greit å observere at vi har vært borti iterasjoner tidligere: i seksjon 4.11 brukte vi egenvektorer og egenverdier til å studere iterasjon av lineære funksjoner. Det er langt verre å få oversikt over oppførselen til de ikke-lineære iterasjonene vi nå skal konsentrere oss om.

Når man tenker på iterasjoner, er det ofte lurt å forestille seg at punktene i følgen representerer tilstander ved forskjellige tidspunkt;  $\mathbf{x}_0$  er tilstanden ved tiden 0,  $\mathbf{x}_1$  er tilstanden ved tiden 1, osv. Funksjonen  $\mathbf{F}$  blir da en mekanisme som oppdaterer tilstanden fra et tidspunkt til det neste (i kapittel 1 tenkte vi på funksjoner gitt av matriser på denne måten). La oss se på et lite eksempel som kanskje gjør tankegangen lettere å forstå.

**Eksempel 1:** To dyreslag, et byttedyr og et rovdyr, lever i det samme området. Dersom det ett år er  $x_n$  byttedyr og  $y_n$  rovdyr i området, tenker man seg at antall dyr året etter er gitt ved

$$\begin{aligned} x_{n+1} &= ax_n - bx_n y_n \\ y_{n+1} &= cy_n + dx_n y_n \end{aligned}$$

der  $a, b, c, d$  er positive tall. Legg merke til logikken; “kryssleddene”  $x_n y_n$  representerer møter mellom byttedyr og rovdyr, og slike møter reduserer veksten av byttedyr, men bidrar til vekst i rovdyrbestanden. Hvis vi innfører funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  ved

$$\mathbf{F} \left( \begin{array}{c} x \\ y \end{array} \right) = \left( \begin{array}{c} ax - bxy \\ cy + dxy \end{array} \right)$$

og lar  $\mathbf{x}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$ , ser vi at systemet ovenfor kan skrives som  $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$ .

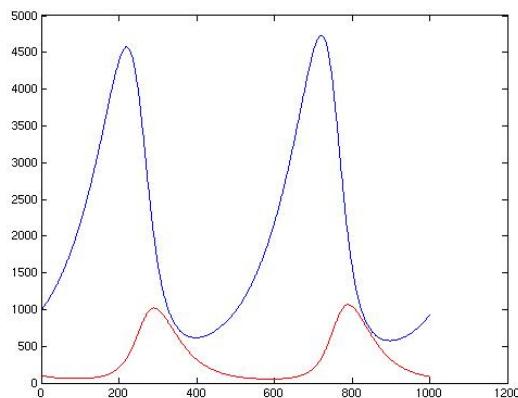
Det er ikke lett å gjette hvordan et system av denne typen vil utvikle seg i tiden, så la oss bruke MATLAB som hjelpemiddel til å se på et spesielt tilfelle. Vi velger  $a = 1.01, b = 3 \cdot 10^{-5}, c = 0.98, d = 10^{-5}$ . Følgende m-fil regner ut utviklingen når vi starter med  $m$  byttedyr og  $k$  rovdyr og gjennomfører  $N$  iterasjoner. Vær oppmerksom på at det er en liten forskjyvning i nummereringen av leddene i følgen; i teoretisk arbeid får vi ofte penest uttrykk om vi begynner iterasjonen med punkt nummer 0 (altså  $\mathbf{x}_0$  som ovenfor), men i programmet nedenfor har vi tatt hensyn til MATLABs forkjærighet for å la startpunktet være nummer 1 (og ikke nummer 0).

```
function [x,y]=byttedyr(m,k,N)
x=[m]; %med disse linjene forteller vi
y=[k]; %MATLAB at iterasjonen starter i punktet (m,k)
for n=1:N % starter løkken som utfører iterasjonene
    x(n+1)=1.01*x(n)-3*10^(-5)*x(n)*y(n);
    y(n+1)=0.98*y(n)+10^(-5)*x(n)*y(n);
end %avslutter for-løkken
```

Dersom vi ønsker å se grafisk på utviklingen når vi starter med 1000 byttedyr og hundre rovdyr, kan vi gi kommandoene

```
>> [x,y]=byttedyr(1000,100,1000);
>> plot(x)
>> hold on
>> plot(y,'r')
```

Vi får dette resultatet:



Figur 1: Utviklingen av byttedyr (øverst) og rovdyr (nederst).

Vi ser at bestandene følger et bølgemønster med klare topper og bunner. Logikken er ikke så vanskelig å forstå; til å begynne med er det relativt få rovdyr, og byttedyrbestanden vokser. Dette fører til gode betingelser for rovdyrbestanden som også begynner å vokse kraftig. Til slutt gjør rovdyrene så kraftig innhogg at byttedyrbestanden begynner å avta. Etter hvert fører dette til dårligere forhold for rovdyrene, og rovdyrbestanden begynner også å avta. Dette gir etter hvert bedre forhold for byttedyrene som begynner å ta seg opp igjen osv.



**Bemerkning:** La oss smette inn en liten bemerkning om effektivitet i MATLAB-beregninger. Programmet ovenfor er kort, og det fungerer utmerket for de små datamengdene vi har i dette eksemplet. Det er imidlertid lite effektivt fordi MATLAB hele tiden må endre dimensjonen på vektorene  $x$  og  $y$ . Det viser seg at MATLAB arbeider mye mer effektivt dersom vi gir disse vektorene den “riktige” dimensjonen helt fra starten av. Det kan vi gjøre ved å endre programmet til

```
function [x,y]=byttedyr(m,k,N)
x=zeros(1,N);
y=zeros(1,N);
x(1)=m;
y(1)=k;
for n=1:N-1
    x(n+1)=1.01*x(n)-3*10^(-5)*x(n)*y(n);
    y(n+1)=0.98*y(n)+10^(-5)*x(n)*y(n);
end
```

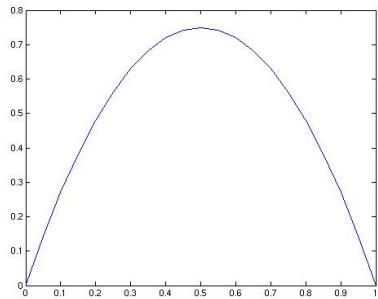
Kommandoen `zeros(1,N)` gir oss et  $N$ -tuppel (radvektor) med bare 0'er, og sørger derfor for at vektorene våre har den riktige dimensjonen helt fra starten av.

Bølgemønsteret ovenfor er bare ett av mange man kan støte på når man itererer en funksjon. For å bli litt kjent med de forskjellige mulighetene skal vi bruke MATLAB til å gjøre noen eksperimenter. Selv om vi i dette heftet hovedsakelig er interessert i funksjoner av flere variable, skal vi gjennomføre disse eksperimentene for en funksjon  $f : \mathbb{R} \rightarrow \mathbb{R}$  av én variabel for å få oversiktlig figurer.

Vi skal stort sett arbeide med funksjoner  $f : [0, 1] \rightarrow [0, 1]$  gitt ved

$$f(x) = bx(1 - x)$$

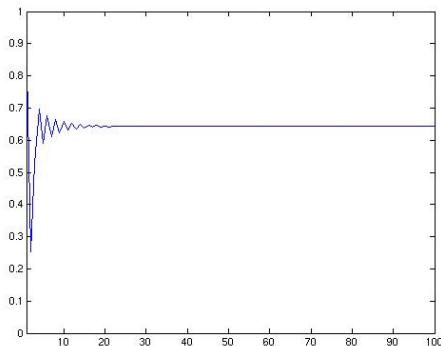
der  $b$  er en konstant mellom 0 og 4 (når  $b$  ligger utenfor dette intervallet, vil ikke  $f$  avbilde intervallet  $[0, 1]$  inn i  $[0, 1]$ ). Funksjoner av denne typen ser ut som grafen på figur 2 (større  $b$ -verdier gir høyere topp).

Figur 2: En graf av typen  $f(x) = bx(1 - x)$ 

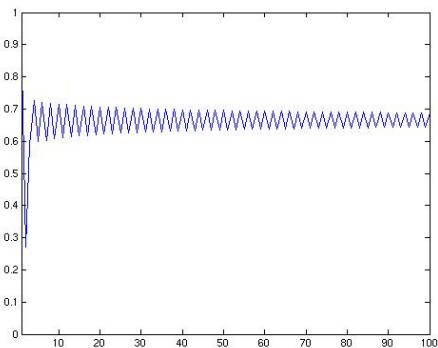
For å gjennomføre iterasjonen, lager vi et lite program `iterasjon.m` (uten å bry oss for mye om beregningseffektivitet):

```
function x=iterasjon(a,b,N)
x=[a];
for n=1:N
    x(n+1)=b*x(n)*(1-x(n));
end
```

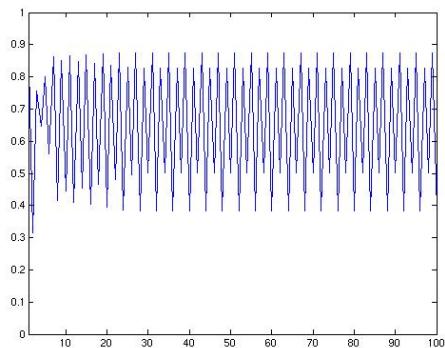
Input-parametrene  $a$ ,  $b$  og  $N$  angir henholdsvis startverdien, parameteren  $b$  i ligningen ovenfor og antall iterasjonen. Gir vi MATLAB kommandoene `>>iterasjon(0.9,2.8,100)` etterfulgt av `>>plot(x)`, får vi figuren nedenfor:

Figur 3:  $b = 2.8$ .

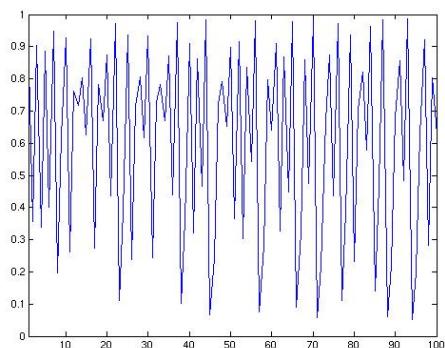
Vi ser at etter noen innledende fluktusjoner, slår følgen seg til ro og nærmer seg en grenseverdi  $x \approx 0.64$ . Dette punktet  $x$  er et likevektspunkt (eller *fikspunkt* som matematikere liker å si) i den forstand at  $f(x) = x$ . La oss nå endre  $b$ -verdien og prøve med  $b = 3.0$ . Vi får da utviklingen i figur 4. I dette tilfellet nærmer vi oss også en slags grenseverdi, men dette er ikke et fast punkt, men en svingbevegelse mellom to faste punkter — vi nærmer oss en “stabil bane med periode 2”.

Figur 4:  $b = 3.0$ .

Går vi et skritt videre og velger  $b = 3.5$ , får vi utviklingen i figur 5. Denne gangen nærmer vi oss en “stabil bane med periode 4”, altså en svingning mellom fire faste punkter.

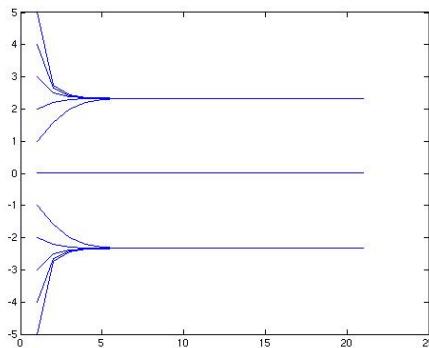
Figur 5:  $b = 3.5$ .

Til slutt setter vi  $b = 3.95$  og får denne figuren:

Figur 6:  $b = 3.95$ .

Her er det tilsynelatende ingen orden i det hele tatt, bare usystematiske svingninger opp og ned.

Eksemplene ovenfor viser noen av de fenomenene vi kan støte på når vi itererer en funksjon. Vær oppmerksom på at oppførselen også kan avhenge av startverdien  $x_0$  (altså parameteren  $a$  i programmet): I figur 6 har vi iterert funksjonen  $g(x) = 2 \arctan x$  med forskjellige startverdier. Vi ser at dersom vi har en positiv startverdi, konvergerer følgen mot et positivt fikspunkt  $y \approx 2.3311$ , men dersom vi starter i et negativt punkt, konvergerer den mot et negativt fikspunkt  $z \approx -2.3311$ . Det finnes også et tredje fikspunkt  $u = 0$ , men det er *frastøtende* i den forstand at følgen bare konvergerer mot det dersom  $x_0$  (og dermed alle  $x_n$ ) er lik 0. På tilsvarende måte kan lengden til en periodisk bane avhenge av startpunktet; ett startpunkt kan f.eks. lede til baner med periode 3, mens et annet startpunkt leder til baner med periode 14 (for samme  $b$ -verdi).



Figur 6: Iterasjon av  $g(x) = 2 \arctan x$  for forskjellige startverdier.

Det viser seg at itererte følger kan ha enda mer komplisert oppførsel enn det vi har sett eksempler på her, spesielt i høyeredimensjonale systemer der vi itererer en funksjon  $\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^m$  for  $m > 1$ . Dette får vi komme tilbake til en annen gang; hensikten med denne seksjonen er bare å gi deg en viss følelse for de fenomenene som kan oppstå ved iterasjon.

### Oppgaver til seksjon 5.4

- En følge  $\{x_n, y_n\}$  i  $\mathbb{R}^2$  er gitt ved  $x_1 = y_1 = 0$  og

$$\begin{aligned} x_{n+1} &= 0.6x_n - 0.6y_n + 0.2 \\ y_{n+1} &= 0.6x_n + 0.6y_n + 1 \end{aligned}$$

Skriv et MATLAB-program som beregner elementer i følgen, og tegn dem inn i en figur. Vis at funksjonen

$$\mathbf{F}(x, y) = \begin{pmatrix} 0.6x - 0.6y + 0, 2 \\ 0.6x + 0.6y + 1 \end{pmatrix}$$

har et fikspunkt.

- 2.** To dyreslag lever i det samme området. Dersom det er hhv.  $x_n$  og  $y_n$  dyr av hvert slag i området ett år, regner man at tallene året etter er gitt ved

$$\begin{aligned}x_{n+1} &= 0.9x_n + 0.01y_n - 10 \\y_{n+1} &= -1.01x_n + y_n + 300\end{aligned}$$

Skriv et MATLAB-program som beregner hvordan bestanden utvikler seg, og plott resultatet når  $x_1 = 20$ ,  $y_1 = 2000$ . Finn et likevektpunkt for bestandene.

- 3.** I denne oppgaven skal vi se på befolkningsutviklingen i to land. Dersom innbyggertallet (målt i millioner) i de to landene er hhv.  $x_n$  og  $y_n$  ett tiår, regner man at de tilsvarende tallene ti år etter er gitt ved

$$\begin{aligned}x_{n+1} &= 1.1x_n + 0.001y_n - 0.5 \\y_{n+1} &= 0.95y_n + 0.002x_n + 0.2\end{aligned}$$

Skriv et MATLAB-program som beregner hvordan innbyggertallene utvikler seg, og plott resultatet når  $x_1 = 50$ ,  $y_1 = 8$ .

- 4.** To bensinstasjoner  $X$  og  $Y$  konkurrerer ved å tilpasse seg hverandres priser. Dersom prisene en uke er hhv.  $x_n$  og  $y_n$ , vil stasjon  $X$  uken etter sette sin pris til  $1.01 \cdot \frac{x_n+y_n}{2}$ , mens stasjon  $Y$  vil velge den prisen som er lavest av  $x_n$  og  $1.1y_n$ . Lag et MATLAB-program som viser hvordan prisene vil utvikle seg. Kjør programmet både med  $x_1 = 8$ ,  $y_1 = 12$  og med  $x_1 = 12$ ,  $y_1 = 8$ . Sammenlign prisutviklingene.

- 5.** To insektstyper konkurrerer om det samme området. Anta at  $x_n$  og  $y_n$  er antall insekter (målt i millioner) i området i år  $n$ . Vi regner at bestanden i år  $n+1$  da er gitt ved

$$\begin{aligned}x_{n+1} &= 2.2x_n(1-x_n) + 0.01x_ny_n \\y_{n+1} &= 3.1y_n(1-y_n) - 0.02x_ny_n\end{aligned}$$

Skriv et MATLAB-program som beregner hvordan bestandene utvikler seg, og plott resultatet når  $x_1 = 0.5$ ,  $y_1 = 0.5$  og når  $x_1 = 0.1$ ,  $y_1 = 0.8$ . Eksperimenter også med andre startverdier i intervallet  $(0, 1)$ .

- 6.** To firmaer konkurrerer i det samme markedet. Dersom prisene på produktet deres (målt i tusen kroner) er hhv.  $p$  og  $q$ , regner firmaene med å selge hhv.

$$E_1(p, q) = 1000e^{-\frac{p}{q} - \alpha(p+q)}$$

og

$$E_2(p, q) = 1000e^{-\frac{q}{p} - \beta(p+q)}$$

eksemplarer, der  $\alpha$  og  $\beta$  er konstanter.

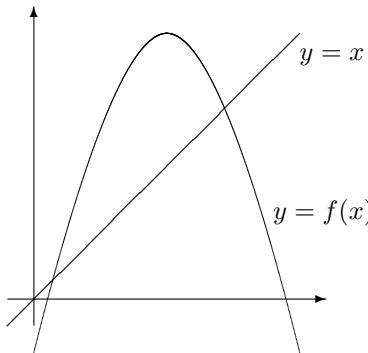
- a) Anta at prisen  $q$  ligger fast. Vis at firma 1 da får størst salgsinntekter ved å selge sitt produkt for  $p^* = \frac{q}{1+\alpha q}$ . Vis tilsvarende at hvis  $p$  ligger fast, så får firma 2 størst salgsinntekter ved å selge sitt produkt for  $q^* = \frac{p}{1+\beta p}$ .

- b) Det første året velger firmaene prisene  $p_1$  og  $q_1$ . De bestemmer seg for at prisen året etter skal være hhv.  $p_2 = 1.1p_1^* = 1.1\frac{q_1}{1+\alpha q_1}$  og  $q_2 = 1.1q_1^* = 1.1\frac{p_1}{1+\beta p_1}$ . Denne politikken holder de fast ved i årene som kommer. Skriv et MATLAB-program som beregner prisutviklingen for de to produktene. Parametrene  $\alpha$  og  $\beta$  skal inngå blant input-variablene.
- c) Kjør programmet med  $\alpha = \beta = 0.05$ . Bruk startverdiene  $(x_1, y_1) = (3, 4)$ ,  $(x_1, y_1) = (4, 3)$ ,  $(x_1, y_1) = (1, 1.3)$ ,  $(x_1, y_1) = (1.3, 1)$  og sammenlign resultatene.
- d) Gjenta punkt c) med  $\alpha = 0.05, \beta = 0.02$ .
- 7.** La  $f : \mathbb{R} \rightarrow \mathbb{R}$  være funksjonen  $f(x) = x^2 + x - 2$ . Vis at  $\bar{x} = \sqrt{2}$  er et flikspunkt for  $f$ . Skriv et MATLAB-program som utfører iterasjonen  $x_{n+1} = f(x_n)$ . Start programmet med  $x_1 = \sqrt{2}$ . Hva skjer (du bør nok utføre ca. 30 iterasjoner før du ser noe)? Forklar!
- 8.** Skriv et MATLAB-program for iterasjon av funksjoner  $f : [0, \pi] \rightarrow \mathbb{R}$  gitt ved  $f(x) = b \sin x$ , der  $b$  er en konstant. Eksperimenter med forskjellige startverdier og forskjellige  $b$ -verdier slik vi gjorde med funksjonen  $f(x) = bx(1-x)$  i teksten.

**9.** I denne oppgaven er  $A = \begin{pmatrix} 1.3 & -0.2 \\ 0.1 & 1 \end{pmatrix}$ .

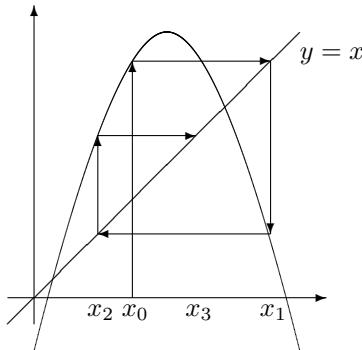
- a) Finn egenverdiene og egenvektorene til  $A$ .
- b) Skriv vektorene  $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$  og  $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$  som lineærkombinasjoner av basisvektorene du fant i a).
- c) Funksjonen  $f : \mathbb{R} \rightarrow \mathbb{R}$  er gitt ved  $f(x) = \lambda x + k$  der  $\lambda$  og  $k$  er konstanter,  $\lambda \neq 1$ . Vis at når vi itererer  $f$  med startpunkt  $x_0$ , så er  $x_n = \lambda^n \left( x_0 - \frac{k}{1-\lambda} \right) + \frac{k}{1-\lambda}$
- d) Anta at  $\mathbf{r}_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$  og at følgen  $\{\mathbf{r}_n\}$  fremkommer ved iterasjonen  $\mathbf{r}_{n+1} = A\mathbf{r}_n + \mathbf{b}$  der  $\mathbf{b} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ . Finn  $\mathbf{r}_n$ .

- 10.** I denne oppgaven skal vi se på en grafisk metode for å studere iterasjon av en kontinuerlig funksjon  $f$  av én variabel. I figur 1 har vi tegnet opp funksjonsgrafen og linjen  $y = x$  i samme koordinatsystem.



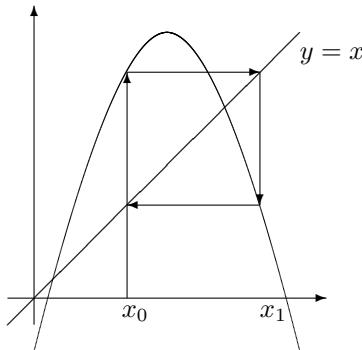
Figur 1

- a) Forklar at fikspunktene til  $f$  er det samme som skjæringspunktene mellom linjen  $y = x$  og grafen til  $f$ .



Figur 2

- b) Figur 2 viser hvordan vi grafisk kan finne punktene  $x_1 = f(x_0)$ ,  $x_2 = f(x_1)$ ,  $x_3 = f(x_2)$  osv. Forklar hvordan og hvorfor metoden virker.  
 c) Figur 3 viser en bane med periode 2. Lag en tilsvarende figur som viser en bane med periode 3 (du kan godt bruke en annen funksjonsgraf  $y = f(x)$ ).



Figur 3

- d) I resten av oppgaven skal vi se på funksjoner  $f : [0, 1] \rightarrow [0, 1]$  gitt ved  $f(x) = bx(1-x)$ , der  $0 \leq b \leq 4$ . Vis at dersom  $b > 1$ , så har  $f$  to fikspunkter  $0$  og  $1 - \frac{1}{b}$ .  
 e) Velg  $b = \frac{7}{2}$ . Vis at banen som starter i  $x_0 = \frac{3}{7}$  er periodisk med periode 2.  
 f) Bruk MATLAB eller en lommeregner til å tegne grafen til  $f^{\circ 2}$ . Vis at grafen skjærer linjen  $y = x$  i punktene  $0$ ,  $\frac{3}{7}$ ,  $\frac{5}{7}$  og  $\frac{6}{7}$ . Forklar.  
 g) (arbeidskrevende) Vis at dersom  $3 < b \leq 4$ , så har funksjonen  $f(x) = bx(1-x)$  baner med periode to som starter i punktene  $\frac{b+1 \pm \sqrt{(b-1)^2-4}}{2b}$ .

## 5.5 Konvergens mot et fikspunkt

I forrige seksjon så vi på noen av de typene oppførsel vi kan få når vi itererer en funksjon  $\mathbf{F}$ . En av de enkleste og viktigste er at følgen konvergerer mot en likevektstilstand  $\mathbf{x}$ , det vil si en tilstand slik at  $\mathbf{F}(\mathbf{x}) = \mathbf{x}$  (se figur 3 og

6 i forrige seksjon). Matematisk kaller vi en slik tilstand for et *fikspunkt* for  $\mathbf{F}$ :

**Definisjon 5.5.1** *Anta at  $A$  er en delmengde av  $\mathbb{R}^m$  og at  $\mathbf{F}$  er en funksjon fra  $A$  til  $\mathbb{R}^m$ . Vi sier at  $\mathbf{x} \in A$  er et fikspunkt for  $\mathbf{F}$  dersom  $\mathbf{F}(\mathbf{x}) = \mathbf{x}$ .*

Det er en del naturlige spørsmål knyttet til iterasjon og fikspunkter: Når har en funksjon  $\mathbf{F}$  et fikspunkt? Dersom  $\mathbf{F}$  har et fikspunkt  $\mathbf{x}$ , når vil en følge dannet ved iterasjon av  $\mathbf{F}$  konvergere mot  $\mathbf{x}$ ? Dersom en funksjon har flere fikspunkter, hvilket av disse vil en iterert følge konvergere mot? Hvorfor vil dette avhenge av hvilket punkt  $\mathbf{x}_0$  vi startet iterasjonen i, altså av begynnelsestilstanden til systemet? Dette er vanskelige spørsmål som vi ikke kan gi utfyllende svar på her, men vi skal i hvert fall bevise et resultat som er nyttig i mange sammenhenger. Dette resultatet gjelder for *kontraksjoner* som er en spesielt enkel type funksjoner å ha med å gjøre. I seksjon 5.7 skal vi se hvordan vi kan bruke fikspunktiterasjon til å få et grep på omvendte funksjoner og implisitt gitte funksjoner.

**Definisjon 5.5.2** *Anta at  $A$  er en ikke-tom delmengde av  $\mathbb{R}^m$ . En funksjon  $\mathbf{F} : A \rightarrow A$  kalles en kontraksjon av mengden  $A$  dersom det finnes et positivt tall  $C < 1$  slik at*

$$|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \leq C|\mathbf{x} - \mathbf{y}|$$

*for alle  $\mathbf{x}, \mathbf{y} \in A$ . Vi kaller  $C$  en kontraksjonsfaktor for  $F$ .*

Siden  $C$  er mindre enn 1, ser vi at kontraksjoner reduserer avstanden mellom punkter — avstanden mellom  $\mathbf{F}(\mathbf{x})$  og  $\mathbf{F}(\mathbf{y})$  er mindre enn avstanden mellom  $\mathbf{x}$  og  $\mathbf{y}$ . Legg også merke til at vi krever at alle verdiene til  $\mathbf{F}$  skal ligge i  $A$ ; funksjonen  $\mathbf{F}$  skal altså avbilde mengden  $A$  inn i  $A$  selv. Når man skal vise at en funksjon  $\mathbf{F}$  er en kontraksjon av en mengde  $A$ , er dette ofte det vanskeligste punktet å sjekke.

Vi trenger litt notasjon før vi går videre. Dersom følgen  $\{\mathbf{x}_n\}$  oppstår når vi itererer  $\mathbf{F}$  med  $\mathbf{x}_0$  som startpunkt, ser vi at

$$\mathbf{x}_1 = \mathbf{F}(\mathbf{x}_0), \mathbf{x}_2 = \mathbf{F}(\mathbf{x}_1) = \mathbf{F}(\mathbf{F}(\mathbf{x}_0)), \mathbf{x}_3 = \mathbf{F}(\mathbf{x}_2) = \mathbf{F}(\mathbf{F}(\mathbf{F}(\mathbf{x}_0))), \dots$$

Generelt er  $\mathbf{x}_n = \mathbf{F}(\mathbf{F}(\dots(\mathbf{F}(\mathbf{x}_0))\dots))$  der vi har  $n$   $\mathbf{F}$ 'er etter hverandre. For å få en mer oversiktlig notasjon, skriver vi  $\mathbf{x}_n = \mathbf{F}^{\circ n}(\mathbf{x}_0)$ ; med andre ord er  $\mathbf{F}^{\circ n}$  den funksjonen vi får når vi setter  $\mathbf{F}$  sammen med seg selv  $n$  ganger.

**Lemma 5.5.3** *Anta at  $\mathbf{F} : A \rightarrow A$  er en kontraksjon med kontraksjonsfaktor  $C$ . For alle  $\mathbf{x}, \mathbf{y} \in A$  og alle  $n \in \mathbb{N}$  er da*

$$|\mathbf{F}^{\circ n}(\mathbf{x}) - \mathbf{F}^{\circ n}(\mathbf{y})| \leq C^n |\mathbf{x} - \mathbf{y}|$$

$\mathbf{F}^{\circ n}$  er altså en kontraksjon med kontraksjonsfaktor  $C^n$ .

*Bevis:* For  $n = 1$  er dette bare definisjonen av kontraksjon. For  $n = 2$  har vi:

$$|\mathbf{F}^{\circ 2}(\mathbf{x}) - \mathbf{F}^{\circ 2}(\mathbf{y})| = |\mathbf{F}(\mathbf{F}(\mathbf{x})) - \mathbf{F}(\mathbf{F}(\mathbf{y}))| \leq C|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \leq C^2|\mathbf{x} - \mathbf{y}|$$

Ved å fortsette på samme måte, får vi resultatet for alle  $n \in \mathbb{N}$  (du kan føre et formelt induksjonsbevis hvis du vil).  $\square$

Legg spesielt merke til at dersom  $\mathbf{x}_k$  og  $\mathbf{x}_{k+1}$  er to ledd som følger etter hverandre i følgen, så er

$$|\mathbf{x}_k - \mathbf{x}_{k+1}| = |\mathbf{F}^{\circ k}(\mathbf{x}_0) - \mathbf{F}^{\circ k}(\mathbf{x}_1)| \leq C^k|\mathbf{x}_0 - \mathbf{x}_1| \quad (5.5.1)$$

Vi er nå klare til å vise hovedresultatet i denne seksjonen. Det har navn etter den polske matematikeren Stefan Banach (1892-1945).

**Teorem 5.5.4 (Banachs fikspunktsteoremet)** *Anta at  $A$  er en ikke-tom, lukket delmengde av  $\mathbb{R}^m$  og at  $\mathbf{F} : A \rightarrow A$  er en kontraksjon av  $A$  med kontraksjonsfaktor  $C$ . Da har  $\mathbf{F}$  nøyaktig ett fikspunkt  $\mathbf{x}$  i  $A$ . Uansett hvilket punkt  $\mathbf{x}_0$  i  $A$  vi starter iterasjonen i, vil følgen  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$  der  $\mathbf{x}_n = \mathbf{F}^{\circ n}(\mathbf{x}_0)$  konvergere mot  $\mathbf{x}$ , og for alle  $n \in \mathbb{N}$  er*

$$|\mathbf{x}_n - \mathbf{x}| \leq \frac{C^n}{1-C}|\mathbf{x}_0 - \mathbf{x}_1|$$

*Bevis:* La oss først vise at  $\mathbf{F}$  kan ha høyst ett fikspunkt. Dersom både  $\mathbf{x}$  og  $\mathbf{y}$  er fikspunkter, har vi nemlig

$$|\mathbf{x} - \mathbf{y}| = |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| \leq C|\mathbf{x} - \mathbf{y}|$$

Det betyr at  $|\mathbf{x} - \mathbf{y}| \leq C|\mathbf{x} - \mathbf{y}|$ , og siden  $C < 1$  er det bare mulig dersom  $|\mathbf{x} - \mathbf{y}| = 0$ , dvs. dersom  $\mathbf{x} = \mathbf{y}$ .

Neste skritt er å vise at følgen  $\{\mathbf{x}_n\}$  er en Cauchy-følge uansett hvilket punkt  $\mathbf{x}_0 \in A$  vi begynner iterasjonen i. Dersom vi har to ledd i følgen  $\mathbf{x}_n$  og  $\mathbf{x}_k$  med  $n < k$ , har vi

$$\begin{aligned} |\mathbf{x}_n - \mathbf{x}_k| &= |(\mathbf{x}_n - \mathbf{x}_{n+1}) + (\mathbf{x}_{n+1} - \mathbf{x}_{n+2}) + \dots + (\mathbf{x}_{k-2} - \mathbf{x}_{k-1}) + (\mathbf{x}_{k-1} - \mathbf{x}_k)| \leq \\ &\leq |\mathbf{x}_n - \mathbf{x}_{n+1}| + |\mathbf{x}_{n+1} - \mathbf{x}_{n+2}| + \dots + |\mathbf{x}_{k-2} - \mathbf{x}_{k-1}| + |\mathbf{x}_{k-1} - \mathbf{x}_k| \leq \\ &\leq C^n|\mathbf{x}_0 - \mathbf{x}_1| + C^{n+1}|\mathbf{x}_0 - \mathbf{x}_1| + \dots + C^{k-2}|\mathbf{x}_0 - \mathbf{x}_1| + C^{k-1}|\mathbf{x}_0 - \mathbf{x}_1| \leq \\ &\leq \left( C^n + C^{n+1} + \dots + C^{k-2} + C^{k-1} + \dots \right) |\mathbf{x}_0 - \mathbf{x}_1| = \frac{C^n}{1-C}|\mathbf{x}_0 - \mathbf{x}_1| \end{aligned}$$

der vi først har brukt trekantulikheten, så ulikheten i formel (5.4.1) ovenfor og til slutt summeformelen for en geometrisk rekke. Siden  $C < 1$ , kan vi få uttrykket  $\frac{C^n}{1-C}|\mathbf{x}_0 - \mathbf{x}_1|$  så lite vi måtte ønske oss ved å velge  $n$  tilstrekkelig stor. Dette betyr at  $\{\mathbf{x}_n\}$  er en Cauchy-følge, og ifølge teorem 5.2.6

konvergerer den mot en grense  $\mathbf{x}$  (siden  $A$  er lukket, ligger  $\mathbf{x}$  i  $A$  — husk setning 5.1.6). Det er lett å vise at grensepunktet  $\mathbf{x}$  er et fikspunkt for  $\mathbf{F}$ ; siden enhver kontraksjon er kontinuerlig, har vi nemlig ifølge setning 5.1.7:

$$\mathbf{x} = \lim_{n \rightarrow \infty} \mathbf{x}_{n+1} = \lim_{n \rightarrow \infty} \mathbf{F}(\mathbf{x}_n) = \mathbf{F}\left(\lim_{n \rightarrow \infty} \mathbf{x}_n\right) = \mathbf{F}(\mathbf{x})$$

Siden vi allerede vet at  $\mathbf{F}$  ikke kan ha mer enn ett fikspunkt, betyr dette at følgen  $\{\mathbf{x}_n\}$  konvegerer mot det samme punktet  $\mathbf{x}$  uansett hvilket startpunkt  $\mathbf{x}_0$  man bruker.

Det gjenstår å vise den siste formelen i teoremet. Ifølge utledningen ovenfor, er

$$|\mathbf{x}_n - \mathbf{x}_k| \leq \frac{C^n}{1-C} |\mathbf{x}_0 - \mathbf{x}_1|$$

når  $k > n$ . Lar vi  $k \rightarrow \infty$ , får vi

$$|\mathbf{x}_n - \mathbf{x}| \leq \frac{C^n}{1-C} |\mathbf{x}_0 - \mathbf{x}_1|$$

og beviset er fullført (tenk gjennom det siste resonnementet!)  $\square$

Legg merke til at beviset ovenfor illustrerer filosofien vår fra seksjon 5.2 — det er mye lettere å vise at en følge konvergerer ved å sjekke at den er en Cauchy-følge enn ved å konstruere grenselementet!

**Bemerkning:** Situasjonen i teoremet ovenfor er den best tenkelige — funksjonen  $\mathbf{F}$  har nøyaktig ett fikspunkt  $\mathbf{x}$ , og itererer vi  $\mathbf{F}$ , vil følgen konvergere mot  $\mathbf{x}$  uansett hvilket punkt  $\mathbf{x}_0$  vi starter i. Det er ett poeng til som er viktig, og det er at vi har kontroll på *hvor fort* følgen  $\{\mathbf{x}_n\}$  konvergerer mot  $\mathbf{x}$ . I praksis er det nemlig sjeldent vi kan finne et fikspunkt helt nøyaktig, alt vi kan gjøre er å iterere funksjonen så mange ganger at  $\mathbf{x}_n$  kommer så nær fikspunktet som vi trenger. Ulikheten

$$|\mathbf{x}_n - \mathbf{x}| \leq \frac{C^n}{1-C} |\mathbf{x}_0 - \mathbf{x}_1|$$

gir oss kontroll over dette — ønsker vi at  $\mathbf{x}_n$  skal gi oss fikspunktet med en nøyaktighet bedre enn  $\epsilon$ , må vi velge  $n$  så stor at

$$\frac{C^n}{1-C} |\mathbf{x}_0 - \mathbf{x}_1| < \epsilon$$

Det betyr at med en gang vi har regnet ut  $\mathbf{x}_1$ , har vi den informasjonen vi trenger for å beregne konvergeringsraten.

Vi skal nå se på et eksempel som viser at vi ikke kan fjerne betingelsen om at  $\mathbf{F}$  er en kontraksjon fra teoremet ovenfor.

**Eksempel 1:** La  $A = \{\mathbf{x} \in \mathbb{R}^2 : 1 \leq |\mathbf{x}| \leq 2\}$  være området mellom to sirkler i planet, og la  $\mathbf{F}$  være avbildningen som dreier hele området  $A$  en vinkel  $\theta < 2\pi$  om origo. Da er  $|\mathbf{x} - \mathbf{y}| = |\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})|$  (siden alle punkter dreies samme vinkel), men det finnes ingen fikspunkter siden alle punkter er rotert i forhold til utgangspunktet.

Legg merke til at dersom vi erstatter  $A$  med området  $A' = \{\mathbf{x} \in \mathbb{R}^2 : |\mathbf{x}| \leq 2\}$ , der ”hullet” i midten er fjernet, så har  $\mathbf{F}$  et fikspunkt, nemlig  $\mathbf{0}$ . Dette fikspunktet kan vi imidlertid ikke nå frem til ved iterasjon; enhver iterasjon av  $\mathbf{F}$  sender punkter i sirkelbaner rundt origo. ♣

For å bruke Banachs fikspunktteorem trenger vi en metode til å vise at  $\mathbf{F}$  er en kontraktsjon. Vi skal arbeide oss frem mot et kriterium som ofte er nyttig, og minner først om en hjelpesetning vi allerede har bevist (se setning 3.2.3):

**Setning 5.5.5 (Middelverdisetning for funksjoner av flere variable)**  
*Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $m$  variable, og at  $f$  er deriverbar i et område som inneholder linjestykket mellom punktene  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ . Da finnes det et punkt  $\mathbf{c}$  på linjestykket fra  $\mathbf{a}$  til  $\mathbf{b}$  slik at*

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \cdot (\mathbf{b} - \mathbf{a})$$

Den neste setningen bringer oss enda et skritt nærmere resultatet vårt:

**Setning 5.5.6** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en funksjon av  $m$  variable, og at  $\mathbf{F}$  er deriverbar i et område som inneholder linjestykket mellom punktene  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ . Da finnes det punkter  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m$  på linjestykket fra  $\mathbf{a}$  til  $\mathbf{b}$ , slik at*

$$|\mathbf{F}(\mathbf{b}) - \mathbf{F}(\mathbf{a})| \leq |\mathbf{b} - \mathbf{a}| \sqrt{\nabla F_1(\mathbf{c}_1)^2 + \dots + \nabla F_m(\mathbf{c}_m)^2}$$

der  $F_1, F_2, \dots, F_m$  er komponentene til  $\mathbf{F}$ .

*Bevis:* Bruker vi forrige setning på den  $i$ -te komponenten  $F_i$ , får vi et punkt  $\mathbf{c}_i$  på linjestykket fra  $\mathbf{a}$  til  $\mathbf{b}$  slik at

$$F_i(\mathbf{b}) - F_i(\mathbf{a}) = \nabla F_i(\mathbf{c}_i) \cdot (\mathbf{b} - \mathbf{a})$$

Ved Schwarz' ulikhet er da

$$|F_i(\mathbf{b}) - F_i(\mathbf{a})| \leq |\nabla F_i(\mathbf{c}_i)| |\mathbf{b} - \mathbf{a}|$$

Dermed er

$$\begin{aligned} |\mathbf{F}(\mathbf{b}) - \mathbf{F}(\mathbf{a})| &= \sqrt{(F_1(\mathbf{b}) - F_1(\mathbf{a}))^2 + \dots + (F_m(\mathbf{b}) - F_m(\mathbf{a}))^2} \leq \\ &\leq \sqrt{|\nabla F_1(\mathbf{c}_1)|^2 |\mathbf{b} - \mathbf{a}|^2 + \dots + |\nabla F_m(\mathbf{c}_m)|^2 |\mathbf{b} - \mathbf{a}|^2} \end{aligned}$$

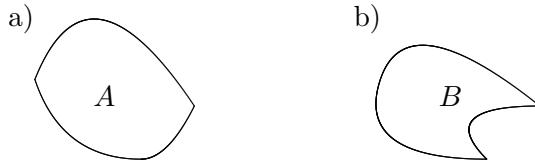
$$= |\mathbf{b} - \mathbf{a}| \sqrt{|\nabla F_1(\mathbf{c}_1)|^2 + \cdots + |\nabla F_m(\mathbf{c}_m)|^2}$$

og setningen er bevist.  $\square$

Denne setningen forteller oss at dersom det finnes et tall  $C < 1$  slik at

$$\sqrt{|\nabla F_1(\mathbf{c}_1)|^2 + \cdots + |\nabla F_m(\mathbf{c}_m)|^2} \leq C$$

for alle punkter  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m$  vi kan komme borti, så er  $\mathbf{F}$  en kontraksjon. For å formulere dette som en setning på en grei måte, trenger vi begrepet *konveks mengde*. En delmengde  $A$  av  $\mathbb{R}^m$  kalles *konveks* dersom det er slik at hver gang  $\mathbf{a}$  og  $\mathbf{b}$  er med i  $A$ , så er hele linjestykket mellom  $\mathbf{a}$  og  $\mathbf{b}$  også med i  $A$ . Intuitivt er en mengde konveks dersom randen “buler utover”. Figuren nedenfor viser en konveks mengde  $A$  og en ikke-konveks mengde  $B$ . Legg merke til at det er lett å finne to punkter i  $B$  slik at ikke hele linjestykket som forbinder dem, ligger i  $B$ .



Figur 1: Konveks mengde  $A$  og ikke-konveks mengde  $B$

**Setning 5.5.7** *Anta at  $A$  er en ikke-tom, lukket, konveks delmengde av  $\mathbb{R}^m$  og at  $\mathbf{F} : A \rightarrow A$  er en avbildning som er deriverbar i  $A$ . Anta at det finnes et tall  $C < 1$  slik at*

$$\sqrt{|\nabla F_1(\mathbf{c}_1)|^2 + \cdots + |\nabla F_m(\mathbf{c}_m)|^2} \leq C$$

*for alle punkter  $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_m \in A$ . Da er  $\mathbf{F}$  en kontraksjon og har et entydig fikspunkt. Vi kan iterere oss frem til fikspunktet ved å starte i et hvilket som helst punkt  $\mathbf{x}_0$  i  $A$ .*

**Bevis:** Dette er bare å kombinere setning 5.5.6 med teorem 5.5.4. Legg merke til at siden mengden  $A$  er konveks, vil ethvert punkt  $\mathbf{c}_i$  på linjestykket mellom to punkter  $\mathbf{a}, \mathbf{b} \in A$  selv ligge i  $A$ .  $\square$

**Bemerkning:**  $C$ -verdien vi får fra setningen ovenfor er ofte langt unna den beste (dvs. den minste) kontraksjonsfaktoren til  $\mathbf{F}$ , og det kan godt hende at  $\mathbf{F}$  er en kontraksjon selv om

$$\sqrt{|\nabla F_1(\mathbf{c}_1)|^2 + \cdots + |\nabla F_m(\mathbf{c}_m)|^2} > 1$$

La oss se på et (ganske dårlig!) eksempel:

**Eksempel 2:** Definer en avbildning  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  ved

$$F(x, y, z) = \begin{pmatrix} \frac{x}{8} - \frac{y}{2} + \frac{z}{4} + 1 \\ \frac{x}{4} + \frac{z}{4} + 2 \\ \frac{x}{2} + \frac{y}{2} - 1 \end{pmatrix}$$

Her er Jacobi-matrisen

$$\mathbf{F}(x, y, z) = \begin{pmatrix} \frac{1}{8} & -\frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Gradienten til  $F_1$  er den første linjen i matrisen, altså

$$\nabla F_1(x, y, z) = \frac{1}{8} \mathbf{i} - \frac{1}{2} \mathbf{j} + \frac{1}{4} \mathbf{k}$$

Tilsvarende er gradientene til  $F_2$  og  $F_3$  gitt av de neste linjene i matrisen

$$\nabla F_2(x, y, z) = \frac{1}{4} \mathbf{i} + \frac{1}{4} \mathbf{k}$$

og

$$\nabla F_3(x, y, z) = \frac{1}{2} \mathbf{i} + \frac{1}{2} \mathbf{j}$$

Uansett hvilke punkter vi evaluerer gradientene i, har vi dermed

$$\begin{aligned} & |\nabla F_1|^2 + |\nabla F_2|^2 + |\nabla F_3|^2 = \\ &= \left(\frac{1}{8}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^2 + 0^2 + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + 0^2 = \\ &= \frac{61}{64} \end{aligned}$$

Dermed er

$$\sqrt{|\nabla F_1|^2 + |\nabla F_2|^2 + |\nabla F_3|^2} = \sqrt{\frac{61}{64}} = \frac{\sqrt{61}}{8} < 1$$

og  $\mathbf{F}$  er altså en kontraksjon og har et entydig fikspunkt.

For å finne (en tilnærmet verdi for) fikspunktet, starter vi en iterasjon. Følgende MATLAB-program `fikspunkt.m` starter med punktet  $\mathbf{x}_0 = (a, b, c)$  og gjennomfører  $N$  iterasjoner. Vær oppmerksom på den vanlige forskyvningen i nummereringen av punktene; i MATLAB-programmet lar vi startpunktet være  $(x_1, y_1, z_1)$ . Husk også at vi kan gjøre programmet mer effektivt ved å gi  $\mathbf{x}$ ,  $\mathbf{y}$  og  $\mathbf{z}$  riktig lengde fra starten av (se bemerkningen etter eksempel 1 i seksjon 5.4).

```

function [x,y,z]=fikspunkt(a,b,c,N)
x=[a]; %med disse linjene forteller vi
y=[b]; %MATLAB at iterasjonen starter i
z=[c]; %punktet (a,b,c)
for n=1:N % starter for-løkken som utfører iterasjonene
    x(n+1)=x(n)/8-y(n)/2+z(n)/4+1;
    y(n+1)=x(n)/4+z(n)/4+2;
    z(n+1)=x(n)/2+y(n)-1;
end %avslutter for-løkken

```

For å regne ut de første 10 verdiene med startpunkt  $\mathbf{x}_0 = (0, 0, 0)$ , gir vi nå kommandoen

```
>> fikspunkt(0,0,0,9)
```

Output er:

$x_0 =$	$x_1 =$	$x_2 =$	$x_3 =$	$x_4 =$	$x_5 =$	$x_6 =$	$x_7 =$	$x_8 =$	$x_9 =$
0.0000	1.0000	-0.1250	0.3594	0.1074	0.2322	0.1696	0.2009	0.1853	0.1931
$y_0 =$	$y_1 =$	$y_2 =$	$y_3 =$	$y_4 =$	$y_5 =$	$y_6 =$	$y_7 =$	$y_8 =$	$y_9 =$
0.000	2.0000	2.0000	2.3438	2.3242	2.4077	2.4025	2.4234	2.4221	2.4273

Vi ser at følgen  $\mathbf{x}_n$  ser ut til å stabilisere seg rundt  $(0.2, 2.4, 1.5)$ , men at det fortsatt er ganske store fluktuasjoner. Vi gir derfor kommandoen

```
>> fikspunkt(0,0,0,19)
```

for også å få de neste ti verdiene. De er:

$x_{10} =$	$x_{11} =$	$x_{12} =$	$x_{13} =$	$x_{14} =$	$x_{15} =$	$x_{16} =$	$x_{17} =$	$x_{18} =$	$x_{19} =$
0.1892	0.1911	0.1902	0.1906	0.1904	0.1905	0.1905	0.1905	0.1905	0.1905
$y_{10} =$	$y_{11} =$	$y_{12} =$	$y_{13} =$	$y_{14} =$	$y_{15} =$	$y_{16} =$	$y_{17} =$	$y_{18} =$	$y_{19} =$
2.4269	2.4282	2.4282	2.4285	2.4285	2.4286	2.4285	2.4286	2.4286	2.4286

Nå ser vi en tydelig konvergens mot et fikspunkt med (tilnærmet) verdi  $(0.1905, 2.4286, 1.5239)$ .

Programmet **fikspunkt.m** er ganske primitivt; vi må bestemme på forhånd hvor mange iterasjoner vi ønsker. I praksis bruker man ofte mer avanserte programmer (gjerne med en **while**-løkke) som avbryter iterasjonen når man har fått en viss nøyaktighet.

Vi innledet med å si at dette er et ganske dårlig eksempel. Grunnen er at vi kunne ha funnet fikspunktet direkte ved å løse ligningssystemet  $\mathbf{F}(x, y, z) = (x, y, z)$  ved regning. Dette er et lineært ligningssystem som er lett å løse med metodene i forrige kapittel. Metoden ovenfor har imidlertid også sine fordeler; vi har vist at fikspunktet er tiltrekkende (dvs. at iterasjoner av  $\mathbf{F}$  alltid konvergerer). Det viser seg også at når man får store lineære ligningssystemer med tusenvis av ligninger og ukjente, så er det ofte raskere å løse dem ved en eller annen form for iterasjon enn ved radoperasjoner. ♣

Et problem med teorien i denne seksjonen er at det ofte kan være vanskelig å vise at den funksjonen  $\mathbf{F}$  som vi arbeider med, virkelig er en kontraktsjon. Ofte er ikke  $\mathbf{F}$  en kontraktsjon i hele  $\mathbb{R}^m$ , men bare i mindre områder i nærheten av fikspunktet, og det kan ofte være vanskelig å finne et slikt område  $A$  (for at det skal fungere, må  $\mathbf{F}$  avbilde  $A$  inn i  $A$ ). I slike tilfeller kan det ofte være lurt å prøve seg med en fikspunktiterasjon uten å være sikker på at  $\mathbf{F}$  er en kontraktsjon. I praktisk arbeid er ofte et fikspunkt funnet på denne måten mer enn godt nok, og i teoretisk arbeid kan informasjon om hvor fikspunktet (sannsynligvis) ligger, gjøre det lettere å lokalisere området der  $\mathbf{F}$  er en kontraktsjon.

Det kan være greit å vite at det finnes mer generelle fikspunktteoremer enn Banachs. *Brouwers fikspunktteorem* sier for eksempel at dersom  $A$  er en ikke-tom, lukket, begrenset, konveks delmengde av  $\mathbb{R}^m$ , så vil enhver kontinuerlig funksjon  $\mathbf{F} : A \rightarrow A$  ha ett eller flere fikspunkt. Det er imidlertid ingen garanti for at disse fikspunktene kan finnes ved iterasjon.

### Oppgaver til seksjon 5.5

**1.** (Fra Oblig 2 i MAT1110, V-07) I denne oppgaven skal vi se på iterasjon av funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  gitt ved

$$\mathbf{F}(x, y) = \left( \frac{1}{2} \sin(x + y), \frac{1}{2} \cos(x - y) \right)$$

- a) Lag et MATLAB-program for å beregne følger  $\mathbf{u}_{n+1} = \mathbf{F}(\mathbf{u}_n)$ .
- b) Kjør programmet ditt med startpunkt  $\mathbf{u}_1 = (x_1, y_1) = (1, -1)$  og et passende antall iterasjoner. La MATLAB tegne følgen  $\{\mathbf{u}_n\}$ .
- c) Kjør programmet 6 ganger til. I hvert tilfelle lar du MATLAB velge et tilfeldig startpunkt  $(x_1, y_1)$  med  $-2.5 \leq x_1, y_1 \leq 2.5$  (kommandoen `>> rand` produserer et tilfeldig tall mellom 0 og 1). Tegn opp sekvensene i hvert sitt subplot, og sett aksene slik at resultatene er lette å sammenligne.

**2.** La  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  være en kontinuerlig funksjon. Vis at dersom det finnes en følge  $\{\mathbf{u}_n\}$  der  $\mathbf{u}_{n+1} = \mathbf{F}(\mathbf{u}_n)$  som konvergerer mot  $\mathbf{u}$ , så er  $\mathbf{u}$  et fikspunkt for  $\mathbf{F}$ .

**3.** Vis at enhver kontinuerlig funksjon  $f : [0, 1] \rightarrow [0, 1]$  har et fikspunkt.

**4.** Du står midt i et landskap og holder et kart over det samme området vannrett foran deg. Forklar at det finnes nøyaktig ett punkt på kartet som ligger loddrett over det tilsvarende punktet i landskapet.

**5.** I denne oppgaven skal vi vise følgende generalisering av Banachs fikspunktteorem:

**Teorem** *Anta at  $A$  er en ikke-tom, lukket delmengde av  $\mathbb{R}^m$ , og at  $\mathbf{F} : A \rightarrow A$  er funksjon slik at  $\mathbf{F}^{\circ k}$  er en kontraktsjon for en  $k \in \mathbb{N}$ . Da har  $\mathbf{F}$  nøyaktig ett fikspunkt  $\mathbf{x}$  i  $A$ .*

Vi skal gjennomføre beviset i to skritt:

- Vis at ethvert fikspunkt for  $\mathbf{F}$  også er et fikspunkt for  $\mathbf{F}^{\circ k}$ . Forklar at dette medfører at  $\mathbf{F}$  kan ha høyst ett fikspunkt.
- Ifølge Banachs fikspunktteorem har  $\mathbf{F}^{\circ k}$  et fikspunkt  $\mathbf{x}$ . Vis at  $\mathbf{F}(\mathbf{x})$  også er et fikspunkt for  $\mathbf{F}^{\circ k}$ . Fullfør beviset for teoremet.
- La  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  være gitt ved  $\mathbf{F}(x, y) = (2y + 1, \frac{x}{3} + 1)$ . Vis at  $\mathbf{F}$  ikke er en kontraktsjon, men at  $\mathbf{F}^{\circ 2}$  er det. Finn fikspunktet til  $\mathbf{F}$ .

**6.** (Fra Oblig 1 i MAT1110, V-07) I denne oppgaven skal vi eksperimentere med en annen type "fikspunkter". Disse "fikspunktene" er *mengder* som ikke forandrer seg når vi anvender en sammensetning av funksjoner på dem.

Vi starter med funksjonene  $\mathbf{T}_1 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ ,  $\mathbf{T}_2 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  og  $\mathbf{T}_3 : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  gitt ved  $\mathbf{T}_1(x, y) = (\frac{x}{2}, \frac{y}{2})$ ,  $\mathbf{T}_2(x, y) = (\frac{x}{2} + \frac{1}{2}, \frac{y}{2})$  og  $\mathbf{T}_3(x, y) = (\frac{x}{2} + \frac{1}{4}, \frac{y}{2} + \frac{\sqrt{3}}{4})$

- La  $S$  være den likesidede trekanten med hjørner i punktene  $(0, 0)$ ,  $(1, 0)$  og  $(\frac{1}{2}, \frac{\sqrt{3}}{2})$ . Beskriv bildene  $\mathbf{T}_1(S)$ ,  $\mathbf{T}_2(S)$  og  $\mathbf{T}_3(S)$  av  $S$  under henholdsvis  $\mathbf{T}_1$ ,  $\mathbf{T}_2$  og  $\mathbf{T}_3$  (husk at generelt er  $\mathbf{T}(S) = \{\mathbf{T}(x, y) \mid (x, y) \in S\}$ ).
- La  $S_1$  være det samlede bildet av  $S$  under  $\mathbf{T}_1$ ,  $\mathbf{T}_2$  og  $\mathbf{T}_3$  (vi har altså  $S_1 = \mathbf{T}_1(S) \cup \mathbf{T}_2(S) \cup \mathbf{T}_3(S)$ ). Beskriv bildet  $S_2$  av  $S_1$  under  $\mathbf{T}_1$ ,  $\mathbf{T}_2$  og  $\mathbf{T}_3$  (vi har altså  $S_2 = \mathbf{T}_1(S_1) \cup \mathbf{T}_2(S_1) \cup \mathbf{T}_3(S_1)$ ). Lag en skisse av  $S_1$  og  $S_2$ . Hvordan tror du  $S_3$  og  $S_4$  blir seende ut? Hvordan går det når du tegner inn  $S_n$  for alle  $n$ ? Mengden  $A$  vi da får kalles *Sierpinski-trekanten*, og den er et slags fikspunkt i den forstand at den ikke endrer seg når vi bruker  $\mathbf{T}_1$ ,  $\mathbf{T}_2$  og  $\mathbf{T}_3$  på den (dvs. at  $A = \mathbf{T}_1(A) \cup \mathbf{T}_2(A) \cup \mathbf{T}_3(A)$ ).

Vi skal nå se på følgende prosedyre: Start med et punkt  $(x_1, y_1)$  i planet. Velg tilfeldig én av funksjonen  $\mathbf{T}_1$ ,  $\mathbf{T}_2$ ,  $\mathbf{T}_3$ , og la  $(x_2, y_2)$  være det punktet vi får når vi bruker denne funksjonen på  $(x_1, y_1)$ . Velg på ny én av funksjonen  $\mathbf{T}_1$ ,  $\mathbf{T}_2$ ,  $\mathbf{T}_3$  på en tilfeldig måte, og la  $(x_3, y_3)$  være det punktet vi får når vi bruker denne funksjonen på  $(x_2, y_2)$ . Fortsetter vi på denne måten, får vi en følge av punkter  $\{(x_n, y_n)\}$  i planet. Det viser seg at denne følgen alltid hoper seg opp på den samme mengden. Vi skal undersøke dette fenomenet nærmere.

- Forklar at følgende MATLAB-program er en implementering av prosedyren ovenfor.

```
function [x,y]=Sierpinski(a,b,N)
x(1)=a;
y(1)=b;
for n=1:N
z=rand;
if z<=1/3
x(n+1)=x(n)/2;
y(n+1)=y(n)/2;
elseif z<=2/3
x(n+1)=x(n)/2+1/2;
y(n+1)=y(n)/2;
else
x(n+1)=x(n)/2+1/4;
y(n+1)=y(n)/2+sqrt(3)/4;
```

```

end
end
plot(x,y,'.')

```

Du må sannsynligvis bruke MATLABs hjelpefunksjon for å finne ut av noen av kommandoene.

- d) Legg programmet ovenfor på en m-fil og kjør det med forskjellige input-verdier  $a$ ,  $b$  og  $N$ . Beskriv det du ser. Velg noen av kjøringene, og presenter dem i samme figur ved hjelp av kommandoen `subplot`. (*Hint:* Du bør velge  $N$  opp til størrelsesorden 10 000 for å få et godt bilde, men det er lærerikt også å se på mindre verdier av  $N$ . Det kan også være lurt å bruke kommandoen `axis('equal')` før du lagrer bildene.)
- e) Undersøk også hvordan programmet nedenfor fungerer (rett opp figurene med `axis('equal')`).

```

function [x,y]=blomst(r,d,s,N)
x(1)=1;
y(1)=0;
for n=1:N
z=rand;
if z<=0.95
x(n+1)=r*(x(n)*cos(pi/d)-y(n)*sin(pi/d));
y(n+1)=r*(x(n)*sin(pi/d)+y(n)*cos(pi/d));
elseif z>.95
x(n+1)=s*x(n)+1;
y(n+1)=s*y(n);
end
end
plot(x,y,'.')

```

Kjør programmet først med  $r = 0.97, d = 7, s = \frac{1}{6}, N = 10000$ , men eksperimenter også med andre verdier. Bruk de to funksjonene  $\mathbf{F}_1(x, y) = r(x \cos \frac{\pi}{d} - y \sin \frac{\pi}{d}, x \sin \frac{\pi}{d} + y \cos \frac{\pi}{d})$  og  $\mathbf{F}_2(x, y) = (sx + 1, sy)$  til å forklare hvordan bildene fremkommer. (*Hint:* I en “typisk” kjøring vil man velge  $\mathbf{F}_1$  mange ganger før  $\mathbf{F}_2$  dukker opp første gang. Deretter kommer sannsynligvis en ny sekvens av  $\mathbf{F}_1$ ’ere osv. Prøv å knytte dette bildet til de geometriske tolkningene av  $\mathbf{F}_1$  og  $\mathbf{F}_2$ .)

## 5.6 Newtons metode i flere variable

I forrige kapittel så vi hvordan vi kan løse lineære ligningssystemer med  $m$  ligninger og  $m$  ukjente:

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\
&\vdots && \vdots & \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mm}x_m &= b_m
\end{aligned}$$

Lineære ligningssystemer er viktige i mange sammenhenger, men de fleste ligningssystemene som dukker opp i praksis, er ikke av denne typen, men inneholder mer kompliserte funksjonsuttrykk. Vi kan f.eks. ha tre mer kompliserte ligninger slik som her:

$$\begin{aligned} 3x^2y + 2e^{z+z} &= 0 \\ 2z^2 \cos(xy^2 + z) + e^x &= 0 \\ x^3(y^2 + z) &= 0 \end{aligned}$$

Slike ligningssystemer er som regel umulig å løse eksakt, og man må derfor bruke datamaskin for å finne tilnærmede løsninger. Vi skal se hvordan Newtons metode, som du kjenner for funksjoner av én variabel (se *Kalkulus*, seksjon 7.3), kan utvides til ligningssystemer.

Vi observerer først at det å løse et ligningssystem med  $m$  ligninger og  $m$  ukjente, er det samme som å finne nullpunktene til en funksjon  $\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ . Lar vi f.eks.  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  være funksjonen

$$\mathbf{F}(x, y, z) = \begin{pmatrix} 3x^2y + 2e^{z+z} \\ 2z^2 \cos(xy^2 + z) + e^x \\ x^3(y^2 + z) \end{pmatrix}$$

ser vi at løsningene til ligningene ovenfor er de samme som nullpunktene til funksjonen  $\mathbf{F}$ . Vi er altså interessert i å finne nullpunkter til funksjoner  $\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ .

La oss anta at vi vet eller mistenker at funksjonen har et nullpunkt i nærheten av et punkt  $\mathbf{x}_0 \in \mathbb{R}^m$ . Vi skal forsøke å finne en bedre tilnærming til nullpunktet enn  $\mathbf{x}_0$ . Husk at i nærheten av  $\mathbf{x}_0$  er  $\mathbf{F}$  godt tilnærmet av sin linearisering i  $\mathbf{x}_0$ , altså av den affine funksjonen

$$T_{\mathbf{x}_0}\mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_0) + \mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

der  $\mathbf{F}'(\mathbf{x}_0)$  er Jacobi-matrisen til  $\mathbf{F}$  i punktet  $\mathbf{x}_0$  (se seksjon 2.9). Istedentfor å løse den kompliserte ligningen  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ , løser vi den enklere ligningen  $T_{\mathbf{x}_0}\mathbf{F}(\mathbf{x}) = \mathbf{0}$ , dvs. ligningen

$$\mathbf{F}(\mathbf{x}_0) + \mathbf{F}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \mathbf{0}$$

Dette er et lineært ligningssystem, og hvis Jacobi-matrisen  $\mathbf{F}'(\mathbf{x}_0)$  er inverterbar, har det løsningen

$$\mathbf{x} = \mathbf{x}_0 - \mathbf{F}'(\mathbf{x}_0)^{-1}\mathbf{F}(\mathbf{x}_0)$$

Som allerede påpekt er dette en løsning av det forenklede ligningssystemet  $T_{\mathbf{x}_0}\mathbf{F}(\mathbf{x}) = \mathbf{0}$  og ikke av det opprinnelige ligningssystemet  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$  som vi egentlig vil løse, men med litt flaks er  $\mathbf{x}$  en bedre tilnærming til løsningen

enn den gjetningen  $\mathbf{x}_0$  som vi startet med. Vi lar derfor  $\mathbf{x}_1 = \mathbf{x}$  være vår andre tilnærming til løsning.

Vi gjentar nå prosedyren med  $\mathbf{x}_1$  som input istedenfor  $\mathbf{x}_0$  og får en ny tilnærming

$$\mathbf{x}_2 = \mathbf{x}_1 - \mathbf{F}'(\mathbf{x}_1)^{-1}\mathbf{F}(\mathbf{x}_1)$$

Deretter bruker vi  $\mathbf{x}_2$  som input og får en ny tilnærming  $\mathbf{x}_3$ :

$$\mathbf{x}_3 = \mathbf{x}_2 - \mathbf{F}'(\mathbf{x}_2)^{-1}\mathbf{F}(\mathbf{x}_2)$$

Slik fortsetter vi å følge mønsteret

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$$

og håper å nærme oss en løsning av det opprinnelige ligningssystemet  $\mathbf{F}(\mathbf{x}) = 0$ .

**Definisjon 5.6.1** *Anta at  $\mathbf{F} : A \rightarrow \mathbb{R}^m$  er en deriverbar funksjon av  $m$  variable. Newtons metode anvendt på  $\mathbf{F}$  med startpunkt  $\mathbf{x}_0$  gir oss følgen  $\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \dots$  der*

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$$

Legg merke til at Newtons metode ikke er noe annet enn fikspunkts-iterasjon av funksjonen

$$\mathbf{G}(\mathbf{x}) = \mathbf{x} - \mathbf{F}'(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$$

Legg også merke til at prosedyren forutsetter at Jacobi-matrisen  $\mathbf{F}'(\mathbf{x}_n)$  er inverterbar for alle  $n$ .

La oss se på et eksempel.

**Eksempel 1:** Vi skal bruke Newtons metode til å finne en løsning av ligningssystemet

$$\begin{aligned} x^2y + 1 &= 0 \\ e^x + y &= 0 \end{aligned}$$

Sagt på en annen måte skal vi finne et nullpunkt for funksjonen

$$\mathbf{F}(x, y) = \begin{pmatrix} x^2y + 1 \\ e^x + y \end{pmatrix}$$

Denne funksjonen har Jacobi-determinant

$$\mathbf{F}'(x, y) = \begin{pmatrix} 2xy & x^2 \\ e^x & 1 \end{pmatrix}$$

Lar vi  $\mathbf{x}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ ,  $\mathbf{x}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ ,  $\mathbf{x}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots$  være en følge som fremkommer når vi bruker Newtons metode på  $\mathbf{F}$ , ser vi at iterasjonsformelen

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$$

kan skrives

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} - \begin{pmatrix} 2x_n y_n & x_n^2 \\ e^{x_n} & 1 \end{pmatrix}^{-1} \begin{pmatrix} x_n^2 y_n + 1 \\ e^{x_n} + y_n \end{pmatrix}$$

For å regne ut punktene i følgen skriver vi et lite MATLAB-program. Legg merke til at variablen  $u$  alltid inneholder det siste punktet  $\mathbf{x}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix}$  vi har regnet ut, mens variablene  $x$  og  $y$  lagrer hele listen av  $x$ -og  $y$ -koordinater. Legg også merke til at vi bruker kommandoen  $A \backslash v$  istedenfor  $A^{-1}v$  for å spare regnearbeid (dette er mer effektivt enn å tvinge MATLAB til å regne ut  $A^{-1}$ ).

```
function [x,y]=newtonfler(a,b,N)
x=zeros(1,N); % sørger for at vektoren x har "riktig" lengde
y=zeros(1,N); % sørger for at vektoren y har "riktig" lengde
x(1)=a; %laster inn x-koordinaten til startpunktet
y(1)=b; %laster inn y-koordinaten til startpunktet
u=[a;b]; %setter u lik startpunktet
for n=1:N
A=[2*x(n)*y(n) x(n)^2;exp(x(n)) 1]; %setter A lik Jacobi-matrisen
v=[x(n)^2*y(n)+1;exp(x(n))+y(n)]; %setter v lik funksjonsverdien
u=u-A\ v; %oppdaterer u
x(n+1)=u(1); %finner x(n+1) som førstekoordinaten til u
y(n+1)=u(2); %finner y(n+1) som annenkoordinaten til u
end
```

For å kjøre programmet med startpunkt  $(-0.7, -2.5)$ , gir vi kommando-en

```
>> [x,y]=newtonfler(-.7,-2.5,20);
```

Vi kan få ut  $x$ - og  $y$ -verdiene ved å skrive henholdsvis `>> x` og `>> y`. Tabellen nedenfor viser de første verdiene:

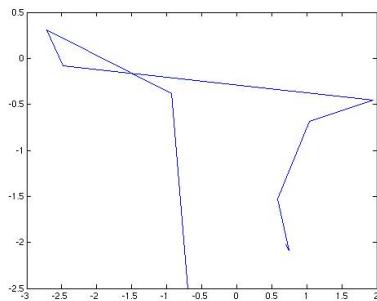
$x_0 =$	$x_1 =$	$x_2 =$	$x_3 =$	$x_4 =$	$x_5 =$	$x_6 =$	$x_7 =$	$x_8 =$	$x_9 =$	$x_{10} =$
-0.7000	-0.9323	-2.7167	-2.4802	1.9425	1.0409	0.5810	0.7475	0.7056	0.7035	0.7035

$y_0 =$	$y_1 =$	$y_2 =$	$y_3 =$	$y_4 =$	$y_5 =$	$y_6 =$	$y_7 =$	$y_8 =$	$y_9 =$	$y_{10} =$
-2.5000	-0.3812	0.3088	-0.0817	-0.4540	-0.6865	-1.5295	-2.0855	-2.0233	-2.0208	-2.0207

Følgen ser altså ut til å konvergere mot et nullpunkt med koordinater tilnærmet lik  $(0.7035, -2.0207)$  (de neste tallene i utskriften bekrefter dette

inntrykket). For å få bedre oversikt hvordan følgen oppfører seg, kan vi plotte den med kommandoen `>> plot(x,y)`.



Figur 1: Konvergens av Newtons metode

Resultatet ser du i figur 1. Legg merke til at de siste skrittene er så små at du ikke kan se dem på figuren. ♣

Programmet i eksemplet ovenfor er ganske primitivt siden det er skreddersydd for akkurat den funksjonen vi skal studere. Skal man bruke Newtons metode i flere sammenhenger (eller som del av et større program), bør man lage en versjon der man kan laste inn de funksjonene man skal arbeide med.

### Konvergens av Newtons metode

I eksemplet ovenfor konvergerte Newtons metode mot et nullpunkt, men i andre eksempler er dette slett ikke tilfellet — vi kan f.eks. få en følge som går mot det uendelig fjerne. Selv i eksemplet ovenfor er vi ikke hundre prosent sikre på at vi virkelig er i nærheten av et nullpunkt; det går an å lage eksempler der det ser ut som vi har konvergens mot et nullpunkt som i virkeligheten ikke finnes. Et tilleggsproblem får vi dersom funksjonen har mer enn ett nullpunkt, og vi ikke er sikre på hvilket nullpunkt Newtons metode konvergerer mot — det kan godt være et helt annet enn det vi er interessert i.

Ser du etter hva som står i *Kalkulus* (seksjon 7.3) eller i de fleste andre lærebøker om Newtons metode, vil du finne en setning av denne typen:

**Setning 5.6.2 (Newtons metode i én variabel)** *Anta  $f : \mathbb{R} \rightarrow \mathbb{R}$  har et nullpunkt i  $a$ . Dersom  $f'(a) \neq 0$ , og  $f''(x)$  eksisterer og er kontinuerlig i en omegn rundt  $a$ , så finnes det en  $\delta > 0$  slik at hvis  $x_0 \in (a - \delta, a + \delta)$ , så konvergerer følgen  $\{x_n\}$  i Newtons metode mot  $a$ .*

Setninger av denne typen (det finnes tilsvarende for funksjoner av flere variable) forteller deg at dersom du starter nær nok et nullpunkt, så vil Newtons metode (så sant betingelsene er oppfylt) konvergere mot dette nullpunktet. Slike setninger er teoretisk beroligende, men de er til liten nytte når

man bruker Newtons metode i praksis — da kjenner man jo ikke nullpunktene, og kan umulig vite om man er “nær nok” eller ikke. Vi skal isteden se på et teorem som har praktisk nytte — det gir nemlig en mulighet til å sjekke konvergens med en gang vi starter beregningene. Teoremet kalles gjerne *Kantorovitsj’ teorem* etter den russiske matematikeren og økonomen Leonid Vitaljevitsj Kantorovitsj (1912-1986) som vant Nobels minnepris i økonomi i 1975.

Røft sett sier Kantorovitsj’ teorem at Newtons metode konvergerer der som det første skrittet  $\mathbf{x}_1 - \mathbf{x}_0$  er “tilstrekkelig lite”, og at nullpunktet den konvergerer mot da ligger innenfor en kule om  $\mathbf{x}_0$  med radius  $2|\mathbf{x}_1 - \mathbf{x}_0|$ . Problemet er at hva som er “tilstrekkelig lite” avhenger av den funksjonen  $\mathbf{F}$  vi arbeider med, og det punktet  $\mathbf{x}_0$  vi starter i. Dess raskere  $\mathbf{F}'$  varierer og dess mindre  $\mathbf{F}'(\mathbf{x}_0)$  er, dess strengere må vi være med kravet til  $\mathbf{x}_1 - \mathbf{x}_0$ .

La oss aller først innføre et mål på hvor fort  $\mathbf{F}'$  varierer. I seksjon 5.3 definerte vi *operatornormen*  $|A|$  til en kvadratisk matrise  $A$  ved

$$|A| = \sup \left\{ \frac{|A\mathbf{x}|}{|\mathbf{x}|} : \mathbf{x} \in \mathbb{R}^m, \mathbf{x} \neq \mathbf{0} \right\}$$

og understreket at dette er et (av flere mulige) mål på hvor stor en matrise er. Når vi i Kantorovitsj’ teorem har forutsetningen

$$|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})| \leq M|\mathbf{u} - \mathbf{v}| \quad \text{for alle } \mathbf{u}, \mathbf{v} \in U_0$$

betyr dette at størrelsen til differensen mellom Jacobi-matrisen i punktet  $\mathbf{u}$  og Jacobi-matrisen i punktet  $\mathbf{v}$  er mindre enn  $M$  ganger størrelsen til differensen mellom  $\mathbf{u}$  og  $\mathbf{v}$  — med andre ord: Jacobi-matrisen endrer seg ikke veldig mye raskere enn avstanden mellom punktene. Som vi snart skal se, vil to ganger deriverbare funksjoner oppfylle en slik betingelse, men kanskje med en  $M$  som er ganske stor.

Før vi skriver opp Kantorovitsj’ teorem, minner vi om at ifølge Newtons metode er

$$\mathbf{x}_1 = \mathbf{x}_0 - \mathbf{F}'(\mathbf{x}_0)^{-1}\mathbf{F}(\mathbf{x}_0),$$

og følgelig er

$$|\mathbf{x}_1 - \mathbf{x}_0| = |\mathbf{F}'(\mathbf{x}_0)^{-1}\mathbf{F}(\mathbf{x}_0)|$$

**Teorem 5.6.3 (Kantorovitsj’ teorem)** *La  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  være en deriverbar funksjon definert på en åpen, konveks delmengde  $U$  av  $\mathbb{R}^m$ , og anta at det finnes en konstant  $M$  slik at*

$$|\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{y})| \leq M|\mathbf{x} - \mathbf{y}| \quad \text{for alle } \mathbf{x}, \mathbf{y} \in U$$

*La  $\mathbf{x}_0$  være et punkt i  $U$ , og anta at Jacobi-matrisen  $\mathbf{F}'(\mathbf{x}_0)$  i  $\mathbf{x}_0$  er inverterbar med*

$$|\mathbf{F}'(\mathbf{x}_0)^{-1}| \leq K$$

Anta til slutt at den lukkede kulen  $\overline{B}(\mathbf{x}_0, \frac{1}{KM})$  er inneholdt i definisjonsmengden  $U$ . Dersom

$$|\mathbf{x}_1 - \mathbf{x}_0| = |\mathbf{F}'(\mathbf{x}_0)^{-1} \mathbf{F}(\mathbf{x}_0)| \leq \frac{1}{2KM},$$

er  $\mathbf{F}'(\mathbf{x})$  inverterbar for alle  $\mathbf{x}$  i den åpne kulen  $B(\mathbf{x}_0, \frac{1}{KM})$ . Starter vi Newtons metode i  $\mathbf{x}_0$ , vil alle punktene  $\mathbf{x}_n$  ligge i  $B(\mathbf{x}_0, \frac{1}{KM})$ , og de vil konvergere mot et punkt  $\mathbf{x} \in \overline{B}(\mathbf{x}_0, \frac{1}{KM})$  der  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ .

Det er altså tallet  $\frac{1}{2KM}$  som avgjør hvor lite det første skrittet  $\mathbf{x}_1 - \mathbf{x}_0$  må være for at vi skal være sikret konvergens. Du har kanskje stusset på at vi bruker ulikheten  $|\mathbf{F}'(\mathbf{x}_0)^{-1}| \leq K$  til å beskrive  $K$ , og ikke bare setter  $K$  lik  $|\mathbf{F}'(\mathbf{x}_0)^{-1}|$ . Grunnen til dette er at  $|\mathbf{F}'(\mathbf{x}_0)^{-1}|$  er vanskelig å beregne eksakt, og at det derfor er greit å ha et kriterium som tillater oss å bruke et estimat  $K$ .

Vær for øvrig oppmerksom på at følgen  $\{\mathbf{x}_n\}$  godt kan konvergere uten at betingelsene i Kantorovitsj' teorem er oppfylt — i mange tilfeller vil betingelsene først være oppfylt etter at vi har gjennomført noen iterasjoner. Selv om betingelsene ikke er oppfylt for  $\mathbf{x}_1 - \mathbf{x}_0$ , bør du derfor teste  $\mathbf{x}_2 - \mathbf{x}_1$ ,  $\mathbf{x}_3 - \mathbf{x}_2$  og noen ledd til før du gir opp.

Før vi går videre, tar vi med to tilleggsresultater — ett om entydighet av nullpunktet og ett om konvergenshastighet. Begge er egentlig del av Kantorovitsj' resultat, men vi velger å presentere dem separat for ikke å få et altfor overlesset teorem. Vi kommer tilbake med bevisene etter at vi har bevist Kantorovitsj' teorem.

**Setning 5.6.4** Anta at betingelsene i Kantorovitsj' teorem er oppfylt. Da har  $\mathbf{F}$  nøyaktig ett nullpunkt i den lukkede kulen  $\overline{B}(\mathbf{x}_0, \frac{1}{KM})$ , nemlig det vi finner ved å starte Newtons metode i  $\mathbf{x}_0$ .

**Setning 5.6.5** Anta at betingelsene i Kantorovitsj' teorem er oppfylt, og la  $\mathbf{x}$  være grensepunktet for følgen  $\{\mathbf{x}_n\}$  gitt av Newtons metode med startpunkt  $\mathbf{x}_0$ . Anta at  $|\mathbf{x}_1 - \mathbf{x}_0| \leq \epsilon \leq \frac{1}{2KM}$ . Da er

$$|\mathbf{x} - \mathbf{x}_n| \leq \frac{1}{KM} \left( \frac{(1 - \sqrt{1 - 2h})^{2^n}}{2^n} \right)$$

der  $h = KM\epsilon \leq \frac{1}{2}$ .

Det siste resultatet ser litt mystisk ut før man har lest beviset for Kantorovitsj's teorem, men observer at når  $h < \frac{1}{2}$ , så er  $1 - \sqrt{1 - 2h} < 1$ . Uttrykket  $(1 - \sqrt{1 - 2h})^{2^n}$  går derfor svært raskt mot 0 (man kaller ofte fenomenet *superkonvergens*).

Før vi går løs på bevisene for Kantorovitsj' teorem og de to tilleggssetningene ovenfor, er det et lite problem vi bør ta oss av — hvordan finner vi i praksis et tall  $M$  slik at  $|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})| \leq M|\mathbf{u} - \mathbf{v}|$  for alle  $\mathbf{u}, \mathbf{v}$  i en mengde  $U$ ? Setningen nedenfor gir oss en nyttig metode, men vær klar over at den ofte gir en  $M$ -verdi som er mye større enn nødvendig. Legg merke til at setningen (og beviset) er en litt mer komplisert variant av setning 5.5.6.

**Setning 5.6.6** *Anta at  $U$  er en åpen, konveks delmengde av  $\mathbb{R}^m$  og at  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  har kontinuerlige annenderiverte. Anta at det for alle tripler  $(i, j, k)$ ,  $1 \leq i, j, k \leq m$ , finnes tall  $m_{i,j,k}$  slik at*

$$\left| \frac{\partial^2 F_i}{\partial x_k \partial x_j}(\mathbf{x}) \right| \leq m_{i,j,k} \quad \text{for alle } \mathbf{x} \in U$$

Sett

$$M = \left( \sum_{1 \leq i, j, k \leq m} m_{i,j,k}^2 \right)^{\frac{1}{2}}$$

Da er

$$|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})| \leq M|\mathbf{u} - \mathbf{v}| \quad \text{for alle } \mathbf{u}, \mathbf{v} \in U$$

*Bevis:* Det  $i, j$ -te elementet i matrisen  $\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})$  er  $\frac{\partial F_i}{\partial x_j}(\mathbf{u}) - \frac{\partial F_i}{\partial x_j}(\mathbf{v})$ . Ifølge middelverdisetningen for funksjoner av flere variable (se setning 3.2.3 eller setning 5.5.5), finnes det et punkt  $\mathbf{c}_{i,j}$  på linjestykket mellom  $\mathbf{u}$  og  $\mathbf{v}$  slik at

$$\frac{\partial F_i}{\partial x_j}(\mathbf{u}) - \frac{\partial F_i}{\partial x_j}(\mathbf{v}) = \nabla \left( \frac{\partial F_i}{\partial x_j} \right) (\mathbf{c}_{i,j}) \cdot (\mathbf{u} - \mathbf{v})$$

Ifølge Schwartz' ulikhet er dermed

$$\left| \frac{\partial F_i}{\partial x_j}(\mathbf{u}) - \frac{\partial F_i}{\partial x_j}(\mathbf{v}) \right| \leq \left| \nabla \left( \frac{\partial F_i}{\partial x_j} \right) (\mathbf{c}_{i,j}) \right| |\mathbf{u} - \mathbf{v}|$$

Bruker vi at

$$\left| \nabla \left( \frac{\partial F_i}{\partial x_j} \right) (\mathbf{c}_{i,j}) \right| = \sqrt{\left( \frac{\partial^2 F_i}{\partial x_1 \partial x_j}(\mathbf{c}_{i,j}) \right)^2 + \cdots + \left( \frac{\partial^2 F_i}{\partial x_m \partial x_j}(\mathbf{c}_{i,j}) \right)^2} \leq \sqrt{\sum_{k=1}^m m_{i,j,k}^2},$$

får vi

$$\left| \frac{\partial F_i}{\partial x_j}(\mathbf{u}) - \frac{\partial F_i}{\partial x_j}(\mathbf{v}) \right| \leq \sqrt{\sum_{k=1}^m m_{i,j,k}^2} |\mathbf{u} - \mathbf{v}|$$

Husk at normen  $\|A\|$  til en  $m \times m$ -matrise  $A$  med komponenter  $a_{ij}$  er gitt ved

$$\|A\| = \sqrt{\sum_{1 \leq i, j \leq m} a_{ij}^2}$$

Dermed er

$$\begin{aligned}\|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})\| &= \sqrt{\sum_{i,j=1}^m \left( \frac{\partial F_i}{\partial x_j}(\mathbf{u}) - \frac{\partial F_i}{\partial x_j}(\mathbf{v}) \right)^2} \leq \\ &\leq \sqrt{\sum_{1 \leq i,j \leq m} \left( \sqrt{\sum_{k=1}^m m_{i,j,k}^2} |\mathbf{u} - \mathbf{v}| \right)^2} = \sqrt{\sum_{1 \leq i,j,k \leq m} m_{i,j,k}^2} |\mathbf{u} - \mathbf{v}| = M |\mathbf{u} - \mathbf{v}|\end{aligned}$$

Siden *operatornormen*  $|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})|$  er mindre enn *normen*  $\|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})\|$  (se seksjon 5.3), følger det at

$$|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})| \leq M |\mathbf{u} - \mathbf{v}|$$

□

### \*Bevis for Kantorovitsj' teorem

Vi skal nå bevise Kantorovitsj' teorem. Før vi begynner, gjør vi oppmerksom på at argumentet er både langt og krevende sammenlignet med de fleste andre bevisene i dette kompendiet. Det gir imidlertid en flott illustrasjon av hvordan teknikkene våre kan brukes til å bevise avanserte matematiske resultater. Før vi kommer til selve beviset for Kantorovitsj' teorem, skal vi bevise tre lemmaer. I tillegg skal vi gjøre viktig bruk av Banachs lemma (teorem 5.3.11), så har du ikke lest det før, bør du gjøre det nå.

Det første lemmaet vårt vil hjelpe oss å utnytte betingelsen  $|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})| \leq M |\mathbf{u} - \mathbf{v}|$  i Kantorovitsj' teorem.

**Lemma 5.6.7** *Anta at  $U$  er en åpen, konveks delmengde av  $\mathbb{R}^m$ , og at  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  er en deriverbar funksjon slik at*

$$|\mathbf{F}'(\mathbf{u}) - \mathbf{F}'(\mathbf{v})| \leq M |\mathbf{u} - \mathbf{v}| \quad \text{for alle } \mathbf{u}, \mathbf{v} \in U$$

*Da er*

$$|\mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{x}) - \mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x})| \leq \frac{M}{2} |\mathbf{y} - \mathbf{x}|^2$$

*for alle  $\mathbf{x}, \mathbf{y} \in U$ .*

*Bevis:* Siden  $U$  er konveks, ligger alle punkter på linjestykket mellom  $\mathbf{x}$  og  $\mathbf{y}$  i  $U$  (dette er definisjonen av konveksitet). Linjestykket er parametrisert ved

$$\mathbf{r}(t) = \mathbf{x} + t(\mathbf{y} - \mathbf{x}) \quad \text{der } t \in [0, 1],$$

og deriverer vi  $\mathbf{G}(t) = \mathbf{F}(\mathbf{r}(t))$ , gir kjerneregelen

$$\mathbf{G}'(t) = \mathbf{F}'(\mathbf{r}(t))(\mathbf{y} - \mathbf{x})$$

som vi kan skrive om til

$$\mathbf{G}'(t) = \mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) + (\mathbf{F}'(\mathbf{r}(t)) - \mathbf{F}'(\mathbf{x}))(\mathbf{y} - \mathbf{x})$$

Integratorer vi på begge sider, får vi

$$\begin{aligned} \mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{x}) &= \mathbf{G}(1) - \mathbf{G}(0) = \int_0^1 \mathbf{G}'(t) dt = \\ &= \int_0^1 (\mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) + (\mathbf{F}'(\mathbf{r}(t)) - \mathbf{F}'(\mathbf{x}))(\mathbf{y} - \mathbf{x})) dt = \\ &= \mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \int_0^1 (\mathbf{F}'(\mathbf{r}(t)) - \mathbf{F}'(\mathbf{x}))(\mathbf{y} - \mathbf{x}) dt \end{aligned}$$

der vi i det siste skrittet har brukt at  $\mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x})$  er konstant. Dette betyr at

$$\begin{aligned} |\mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{x}) - \mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x})| &= \left| \int_0^1 (\mathbf{F}'(\mathbf{r}(t)) - \mathbf{F}'(\mathbf{x}))(\mathbf{y} - \mathbf{x}) dt \right| \leq \\ &\leq \int_0^1 |(\mathbf{F}'(\mathbf{r}(t)) - \mathbf{F}'(\mathbf{x}))| |\mathbf{y} - \mathbf{x}| dt \leq M \int_0^1 |\mathbf{r}(t) - \mathbf{x}| |\mathbf{y} - \mathbf{x}| dt \end{aligned}$$

Bruker vi at  $\mathbf{r}(t) - \mathbf{x} = (\mathbf{x} + t(\mathbf{y} - \mathbf{x})) - \mathbf{x} = t(\mathbf{y} - \mathbf{x})$ , får vi

$$|\mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{x}) - \mathbf{F}'(\mathbf{x})(\mathbf{y} - \mathbf{x})| \leq M \int_0^1 t |\mathbf{y} - \mathbf{x}|^2 dt = \frac{M}{2} |\mathbf{y} - \mathbf{x}|^2 \quad \square$$

Vi skal bevise Kantorovitsj' teorem ved å sammenligne følgen  $\{\mathbf{x}_n\}$  med en atskillig enklere tallfølge. Her er det grunnleggende sammenligningsprinsippet:

**Lemma 5.6.8** *La  $\{\mathbf{x}_n\}$  være en følge av punkter i  $\mathbb{R}^m$ . Anta at det finnes en voksende følge  $\{t_n\}$  av punkter på tallinjen som konvergerer mot et tall  $t$ , og som er slik at*

$$|\mathbf{x}_{n+1} - \mathbf{x}_n| \leq t_{n+1} - t_n$$

*for alle  $n$ . Da konvergerer  $\{\mathbf{x}_n\}$  mot et punkt  $\mathbf{x} \in \mathbb{R}^m$  og*

$$|\mathbf{x} - \mathbf{x}_n| \leq t - t_n$$

*for alle  $n$ .*

*Bevis:* Vi begynner med å vise at  $\{\mathbf{x}_n\}$  er en Cauchy-følge. Hvis  $k > n$ , har vi

$$\begin{aligned} |\mathbf{x}_k - \mathbf{x}_n| &= |(\mathbf{x}_k - \mathbf{x}_{k-1}) + (\mathbf{x}_{k-1} - \mathbf{x}_{k-2}) + \cdots + (\mathbf{x}_{n+1} - \mathbf{x}_n)| \leq \\ &\leq |\mathbf{x}_k - \mathbf{x}_{k-1}| + |\mathbf{x}_{k-1} - \mathbf{x}_{k-2}| + \cdots + |\mathbf{x}_{n+1} - \mathbf{x}_n| \leq \end{aligned}$$

$$\leq (t_k - t_{k-1}) + (t_{k-1} - t_{k-2}) + \cdots + (t_{n+1} - t_n) = t_k - t_n$$

Siden følgen  $\{t_n\}$  konvergerer, er den en Cauchy-følge, og følgelig kan vi få  $t_k - t_n$  så liten vi vil ved å velge  $n$  og  $k$  store nok. Men dermed kan vi også få  $|\mathbf{x}_k - \mathbf{x}_n|$  så liten vi vil, og følgelig er  $\{\mathbf{x}_n\}$  en Cauchy-følge og må konvergere mot et punkt  $\mathbf{x}$ .

Fra ulikhettene ovenfor vet vi at  $|\mathbf{x}_k - \mathbf{x}_n| \leq t_k - t_n$ . Holder vi  $n$  fast og lar  $k \rightarrow \infty$ , får vi  $|\mathbf{x} - \mathbf{x}_n| \leq t - t_n$ .  $\square$

I lemmaet ovenfor sier vi at tallfølgen  $\{t_n\}$  majoriserer den opprinnelige følgen  $\{\mathbf{x}_n\}$ . I beviset for Kantorovitsj' teorem skal vi majorisere den opprinnelige følgen  $\{\mathbf{x}_n\}$  ved hjelp av en tallfølge  $\{t_n\}$  som fremkommer ved å bruke Newtons metode på en enkel annengradsfunksjon  $P(t)$  på tallingen. Det siste lemmaet (som ser atskillig verre ut enn det er) vil gi oss de nødvendige verktøyene for å gjennomføre denne majoriseringen.

**Lemma 5.6.9** *Anta at  $a, b$  og  $c$  er positive reelle tall, og la  $P : \mathbb{R} \rightarrow \mathbb{R}$  være annengradspolynomet*

$$P(t) = at^2 - bt + c$$

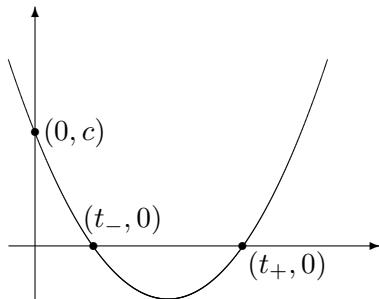
*Anta at  $P$  har reelle røtter*

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

*og la  $t_-$  være den minste roten  $t_- = \frac{b - \sqrt{b^2 - 4ac}}{2a}$ . Anta at  $\{t_n\}$  er den følgen vi får når vi bruker Newtons metode på  $P(t)$  med startverdi  $t_0 = 0$ . Da er  $\{t_n\}$  en voksende følge som konvergerer mot  $t_-$ . Punktene i følgen tilfredstiller ligningen*

$$t_{n+1} - t_n = \frac{a(t_n - t_{n-1})^2}{b - 2at_n} \quad (5.6.1)$$

*Bevis:* Siden  $a, b$  og  $c$  er positive og polynomet har reelle røtter, er det lett å sjekke at grafen må se ut som på figur 2. Det er også lett å se grafisk at følgen vi får ved å starte Newtons metode i  $t_0 = 0$ , må være voksende og konvergere mot  $t_-$  (se oppgave 5.6.8).



Figur 2: En parabel  $P(t) = at^2 - bt + c$

Det gjenstår å vise ligning (5.6.1). Her trenger vi et litt fiffig regnestykke. Siden  $P'(t) = 2at - b$ , er Newtons metode gitt ved

$$t_n = t_{n-1} - \frac{P(t_{n-1})}{P'(t_{n-1})} = t_{n-1} - \frac{at_{n-1}^2 - bt_{n-1} + c}{2at_{n-1} - b} = \frac{at_{n-1}^2 - c}{2at_{n-1} - b}$$

Ganger vi denne likheten  $t_n = \frac{at_{n-1}^2 - c}{2at_{n-1} - b}$  med  $2at_{n-1} - b$ , får vi

$$2at_n t_{n-1} - bt_n = at_{n-1}^2 - c$$

som igjen gir

$$-bt_n + c = -2at_n t_{n-1} + at_{n-1}^2 \quad (5.6.2)$$

Etter Newtons metode er også

$$t_{n+1} - t_n = -\frac{P(t_n)}{P'(t_n)} = -\frac{at_n^2 - bt_n + c}{2at_n - b}$$

Bruker vi (5.6.2) og annen kvadratsetning, kan dette skrives som

$$t_{n+1} - t_n = -\frac{at_n^2 - 2at_n t_{n-1} + at_{n-1}^2}{2at_n - b} = \frac{a(t_n - t_{n-1})^2}{b - 2at_n}$$

□

Vi er nå klare til å bevise Kantorovitsj' teorem og dets to kompanjonger, setningene 5.6.4 og 5.6.5. I disse bevisene får vi bruk for to størrelser  $\epsilon$  og  $h$ . Den første av disse er bare et tall slik at

$$|\mathbf{x}_1 - \mathbf{x}_0| \leq \epsilon \leq \frac{1}{2KM}$$

Når  $\epsilon$  er valgt, setter vi

$$h = KM\epsilon$$

Legg merke til at siden  $\epsilon \leq \frac{1}{2KM}$ , er  $h \leq \frac{1}{2}$ . Det kan virke litt merkelig at vi ikke spesifiserer verdien til  $\epsilon$ , men bare lar den være et fritt tall mellom  $|\mathbf{x}_1 - \mathbf{x}_0|$  og  $\frac{1}{2KM}$ . Dette skyldes at det ikke alltid er den samme verdien av  $\epsilon$  som er mest effektiv å bruke, og at vi derfor vil ha mulighet til å bytte  $\epsilon$ -verdi etter behov.

For å gjøre det enklere å lese beviset, skrive vi opp Kantorovitsj' teorem på nytt.

**Teorem 5.6.10 (Kantorovitsj' teorem)** *La  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  være en deriverbar funksjon definert på en åpen, konveks delmengde  $U$  av  $\mathbb{R}^m$ , og anta at det finnes en konstant  $M$  slik at*

$$|\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{y})| \leq M|\mathbf{x} - \mathbf{y}| \quad \text{for alle } \mathbf{x}, \mathbf{y} \in U$$

La  $\mathbf{x}_0$  være et punkt i  $U$ , og anta at Jacobi-matrisen  $\mathbf{F}'(\mathbf{x}_0)$  i  $\mathbf{x}_0$  er inverterbar med

$$|\mathbf{F}'(\mathbf{x}_0)^{-1}| \leq K$$

Anta til slutt at den lukkedekulen  $\overline{B}(\mathbf{x}_0, \frac{1}{KM})$  er inneholdt i definisjonsmengden  $U$ . Dersom

$$|\mathbf{x}_1 - \mathbf{x}_0| = |\mathbf{F}'(\mathbf{x}_0)^{-1} \mathbf{F}(\mathbf{x}_0)| \leq \frac{1}{2KM},$$

er  $\mathbf{F}'(\mathbf{x})$  inverterbar for alle  $\mathbf{x}$  i den åpnekulen  $B(\mathbf{x}_0, \frac{1}{KM})$ . Starter vi Newtons metode i  $\mathbf{x}_0$ , vil alle punktene  $\mathbf{x}_n$  ligge i  $B(\mathbf{x}_0, \frac{1}{KM})$ , og de vil konvergere mot et punkt  $\mathbf{x} \in \overline{B}(\mathbf{x}_0, \frac{1}{KM})$  der  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ .

*Bevis:* Ifølge Banachs lemma (teorem 5.3.11) er  $\mathbf{F}'(\mathbf{x})$  inverterbar dersom  $|\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{x}_0)| < |\mathbf{F}'(\mathbf{x}_0)^{-1}|^{-1}$ . Siden  $|\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{x}_0)| \leq M|\mathbf{x} - \mathbf{x}_0|$  og  $K^{-1} \leq |\mathbf{F}'(\mathbf{x}_0)^{-1}|^{-1}$ , er denne betingelsen oppfylt dersom  $M|\mathbf{x} - \mathbf{x}_0| < K^{-1}$ , dvs. når  $|\mathbf{x} - \mathbf{x}_0| < \frac{1}{KM}$ . Altså er  $\mathbf{F}'(\mathbf{x})$  inverterbar når  $\mathbf{x} \in B(\mathbf{x}_0, \frac{1}{KM})$ . Fra Banachs lemma vet vi at da er

$$|\mathbf{F}'(\mathbf{x})^{-1}| \leq \frac{|\mathbf{F}'(\mathbf{x}_0)^{-1}|}{1 - |\mathbf{F}'(\mathbf{x}_0)^{-1}| |\mathbf{F}'(\mathbf{x}) - \mathbf{F}'(\mathbf{x}_0)|} \leq \frac{K}{1 - KM|\mathbf{x} - \mathbf{x}_0|}$$

Dersom  $\mathbf{x}_n \in B(\mathbf{x}_0, \frac{1}{KM})$  (dette vil følge fra argumenter senere i beviset), har vi dermed

$$|\mathbf{x}_{n+1} - \mathbf{x}_n| = |-\mathbf{F}'(\mathbf{x}_n)^{-1} \mathbf{F}(\mathbf{x}_n)| \leq |\mathbf{F}'(\mathbf{x}_n)^{-1}| |\mathbf{F}(\mathbf{x}_n)| \leq \frac{K |\mathbf{F}(\mathbf{x}_n)|}{1 - KM|\mathbf{x}_n - \mathbf{x}_0|}$$

Bruker vi at  $\mathbf{F}(\mathbf{x}_{n-1}) + \mathbf{F}'(\mathbf{x}_{n-1})(\mathbf{x}_n - \mathbf{x}_{n-1}) = \mathbf{0}$  (dette er selve utgangspunktet for Newtons metode), ser vi at

$$|\mathbf{F}(\mathbf{x}_n)| = |\mathbf{F}(\mathbf{x}_n) - \mathbf{F}(\mathbf{x}_{n-1}) - \mathbf{F}'(\mathbf{x}_{n-1})(\mathbf{x}_n - \mathbf{x}_{n-1})| \leq \frac{M}{2} |\mathbf{x}_n - \mathbf{x}_{n-1}|^2$$

ifølge lemma 5.6.7. Setter vi dette inn i ulikheten ovenfor, får vi

$$|\mathbf{x}_{n+1} - \mathbf{x}_n| \leq \frac{KM|\mathbf{x}_n - \mathbf{x}_{n-1}|^2}{2(1 - KM|\mathbf{x}_n - \mathbf{x}_0|)}$$

Denne ulikheten minner om likheten (5.6.1) i lemma 5.6.6, og det er denne observasjonen som er utgangspunktet for vårt majoriseringstriks.

La  $P(t) = \frac{KM}{2}t^2 - t + \epsilon$  der  $\epsilon$  (som ovenfor) er et tall slik at  $|\mathbf{x}_1 - \mathbf{x}_0| \leq \epsilon \leq \frac{1}{2KM}$ . Løser vi annengradsligningen  $P(t) = 0$ , får vi

$$t = \frac{1 \pm \sqrt{1 - 4\frac{KM}{2}\epsilon}}{2\frac{KM}{2}} = \frac{1 \pm \sqrt{1 - 2h}}{KM}$$

der  $h = KM\epsilon \leq \frac{1}{2}$ . Dette betyr at betingelsene i lemma 5.6.6 er oppfylt, og vi vet dermed at hvis vi bruker Newtons metode på  $P(t)$  med startpunkt  $t_0 = 0$ , så vil følgen  $\{t_n\}$  vokse mot grenseverdien

$$t_- = \frac{1 - \sqrt{1 - 2h}}{KM}$$

Legg merke til at  $t_- \leq \frac{1}{KM}$ .

Fra lemma 5.6.6 vet vi også at

$$t_{n+1} - t_n = \frac{KM(t_n - t_{n-1})^2}{2(1 - KMt_n)}$$

Ved induksjon er det nå lett å vise at  $|\mathbf{x}_{n+1} - \mathbf{x}_n| \leq t_{n+1} - t_n$ . For  $n = 0$  følger dette av at  $t_1 - t_0 = -\frac{P(0)}{P'(0)} = -\frac{\epsilon}{-1} = \epsilon$ , mens  $|\mathbf{x}_1 - \mathbf{x}_0| \leq \epsilon$  per definisjon av  $\epsilon$ . For å gjennomføre induksjonstrinnet antar vi at  $|\mathbf{x}_{k+1} - \mathbf{x}_k| \leq t_{k+1} - t_k$  for alle  $k < n$ . Da er  $|\mathbf{x}_n - \mathbf{x}_0| \leq t_n - t_0 = t_n$ , og vi har

$$|\mathbf{x}_{n+1} - \mathbf{x}_n| \leq \frac{KM|\mathbf{x}_n - \mathbf{x}_{n-1}|^2}{2(1 - KM|\mathbf{x}_n - \mathbf{x}_0|)} \leq \frac{KM(t_n - t_{n-1})^2}{2(1 - KMt_n)} = t_{n+1} - t_n$$

Dette viser at  $\{t_n\}$  majoriserer  $\{\mathbf{x}_n\}$ . Legg merke til at siden  $|\mathbf{x}_n - \mathbf{x}_0| \leq t_n < t_- < \frac{1}{KM}$ , er følgen  $\{\mathbf{x}_n\}$  hele tiden innenfor denkulen  $B(\mathbf{x}_0, \frac{1}{KM})$  der vi vet at  $\mathbf{F}'(\mathbf{x})$  er inverterbar, og vi har derfor ingen problemer med å utføre Newton-iterasjonen uendelig mange ganger. Siden  $\{t_n\}$  majoriserer  $\{\mathbf{x}_n\}$ , konvergerer  $\mathbf{x}_n$  mot et punkt  $\mathbf{x}$  ifølge lemma 5.6.5. Siden  $\mathbf{x}_n \in B(\mathbf{x}_0, \frac{1}{KM})$ , må grensepunktet  $\mathbf{x}$  ligge i den lukkedekulen  $\overline{B}(\mathbf{x}_0, \frac{1}{KM})$  (som regel vil det ligge i den åpnekulen  $B(\mathbf{x}_0, \frac{1}{KM})$ ).

Det gjenstår å vise at  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ . Vi har allerede vist at  $|\mathbf{F}(\mathbf{x}_n)| \leq \frac{M}{2}|\mathbf{x}_n - \mathbf{x}_{n-1}|^2$ . Siden  $\{\mathbf{x}_n\}$  konvergerer, går uttrykket på høyre side mot null, og følgelig er  $\lim_{n \rightarrow \infty} |\mathbf{F}(\mathbf{x}_n)| = 0$ . Siden  $\mathbf{F}$  er kontinurlig, følger det at  $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ .  $\square$

*Bevis for setning 5.6.4:* Fra Kantorovitsj' teorem vet vi at  $\mathbf{F}$  har minst ett nullpunkt  $\mathbf{x}$  som er grensen for følgen  $\{\mathbf{x}_n\}$  generert av Newtons metode med startpunkt  $\mathbf{x}_0$ . Vi skal vise at dersom  $\mathbf{y}$  er et vilkårlig nullpunkt for  $\mathbf{F}$  i  $\overline{B}(\mathbf{x}_0, \frac{1}{KM})$ , så vil  $|\mathbf{y} - \mathbf{x}_n| \rightarrow 0$ . Dette medfører åpenbart at  $\mathbf{x} = \mathbf{y}$ , og følgelig finnes det bare ett nullpunkt.

I dette beviset gjelder det å ha  $\epsilon$  så stor som mulig. Vi setter derfor  $\epsilon = \frac{1}{2KM}$  og får  $h = KM\epsilon = \frac{1}{2}$ . Fra lemma 5.6.6 vet vi at hvis

$$\mathbf{r}_n = \mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{x}_n) - \mathbf{F}'(\mathbf{x}_n)(\mathbf{y} - \mathbf{x}_n)$$

så er

$$|\mathbf{r}_n| \leq \frac{M}{2}|\mathbf{y} - \mathbf{x}_n|^2$$

Siden  $\mathbf{F}(\mathbf{y}) = \mathbf{0}$ , får vi også fra ligningen ovenfor at

$$\mathbf{y} - \mathbf{x}_n + \mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n) = -\mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{r}_n$$

Siden  $\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{F}(\mathbf{x}_n)$ , gir dette

$$\mathbf{y} - \mathbf{x}_{n+1} = -\mathbf{F}'(\mathbf{x}_n)^{-1}\mathbf{r}_n$$

Fra beviset for Kantorovitsj' teorem vet vi at

$$|\mathbf{F}'(\mathbf{x}_n)^{-1}| \leq \frac{K}{1 - KM|\mathbf{x}_n - \mathbf{x}_0|}$$

og dermed er

$$|\mathbf{y} - \mathbf{x}_{n+1}| \leq \frac{KM|\mathbf{y} - \mathbf{x}_n|^2}{2(1 - KM|\mathbf{x}_n - \mathbf{x}_0|)} \leq \frac{KM|\mathbf{y} - \mathbf{x}_n|^2}{2(1 - KM|t_n|)}$$

der  $\{t_n\}$  er den majoriserende tallfølgen fra beviset for Kantorovitsj' teorem.  
Siden

$$t_{n+1} - t_n = \frac{KM(t_n - t_{n-1})^2}{2(1 - KMt_n)}$$

er tanken nå å vise ved induksjon at  $|\mathbf{y} - \mathbf{x}_n| \leq t_n - t_{n-1}$  for alle  $n$ . Siden  $(t_n - t_{n-1}) \rightarrow 0$ , vil da  $|\mathbf{y} - \mathbf{x}_n| \rightarrow 0$ . Induksjonen er enkel hvis vi bare kan få startbetingelsene til å stemme. Ved å bruke Newtons metode baklengs, ser vi at  $t_{-1} = -\sqrt{\frac{2\epsilon}{KM}} = -\frac{1}{\sqrt{KM}}$  (husk at vi har valgt  $\epsilon$  slik at  $KM\epsilon = \frac{1}{2}$ ). Dermed er  $t_0 - t_{-1} = \frac{1}{\sqrt{KM}}$ . Siden  $\mathbf{y} \in \overline{B}(\mathbf{x}_0, \frac{1}{\sqrt{KM}})$ , er  $|\mathbf{y} - \mathbf{x}_0| \leq \frac{1}{\sqrt{KM}}$ . Med dette har vi  $|\mathbf{y} - \mathbf{x}_0| \leq t_0 - t_{-1}$ , og siden resten av induksjonen går greit, er setningen bevist  $\square$

Før vi beviser setning 5.6.5, skal vi se på et spesialtilfelle vi får bruk for underveis.

**Lemma 5.6.11** La  $P(t)$  være polynomet  $\frac{KM}{2}t^2 - t + \epsilon$  i det tilfellet der  $\epsilon = \frac{1}{2KM}$  og  $h = KM\epsilon = \frac{1}{2}$ , dvs. at

$$P(t) = \frac{KM}{2}t^2 - t + \frac{1}{2KM}$$

La  $\{t_n\}$  være følgen vi får når vi bruker Newtons metode på  $P(t)$  med  $t_0 = 0$ , og la  $t_-$  være nullpunktet denne følgen konvergerer mot. Da er  $t_- - t_n = \frac{2^{-n}}{KM}$ .

*Bevis:* Siden

$$P(t) = \frac{KM}{2}t^2 - t + \frac{1}{2KM} = \frac{KM}{2} \left( t - \frac{1}{KM} \right)^2$$

er  $t_- = \frac{1}{KM}$ . Vi har også  $P'(t) = KM(t - \frac{1}{KM})$ , så Newtons metode gir

$$t_{n+1} = t_n - \frac{\frac{KM}{2}(t_n - \frac{1}{KM})^2}{KM(t_n - \frac{1}{KM})} = \frac{1}{2}t_n + \frac{1}{2KM}$$

En liten omforming leder til

$$\frac{1}{KM} - t_{n+1} = \frac{1}{2} \left( \frac{1}{KM} - t_n \right)$$

som ved induksjon gir

$$\frac{1}{KM} - t_n = 2^{-n} \left( \frac{1}{KM} - t_0 \right) = \frac{2^{-n}}{KM}$$

der vi har brukt at  $t_0 = 0$ . Siden  $t_- = \frac{1}{KM}$ , følger setningen.  $\square$

*Bevis for setning 5.6.5:* La  $\{t_n\}$  være tallfølgen som majoriserer  $\{\mathbf{x}_n\}$  i beviset for Kantorovitsj' teorem. Ifølge lemma 5.6.7 er  $|\mathbf{x} - \mathbf{x}_n| \leq t_- - t_n$ , så det holder å vise at

$$t_- - t_n \leq \frac{1}{KM} \left( \frac{(1 - \sqrt{1 - 2h})^{2^n}}{2^n} \right)$$

Vi begynner med å utlede identiteten

$$(t_- - t_{n+1}) = \frac{KM(t_- - t_n)^2}{2(1 - KMt_n)}$$

Bruker vi formelen for Newtons metode på  $P(t) = \frac{KM}{2}t^2 - t + \epsilon$ , får vi

$$t_- - t_{n+1} = t_- - t_n + \frac{\frac{KM}{2}t_n^2 - t_n + \epsilon}{KMt_n - 1} = \frac{KMt_-t_n - t_- + \frac{KM}{2}t_n^2 - t_n + \epsilon}{KMt_n - 1}$$

Siden  $t_-$  er en løsning av ligningen  $\frac{KM}{2}t^2 - t + \epsilon = 0$ , er

$$t_- = \frac{KM}{2}t_-^2 + \epsilon$$

Setter vi dette inn for (den ene forekomsten av)  $t_-$  i formelen ovenfor, får vi

$$\begin{aligned} t_- - t_{n+1} &= \frac{KMt_-t_n - t_- + \frac{KM}{2}t_n^2 - t_n + \epsilon}{KMt_n - 1} = \\ &= \frac{KMt_-t_n - \frac{KM}{2}t_-^2 - \epsilon + \frac{KM}{2}t_n^2 - t_n + \epsilon}{KMt_n - 1} = \frac{KM(t_- - t_n)^2}{2(1 - KMt_n)} \end{aligned}$$

som er identiteten vi skulle vise.

Neste skritt er å kontrollere faktoren  $1 - KMt_n$  i nevneren. Når  $\epsilon = \frac{1}{2KM}$  (som i lemmaet ovenfor), er  $t_n = \frac{1}{KM} (1 - \frac{1}{2^n})$ , og følgelig  $1 - KMt_n = 2^{-n}$  i dette tilfellet. Vi skal snart vise at dette er den minste verdien  $1 - KMt_n$  kan ha (vi har altså  $1 - KMt_n \geq 2^{-n}$ ), og dermed får vi

$$t_- - t_{n+1} = \frac{KM(t_- - t_n)^2}{2(1 - KMt_n)} \leq KM2^{n-1}(t - t_n)^2$$

Det er lett å vise ved induksjon at dersom  $\{r_n\}$  er en tallfølge slik at

$$r_{n+1} \leq c2^{n-1}r_n^2$$

for alle  $n \geq 0$ , så er  $r_n \leq \frac{(cr_0)^{2^n}}{c2^n}$ . Bruker vi dette på tallfølgen  $r_n = t_- - t_n$  (og husker at  $r_0 = t_- - 0 = \frac{1-\sqrt{1-2h}}{KM}$ ), får vi

$$t_- - t_n \leq \frac{\left(KM \frac{1-\sqrt{1-2h}}{KM}\right)^{2^n}}{KM2^n} = \frac{1}{KM} \frac{(1-\sqrt{1-2h})^{2^n}}{2^n}$$

som er formelen vi skulle frem til.

Det gjenstår å vise at vi virkelig har ulikheten  $1 - KMt_n \geq 2^{-n}$  for alle valg av  $\epsilon$  (og ikke bare for  $\epsilon = \frac{1}{2KM}$ ). Vi tenker oss derfor at vi bruker Newtons metode med  $t_0 = 0$  for forskjellige valg av  $\epsilon \in [0, \frac{1}{2KM}]$ , og at vi lar  $t_n(\epsilon)$  være det  $n$ -te punktet i iterasjonen. Ser vi grafisk på hvordan Newtons metode fungerer, virker det naturlig at  $t_n(\epsilon)$  er voksende som funksjon av  $\epsilon$ . Kan vi bevise dette, er vi ferdig fordi vi da har  $1 - KMt_n(\epsilon) \geq 1 - KMt_n(\frac{1}{2KM}) = 2^{-n}$ .

Siden  $t_0(\epsilon) = 0, t_1(\epsilon) = \epsilon$ , ser vi at både  $t_0(\epsilon), t_1(\epsilon)$  og  $t_1(\epsilon) - t_0(\epsilon)$  er voksende i  $\epsilon$ . Ifølge formel (5.6.1) i lemma 5.6.8 er

$$t_{n+1}(\epsilon) - t_n(\epsilon) = \frac{KM(t_n(\epsilon) - t_{n-1}(\epsilon))}{2(1 - KMt_n(\epsilon))}$$

Vi ser at hvis både  $t_n(\epsilon)$  og  $t_n(\epsilon) - t_{n-1}(\epsilon)$  er voksende i  $\epsilon$ , så forteller denne likheten oss at  $t_{n+1}(\epsilon) - t_n(\epsilon)$  er voksende i  $\epsilon$ , noe som igjen medfører at  $t_{n+1}(\epsilon)$  er voksende. Ved induksjon må da  $t_n(\epsilon)$  være voksende i  $\epsilon$  for alle  $n$ , og setningen er bevist.  $\square$

## Oppgaver til seksjon 5.6

**1.** La  $f : \mathbb{R} \rightarrow \mathbb{R}$  være funksjonen  $f(x) = x^3 - x$ .

- a) Vis at  $f$  har tre nullpunkter.
- b) Skriv et MATLAB-program for å finne nullpunktene til  $f$  ved hjelp av Newtons metode.
- c) Kjør programmet ditt med startverdier 0.4, 0.5 og 0.6. Prøv å gjette hva som skjer før du kjører programmet.

- d) Prøv å forklare resultatene i c) geometrisk ved å studere hvordan Newtons metode virker.

**2.** Funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} x^2 - y \\ x^2 + y^2 - 1 \end{pmatrix}$$

- a) Skriv et MATLAB-program for å finne nullpunktene til  $\mathbf{F}$  ved hjelp av Newtons metode.
- b)  $\mathbf{F}$  har to nullpunkter. Finn begge to ved å velge forskjellige startverdier.
- c) Finn nullpunktene til  $\mathbf{F}$  ved regning og sjekk at svarene stemmer med det du fant i b).

**3.** Funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} \frac{1}{2} \sin(x + y) \\ \frac{1}{2} \cos(x - y) \end{pmatrix}$$

- a) Beskriv de (uendelig mange) nullpunktene til  $\mathbf{F}$ .
- b) Skriv et MATLAB-program for å finne nullpunktene til  $\mathbf{F}$  ved hjelp av Newtons metode.
- c) Eksperimenter med forskjellige startverdier og se hvordan Newtons metode finner frem til forskjellige nullpunkter.

**4.** Funksjonen  $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er gitt ved

$$\mathbf{F}(x, y) = \begin{pmatrix} y^2 - x - 3 \\ x^2 - y + \sin(x - y) - 2 \end{pmatrix}$$

Skriv et MATLAB-program for å finne nullpunktene til  $\mathbf{F}$  ved hjelp av Newtons metode. Bruk programmet til å finne i hvert fall to (tilnærmede) nullpunkter.

**5.** Skriv et MATLAB-program for å løse ligningssystemet

$$\begin{aligned} e^{x+y} &= \sin(x - y) \\ y^2 - x^2 &= 1 \end{aligned}$$

Finn en (tilnærmet) løsning av systemet.

**6.** Skriv et MATLAB-program for å løse ligningssystemet

$$\begin{aligned} x^2 + y^2 + z^2 &= 9 \\ x^2 - y^2 + 2z^2 &= 1 \\ x + y + 10z &= 1 \end{aligned}$$

Finn minst to (tilnærmede) løsninger av systemet.

**7.** I denne oppgaven skal vi se på konvergensraten i setning 5.6.5. Bruk MATLAB til å regne ut noen ledd av følgen

$$z_n = \frac{(1 - \sqrt{1 - 2h})^{2^n}}{2^n}$$

for  $h = \frac{1}{2}$ ,  $h = 0.4$  og  $h = 0.3$ . Sammenlign konvergenshastighetene.

**8.** (Denne oppgaven fyller inn detaljer i beviset for lemma 5.6.9, men den kan godt gjøres uavhengig av lemmaet.) Anta at  $a$ ,  $b$  og  $c$  er positive reelle tall, og la  $P : \mathbb{R} \rightarrow \mathbb{R}$  være annengradspolynomet

$$P(t) = at^2 - bt + c$$

og forutsett at  $P$  har reelle røtter

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

Anta at  $\{t_n\}$  er den følgen vi får når vi starter Newtons metode for  $P$  med  $t_0 = 0$ . Vis at  $\{t_n\}$  er en voksende følge og at  $t_n$  er mindre enn den minste rotten  $t_- = \frac{b - \sqrt{b^2 - 4ac}}{2a}$  for alle  $n$  (bruk induksjon). Forklar at dette betyr at  $\{t_n\}$  konvergerer mot en grenseverdi  $t^*$ . Vis at  $t^* = t_-$ .

**9.** La  $\mathbf{F}$  være en funksjon fra  $\mathbb{R}^m$  til  $\mathbb{R}^m$ . Et fikspunkt for  $\mathbf{F}$  er det samme som et nullpunkt for funksjonen

$$\mathbf{G}(\mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{x}$$

a) Vis at når vi bruker Newtons metode på  $\mathbf{G}$ , så får vi iterasjonen

$$\mathbf{x}_{n+1} = (\mathbf{F}'(\mathbf{x}_n) - I_n)^{-1}(\mathbf{F}'(\mathbf{x}_n)\mathbf{x}_n - \mathbf{F}(\mathbf{x}_n))$$

- b) La  $f : \mathbb{R} \rightarrow \mathbb{R}$  være funksjonen gitt ved  $f(x) = x^3 + 3x + 1$ . Vis at funksjonen har nøyaktig ett fikspunkt og at det ligger mellom  $-1$  og  $0$ .
- c) Prøv å finne fikspunktet ved fikspunktiterasjon. Velg  $-\frac{1}{2}$  som startpunkt.
- d) Prøv å finne fikspunktet ved å bruke Newtons metode på funksjonen  $g(x) = f(x) - x$ . Velg startpunkt  $-\frac{1}{2}$ .

**10.** I denne oppgaven skal vi se på et enkelt eksempel på hvordan Kantorovitsj' teorem kan brukes i praksis. Eksemplet er så enkelt at det å bruke Kantorovitsj' teorem er å skyte spurver med kanoner, men hensikten er å illustrere bruken av teoremet i en situasjon der de regnetekniske komplikasjonene forsvinner.

Vi skal bruke Newtons metode til å finne et nullpunkt for funksjonen  $f(x) = x - \cos x$  (ifølge skjæringssetningen må det finnes et nullpunkt). Vi starter iterasjonen med  $x_0 = 0$ .

- a) Vis at  $|f'(x)| \leq 2$  for alle  $x$ . Forklar at vi dermed kan velge  $M = 2$  i Kantorovitsj' teorem.
- b) Vis at  $f'(x_0) = 1$ . Forklar at vi kan bruke  $K = 1$  i Kantorovitsj' teorem.
- c) Sjekk at  $f'(x_0)^{-1}f(x_0) = -1$ , og forklar at betingelsen i Kantorovitsj' teorem ikke er oppfylt for startpunktet  $x_0 = 0$  og de verdiene vi har funnet for  $K$  og  $M$ .

- d) Bruk allikevel Newtons metode til å regne ut at  $x_1 = 1$ .
- e) Vi skal nå bruke Kantorovitsj' teorem med  $x_1$  som utgangspunkt. Forklar at vi nå kan bruke  $M = 2$ ,  $K = 0.6$  og vis at  $|x_2 - x_1| \leq 0.25$ . Konkluder med at Kantorovitsj' teorem garanterer at iterasjonen konvergerer mot et nullpunkt for  $f$ .

## 5.7 Omvendte og implisitte funksjoner

Det første temaet vi skal ta opp i denne seksjonen, er omvendte (inverse) funksjoner. Fra teorien om funksjoner av én variabel husker du sikkert at hvis en funksjon  $f$  tar oss fra en  $x$ -verdi til en  $y$ -verdi, så bringer den omvendte funksjonen  $f^{-1}$  oss tilbake fra  $y$ -verdien til  $x$ -verdien — er  $y = f(x)$ , så er altså  $x = f^{-1}(y)$ . Du husker sikkert også at for å få teorien til å fungere ordentlig, må vi anta at funksjonen  $f$  er *injektiv*, dvs. at det til hver  $y$  finnes høyst én  $x$  slik at  $y = f(x)$ . I noen tilfeller må vi sørge for at funksjonen blir injektiv ved å redusere definisjonsområdet; dette var trikset vi brukte for å få definert de omvendte trigonometriske funksjonene arcsin, arccos og arctan.

Teorien for omvendte funksjoner av én variabel er vanskelig nok, men den har én stor fordel; siden kontinuerlige, injektive funksjoner på et intervall er strengt monotone, er det som regel lett å avgjøre om en funksjon er injektiv eller ikke. Det er også lett å se hvor mye vi må innskrenke definisjonsområdet for å gjøre en ikke-injektiv funksjon injektiv. I høyere dimensjoner er det flere geometriske muligheter, og det er slett ikke lett å få oversikt over når en funksjon er injektiv. Heldigvis finnes det et teorem som sier at dersom Jacobi-matrisen til  $\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^m$  er inverterbar i punktet  $\bar{\mathbf{x}}$ , så er  $\mathbf{F}$  injektiv når vi innskrenker den til en (tilstrekkelig liten) omegn rundt  $\bar{\mathbf{x}}$ , og den har en invers funksjon  $\mathbf{G}$  som er definert i en omegn rundt punktet  $\bar{\mathbf{y}} = \mathbf{F}(\bar{\mathbf{x}})$ . Vi kan også regne ut den deriverte til den omvendte funksjonen  $\mathbf{G}$  dersom vi kjenner den deriverte til  $\mathbf{F}$ ; akkurat som i det endimensjonale tilfellet har vi

$$\mathbf{G}'(\bar{\mathbf{y}}) = \mathbf{F}'(\bar{\mathbf{x}})^{-1}$$

Alt dette kalles *omvendt funksjonsteorem*, og dette teoremet er vårt første mål i denne seksjonen.

La oss begynne med noen definisjoner. Når vi arbeider med inverse funksjoner, er det viktigere enn ellers å holde styr på definisjonsmengder og verdimengder. Vi lar  $D_{\mathbf{F}}$  betegne *definisjonsmengden* til  $\mathbf{F}$  (dvs. de  $\mathbf{x}$  som  $\mathbf{F}(\mathbf{x})$  er definert for), og vi lar

$$V_{\mathbf{F}} = \{\mathbf{F}(\mathbf{x}) : \mathbf{x} \in D_{\mathbf{F}}\}$$

være *verdimengden* til  $\mathbf{F}$ .

**Definisjon 5.7.1** Funksjonen  $\mathbf{F} : D_{\mathbf{F}} \rightarrow V_{\mathbf{F}}$  kalles injektiv dersom det til hver  $\mathbf{y} \in V_{\mathbf{F}}$  bare finnes én  $\mathbf{x} \in D_{\mathbf{F}}$  slik at  $\mathbf{y} = \mathbf{F}(\mathbf{x})$ . I så fall er den omvendte funksjonen  $\mathbf{G} : V_{\mathbf{F}} \rightarrow D_{\mathbf{F}}$  definert ved

$$\mathbf{G}(\mathbf{y}) = \mathbf{x} \quad \text{dersom} \quad \mathbf{F}(\mathbf{x}) = \mathbf{y}$$

Den omvendte funksjonen  $\mathbf{G}$  betegnes ofte med  $\mathbf{F}^{-1}$ .

Anta at vi har en funksjon  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  og at  $U_0$  er en delmengde av  $U$ . Med restriksjonen av  $\mathbf{F}$  til  $U_0$  mener vi den funksjonen vi får når vi innskrenker definisjonsområdet til  $\mathbf{F}$  til å være  $U_0$  (men ellers ikke gjør noen endringer). Vi sier at  $U_0$  er en *omegn* om punktet  $\mathbf{x}$  dersom  $\mathbf{x}$  er et indre punkt i  $U_0$ , dvs. dersom  $U_0$  inneholder en åpen kule med sentrum i  $\mathbf{x}$ . Vi kan nå gi en presis formulering av resultatet vi er på jakt etter:

**Teorem 5.7.2 (Omvendt funksjonsteorem)** Anta at  $U$  er en åpen mengde i  $\mathbb{R}^m$  og at  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  har kontinuerlige partiellderiverte. Anta at  $\bar{\mathbf{x}} \in U$  og at Jacobi-matrisen  $\mathbf{F}'(\bar{\mathbf{x}})$  er inverterbar. Da finnes det en omegn  $U_0 \subset U$  om  $\bar{\mathbf{x}}$  slik at  $\mathbf{F}$  restriktert til  $U_0$  er injektiv. Verdimengden  $V$  til denne restriksjonen er en omegn om  $\bar{\mathbf{y}} = \mathbf{F}(\bar{\mathbf{x}})$ , og den omvendte funksjonen  $\mathbf{G} : V \rightarrow U_0$  er deriverbar i  $\bar{\mathbf{y}}$  med Jacobi-matrise

$$\mathbf{G}'(\bar{\mathbf{y}}) = \mathbf{F}'(\bar{\mathbf{x}})^{-1}$$

Beviset for omvendt funksjonsteorem er ganske vanskelig, og vi skal utsette det til slutten av seksjonen. Det vanskeligste punktet er å vise at det i det hele tatt finnes en omvendt funksjon. Legg merke til at dette er det samme som å vise at ligningen  $\mathbf{y} = \mathbf{F}(\mathbf{x})$  (med  $\mathbf{x}$  som ukjent) har en løsning når  $\mathbf{y}$  er i nærheten av  $\bar{\mathbf{y}}$ . Vi skal bruke Banachs fikspunktteorem til å vise at dette alltid er tilfellet.

Et annet tidkrevende punkt i beviset er å vise at den omvendte funksjonen  $\mathbf{G}$  er deriverbar i punktet  $\bar{\mathbf{y}}$ . Når dette er vist, er det imidlertid ikke vanskelig å finne ut hva den deriverte er. Siden  $\mathbf{F}$  og  $\mathbf{G}$  er omvendte funksjoner, har vi nemlig

$$\mathbf{G}(\mathbf{F}(\mathbf{x})) = \mathbf{x}$$

for alle  $\mathbf{x}$  i nærheten av  $\bar{\mathbf{x}}$ . Deriverer vi venstresiden av dette uttrykket, får vi ved kjerneregelen

$$\mathbf{G}'(\mathbf{F}(\mathbf{x}))\mathbf{F}'(\mathbf{x})$$

mens den deriverte av høyresiden er identitetsmatrisen  $I_m$  (hvorfor det?). Dermed er

$$\mathbf{G}'(\mathbf{F}(\mathbf{x}))\mathbf{F}'(\mathbf{x}) = I_m$$

og følgelig er

$$\mathbf{G}'(\mathbf{F}(\mathbf{x})) = \mathbf{F}'(\mathbf{x})^{-1}$$

Setter vi inn  $\mathbf{x} = \bar{\mathbf{x}}$  og bruker at  $\mathbf{F}(\bar{\mathbf{x}}) = \bar{\mathbf{y}}$ , får vi formelen i teoremet.

La oss se på et eksempel.

**Eksempel 1:** Vi skal vise at funksjonen

$$\mathbf{F}(x_1, x_2) = \begin{pmatrix} e^{x_1} + x_2 \\ x_2 \cos x_1 \end{pmatrix}$$

er injektiv når den restriktører til en passende omegn om  $(0, 0)$ , og at denne restriksjonen har en omvendt funksjon  $\mathbf{G}$  definert i en omegn om  $(1, 0)$ . Vi skal også finne de partielle deriverte til  $\mathbf{G}$  i punktet  $(1, 0)$ .

Vi observerer først at

$$\mathbf{F}(0, 0) = \begin{pmatrix} e^0 + 0 \\ 0 \cos 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Videre er

$$\mathbf{F}'(x_1, x_2) = \begin{pmatrix} e^{x_1} & 1 \\ -x_2 \sin x_1 & \cos x_1 \end{pmatrix}$$

som gir

$$\mathbf{F}'(0, 0) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Siden  $\det(\mathbf{F}'(0, 0)) = 1$ , er  $\mathbf{F}'(0, 0)$  inverterbar, og følgelig er  $\mathbf{F}$  injektiv i en omegn om  $(0, 0)$  og har en omvendt funksjon  $\mathbf{G}$  definert i en omegn om  $\mathbf{F}(0, 0) = (1, 0)$ . Siden

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

er

$$\mathbf{G}'(1, 0) = \mathbf{F}'(0, 0)^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

Det betyr at  $\frac{\partial G_1}{\partial y_1}(1, 0) = 1$ ,  $\frac{\partial G_1}{\partial y_2}(1, 0) = -1$ ,  $\frac{\partial G_2}{\partial y_1}(1, 0) = 0$  og  $\frac{\partial G_2}{\partial y_2}(1, 0) = 1$ . ♣

Omvendt funksjonsteorem brukes mye både i teori og anvendelser. I anvendelser har man ofte behov for å bytte om på hva som er uavhengige og avhengige størrelser — istedenfor å tenke på etterspørsmålet som en funksjon av prisene, har man plutselig lyst til å tenke på prisene som funksjon av etterspørsmålet. Omvendt funksjonsteorem (og særlig formelen med de deriverte) gjør det enkelt å bytte synsvinkel på denne måten. I teoretisk arbeid er det ofte betingelsene som er viktige — vi trenger garantier for at de funksjonene vi skal arbeide med, virkelig finnes og har de egenskapene vi ønsker oss.

**Bemerkning:** At  $\mathbf{F} : D_{\mathbf{F}} \rightarrow V_{\mathbf{F}}$  har en omvendt funksjon betyr at ligningsystemet  $\mathbf{y} = \mathbf{F}(\mathbf{x})$  har en entydig løsning  $\mathbf{x} \in D_{\mathbf{F}}$  for alle  $\mathbf{y} \in V_{\mathbf{F}}$ . Omvendt

funksjonsteorem kan derfor oppfattes som et result som forteller oss hvordan løsningen  $\mathbf{x}$  av ligningssystemet  $\mathbf{y} = \mathbf{F}(\mathbf{x})$  avhenger av  $\mathbf{y}$ . Det er derfor ikke så rart at det beviset vi skal gi for omvendt funksjonsteorem (se slutten av seksjonen), bygger på det vi vet om løsbarhet av ligningssystemer (nærmere bestemt på Banachs fikspunktteorem 5.5.4).

### Implisitt funksjonsteorem

Vi skal nå se på en viktig konsekvens av omvendt funksjonsteorem, nemlig det som kalles *implisitt funksjonsteorem*. Det kan være lurt å ta utgangspunkt i et konkret eksempel. Vi vet at kuleflaten med radius 1 er beskrevet av ligningen

$$x^2 + y^2 + z^2 = 1$$

Denne beskrivelsen av kuleflaten er nyttig for en del formål, men for andre hadde det vært mer effektivt å tenke på kuleflaten som en funksjonsgraf. Dette er ikke mulig hvis vi vil ha med både øvre og nedre halvkule, men dersom vi er fornøyd med (for eksempel) øvre halvkule, kan vi løse ligningen ovenfor for  $z$  og få

$$z = \sqrt{1 - x^2 - y^2}$$

Dette viser at i hvert fall lokalt (vi innskrenker oss til øvre halvkule) vil ligningen  $x^2 + y^2 + z^2 = 1$  definere  $z$  som en funksjon  $z = g(x, y) = \sqrt{1 - x^2 - y^2}$  av  $x$  og  $y$ . Dette eksemplet kan vi generalisere. Anta at vi har en funksjon  $f(x_1, x_2, \dots, x_m, y)$  av  $m+1$ -variable, og at vi er interessert i mengden av punkter  $(x_1, x_2, \dots, x_m, y)$  som tilfredsstiller ligningen

$$f(x_1, x_2, \dots, x_m, y) = 0$$

(det er lurt å tenke på dette som en generalisert flate). Akkurat som ovenfor kan vi tenke oss at vi løser denne ligningen for  $y$  og får et uttrykk

$$y = g(x_1, x_2, \dots, x_m)$$

Da har vi beskrevet den generaliserte flaten vår som grafen til en funksjon  $g$  med  $m$  variable. Vi kaller en slik funksjon  $g$  en *implisitt gitt* (eller bare *implisitt*) funksjon.

Det er flere grunner til at denne planen kanskje ikke lar seg gjennomføre. Dersom  $f$  er funksjonen

$$f(x_1, x_2, y) = x_1^2 + x_2^2 + y^2 + 1$$

så har for eksempel ikke ligningen  $f(x_1, x_2, y) = 0$  løsninger i det hele tatt. Et annet problem er at ligningen kan ha flere løsninger. I halvkuleeksemplet ovenfor har vi to løsninger

$$z = \pm \sqrt{1 - x^2 - y^2}$$

Disse er greie å skille mellom dersom vi vet at vi vil være på enten øvre eller nedre halvkule, men det finnes andre eksempler der det er vanskeligere å skjøte ting sammen. Et tredje problem er at det kan være vanskelig å løse ligningen  $f(x_1, x_2, \dots, x_m, y) = 0$  for  $y$  selv i de tilfellene der det finnes en løsning. Det er i den siste situasjonen vi virkelig får bruk for teoremet nedenfor. Det forteller oss når ligningen har en løsning, og gir oss viktig informasjon om løsningsfunksjonen i de tilfellene der vi ikke greier å regne den ut.

**Teorem 5.7.3 (Implisitt funksjonsteorem)** *Anta at  $U$  er en åpen delmengde av  $\mathbb{R}^{m+1}$  og la  $f : U \rightarrow \mathbb{R}$  være en funksjon med kontinuerlige partiellderiverte. Anta at  $(\bar{\mathbf{x}}, \bar{y}) = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m, \bar{y})$  er et punkt i  $U$  der  $f(\bar{\mathbf{x}}, \bar{y}) = 0$ . Anta videre at  $\frac{\partial f}{\partial y}(\bar{\mathbf{x}}, \bar{y}) \neq 0$ . Da finnes det en omegn  $U_0$  om  $\bar{\mathbf{x}}$ , og en deriverbar funksjon  $g : U_0 \rightarrow \mathbb{R}$  slik at  $\mathbf{g}(\bar{\mathbf{x}}) = \bar{y}$  og*

$$f(\mathbf{x}, g(\mathbf{x})) = 0$$

for alle  $\mathbf{x} \in U_0$ . Den deriverte til  $g$  er gitt ved

$$\frac{\partial g}{\partial x_i}(\bar{\mathbf{x}}) = -\frac{\frac{\partial f}{\partial x_i}(\bar{\mathbf{x}}, \bar{y})}{\frac{\partial f}{\partial y}(\bar{\mathbf{x}}, \bar{y})}$$

*Bevis:* La oss begynne med det som kanskje ser vanskeligst ut, men som faktisk er lettest, nemlig formelen for  $\frac{\partial g}{\partial x_i}(\bar{\mathbf{x}})$ . Anta derfor at vi har greid å vise at det finnes en deriverbar funksjon  $g$  slik at

$$f(\mathbf{x}, g(\mathbf{x})) = 0$$

i en omegn om  $\bar{\mathbf{x}}$ . Partiellderiverer vi begge sider mhp.  $x_i$ , får vi ifølge kjerneregelen

$$\frac{\partial f}{\partial x_i}(\mathbf{x}, g(\mathbf{x})) + \frac{\partial f}{\partial y}(\mathbf{x}, g(\mathbf{x})) \frac{\partial g}{\partial x_i}(\mathbf{x}) = 0$$

Setter vi inn  $\mathbf{x} = \bar{\mathbf{x}}$ ,  $\bar{y} = g(\bar{\mathbf{x}})$  og bruker at  $\frac{\partial f}{\partial y}(\bar{\mathbf{x}}, \bar{y}) \neq 0$ , får vi

$$\frac{\partial g}{\partial x_i}(\bar{\mathbf{x}}) = -\frac{\frac{\partial f}{\partial x_i}(\bar{\mathbf{x}}, \bar{y})}{\frac{\partial f}{\partial y}(\bar{\mathbf{x}}, \bar{y})}$$

som er den formelen vi skulle vise.

For å vise at funksjonen  $g$  virkelig finnes og er deriverbar, bruker vi et triks — vi anvender omvendt funksjonsteorem på funksjonen  $\mathbf{F} : \mathbb{R}^{m+1} \rightarrow \mathbb{R}^{m+1}$  gitt ved

$$\mathbf{F}(x_1, x_2, \dots, x_m, y) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ f(x_1, x_2, \dots, x_m, y) \end{pmatrix}$$

Denne funksjonen har Jacobi-matrise

$$\mathbf{F}'(x_1, x_2, \dots, x_m, y) = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ \frac{\partial f}{\partial x_1}(\mathbf{x}, y) & \frac{\partial f}{\partial x_2}(\mathbf{x}, y) & \dots & \frac{\partial f}{\partial x_m}(\mathbf{x}, y) & \frac{\partial f}{\partial y}(\mathbf{x}, y) \end{pmatrix}$$

Regner vi ut determinanten til Jacobi-matrissen, får vi

$$\det(\mathbf{F}'(x_1, x_2, \dots, x_m, y)) = \frac{\partial f}{\partial y}(\mathbf{x}, y)$$

Setter vi inn  $(\bar{\mathbf{x}}, \bar{y})$ , ser vi at

$$\det(\mathbf{F}'(\bar{\mathbf{x}}, \bar{y})) = \frac{\partial f}{\partial y}(\bar{\mathbf{x}}, \bar{y}) \neq 0$$

ifølge antagelsen i teoremet. Jacobi-matrissen  $\mathbf{F}'(\bar{\mathbf{x}}, \bar{y})$  er dermed inverterbar, og vi kan bruke omvendt funksjonsteorem på  $\mathbf{F}$ : Restriktører vi  $\mathbf{F}$  til en tilstrekkelig liten omegn om  $(\bar{\mathbf{x}}, \bar{y})$ , har den en omvendt funksjon  $\mathbf{G}$  definert på en åpen mengde  $V \subset \mathbb{R}^{m+1}$  som inneholder  $\mathbf{F}(\bar{\mathbf{x}}, \bar{y}) = (\bar{\mathbf{x}}, f(\bar{\mathbf{x}}, \bar{y})) = (\bar{\mathbf{x}}, 0)$ . På grunn av den spesielle formen til  $\mathbf{F}$ , er det lett å se at  $\mathbf{G}$  må ha formen

$$\mathbf{G}(x_1, x_2, \dots, x_m, z) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ h(x_1, x_2, \dots, x_m, z) \end{pmatrix}$$

der  $h$  er en deriverbar funksjon. Siden  $G$  er den omvendte funksjonen til  $\mathbf{F}$ , har vi videre

$$\mathbf{F}(x_1, x_2, \dots, x_m, h(x_1, x_2, \dots, x_m, z)) = \mathbf{F}(\mathbf{G}(x_1, x_2, \dots, x_m, z)) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ z \end{pmatrix}$$

På den annen side vet vi fra definisjonen av  $\mathbf{F}$  at

$$\mathbf{F}(x_1, x_2, \dots, x_m, h(x_1, x_2, \dots, x_m, z)) =$$

$$= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \\ f(x_1, x_2, \dots, x_m, h(x_1, x_2, \dots, x_m, z)) \end{pmatrix}$$

Sammenligner vi disse to formlene, ser vi at

$$z = f(x_1, x_2, \dots, x_m, h(x_1, x_2, \dots, x_m, z))$$

Setter vi inn  $z = 0$ , har vi dermed

$$f(x_1, x_2, \dots, x_m, h(x_1, x_2, \dots, x_m, 0)) = 0$$

Vi lar nå  $U_0 = \{(x_1, x_2, \dots, x_m) \mid (x_1, x_2, \dots, x_m, 0) \in V\}$ , og definerer  $g : U_0 \rightarrow \mathbb{R}$  ved  $g(x_1, x_2, \dots, x_m) = h(x_1, x_2, \dots, x_m, 0)$ . Det følger fra utledningene ovenfor at  $g(\bar{\mathbf{x}}) = \bar{y}$ , at  $f(\mathbf{x}, g(\mathbf{x})) = 0$  for alle  $\mathbf{x} \in U_0$ , og at  $g$  er deriverbar i  $\bar{\mathbf{y}}$  med

$$\frac{\partial g}{\partial x_i}(\bar{\mathbf{x}}) = -\frac{\frac{\partial f}{\partial x_i}(\bar{\mathbf{x}}, \bar{y})}{\frac{\partial f}{\partial y}(\bar{\mathbf{x}}, \bar{y})}$$

Det gjenstår én liten detalj. Teoremet påstår at  $g$  er deriverbar i hele omegnen  $U_0$ , men vi har bare vist deriverbarhet i punktet  $\bar{\mathbf{x}}$ . Dette er lett å fikse med et lite triks. Vi bruker argumentene ovenfor på nytt, men erstatter punktet  $(\bar{\mathbf{x}}, \bar{y})$  med et fritt valgt punkt  $(\mathbf{x}, g(\mathbf{x}))$  der  $\mathbf{x} \in U_0$  (det kan hende vi må innsnevre  $U_0$  noe for å være sikre på at  $\frac{\partial f}{\partial y}(\mathbf{x}, g(\mathbf{x})) \neq 0$ ).  $\square$

**Eksempel 2:** La

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

Vi skal vise at det finnes en funksjon  $g(x)$  definert i en omegn om 0 slik at  $g(0) = 0$  og  $f(x, g(x)) = 0$ . Vi skal også regne ut  $g'(0)$ .

Vi observerer først at  $f(0, 0) = e^{0+0} + 0 - 1 = 0$ . Videre er  $\frac{\partial f}{\partial y}(x, y) = e^{x+y} + 1$ , så  $\frac{\partial f}{\partial y}(0, 0) = e^{0+0} + 1 = 2 \neq 0$ . Dette betyr at det finnes en (implisitt definert) funksjon  $g$  slik at  $g(0) = 0$  og  $f(x, g(x)) = 0$  i en omegn om 0. Den deriverte til  $g$  er gitt ved

$$g'(0) = -\frac{\frac{\partial f}{\partial x}(0, 0)}{\frac{\partial f}{\partial y}(0, 0)} = -\frac{e^{0+0}}{e^{0+0} + 1} = -\frac{1}{2}$$



Ofte kan det være lønnsomt å bruke implisitt derivasjon også i situasjoner der vi faktisk *kan* finne et uttrykk for den implisitt gitte funksjonen:

**Eksempel 3:** Vi skal finne stigningstallene i  $x$ - og  $y$ -retning i et punkt  $(x, y, z)$  på kuleflaten

$$x^2 + y^2 + z^2 = R^2$$

Istedentfor å referere eksplisitt til teoremet ovenfor, skal vi bruke implisitt derivasjon slik det som regel gjøres i praksis. Vi tenker på  $z$  som en funksjon

av  $x$  og  $y$  og deriverer ligningen ved hjelp av kjerneregelen. Deriverer vi mhp.  $x$ , får vi

$$2x + 2z \frac{\partial z}{\partial x} = 0$$

og deriverer vi mhp.  $y$ , får vi

$$2y + 2z \frac{\partial z}{\partial y} = 0$$

Dette gir  $\frac{\partial z}{\partial x} = -\frac{x}{z}$  og  $\frac{\partial z}{\partial y} = -\frac{y}{z}$  (forutsatt at  $z \neq 0$ ). Du kan selv bruke implisitt funksjonsteorem til å rettferdiggjøre disse regningene. ♣

Vi tar også med et eksempel som viser hvordan implisitt derivasjon dukker opp i andre fag.

**Eksempel 4:** Når man studerer gasser, er det tre naturlige variable; trykket  $p$ , temperaturen  $T$  og volumet  $V$ . I gassmodeller er de knyttet sammen gjennom en ligning  $f(p, V, T) = 0$ . Funksjonen  $f$  varierer fra en modell til en annen — i teorien for såkalte *ideelle gasser* er f.eks.  $f(p, v, T) = pV - kT$  der  $k$  er en konstant. Siden variablene er knyttet sammen gjennom en ligning, tenker man ofte på hver av disse størrelsene  $p$ ,  $V$  og  $T$  som en funksjon av de to andre. Tenker vi f.eks. på trykket som en funksjon av volumet og temperaturen, har vi en funksjon  $p(V, T)$  som oppfyller ligningen

$$f(p(V, T), V, T) = 0$$

altså en implisitt gitt funksjon. Vi kan finne uttrykk for de partiellderiverte til  $p$  ved å derivere implisitt. Deriverer vi først mhp.  $V$ , får vi

$$\frac{\partial f}{\partial p} \frac{\partial p}{\partial V} + \frac{\partial f}{\partial V} = 0,$$

mens derivasjon mhp.  $T$  gir

$$\frac{\partial f}{\partial p} \frac{\partial p}{\partial T} + \frac{\partial f}{\partial T} = 0,$$

Løser vi for de partiellderiverte til  $p$ , ser vi at

$$\frac{\partial p}{\partial V} = -\frac{\frac{\partial f}{\partial V}}{\frac{\partial f}{\partial p}} \quad \text{og} \quad \frac{\partial p}{\partial T} = -\frac{\frac{\partial f}{\partial T}}{\frac{\partial f}{\partial p}}$$

Regninger av denne typen (gjerne litt mer innfløkt enn disse) spiller en sentral rolle i mange anvendelser. Som oftest argumenterer man uformelt slik som her; man deriverer ved bruk av kjerneregelen uten å bry seg for mye om deriverbarhet og andre matematiske finurligheter. Ønsker man bedre matematiske begrunnelser, er implisitt funksjonsteorem det riktige redskapet. ♣

Det finnes også en vektorvaluert versjon av implisitt funksjonsteorem. Her tenker vi oss at vi skal finne  $k$  variable  $y_1, y_2, \dots, y_k$  uttrykt ved  $m$  andre variable  $x_1, x_2, \dots, x_m$ . Siden vi har  $k$  ukjente, trenger vi også  $k$  ligninger, så vi starter med en funksjon  $\mathbf{F} : \mathbb{R}^{m+k} \rightarrow \mathbb{R}^k$  og ønsker å løse ligningssystemet

$$\mathbf{F}(x_1, \dots, x_m, y_1, \dots, y_k) = \mathbf{0}$$

for  $y_1, \dots, y_k$  (legg merke til at ligningssystemet har  $k$  ligninger og  $k$  ukjente). Sagt med andre ord ønsker vi å finne funksjoner  $g_1(x_1, \dots, x_m), \dots, g_k(x_1, \dots, x_m)$  slik at

$$\mathbf{F}(x_1, \dots, x_m, g_1(x_1, \dots, x_m), \dots, g_k(x_1, \dots, x_m)) = \mathbf{0}$$

Disse uttrykkene blir enklere om vi bruker vektornotasjon. Skriver vi  $\mathbf{x} = (x_1, \dots, x_m)$ ,  $\mathbf{y} = (y_1, \dots, y_k)$  og  $\mathbf{G}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_k(\mathbf{x}))$ , kan formelen ovenfor skrives

$$\mathbf{F}(\mathbf{x}, \mathbf{G}(\mathbf{x})) = \mathbf{0}$$

Før vi presenterer den vektorvaluerte versjonen av implisitt funksjonssteorem, trenger vi litt notasjon. Anta at vi har en funksjon  $\mathbf{F} : \mathbb{R}^{m+k} \rightarrow \mathbb{R}^k$  som ovenfor, og at vi har delt inn variablene i to grupper  $\mathbf{x} = (x_1, \dots, x_m)$  og  $\mathbf{y} = (y_1, \dots, y_k)$ . Vi lar  $\frac{\partial \mathbf{F}}{\partial \mathbf{x}}$  være den Jacobi-matrisen vi får når vi bare tenker på  $\mathbf{F}$  som en funksjon av  $x$ -variablene, dvs.

$$\frac{\partial \mathbf{F}}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \cdots & \frac{\partial F_1}{\partial x_m} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \cdots & \frac{\partial F_2}{\partial x_m} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_k}{\partial x_1} & \frac{\partial F_k}{\partial x_2} & \cdots & \frac{\partial F_k}{\partial x_m} \end{pmatrix}$$

Tilsvarende lar vi  $\frac{\partial \mathbf{F}}{\partial \mathbf{y}}$  være den Jacobi-matrisen vi får, når vi bare tenker på  $\mathbf{F}$  som en funksjon av  $y$ -variablene, dvs.

$$\frac{\partial \mathbf{F}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_k} \\ \frac{\partial F_2}{\partial y_1} & \frac{\partial F_2}{\partial y_2} & \cdots & \frac{\partial F_2}{\partial y_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_k}{\partial y_1} & \frac{\partial F_k}{\partial y_2} & \cdots & \frac{\partial F_k}{\partial y_k} \end{pmatrix}$$

**Teorem 5.7.4 (Vektorvaluert versjon av implisitt funksjonsteorem)**  
Anta at  $U$  er en åpen delmengde av  $\mathbb{R}^{m+k}$  og la  $\mathbf{F} : U \rightarrow \mathbb{R}^k$  være en funksjon med kontinuerlige partiellderiverte. Anta at  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  er et punkt i  $U$  der

$\mathbf{F}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) = \mathbf{0}$ . Anta videre at  $k \times k$ -matrisen  $\frac{\partial \mathbf{F}}{\partial \mathbf{y}}(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  er inverterbar. Da finnes det en omegn  $U_0$  om  $\bar{\mathbf{x}}$ , og en deriverbar funksjon  $\mathbf{G} : U_0 \rightarrow \mathbb{R}^k$  slik at  $\mathbf{g}(\bar{\mathbf{x}}) = \bar{\mathbf{y}}$  og

$$\mathbf{F}(\mathbf{x}, \mathbf{G}(\mathbf{x})) = \mathbf{0}$$

for alle  $\mathbf{x} \in U_0$ . Jacobi-matrisen til  $\mathbf{G}$  er gitt ved

$$\mathbf{G}'(\bar{\mathbf{x}}) = - \left( \frac{\partial \mathbf{F}}{\partial \mathbf{y}}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \right)^{-1} \left( \frac{\partial \mathbf{F}}{\partial \mathbf{x}}(\bar{\mathbf{x}}, \bar{\mathbf{y}}) \right)$$

Vi overlater beviset til leserne. Ideen er akkurat den samme som i det skalarvaluerte tilfellet, men det er litt flere partiellderiverte å holde styr på.

**Bemerkning:** Som du sikkert har oppdaget allerede, kan vi oppfatte implisitt funksjonsteorem som et teorem om ligningssystemer — vi ønsker å løse ligningssystemet  $\mathbf{F}(x_1, \dots, x_m, y_1, \dots, y_k) = \mathbf{0}$  med hensyn på variablene  $y_1, \dots, y_k$ , og teoremet gir oss informasjon om hvordan disse løsningene avhenger av de øvrige variablene  $x_1, \dots, x_m$ .

### \*Bevis for omvendt funksjonsteorem

Vi skal dele beviset opp i lemmaer, og det er ikke så lett å se hvordan alt passer sammen før du har vært gjennom hele resonnementet. Det første lemmaet er enkelt, men inneholder det grunnleggende eksistensresultatet (det som sikrer at en omvendt funksjon finnes). Utgangspunktet for lemmaet er en selvfølgelighet, nemlig at identitetsavbildningen  $I(\mathbf{x}) = \mathbf{x}$  avbilder kulen  $\overline{B}(\mathbf{x}, r)$  injektivt på  $\overline{B}(\mathbf{x}, r)$ . Lemmaet viser at en avbildning som ikke avviker for mye fra  $I$ , har en lignende egenskap. Husk at

$$\overline{B}(\mathbf{a}, r) = \{\mathbf{x} \in \mathbb{R}^m : |\mathbf{x} - \mathbf{a}| \leq r\}$$

betegner en *lukket* kule i  $\mathbb{R}^m$ .

**Lemma 5.7.5 (Perturbasjonslemma)** La  $\overline{B}(\mathbf{0}, r)$  være en lukket kule i  $\mathbb{R}^m$ , og anta at funksjonen  $\mathbf{H} : \overline{B}(\mathbf{0}, r) \rightarrow \mathbb{R}^m$  er slik at  $\mathbf{H}(\mathbf{0}) = \mathbf{0}$  og

$$|\mathbf{H}(\mathbf{u}) - \mathbf{H}(\mathbf{v})| \leq \frac{1}{2}|\mathbf{u} - \mathbf{v}| \quad \text{for alle } \mathbf{u}, \mathbf{v} \in \overline{B}(\mathbf{0}, r)$$

Da er funksjonen  $\mathbf{L} : \overline{B}(\mathbf{0}, r) \rightarrow \mathbb{R}^m$  definert ved  $\mathbf{L}(\mathbf{x}) = \mathbf{x} + \mathbf{H}(\mathbf{x})$  injektiv, og kulen  $\overline{B}(\mathbf{0}, \frac{r}{2})$  er inneholdt i verdimengden til  $\mathbf{L}$ .

*Bevis:* La oss først vise at  $\mathbf{L}$  er injektiv. Vi antar at  $\mathbf{L}(\mathbf{x}) = \mathbf{L}(\mathbf{y})$ , og må vise at  $\mathbf{x} = \mathbf{y}$ . Per definisjon av  $\mathbf{L}$  er

$$\mathbf{x} + \mathbf{H}(\mathbf{x}) = \mathbf{y} + \mathbf{H}(\mathbf{y})$$

dvs.

$$\mathbf{x} - \mathbf{y} = \mathbf{H}(\mathbf{y}) - \mathbf{H}(\mathbf{x})$$

som gir

$$|\mathbf{x} - \mathbf{y}| = |\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{y})|$$

Ifølge antagelsene er  $|\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{y})| \leq \frac{1}{2}|\mathbf{x} - \mathbf{y}|$ , så likheten ovenfor er bare mulig hvis  $|\mathbf{x} - \mathbf{y}| = 0$ , dvs. hvis  $\mathbf{x} = \mathbf{y}$ .

Det gjenstår å vise at  $\overline{B}(\mathbf{0}, \frac{r}{2})$  er inneholdt i verdimengden til  $\mathbf{L}$ . Vi må da vise at for alle  $\mathbf{y} \in \overline{B}(\mathbf{0}, \frac{r}{2})$ , har ligningen  $\mathbf{L}(\mathbf{x}) = \mathbf{y}$  en løsning i  $\overline{B}(\mathbf{0}, r)$ . Denne ligningen kan skrives

$$\mathbf{x} = \mathbf{y} - \mathbf{H}(\mathbf{x}),$$

så det er nok å vise at funksjonen  $\mathbf{K}(\mathbf{x}) = \mathbf{y} - \mathbf{H}(\mathbf{x})$  har et fikspunkt i  $\overline{B}(\mathbf{0}, r)$ . Dette vil følge av Banachs fikspunktteorem dersom vi kan vise at  $\mathbf{K}$  er en kontraksjon av  $\overline{B}(\mathbf{0}, r)$ . La oss først vise at  $\mathbf{K}$  avbilder  $\overline{B}(\mathbf{0}, r)$  inn i  $\overline{B}(\mathbf{0}, r)$ . Det følger av at

$$|\mathbf{K}(\mathbf{x})| = |\mathbf{y} - \mathbf{H}(\mathbf{x})| \leq |\mathbf{y}| + |\mathbf{H}(\mathbf{x})| \leq \frac{r}{2} + \frac{r}{2} = r$$

der vi har brukt at ifølge betingelsene på  $\mathbf{H}$  er

$$|\mathbf{H}(\mathbf{x})| = |\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{0})| \leq \frac{1}{2}|\mathbf{x} - \mathbf{0}| \leq \frac{r}{2}$$

Til slutt sjekker vi kontraksjonsbetingelsen:

$$|\mathbf{K}(\mathbf{u}) - \mathbf{K}(\mathbf{v})| = |\mathbf{H}(\mathbf{u}) - \mathbf{H}(\mathbf{v})| \leq \frac{1}{2}|\mathbf{u} - \mathbf{v}|$$

Dermed har vi vist at  $\mathbf{K}$  er en kontraksjon, og følgelig har den et entydig fikspunkt i  $\overline{B}(\mathbf{0}, r)$ .  $\square$

I det neste lemmaet skal vi vise at omvendt funksjonsteoremet gjelder for funksjoner  $\mathbf{L}$  slik at  $\mathbf{L}(\mathbf{0}) = \mathbf{0}$  og  $\mathbf{L}'(\mathbf{0}) = I_m$ . Dette kan høres veldig spesielt ut, men det viser seg at det generelle tilfellet følger ved et enkelt variabelskifte.

**Lemma 5.7.6** *Anta  $U$  er et område i  $\mathbb{R}^m$  som inneholder  $\mathbf{0}$  og at  $\mathbf{L} : U \rightarrow \mathbb{R}^m$  er en funksjon med kontinuerlige partiellderiverte slik at  $\mathbf{L}(\mathbf{0}) = \mathbf{0}$  og  $\mathbf{L}'(\mathbf{0}) = I_m$ . Da finnes det en  $r > 0$  slik at  $\mathbf{L}$  er injektiv når den restrikteres til  $\overline{B}(\mathbf{0}, r)$  og har en omvendt funksjon  $\mathbf{M}$  definert på et område som inneholder  $\overline{B}(\mathbf{0}, \frac{r}{2})$ . Den omvendte funksjonen  $\mathbf{M}$  er deriverbar i  $\mathbf{0}$  og har Jacobi-matrise  $\mathbf{M}'(\mathbf{0}) = I_m$ .*

*Bevis:* La  $\mathbf{H}(\mathbf{x}) = \mathbf{L}(\mathbf{x}) - \mathbf{x}$ . Vi skal først vise at  $\mathbf{H}$  tilfredsstiller betingelsene i lemmaet over. Observer først at siden  $\mathbf{L}(\mathbf{0}) = \mathbf{0}$  og  $\mathbf{L}'(\mathbf{0}) = I_m$ , så er  $\mathbf{H}(\mathbf{0}) = \mathbf{0}$  og alle de partiellderiverte  $\frac{\partial H_i}{\partial x_j}(\mathbf{0})$  er lik 0. Spesielt er  $\nabla H_i(\mathbf{0}) = \mathbf{0}$  for alle  $i$ . Bruker vi middelverdisetningen for funksjoner av flere variable (setning 3.2.3 alias setning 5.5.5) på  $H_i$ , får vi dermed

$$H_i(\mathbf{x}) - H_i(\mathbf{y}) = \nabla H_i(\mathbf{c}_i) \cdot (\mathbf{x} - \mathbf{y}) = (\nabla H_i(\mathbf{c}_i) - \nabla H_i(\mathbf{0})) \cdot (\mathbf{x} - \mathbf{y})$$

Siden de partiellderiverte til  $\mathbf{H}$  er kontinuerlige, kan vi få  $(\nabla H_i(\mathbf{c}_i) - \nabla H_i(\mathbf{0}))$  så liten vi måtte ønske ved å velge  $\mathbf{x}$  og  $\mathbf{y}$  tilstrekkelig nær  $\mathbf{0}$ . Spesielt finnes det en  $r > 0$  slik at hvis  $|\mathbf{x}|, |\mathbf{y}| \leq r$ , så er  $|\nabla H_i(\mathbf{c}_i) - \nabla H_i(\mathbf{0})| \leq \frac{1}{2\sqrt{m}}$  for alle  $i$ . Bruker vi Schwarz' ulikhet, ser vi at

$$|H_i(\mathbf{x}) - H_i(\mathbf{y})| \leq |\nabla H_i(\mathbf{c}_i) - \nabla H_i(\mathbf{0})| |\mathbf{x} - \mathbf{y}| \leq \frac{1}{2\sqrt{m}} |\mathbf{x} - \mathbf{y}|$$

og

$$\begin{aligned} |\mathbf{H}(\mathbf{x}) - \mathbf{H}(\mathbf{y})| &= \sqrt{(H_1(\mathbf{x}) - H_1(\mathbf{y}))^2 + \dots + (H_m(\mathbf{x}) - H_m(\mathbf{y}))^2} \leq \\ &\leq \sqrt{m \left( \frac{1}{2\sqrt{m}} |\mathbf{x} - \mathbf{y}| \right)^2} = \frac{1}{2} |\mathbf{x} - \mathbf{y}| \end{aligned}$$

Dermed har vi vist at  $\mathbf{H}$  tilfredsstiller betingelsene i forrige lemma, og siden

$$\mathbf{L}(\mathbf{x}) = \mathbf{x} + \mathbf{H}(\mathbf{x}),$$

vet vi fra lemmaet at  $\mathbf{L}$  restriktert til  $\overline{B}(\mathbf{0}, r)$  er injektiv og at verdimengden inneholder  $\overline{B}(\mathbf{0}, \frac{r}{2})$ . Dette betyr at  $\mathbf{L}$  (restriktert til  $\overline{B}(\mathbf{0}, r)$ ) har en omvendt funksjon  $\mathbf{M}$  som er definert på et område som omfatter  $\overline{B}(\mathbf{0}, \frac{r}{2})$ .

Det gjenstår å vise at  $\mathbf{M}$  er deriverbar i  $\mathbf{0}$  med Jacobi-matrise  $I_m$ , men før vi går løs på deriverbarheten, trenger vi et lite estimat. Ifølge trekantulikheten har vi

$$|\mathbf{x}| = |\mathbf{L}(\mathbf{x}) - \mathbf{H}(\mathbf{x})| \leq |\mathbf{L}(\mathbf{x})| + |\mathbf{H}(\mathbf{x})| \leq |\mathbf{L}(\mathbf{x})| + \frac{1}{2} |\mathbf{x}|$$

som gir

$$\frac{1}{2} |\mathbf{x}| \leq |\mathbf{L}(\mathbf{x})|$$

når vi flytter over.

Vi er nå klar til å vise at den omvendte funksjonen  $\mathbf{M}$  til  $\mathbf{L}$  er deriverbar i  $\mathbf{0}$  med Jacobi-matrise  $I_m$ . Ifølge Setning 2.6.4 er det nok å vise at

$$\lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{M}(\mathbf{y}) - \mathbf{M}(\mathbf{0}) - I_m \mathbf{y}}{|\mathbf{y}|} = \mathbf{0}$$

Siden  $\mathbf{M}(\mathbf{0}) = \mathbf{0}$  og  $I_m \mathbf{y} = \mathbf{y}$ , er dette det samme som

$$\lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{M}(\mathbf{y}) - \mathbf{y}}{|\mathbf{y}|} = \mathbf{0}$$

Siden vi er interessert i grensen når  $|\mathbf{y}| \rightarrow 0$ , kan vi nøye oss med å betrakte  $\mathbf{y} \in \overline{B}(\mathbf{0}, \frac{r}{2})$ . For hver slik  $\mathbf{y}$  vet vi at det finnes en entydig bestemt  $\mathbf{x}$  i  $\overline{B}(\mathbf{0}, r)$  slik at  $\mathbf{y} = \mathbf{L}(\mathbf{x})$  og  $\mathbf{x} = \mathbf{M}(\mathbf{y})$ . Setter vi dette inn i uttrykket ovenfor, får vi

$$\lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{M}(\mathbf{y}) - \mathbf{y}}{|\mathbf{y}|} = \lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{x} - \mathbf{L}(\mathbf{x})}{|\mathbf{y}|} = - \lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{L}(\mathbf{x}) - I_m \mathbf{x}}{|\mathbf{x}|} \cdot \frac{|\mathbf{x}|}{|\mathbf{y}|}$$

Siden  $\frac{1}{2}|\mathbf{x}| \leq |\mathbf{L}(\mathbf{x})| = |\mathbf{y}|$ , vil  $|\mathbf{x}|$  gå mot null når  $|\mathbf{y}|$  går mot null. Siden  $\mathbf{L}$  er deriverbar i  $\mathbf{0}$  med Jacobi-matrise  $I_m$ , er derfor

$$\lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{L}(\mathbf{x}) - I_m \mathbf{x}}{|\mathbf{x}|} = \lim_{|\mathbf{x}| \rightarrow 0} \frac{\mathbf{L}(\mathbf{x}) - \mathbf{L}(\mathbf{0}) - I_m \mathbf{x}}{|\mathbf{x}|} = \mathbf{0}$$

Siden faktoren  $\frac{|\mathbf{x}|}{|\mathbf{y}|} = \frac{|\mathbf{x}|}{|\mathbf{L}(\mathbf{x})|} \leq \frac{|\mathbf{x}|}{\frac{1}{2}|\mathbf{x}|} = 2$  er begrenset, følger det at

$$\lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{M}(\mathbf{y}) - \mathbf{y}}{|\mathbf{y}|} = - \lim_{|\mathbf{y}| \rightarrow 0} \frac{\mathbf{L}(\mathbf{x}) - I_m \mathbf{x}}{|\mathbf{x}|} \cdot \frac{|\mathbf{x}|}{|\mathbf{y}|} = \mathbf{0}$$

Dermed er lemmaet bevist.  $\square$

Vi kan nå bevise omvendt funksjonsteorem. For at du skal slippe å bla for mye, skriver vi det opp på nytt.

**Teorem 5.7.7 (Omvendt funksjonsteorem)** *Anta at  $U$  er en åpen mengde i  $\mathbb{R}^m$  og at  $\mathbf{F} : U \rightarrow \mathbb{R}^m$  har kontinuerlige partiellderiverte. Anta at  $\bar{\mathbf{x}} \in U$  og at Jacobi-matrisen  $\mathbf{F}'(\bar{\mathbf{x}})$  er inverterbar. Da finnes det en omegn  $U_0 \subset U$  om  $\bar{\mathbf{x}}$  slik at  $\mathbf{F}$  restriktert til  $U_0$  er injektiv. Verdimengden  $V$  til denne restriksjonen er en omegn om  $\bar{\mathbf{y}} = \mathbf{F}(\bar{\mathbf{x}})$ , og den omvendte funksjonen  $\mathbf{G} : V \rightarrow U_0$  er deriverbar i  $\bar{\mathbf{y}}$  med Jacobi-matrise*

$$\mathbf{G}'(\bar{\mathbf{y}}) = \mathbf{F}'(\bar{\mathbf{x}})^{-1}$$

*Bevis:* Planen er å omdanne  $\mathbf{F}$  til en funksjon  $\mathbf{L}$  som tilfredsstiller betingelsene i foregående lemma. Denne funksjonen  $\mathbf{L}$  har da en omvendt funksjon  $\mathbf{M}$  som vi kan omdanne til en omvendt funksjon  $\mathbf{G}$  for  $\mathbf{F}$ . Når vi har funnet  $\mathbf{G}$ , er det lett å sjekke at den har de egenskapene som teoremet angir.

Vi begynner med å definere funksjonen  $\mathbf{L}$  ved

$$\mathbf{L}(\mathbf{x}) = A(\mathbf{F}(\mathbf{x} + \bar{\mathbf{x}}) - \bar{\mathbf{y}})$$

der  $A = \mathbf{F}'(\bar{\mathbf{x}})^{-1}$ . Siden  $\mathbf{F}$  er definert i en omegn rundt  $\bar{\mathbf{x}}$ , er  $\mathbf{L}$  definert i en omegn rundt  $\mathbf{0}$ . Vi ser også at

$$\mathbf{L}(\mathbf{0}) = A(\mathbf{F}(\bar{\mathbf{x}}) - \bar{\mathbf{y}}) = \mathbf{0}$$

siden  $\mathbf{F}(\bar{\mathbf{x}}) = \bar{\mathbf{y}}$ . Videre gir kjerneregelen

$$\mathbf{L}'(\mathbf{x}) = A\mathbf{F}'(\mathbf{x} + \bar{\mathbf{x}}),$$

så

$$\mathbf{L}'(\mathbf{0}) = A\mathbf{F}'(\bar{\mathbf{x}}) = I_m$$

siden  $A = \mathbf{F}'(\bar{\mathbf{x}})^{-1}$ .

Dette betyr at  $\mathbf{L}$  oppfyller betingelsene i lemmaet ovenfor, og at  $\mathbf{L}$  restriktert til en kule  $\bar{B}(\mathbf{0}, r)$  har en invers funksjon  $\mathbf{M}$  definert på en mengde som inneholder  $\bar{B}(\mathbf{0}, \frac{r}{2})$ . For å finne en invers funksjon til  $\mathbf{F}$  observerer vi at dersom vi snur litt på ligningen  $\mathbf{L}(\mathbf{x}) = A(\mathbf{F}(\mathbf{x} + \bar{\mathbf{x}}) - \bar{\mathbf{y}})$ , får vi

$$\mathbf{F}(\mathbf{x}) = A^{-1}\mathbf{L}(\mathbf{x} - \bar{\mathbf{x}}) + \bar{\mathbf{y}}$$

for alle  $\mathbf{x} \in \bar{B}(\bar{\mathbf{x}}, r)$ . Siden  $\mathbf{L}$  er injektiv og  $A^{-1}$  er inverterbar, følger det at  $\mathbf{F}$  er injektiv på  $\bar{B}(\bar{\mathbf{x}}, r)$ . For å finne den omvendte funksjonen, løser vi ligningen

$$\mathbf{y} = A^{-1}\mathbf{L}(\mathbf{x} - \bar{\mathbf{x}}) + \bar{\mathbf{y}}$$

med hensyn på  $\mathbf{y}$  og får

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{M}(A(\mathbf{y} - \bar{\mathbf{y}}))$$

Dette betyr at  $\mathbf{F}$  restriktert til  $\mathbf{x} \in \bar{B}(\bar{\mathbf{x}}, r)$  har en invers funksjon  $\mathbf{G}$  definert ved

$$\mathbf{G}(\mathbf{y}) = \bar{\mathbf{x}} + \mathbf{M}(A(\mathbf{y} - \bar{\mathbf{y}}))$$

Siden definisjonsmengden til  $\mathbf{M}$  omfatter hele  $\bar{B}(\mathbf{0}, \frac{r}{2})$ , vil definisjonsmengden til  $\mathbf{G}$  omfatte alle  $\mathbf{y}$  slik at  $|A(\mathbf{y} - \bar{\mathbf{y}})| \leq \frac{r}{2}$ . Siden  $|A(\mathbf{y} - \bar{\mathbf{y}})| \leq |A||\mathbf{y} - \bar{\mathbf{y}}|$ , inkluderer dette  $\bar{B}(\bar{\mathbf{y}}, \frac{r}{2|A|})$ , og følgelig er  $\mathbf{G}$  definert i en omegn om  $\bar{\mathbf{y}}$ .

Resten er lett. Siden  $\mathbf{M}$  er deriverbar og  $\mathbf{G}(\mathbf{y}) = \bar{\mathbf{x}} + \mathbf{M}(A(\mathbf{y} - \bar{\mathbf{y}}))$ , forteller kjerneregelen oss at  $\mathbf{G}$  er deriverbar med Jacobi-matrise

$$\mathbf{G}'(\mathbf{y}) = \mathbf{M}'(A(\mathbf{y} - \bar{\mathbf{y}}))A$$

Setter vi inn  $\mathbf{y} = \bar{\mathbf{y}}$  og bruker at  $\mathbf{M}'(\mathbf{0}) = I_m$ , får vi

$$\mathbf{G}'(\bar{\mathbf{y}}) = I_m A = \mathbf{F}'(\bar{\mathbf{x}})^{-1}$$

siden  $A$  per definisjon er lik  $\mathbf{F}'(\bar{\mathbf{x}})^{-1}$ . □

### Oppgaver til seksjon 5.7

**1.** Vis at funksjonen  $\mathbf{F}(x, y) = \begin{pmatrix} x^2 + y + 1 \\ x - y - 2 \end{pmatrix}$  har en omvendt funksjon  $\mathbf{G}$  definert i en omegn rundt  $(1, -2)$  slik at  $\mathbf{G}(1, -2) = (0, 0)$ . Finn den deriverte til  $\mathbf{G}$  i punktet  $(1, -2)$ . Vis at  $\mathbf{F}$  også har en omvendt funksjon  $\mathbf{H}$  definert i en omegn om  $(1, -2)$  slik at  $\mathbf{H}(1, -2) = (-1, -1)$ . Finn  $\mathbf{H}'(1, -2)$ .

**2.** Vis at funksjonen  $\mathbf{F}(x, y) = \begin{pmatrix} e^{x+y^2-1} \\ x-y \end{pmatrix}$  har en omvendt funksjon  $\mathbf{G}$  definert i en omegn rundt  $(1, -1)$  slik at  $\mathbf{G}(1, -1) = (0, 1)$ . Finn  $\mathbf{G}'(1, -1)$ . Vis at  $\mathbf{F}$  har en annen omvendt funksjon  $\mathbf{H}$  definert i en omegn om  $(1, -1)$  slik at  $\mathbf{H}(1, -1) = (-3, -2)$ . Finn  $\mathbf{H}'(1, -1)$

**3.** Vis at gjennom ethvert punkt  $(x_0, y_0)$  på kurven med ligning  $x^3 + y^3 + y = 1$  går det en funksjon  $y = f(x)$  som tilfredsstiller ligningen. Finn  $f'(x_0)$ .

**4.** La  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  være funksjonen  $f(x, y, z) = xy^2e^z + z$ . Vis at det finnes en funksjon  $g(x, y)$  definert i en omegn om  $(-1, 2, 0)$  slik at  $g(-1, 2) = 0$  og  $f(x, y, g(x, y)) = -4$ . Finn  $\frac{\partial g}{\partial x}(-1, 2)$  og  $\frac{\partial g}{\partial y}(-1, 2)$ .

**5.** (Eksamensoppgave fra MAT 1110 13/6, 2008)

a) Finn den inverse matrisen til

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & 0 & -2 \end{pmatrix}$$

b) Finn Jacobi-matrisen til funksjonen  $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  når

$$\mathbf{F}(x, y, z) = \begin{pmatrix} x+z \\ x^2 + \frac{1}{2}y^2 + z \\ x+z^2 \end{pmatrix}$$

Vis at  $\mathbf{F}$  har en omvendt funksjon  $\mathbf{G}$  definert i en omegn rundt  $(0, \frac{1}{2}, 2)$  slik at  $\mathbf{G}(0, \frac{1}{2}, 2) = (1, 1, -1)$ . Finn  $\mathbf{G}'(0, \frac{1}{2}, 2)$ .

**6.** Finn stigningstallet til tangenten til ellipsen  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$  i punktet  $(x_0, y_0)$ ,  $y_0 \neq 0$ .

**7.** Finn stigningstallet til tangenten til hyperbelen  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$  i punktet  $(x_0, y_0)$ ,  $y_0 \neq 0$ .

**8.** Finn stigningstallet til tangenten til parabelen  $y^2 = 4ax$  i punktet  $(x_0, y_0)$ ,  $y_0 \neq 0$ .

**9.** Når man løser differensialligninger, finner man ofte ut at løsningene tilfredsstiller en ligning av typen  $\phi(x, y(x)) = C$  der  $C$  er en konstant. Vis at  $y'(x) = -\frac{\frac{\partial \phi}{\partial x}(x, y(x))}{\frac{\partial \phi}{\partial y}(x, y(x))}$  forutsatt at de partiellederiverte eksisterer og  $\frac{\partial \phi}{\partial y}(x, y(x)) \neq 0$ .

**10.** En funksjon  $z(x, y)$  tilfredsstiller ligningen

$$x + y^2 + z^3 = 3xyz$$

Finn  $\frac{\partial z}{\partial x}$  og  $\frac{\partial z}{\partial y}$ .

**11.** En funksjon  $z(x, y)$  tilfredsstiller ligningen

$$2x^2 + 3y^2 + z^2 = e^{-z}$$

Finn  $\frac{\partial z}{\partial x}$  og  $\frac{\partial z}{\partial y}$ .

**12.** Vis at det finnes funksjoner  $u(x, y)$  og  $v(x, y)$  definert i et område rundt  $(2, -1)$  som tilfredsstiller ligningene

$$x^2 - y^2 - u^2 + v^2 = 0$$

$$2xy + y^2 - 2u^2 + 3v^4 + 8 = 0$$

og  $u(2, -1) = 2$ ,  $v(2, -1) = 1$ . Finn  $\frac{\partial u}{\partial x}(2, -1)$  og  $\frac{\partial v}{\partial x}(2, -1)$ .

**13.** I oppgaver om “koblede hastigheter” (*Kalkulus*, seksjon 7.2) møter vi denne situasjon: Vi vet hvor fort én størrelse  $y$  endrer seg (dvs. vi kjenner  $y'(t)$ ) og vi ønsker å regne ut hvor fort en annen størrelse  $x$  endrer seg (dvs. vi vil finne  $x'(t)$ ). De to størrelsene er bundet sammen med en ligning  $\phi(x(t), y(t)) = 0$ .

a) Vis at  $x'(t) = -\frac{\frac{\partial \phi}{\partial y}(x(t), y(t))}{\frac{\partial \phi}{\partial x}(x(t), y(t))}y'(t)$ . Hvilke forutsetninger har du gjort i regnestykket ditt?

b) I noen oppgaver kjenner vi *to* hastigheter  $y'(t)$  og  $z'(t)$ , og har en ligning  $\phi(x(t), y(t), z(t)) = 0$ . Finn et uttrykk for  $x'(t)$  i dette tilfellet.

**14.** Anta at  $\phi(x, y, z)$  er en deriverbar funksjon og at det finnes deriverbare funksjoner  $X(y, z)$ ,  $Y(x, z)$  og  $Z(y, z)$  slik at

$$\phi(X(y, z), y, z) = 0 \quad \phi(x, Y(x, z), z) = 0 \quad \text{og} \quad \phi(x, y, Z(x, y)) = 0$$

Vis at vi har (under passende betingelser)

$$\frac{\partial X}{\partial y} \cdot \frac{\partial Y}{\partial z} \cdot \frac{\partial Z}{\partial x} = -1$$

Denne relasjonen skrives ofte med små bokstaver:

$$\frac{\partial x}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial x} = -1$$

og er da til skrek og advarsel for folk som liker å forkorte  $\partial x$ ,  $\partial y$  og  $\partial z$ .

## 5.8 Ekstremalverdisetningen

I resten av kapitlet skal vi konsentrere oss om maksimums- og minimumsproblemer for funksjoner av flere variable. Fra teorien for funksjoner av én

variabel husker vi *ekstremalverdisetningen* (se *Kalkulus*, seksjon 5.3) som sier at en kontinuerlig funksjon definert på et lukket, begrenset intervall er begrenset og har maksimums- og minimumspunkter. I denne seksjonen skal vi bevise et tilsvarende resultat for kontinuerlige funksjoner av flere variable definert på lukkede, begrensede mengder. Vi begynner med noen definisjoner.

**Definisjon 5.8.1** *Anta  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $m$  variable. Vi sier at  $f$  er begrenset dersom det finnes tall  $K, M$  slik at*

$$K \leq f(\mathbf{x}) \leq M \quad \text{for alle } \mathbf{x} \in A$$

*Vi sier at  $\mathbf{c} \in A$  er et (globalt) maksimumspunkt for  $f$  dersom*

$$f(\mathbf{c}) \geq f(\mathbf{x}) \quad \text{for alle } \mathbf{x} \in A$$

*og vi sier at  $\mathbf{d} \in A$  er et (globalt) minimumspunkt for  $f$  dersom*

$$f(\mathbf{d}) \leq f(\mathbf{x}) \quad \text{for alle } \mathbf{x} \in A$$

Dersom en funksjon har maksimums- og minimumspunkter, er den åpenbart begrenset, men det finnes mange eksempler på funksjoner som er begrenset, men ikke har maksimums- og/eller minimumspunkter. Det er heller ikke uvanlig at en funksjon er ubegrenset til tross for at den er definert på en begrenset mengde. For kontinuerlige funksjoner definert på lukkede, begrensede mengder er det imidlertid orden i sakene:

**Setning 5.8.2 (Ekstremalverdisetningen)** *Anta at  $A$  er en lukket, begrenset delmengde av  $\mathbb{R}^m$  og at  $f : A \rightarrow \mathbb{R}$  er kontinuerlig. Da har  $f$  minimumspunkter og maksimumspunkter og er følgelig begrenset.*

*Bevis:* Vi skal vise at  $f$  har et maksimumspunkt. Beviset for minimumspunkt er helt likt og overlates til leserne. La

$$M = \sup\{f(\mathbf{x}) : \mathbf{x} \in A\}$$

der vi er enig om å sette  $M = \infty$  dersom  $f$  ikke er oppad begrenset. Velg en følge  $\{\mathbf{x}_n\}$  i  $A$  slik at  $f(\mathbf{x}_n) \rightarrow M$  når  $n \rightarrow \infty$  (dette er mulig uansett om  $M$  er endelig eller uendelig). Siden  $A$  er lukket og begrenset, har  $\{\mathbf{x}_n\}$  en konvergent delfølge  $\{\mathbf{x}_{n_k}\}$  ifølge teorem 5.2.3 (Bolzano-Weierstrass' teorem). Siden  $A$  er lukket, ligger grensepunktet  $\mathbf{c}$  til denne delfølgen i  $A$  (setning 5.1.6), og ifølge setning 5.1.7 er

$$\lim_{k \rightarrow \infty} f(\mathbf{x}_{n_k}) = f(\mathbf{c})$$

På den annen side er  $\lim_{k \rightarrow \infty} f(\mathbf{x}_{n_k}) = M$  (siden  $f(\mathbf{x}_n) \rightarrow M$  når  $n \rightarrow \infty$ ). Dermed må

$$f(\mathbf{c}) = M$$

Dette viser at  $M$  er endelig ( $f$  kan ikke ha verdien  $\infty$  i et punkt  $\mathbf{c}$ ) og at  $\mathbf{c}$  er et maksimumspunkt for  $f$ .  $\square$

**Bemerkning:** Beviset ovenfor er det samme som du finner for det én-dimensjonale tilfellet i *Kalkulus*, men siden vi har gjort en del forarbeid (spesielt teorem 5.2.3), er det atskillig kortere.

### Oppgaver til seksjon 5.8

1. Anta at  $A$  er en lukket, begrenset delmengde av  $\mathbb{R}^m$  og at  $\mathbf{F} : A \rightarrow \mathbb{R}^k$  er kontinuerlig. Vis at det finnes et tall  $K$  slik at  $|\mathbf{F}(\mathbf{x})| \leq K$  for alle  $\mathbf{x} \in A$ .
2. Anta at  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  er en positiv funksjon slik at  $\lim_{|\mathbf{x}| \rightarrow \infty} f(\mathbf{x}) = 0$ . Vis at  $f$  har et maksimumspunkt.
3. La  $A$  være en lukket og begrenset delmengde av  $\mathbb{R}^m$  og anta at  $\mathbf{F} : A \rightarrow A$  er en kontinuerlig funksjon.
  - a) Vis at funksjonen  $f : A \rightarrow \mathbb{R}$  gitt ved  $f(\mathbf{x}) = |\mathbf{x} - \mathbf{F}(\mathbf{x})|$  er kontinuerlig. Forklar at  $f$  har et minimumspunkt.
  - b) Anta at  $|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})| < |\mathbf{x} - \mathbf{y}|$  for alle  $\mathbf{x}, \mathbf{y} \in A$ ,  $\mathbf{x} \neq \mathbf{y}$ . Vis at  $\mathbf{F}$  har et entydig fikspunkt. Hint: Bruk minimumspunktet fra a).
  - c) Vis ved et eksempel at dersom vi dropper betingelsen om at  $A$  er lukket og begrenset, så behøver ikke  $\mathbf{F}$  ha et fikspunkt.

## 5.9 Maksimums- og minimumspunkter

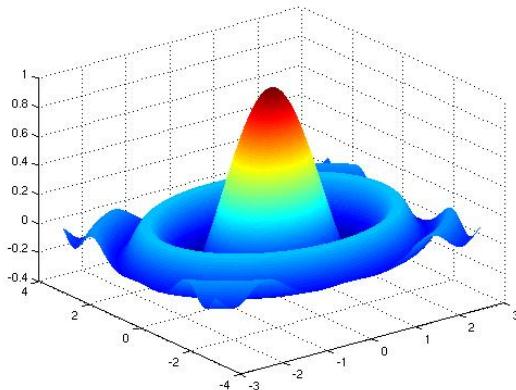
Hvordan finner vi maksimums- og minimumsverdier for funksjoner av flere variable? For funksjoner av én variabel vet vi at vi først må finne de punktene der den førstederiverte er null, og deretter undersøke hva slags punkter dette er ved enten å se på fortegnskiftet til den førstederiverte eller på fortegnet til den annenderiverte. I denne seksjonen skal vi se at det er en tilsvarende teori for funksjoner av flere variable. Hovedideene er de samme som i det envariable tilfellet, men siden geometrien er rikere, er det flere muligheter å holde styr på i den flervariabel teorien.

I forrige seksjon definerte vi (globale) maksimums- og minimumspunkter for funksjoner av flere variable. Vi kan imidlertid ikke regne med å finne de globale ekstremalpunktene direkte, men må gå veien om lokale maksima og minima. Før vi skriver opp definisjonen, minner vi om at snittet  $C \cap D$  av to mengder  $C$  og  $D$  består av de punktene som er med i både  $C$  og  $D$ .

**Definisjon 5.9.1** La  $f : A \rightarrow \mathbb{R}$  være en funksjon av  $m$  variable. Vi sier at  $\mathbf{a} \in A$  er et lokalt maksimumspunkt for  $f$  dersom det finnes en kule  $B(\mathbf{a}, r)$  med sentrum i  $\mathbf{a}$  slik at  $f(\mathbf{a}) \geq f(\mathbf{y})$  for alle  $\mathbf{y} \in B(\mathbf{a}, r) \cap A$ . Tilsvarende kalles  $\mathbf{a}$  et lokalt minimumspunkt dersom det finnes en kule  $B(\mathbf{a}, r)$  slik at

$f(\mathbf{a}) \leq f(\mathbf{y})$  for alle  $\mathbf{y} \in B(\mathbf{a}, r) \cap A$ . Vi bruker lokale ekstremalpunkter som et fellesnavn på lokale maksimal- og minimalpunkter.

Lokale maksimumspunkter ser litt forskjellig ut ettersom de er indre punkter eller randpunkter. Figur 1 viser noen av mulighetene. Et lokalt maksimum i det indre kan f.eks. være en “fjelltopp” som den høyeste toppen på figuren, eller det kan være et punkt på en “åskam” som de andre lokale maksimumspunktene i det indre. I begge disse tilfellene er det lett å forestille seg at alle de partiellderiverte i punktet er 0. Dette behøver imidlertid ikke være tilfellet for lokale maksimumspunkter på randen. Grafen i figur 1 har lokale maksimumspunkter i hjørnene av definisjonsområdet (de fire “flippene” i kanten av figuren), men de partiellderiverte i disse punktene er ikke 0 — punktene ligger i en “skråning” der funksjonen hadde fortsatt å stige hvis den var blitt forlenget på naturlig måte utover definisjonsområdet sitt.



Figur 1: Lokale ekstremalpunkter i det indre og på randen

I denne seksjonen skal vi stort sett konsentrere oss om jakten på lokale ekstremalpunkter i det indre av definisjonsområdet. I neste seksjon skal vi se på en teknikk som (blant annet) kan brukes til å finne mulige ekstremalpunkter på randen. La oss aller først bevise at det virkelig er slik at de partiellderiverte er null i alle lokale ekstremalpunkter i det indre.

**Setning 5.9.2** *Anta at en funksjon  $f: A \rightarrow \mathbb{R}$  har et lokalt maksimum eller minimum i et indre punkt  $\mathbf{a}$ . Dersom  $f$  er deriverbar i  $\mathbf{a}$ , må  $\nabla f(\mathbf{a}) = 0$ , dvs. at  $\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$  for alle  $i$ .*

*Bevis:* Vi fører resultatet tilbake til det tilsvarende resultatet for funksjoner av én variabel (*Kalkulus*, setning 6.2.1). Anta at  $f$  har et lokalt maksimum i  $\mathbf{a} = (a_1, a_2, \dots, a_m)$  (beviset for et lokalt minimum er helt tilsvarende). La  $g$  være funksjonen av én variabel definert ved

$$g(x_i) = f(a_1, a_2, \dots, x_i, \dots, a_m)$$

( $g$  er altså funksjonen vi får når vi “fryser” alle variablene i  $f$  unntatt den  $i$ -te). Da må  $g$  ha et lokalt maksimum for  $x_i = a_i$ , og følgelig er  $g'(a_i) = 0$ . Per definisjon av partiellderiverte er  $\frac{\partial f}{\partial x_i}(\mathbf{a}) = g'(a_i)$ , og dermed er  $\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$ .  $\square$

Ved hjelp av setningen ovenfor kan vi innskrenke jakten på mulige maksimums- og minimumspunkter betraktelig.

**Eksempel 1:** La oss forsøke å lokalisere eventuelle maksimums- og minimumspunkter for funksjonen

$$f(x, y) = 3xy - 3x + 9y$$

Vi deriverer:

$$\frac{\partial f}{\partial x} = 3y - 3 \quad \text{og} \quad \frac{\partial f}{\partial y} = 3x + 9$$

Ifølge setningen ovenfor bør vi se etter punkter hvor begge de partiellderiverte er null. Dette gir ligningssystemet

$$3y - 3 = 0 \quad \text{og} \quad 3x + 9 = 0$$

som har løsningen  $x = -3$ ,  $y = 1$ . Dette betyr at det eneste mulige maksimums- eller minimumspunktet til  $f$  er  $(-3, 1)$ , og at den tilsvarende funksjonsverdien er  $f(-3, 1) = 9$ .

Neste spørsmål er om  $(-3, 1)$  virkelig er et lokalt maksimums- eller minimumspunkt. For å avgjøre dette skal vi bruke et triks som av og til er nyttig. Vi innfører nye variable  $x'$  og  $y'$  slik at punktet  $(-3, 1)$  blir det nye origo, det vil si at vi setter

$$\begin{aligned} x' &= x - (-3) = x + 3 \\ y' &= y - 1 \end{aligned}$$

Legg merke til at  $(x', y') \rightarrow (0, 0)$  når  $(x, y) \rightarrow (-3, 1)$ . Siden  $x = x' - 3$ ,  $y = y' + 1$ , ser vi at

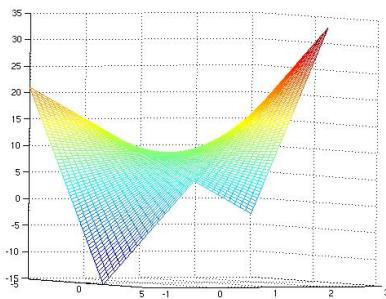
$$\begin{aligned} f(x, y) &= 3xy - 3x + 9y = 3(x' - 3)(y' + 1) - 3(x' - 3) + 9(y' + 1) \\ &= 3x'y' + 3x' - 9y' - 9 - 3x' + 9 + 9y' + 9 \\ &= 9 + 3x'y' \end{aligned}$$

Vi ser at hvis  $x'$  og  $y'$  har samme fortegn, så vil  $f(x, y)$  være større enn  $f(-3, 1) = 9$ . Har derimot  $x'$  og  $y'$  motsatt fortegn, vil  $f(x, y)$  være mindre enn 9. Altså kan  $(-3, 1)$  hverken være et lokalt maksimum eller et lokalt minimum.

La oss til slutt bruke MATLAB til å få et bedre inntrykk av funksjonen. Skriver vi

```
>> x=-5:0.1:1;
>> y=-1:0.1:3;
>> [x,y]=meshgrid(x,y);
>> z=3.*x.*y-3*x+9*y;
>> mesh(x,y,z)
```

svarer MATLAB med figur 2 (etter at vi har rotert litt på aksene for å få et oversiktlig bilde).



Figur 2: Et sadelpunkt

Legg merke til at grafen minner litt om en sal (på en hest), og at “vårt punkt”  $(-3, 1)$  ligger på det stedet der man naturlig sitter på salen. Som vi snart skal komme tilbake til, kalles slike punkter “sadelpunkter”. ♣

Eksemplet ovenfor peker på det som skal være hovedproblemstillingen i resten av denne seksjonen: Hvis  $\nabla f(\mathbf{a}) = 0$ , hvordan avgjør vi da på en effektiv måte om  $\mathbf{a}$  er et lokalt maksimum, minimum eller ingen av delene? Teknikken med å skifte variable er nyttig i en del enkle tilfeller, men vi trenger tyngre skyts for å kunne behandle mer kompliserte uttrykk.

La oss begynne med å innføre litt terminologi. Et punkt  $\mathbf{a}$  der  $\nabla f(\mathbf{a}) = 0$  vil vi kalle et *stasjonært* punkt for funksjonen  $f$ . Et stasjonært punkt som hverken er et lokalt maksimum eller et lokalt minimum, vil vi kalle et *sadelpunkt* (se figur 2 ovenfor). Som vi allerede har vært inne på, er det ikke vanskelig å forstå hvor det siste navnet kommer fra – det punktet du sitter på når du rir på en hest, er et typisk eksempel på et sadelpunkt; det er et minimum når du beveger deg i hestens lengderetning og et maksimum nå du beveger deg på tvers av hesten.

Vi tar med et eksempel til på hvordan man finner stasjonære punkter.

**Eksempel 2:** Finn de stasjonære punktene til

$$f(x, y) = x^2 - y^2 + 4xy - 7x + 3y$$

Vi deriverer:

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2x + 4y - 7 \\ \frac{\partial f}{\partial y} &= -2y + 4x + 3\end{aligned}$$

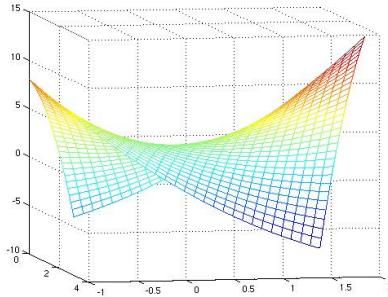
Dette gir ligningene

$$2x + 4y = 7$$

$$4x - 2y = -3$$

Løser vi dette ligningssystemet, får vi  $x = \frac{1}{10}$ ,  $y = \frac{17}{10}$ . Det betyr at punktet  $(\frac{1}{10}, \frac{17}{10})$  er et stasjonært punkt for  $f$ .

Kjører vi MATLAB på samme måte som i forrige eksempel, får vi grafen på figur 3.



Figur 3: Grafen til  $f(x, y) = x^2 - y^2 + 4xy - 7x + 3y$

Den viser at vi også i dette tilfellet har et sadelpunkt. ♣

**Bemerkning:** Legg merke til at når vi skal finne de stasjonære punktene til en funksjon  $f$  av  $m$  variable  $x_1, x_2, \dots, x_m$ , må vi løse et ligningssystem med  $m$  ukjente og  $m$  ligninger:

$$\begin{aligned}\frac{\partial f}{\partial x_1}(x_1, x_2, \dots, x_m) &= 0 \\ \frac{\partial f}{\partial x_2}(x_1, x_2, \dots, x_m) &= 0 \\ &\vdots \quad \vdots \quad = \quad \vdots \\ \frac{\partial f}{\partial x_m}(x_1, x_2, \dots, x_m) &= 0\end{aligned}$$

Dette er et system av den typen som vi brukte Newtons metode til å løse i seksjon 5.6. Når vi får et maks/min-problem i flere variable som går ut over det aller enkleste, må vi regne med å bruke numeriske metoder for å finne løsningen. Newton's metode er én mulighet, men som vi skal se senere (seksjon 5.11), finnes det også mer direkte metoder.

### Taylors formel

Når vi arbeider med funksjoner av to variable slik som i eksemplene ovenfor, kan vi ofte bruke MATLAB eller et lignende verktøy til å undersøke om de stasjonære punktene våre er minimumspunkter, maksimumspunkter eller sadelpunkter. Hvis funksjonsgrafen er svært flat i området rundt det stasjonære punktet, kan det imidlertid være vanskelig å avgjøre visuelt hva slags punkt vi har med å gjøre. Arbeider vi med funksjoner av flere enn to variable, er det atskillig verre å bruke visuelle hjelpe middler. Vi trenger derfor en teori som kan hjelpe oss i klassifiseringen av stasjonære punkter.

For funksjoner av én variabel har vi et slikt hjelpe middel, nemlig *annenderivertesten*. Den sier at hvis  $f$  er en funksjon av én variabel med  $f'(a) = 0$ , så er  $a$  et et lokalt minimum dersom  $f''(a) > 0$  og at  $a$  er et lokalt maksimum dersom  $f''(a) < 0$ . Når  $f''(a) = 0$ , gir testen ingen konklusjon. Vårt mål er å lage en tilsvarende test for funksjoner av flere variable. Dette arbeidet er ganske komplisert fordi en funksjon av flere variable har så mange forskjellige annenderiverte, og de må kombineres på riktig måte for å få en test som virker. Heldigvis skal vi få hjelp av det vi vet om lineær algebra og baserer av egenverdier.

Dersom  $f(x_1, \dots, x_m)$  er en to ganger deriverbar funksjon av  $m$  variable, kan vi skrive opp alle de annenordens partiellderiverte som en  $m \times m$  matrise:

$$Hf(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_m}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_m}(\mathbf{a}) \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_m \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_m \partial x_2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m^2}(\mathbf{a}) \end{pmatrix}$$

Vi kaller dette *Hesse-matrisen* til  $f$  i punktet  $\mathbf{a}$  (ikke bland Hesse-matrisen, som er en matrise av annenderiverte til et skalarfelt, sammen med Jacobi-matrisen, som er en matrise av førstederiverte til en vektorvaluert funksjon!). Foreløpig er Hesse-matrisen bare en grei måte å skrive opp de annenordens partiellderiverte på, men vi skal snart se at den også har matematisk betydning.

Vi husker fra seksjon 2.5 at dersom de blandede partiellderiverte  $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$  og  $\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$  er kontinuerlige, så er de like. Det betyr at Hesse-matrisen  $Hf(\mathbf{a})$  er symmetrisk. Fra lineær algebra (spektralteoremet 4.10.6) vet vi at  $Hf(\mathbf{a})$  da må ha  $m$  reelle egenverdier  $\lambda_1(\mathbf{a}), \lambda_2(\mathbf{a}), \dots, \lambda_m(\mathbf{a})$  (flere av dem kan være like). Det viser seg at det er fortegnet til disse egenverdiene som avgjør om et stasjonært punkt er et lokalt maksimum, et lokalt minimum eller et sadelpunkt. Dersom alle egenverdiene er positive, så er  $\mathbf{a}$  et lokalt minimum, dersom alle er negative, så er  $\mathbf{a}$  et lokalt maksimum, og dersom det finnes egenverdier med motsatte fortegn, så er  $\mathbf{a}$  et sadelpunkt.

For å forstå hva Hesse-matrisen har å gjøre med lokale maksimums- og minimumspunkter, må vi først studere Taylors formel for funksjoner av flere variable.

**Setning 5.9.3 (Taylors formel)** *La  $f$  være en funksjon av  $m$  variable. Anta at de annenordens partiellderiverte til  $f$  er kontinuerlige i en kule  $B(\mathbf{a}, r)$  om  $\mathbf{a}$ . For enhver  $\mathbf{y} \in \mathbb{R}^m$  med  $|\mathbf{y}| < r$  finnes det et tall  $c \in (0, 1)$  slik at*

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a} + c\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$$

I dette uttrykket er  $Hf(\mathbf{a} + c\mathbf{y})\mathbf{y}$  matriseproduktet av matrisen  $Hf(\mathbf{a} + c\mathbf{y})$  og (søylevektoren)  $\mathbf{y}$ , mens  $(Hf(\mathbf{a} + c\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$  er skalarproduktet mellom to vektorer.

*Bevis:* Definer en funksjon  $g$  av én variabel ved

$$g(t) = f(\mathbf{a} + t\mathbf{y}).$$

Bruker vi kjerneregelen for funksjoner av flere variable, ser vi at

$$g'(t) = \sum_{i=1}^m \frac{\partial f}{\partial x_i}(\mathbf{a} + t\mathbf{y})y_i = \nabla f(\mathbf{a} + t\mathbf{y}) \cdot \mathbf{y}$$

Bruker vi kjerneregelen på nytt, får vi

$$g''(t) = \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a} + t\mathbf{y})y_i y_j$$

Dette uttrykket kan også skrives på matriseform

$$g''(t) = (Hf(\mathbf{a} + t\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$$

(skriv opp uttrykket til høyre på koordinatform og sjekk at dette stemmer).

Taylors formel for funksjoner av én variabel (se *Kalkulus*, seksjon 11.2) forteller oss at det finnes et tall mellom 0 og 1 slik at

$$g(1) = g(0) + g'(0) + \frac{1}{2}g''(c)$$

Setter vi inn uttrykkene ovenfor, får vi

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a} + c\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$$

□

Taylors formel forteller oss at i et stasjonært punkt  $\mathbf{a}$  er det fortegnet til  $(Hf(\mathbf{a} + c\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$  som avgjør om  $f(\mathbf{a} + \mathbf{y})$  er større enn eller mindre enn  $f(\mathbf{a})$  (husk at i et stasjonært punkt er  $\nabla f(\mathbf{a}) = 0$ ). Det hadde vært fint om vi kunne ha byttet ut  $(Hf(\mathbf{a} + c\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$  med  $(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y}$  i dette utsagnet, siden vi har mye bedre kontroll over punktet  $\mathbf{a}$  enn over  $\mathbf{a} + c\mathbf{y}$ . Den neste versjonen av Taylors formel sier at vi kan gjøre dette byttet uten å måtte betale altfor mye.

**Setning 5.9.4 (Taylors formel, versjon 2)** La  $f$  være en funksjon av  $m$  variable. Anta at de annenordens partiellederiverte til  $f$  er kontinuerlige i en omegn om  $\mathbf{a}$ . Da finnes det en funksjon  $\varepsilon$  av  $m$  variable slik at

$$\lim_{\mathbf{y} \rightarrow \mathbf{0}} \varepsilon(\mathbf{y}) = 0$$

og

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + \varepsilon(\mathbf{y})|\mathbf{y}|^2$$

for alle tilstrekkelige små  $\mathbf{y} \in \mathbb{R}^m$ .

**Bemerkning:** Setningen ovenfor forteller oss at hvis vi tilnærmer  $f(\mathbf{a} + \mathbf{y})$  med  $f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y}$ , så gjør vi (for små  $\mathbf{y}$ ) en feil som er liten sammenlignet med  $|\mathbf{y}|^2$ . Siden  $\frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y}$  typisk er av samme størrelse som  $|\mathbf{y}|^2$ , betyr dette at det nest siste ledet vil dominere over feilreddet  $\varepsilon(\mathbf{y})|\mathbf{y}|^2$ .

*Bevis for versjon 2 av Taylors formel:* Tar vi utgangspunkt i den første versjonen av Taylors formel og legger til og trekker fra  $\frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y}$  på høyresiden, får vi

$$\begin{aligned} f(\mathbf{a} + \mathbf{y}) &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} \\ &\quad + \frac{1}{2}[(Hf(\mathbf{a} + c\mathbf{y}) - Hf(\mathbf{a})) \cdot \mathbf{y}] \mathbf{y} \end{aligned}$$

Vi forenkler notasjonen ved å sette

$$A(\mathbf{y}) = \frac{1}{2}(Hf(\mathbf{a} + c\mathbf{y}) - Hf(\mathbf{a}))$$

Legg merke til at  $A(\mathbf{y})$  er en  $m \times m$ -matrise der koeffisientene  $a_{ij}(\mathbf{y})$  går mot null når  $\mathbf{y}$  går mot  $\mathbf{0}$ .

Uttrykket for  $f(\mathbf{a} + \mathbf{y})$  kan nå skrives

$$f(\mathbf{a} + \mathbf{y}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + (A(\mathbf{y})\mathbf{y}) \cdot \mathbf{y}$$

Sammenligner vi dette med formelen i setningen, ser vi at vi må sette

$$\varepsilon(\mathbf{y}) = \frac{(A(\mathbf{y})\mathbf{y}) \cdot \mathbf{y}}{|\mathbf{y}|^2} = \left( A(\mathbf{y}) \frac{\mathbf{y}}{|\mathbf{y}|} \right) \cdot \frac{\mathbf{y}}{|\mathbf{y}|}$$

Det gjenstår å vise at  $\lim_{\mathbf{y} \rightarrow \mathbf{0}} \varepsilon(\mathbf{y}) = 0$ , men det er lett:

$$\begin{aligned} \lim_{\mathbf{y} \rightarrow \mathbf{0}} |\varepsilon(\mathbf{y})| &= \lim_{\mathbf{y} \rightarrow \mathbf{0}} \left| \sum_{i,j=1}^m a_{ij}(\mathbf{y}) \frac{y_i}{|\mathbf{y}|} \cdot \frac{y_j}{|\mathbf{y}|} \right| \leq \\ &\leq \lim_{\mathbf{y} \rightarrow \mathbf{0}} \sum_{i,j=1}^m |a_{ij}(\mathbf{y})| = 0 \end{aligned}$$

hvor vi har brukt at  $\frac{|y_i|}{|\mathbf{y}|}, \frac{|y_j|}{|\mathbf{y}|} \leq 1$  og at  $\lim_{\mathbf{y} \rightarrow \mathbf{0}} a_{ij}(\mathbf{y}) = 0$ .  $\square$

For å utnytte Taylors formel trenger vi et enkelt resultat fra lineær algebra.

**Lemma 5.9.5** *La  $A$  være en symmetrisk  $m \times m$ -matrise.*

- a) *Anta at alle egenverdiene til  $A$  er positive:  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ . Da er*

$$(A\mathbf{y}) \cdot \mathbf{y} \geq \lambda_1 |\mathbf{y}|^2$$

*for alle  $\mathbf{y} \in \mathbb{R}^m$*

- b) *Anta at alle egenverdiene til  $A$  er negative:  $0 > \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ . Da er*

$$(A\mathbf{y}) \cdot \mathbf{y} \leq \lambda_1 |\mathbf{y}|^2$$

*for alle  $\mathbf{y} \in \mathbb{R}^m$*

*Bevis:* a) La  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  være en ortonormal basis av egenvektorer for  $A$  slik at  $\mathbf{v}_i$  har egenverdi  $\lambda_i$ . Observer først at hvis  $\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m = \sum_{i=1}^m c_i\mathbf{v}_i$ , så er

$$\begin{aligned} |\mathbf{y}|^2 &= \mathbf{y} \cdot \mathbf{y} = \left( \sum_{i=1}^m c_i \mathbf{v}_i \right) \cdot \left( \sum_{j=1}^m c_j \mathbf{v}_j \right) = \\ &= \sum_{i,j=1}^m c_i c_j (\mathbf{v}_i \cdot \mathbf{v}_j) = c_1^2 + c_2^2 + \dots + c_m^2 \end{aligned}$$

der vi har brukt at  $\mathbf{v}_i \cdot \mathbf{v}_j$  er 1 dersom  $i = j$  og 0 ellers. Videre er

$$A\mathbf{y} = A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m) = c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \dots + c_m\lambda_m\mathbf{v}_m$$

Ved en tilsvarende regning som ovenfor ser vi at

$$\begin{aligned} (A\mathbf{y}) \cdot \mathbf{y} &= (c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \dots + c_m\lambda_m\mathbf{v}_m) \cdot (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_m\mathbf{v}_m) \\ &= c_1^2\lambda_1 + c_2^2\lambda_2 + \dots + c_m^2\lambda_m \geq \lambda_1(c_1^2 + c_2^2 + \dots + c_m^2) \geq \lambda_1 |\mathbf{y}|^2 \end{aligned}$$

- b) Bruk punkt a) på matrisen  $(-A)$ .  $\square$

### Annenderiverttesten

Vi er nå klare til å vise hovedresultatet i dette kapitlet.

**Teorem 5.9.6 (Annenderiverttesten)** *La  $\mathbf{a}$  være et stasjonært punkt for en funksjon  $f$  av  $m$  variable. Anta at de annenordens partiellderiverte til  $f$  er kontinuerlige i en omegn om  $\mathbf{a}$ . Da gjelder:*

- a) Hvis alle egenverdiene til  $Hf(\mathbf{a})$  er (strent) positive, så er  $\mathbf{a}$  et lokalt minimumspunkt.
- b) Hvis alle egenverdiene til  $Hf(\mathbf{a})$  er (strent) negative, så er  $\mathbf{a}$  et lokalt maksimumspunkt.
- c) Hvis  $Hf(\mathbf{a})$  har både (strent) positive og (strent) negative egenverdier, så er  $\mathbf{a}$  et sadelpunkt.

Dersom noen av egenverdiene til  $Hf(\mathbf{a})$  er null og de andre har samme fortegn, så gir testen ingen konklusjon.

*Bevis:* a) La  $\lambda_1$  være den minste egenverdien til  $Hf(\mathbf{a})$ . Ifølge versjon 2 av Taylors formel er da (husk at  $\nabla f(\mathbf{a}) = \mathbf{0}$  siden  $\mathbf{a}$  er et stasjonært punkt):

$$\begin{aligned} f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) &= \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \\ &= \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \\ &\geq \frac{1}{2}\lambda_1|\mathbf{y}|^2 + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \end{aligned}$$

der vi har brukt lemma 5.9.5 i den siste overgangen. Siden  $\lambda_1$  er positiv og  $\varepsilon(\mathbf{y}) \rightarrow 0$ , vil også  $\frac{1}{2}\lambda_1 + \varepsilon(\mathbf{y})$  være positiv når  $\mathbf{y}$  er tilstrekkelig liten. Dermed er

$$f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) \geq (\frac{1}{2}\lambda_1 + \varepsilon(\mathbf{y}))|\mathbf{y}|^2 \geq 0$$

som viser at  $f(\mathbf{a})$  er et lokalt minimum.

b) La  $\lambda_1$  være den største egenverdien til  $Hf(\mathbf{a})$ . Da er

$$\begin{aligned} f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) &= \nabla f(\mathbf{a}) \cdot \mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \\ &= \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \\ &\leq \frac{1}{2}\lambda_1|\mathbf{y}|^2 + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \end{aligned}$$

Siden  $\lambda_1$  er negativ og  $\varepsilon(\mathbf{y}) \rightarrow 0$ , vil også  $\frac{1}{2}\lambda_1 + \varepsilon(\mathbf{y})$  være negativ når  $\mathbf{y}$  er tilstrekkelig liten. Dermed er

$$f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) \leq (\frac{1}{2}\lambda_1 + \varepsilon(\mathbf{y}))|\mathbf{y}|^2 \leq 0$$

som viser at  $f(\mathbf{a})$  er et lokalt maksimum.

c) For å vise at  $\mathbf{a}$  ikke er et lokalt maksimum, lar vi  $\mathbf{y}$  være en egenvektor for  $Hf(\mathbf{a})$  med en positiv egenverdi  $\lambda$ . Da er

$$\begin{aligned} f(\mathbf{a} + \mathbf{y}) - f(\mathbf{a}) &= \nabla f(\mathbf{a})\mathbf{y} + \frac{1}{2}(Hf(\mathbf{a})\mathbf{y}) \cdot \mathbf{y} + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \\ &= \frac{1}{2}\lambda|\mathbf{y}|^2 + \varepsilon(\mathbf{y})|\mathbf{y}|^2 \end{aligned}$$

Har vi valgt  $\mathbf{y}$  tilstrekkelig liten, vil  $\frac{1}{2}\lambda + \varepsilon(\mathbf{y}) > 0$ , og følgelig er  $f(\mathbf{a} + \mathbf{y}) > f(\mathbf{a})$ . Dette viser at  $\mathbf{a}$  ikke er et lokalt maksimum.

For å vise at  $\mathbf{a}$  heller ikke er et lokalt minimum, lar vi isteden  $\mathbf{y}$  være en egenvektor med en *negativ* egenverdi og gjennomfører det samme resonnementet.  $\square$

For funksjoner av to variable har annenderiverttesten også en annen form som er enklere å bruke i praksis.

**Korollar 5.9.7 (Annenderiverttesten i to variable)** *La  $\mathbf{a}$  være et stationært punkt for en funksjon  $f$  av to variable. Anta at de annenordens partiellderiverte er kontinuerlige i en omegn om  $\mathbf{a}$ . La*

$$A = \frac{\partial^2 f}{\partial x^2}(\mathbf{a}), \quad B = \frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) = \frac{\partial^2 f}{\partial y \partial x}(\mathbf{a}), \quad C = \frac{\partial^2 f}{\partial y^2}(\mathbf{a})$$

og la  $D$  være determinanten til Hesse-matrisen:  $D = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = AC - B^2$ . Da gjelder

- (i) *Hvis  $D < 0$ , så er  $\mathbf{a}$  et sadelpunkt.*
- (ii) *Hvis  $D > 0$  og  $A > 0$ , så er  $\mathbf{a}$  et lokalt minimum.*
- (iii) *Hvis  $D > 0$  og  $A < 0$ , så er  $\mathbf{a}$  et lokalt maksimum.*

*Hvis  $D = 0$ , gir testen ingen konklusjon.*

*Bevis:* La  $\lambda_1$  og  $\lambda_2$  være de to egenverdiene til Hesse-matrisen. Vi vet fra lineær algebra (korollar 4.10.10) at determinanten er produktet av egenverdiene, så  $D = \lambda_1 \lambda_2$ .

(i) Hvis  $D < 0$ , må  $\lambda_1$  og  $\lambda_2$  ha motsatt fortegn. Ifølge den generelle annenderiverttesten er  $\mathbf{a}$  et sadelpunkt.

(ii) Hvis  $D > 0$ , har de to egenverdiene samme fortegn, og ifølge den generelle annenderiverttesten må  $\mathbf{a}$  enten være et lokalt maksimum eller minimum. Siden  $A = \frac{\partial^2 f}{\partial x^2}(\mathbf{a}) > 0$ , har  $f$  et lokalt minimum i  $\mathbf{a}$  når vi ser på den som en funksjon av  $x$  alene. Men da må det være et lokalt minimum  $f$  har i  $\mathbf{a}$ .

(iii) Helt analogt til (ii).  $\square$

La oss se hvordan annenderiverttesten virker på noen eksempler. Vi tar først for oss funksjonen fra eksempel 1 på nytt.

**Eksempel 3:** Vi ser altså på funksjonen  $f(x, y) = 3xy - 3x + 9y$  som vi allerede har vist har partiellderiverte

$$\frac{\partial f}{\partial x} = 3y - 3 \quad \text{og} \quad \frac{\partial f}{\partial y} = 3x + 9$$

Vi vet også at begge de partiellderiverte er null i punktet  $(-3, 1)$ . For å bruke annenderiverttesten regner vi ut

$$A = \frac{\partial^2 f}{\partial x^2} = 0, \quad B = \frac{\partial^2 f}{\partial x \partial y} = 3, \quad C = \frac{\partial^2 f}{\partial y^2} = 0$$

som gir

$$D = \begin{vmatrix} A & B \\ B & C \end{vmatrix} = \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} = -9.$$

Ifølge annenderiverttesten er  $(-3, 1)$  et sadelpunkt. ♣

La oss se på et litt mer komplisert eksempel:

**Eksempel 4:** Vi skal finne de stasjonære punktene til

$$f(x, y) = xye^{x-y^2}$$

og avgjøre om de er lokale maksimums-, minimums- eller sadelpunkter.

Derivasjon gir

$$\begin{aligned} \frac{\partial f}{\partial x} &= 1 \cdot ye^{x-y^2} + xye^{x-y^2} = y(1+x)e^{x-y^2} \\ \frac{\partial f}{\partial y} &= x \cdot 1 \cdot e^{x-y^2} + xye^{x-y^2}(-2y) = x(1-2y^2)e^{x-y^2}. \end{aligned}$$

Siden  $e^{x-y^2}$  ikke kan være null, er det nok å løse ligningene

$$\begin{aligned} y(1+x) &= 0 \\ x(1-2y^2) &= 0 \end{aligned}$$

for å finne de stasjonære punktene. Den første ligningen har to løsninger  $x = -1$  og  $y = 0$ . Setter vi  $x = -1$  inn i den andre ligningen, får vi  $y = \pm \frac{1}{\sqrt{2}} = \pm \frac{\sqrt{2}}{2}$ . Setter vi  $y = 0$  inn i den andre ligningen, får vi  $x = 0$ . Vi har altså tre stasjonære punkter  $(-1, \frac{\sqrt{2}}{2})$ ,  $(-1, -\frac{\sqrt{2}}{2})$  og  $(0, 0)$ .

Neste skritt er å regne ut de annenderiverte:

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= y \cdot 1 \cdot e^{x-y^2} + y(1+x)e^{x-y^2} = y(2+x)e^{x-y^2} \\ \frac{\partial^2 f}{\partial x \partial y} &= 1 \cdot (1+x) \cdot e^{x-y^2} + y(1+x)e^{x-y^2}(-2y) = (1+x)(1-2y^2)e^{x-y^2} \\ \frac{\partial^2 f}{\partial y^2} &= x(-4y)e^{x-y^2} + x(1-2y^2)e^{x-y^2}(-2y) = -2xy(3-2y^2)e^{x-y^2} \end{aligned}$$

Vi må undersøke de stasjonære punktene hver for seg.

**Det stasjonære punktet  $(0, 0)$ :** Her er

$$\begin{aligned} A &= \frac{\partial^2 f}{\partial x^2}(0, 0) = 0(2 + 0)e^{0-0^2} = 0 \\ B &= \frac{\partial^2 f}{\partial x \partial y}(0, 0) = (1 + 0)(1 - 2 \cdot 0^2)e^{0-0^2} = 1 \\ C &= \frac{\partial^2 f}{\partial y^2}(0, 0) = -2 \cdot 0 \cdot 0(3 - 2 \cdot 0^2)e^{0-0^2} = 0. \end{aligned}$$

Dette gir  $D = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 0^2 - 1^2 = -1$ . Altså er  $(0, 0)$  et sadelpunkt.

**Det stasjonære punktet  $(-1, \frac{\sqrt{2}}{2})$ :** Her er

$$\begin{aligned} A &= \frac{\partial^2 f}{\partial x^2}\left(-1, \frac{\sqrt{2}}{2}\right) = \frac{\sqrt{2}}{2}(2 + (-1))e^{-1-(\frac{\sqrt{2}}{2})^2} = \frac{\sqrt{2}}{2}e^{-\frac{3}{2}} \\ B &= \frac{\partial^2 f}{\partial y \partial x}\left(-1, \frac{\sqrt{2}}{2}\right) = (1 + (-1))\left(1 - 2\left(\frac{\sqrt{2}}{2}\right)^2\right)e^{-1-(\frac{\sqrt{2}}{2})^2} = 0 \\ C &= \frac{\partial^2 f}{\partial y^2}\left(-1, \frac{\sqrt{2}}{2}\right) = -2(-1)\frac{\sqrt{2}}{2}\left(3 - 2\left(\frac{\sqrt{2}}{2}\right)^2\right)e^{-1-(\frac{\sqrt{2}}{2})^2} = 2\sqrt{2}e^{-\frac{3}{2}} \end{aligned}$$

Dette gir

$$D = \begin{vmatrix} \frac{\sqrt{2}}{2}e^{-3/2} & 0 \\ 0 & 2\sqrt{2}e^{-3/2} \end{vmatrix} = 2e^{-3}$$

Siden  $D > 0$ ,  $A > 0$ , forteller annenderiverttesten oss at  $(-1, \frac{\sqrt{2}}{2})$  er et lokalt minimum.

**Det stasjonære punktet  $(-1, -\frac{\sqrt{2}}{2})$ :** Her er

$$\begin{aligned} A &= \frac{\partial^2 f}{\partial x^2}\left(-1, -\frac{\sqrt{2}}{2}\right) = -\frac{\sqrt{2}}{2}(2 + (-1))e^{-1-(-\frac{\sqrt{2}}{2})^2} = -\frac{\sqrt{2}}{2}e^{-\frac{3}{2}} \\ B &= \frac{\partial^2 f}{\partial y \partial x}\left(-1, -\frac{\sqrt{2}}{2}\right) = (1 + (-1))\left(1 - 2\left(-\frac{\sqrt{2}}{2}\right)^2\right)e^{-1-(-\frac{\sqrt{2}}{2})^2} = 0 \\ C &= \frac{\partial^2 f}{\partial y^2}\left(-1, -\frac{\sqrt{2}}{2}\right) = -2(-1)\left(-\frac{\sqrt{2}}{2}\right)\left(3 - 2\left(-\frac{\sqrt{2}}{2}\right)^2\right)e^{-1-(-\frac{\sqrt{2}}{2})^2} = -2\sqrt{2}e^{-\frac{3}{2}} \end{aligned}$$

Dette gir

$$D = \begin{vmatrix} -\frac{\sqrt{2}}{2}e^{-3/2} & 0 \\ 0 & -2\sqrt{2}e^{-3/2} \end{vmatrix} = 2e^{-3}$$

Siden  $D > 0$ ,  $A < 0$ , må  $(-1, -\frac{\sqrt{2}}{2})$  være et lokalt maksimum. ♣

## Uoppstilte problemer

Til slutt i denne seksjonen skal vi se på noen eksempler på uoppstilte minimums- og maksimumsproblemer.

**Eksempel 5:** Vi har en 1 meter lang ståltråd som skal deles i maksimalt tre biter. Hver bit skal så bøyes sammen til et kvadrat. Hva er det største og minste samlede areal disse rektanglene kan ha?

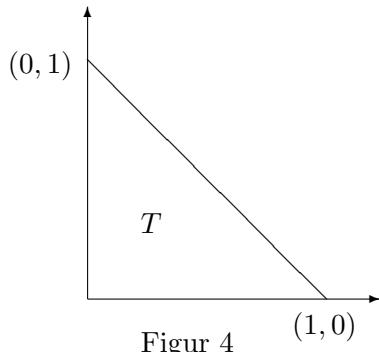
Hvis vi sier at de to første bitene har lengde  $x$  og  $y$ , må den tredje ha lengde  $1-x-y$ . Det totale arealet er dermed (husk at vi må dele lengdene på 4 for å få sidekantene i kvadratene):

$$A(x, y) = \left(\frac{x}{4}\right)^2 + \left(\frac{y}{4}\right)^2 + \left(\frac{1-x-y}{4}\right)^2 = \frac{1}{16}(x^2 + y^2 + (1-x-y)^2)$$

Det er fristende å sette igang å derivere med en gang, men la oss først se hvilke verdier  $x$  og  $y$  kan ha. Vi må åpenbart ha  $x \geq 0$ ,  $y \geq 0$  og  $x+y \leq 1$  (legg merke til at vi godt kan ha  $x=0$ ,  $y=0$  eller  $x+y=1$  — det svarer bare til at vi deler opp stålträden i færre enn tre biter). Dette betyr at vi ønsker å maksimere og minimere funksjonen  $A$  på området

$$T = \{(x, y) : x \geq 0, y \geq 0, x+y \leq 1\}$$

Dette er trekanten med hjørner  $(0,0)$ ,  $(1,0)$  og  $(0,1)$  som vist på figur 4.



Figur 4

Siden  $T$  er en lukket mengde, vet vi fra ekstremalverdisetningen at  $A$  har en maksimums- og en minimumsverdi.

Partiellderiverer vi  $A$ , får vi

$$\begin{aligned} \frac{\partial A}{\partial x}(x, y) &= \frac{1}{16}(2x + 2(1-x-y)(-1)) = \frac{1}{16}(4x + 2y - 2) \\ \frac{\partial A}{\partial y}(x, y) &= \frac{1}{16}(2y + 2(1-x-y)(-1)) = \frac{1}{16}(2x + 4y - 2) \end{aligned}$$

Ligningssystemet  $4x + 2y - 2 = 0$ ,  $2x + 4y - 2 = 0$  er lett å løse og gir  $x = y = \frac{1}{3}$ . Bruker vi annenderiverttesten (gjør det!), ser vi at  $(\frac{1}{3}, \frac{1}{3})$  er et lokalt minimum.

Dette betyr at det eneste potensielle ekstremalpunktet vi har i det indre av  $T$ , er et lokalt minimum i  $(\frac{1}{3}, \frac{1}{3})$ . For å finne andre kandidater må vi se på randen til  $T$ . Den faller naturlig i tre deler, og vi ser på hver del for seg.

Linjestykket fra  $(0, 0)$  til  $(1, 0)$ : På dette linjestykket er  $y = 0$ , og vi får

$$A(x, 0) = \frac{1}{16}(x^2 + (1-x)^2) \quad \text{for } 0 \leq x \leq 1$$

Drøfter vi dette uttrykket som en vanlig funksjon av én variabel, finner vi et minimum for  $x = \frac{1}{2}$  og maksima for  $x = 0$  og  $x = 1$ . Vi har dermed et mulig minimumspunkt i  $(\frac{1}{2}, 0)$  og mulige maksimumspunkter i  $(0, 0)$  og  $(1, 0)$ .

Linjestykket fra  $(0, 0)$  til  $(0, 1)$ : På dette linjestykket er  $x = 0$ , og vi får

$$A(0, y) = \frac{1}{16}(y^2 + (1-y)^2) \quad \text{for } 0 \leq y \leq 1$$

Dette er samme uttrykk som ovenfor bare med  $x$  byttet ut med  $y$ . Vi får derfor et mulig minimumspunkt i  $(0, \frac{1}{2})$  og mulige maksimumspunkter i  $(0, 0)$  og  $(0, 1)$ .

Linjestykket fra  $(1, 0)$  til  $(0, 1)$ : På dette linjestykket er  $y = 1 - x$ , og vi får

$$A(x, 1-x) = \frac{1}{16}(x^2 + (1-x)^2) \quad \text{for } 0 \leq x \leq 1$$

Dette er samme uttrykk som i det første punktet ovenfor, og vi finner et minimum for  $x = \frac{1}{2}$  og maksima for  $x = 0$  og  $x = 1$ . Vi har dermed et mulig minimumspunkt i  $(\frac{1}{2}, \frac{1}{2})$  og mulige maksimumspunkter i  $(1, 0)$  og  $(0, 1)$ .

La oss ta en liten oppsummering: Vi har potensielle minimumspunkter i  $(\frac{1}{3}, \frac{1}{3})$ ,  $(\frac{1}{2}, 0)$ ,  $(0, \frac{1}{2})$  og  $(\frac{1}{2}, \frac{1}{2})$ . Regner vi ut funksjonsverdiene, får vi  $A(\frac{1}{3}, \frac{1}{3}) = \frac{1}{48}$  og  $A(\frac{1}{2}, 0) = A(0, \frac{1}{2}) = A(\frac{1}{2}, \frac{1}{2}) = \frac{1}{32}$ . Dette betyr at minimumspunktet er  $(\frac{1}{3}, \frac{1}{3})$  og minimumsverdien  $\frac{1}{48}$ . Vi får altså mimimalt areal når vi deler opp stålträden i tre like store deler.

På tilsvarende måte har vi de potensielle maksimumspunktene  $(0, 0)$ ,  $(1, 0)$  og  $(0, 1)$ . Regner vi ut funksjonsverdiene, får vi  $A(0, 0) = A(1, 0) = A(1, 1) = \frac{1}{16}$ . Maksimumsverdiene er altså  $\frac{1}{16}$  oppnådd i hjørnene  $A(0, 0) = A(1, 0) = A(1, 1)$ . Geometrisk representerer disse hjørnene den samme løsningen — vi deler ikke opp stålträden i det hele tatt, men bøyer den sammen til ett stort kvadrat. ♣

Eksemplet ovenfor viser at vi ikke kan neglisjere punktene på randen av definisjonsområdet — det kan hende at det er der det interessante foregår! Det neste eksemplet illustrerer (blant annet) de problemene vi kan støte på

når definisjonsområdet *ikke* er begrenset.

**Eksempel 6:** Vi skal lage en boks med volum  $V$ . Hvordan skal vi ordne oss for at overflatearealet  $A$  skal bli minst mulig?

Kaller vi sidekantene  $x, y$  og  $z$  som vist på figur 5, ser vi at arealet blir

$$A = 2xy + 2xz + 2yz$$

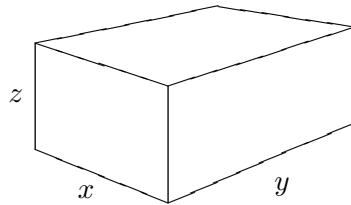
Siden volumet  $V = xyz$  er gitt, kan vi eliminere en av variablene

$$z = \frac{V}{xy}$$

Dette gir

$$A(x, y) = 2xy + \frac{2V}{y} + \frac{2V}{x}$$

I dette uttrykket kan  $x$  og  $y$  være vilkårlige, positive tall.



Figur 5

Vi deriverer  $A$ :

$$\frac{\partial A}{\partial x} = 2y - \frac{2V}{x^2}$$

$$\frac{\partial A}{\partial y} = 2x - \frac{2V}{y^2}$$

For å finne de stasjonære punktene, må vi løse ligningen

$$y = \frac{V}{x^2} \quad \text{og} \quad x = \frac{V}{y^2}.$$

Setter vi det første uttrykket inn i det andre, ser vi at

$$x = \frac{V}{(\frac{V}{x^2})^2} = \frac{x^4}{V}$$

som gir  $x = \sqrt[3]{V}$ . Dette gir  $y = \frac{V}{x^2} = \frac{V}{(\sqrt[3]{V})^2} = \sqrt[3]{V}$ . Vi har altså ett stasjonært punkt  $(\sqrt[3]{V}, \sqrt[3]{V})$ .

La oss regne ut de annenderiverte:

$$\frac{\partial^2 A}{\partial x^2} = \frac{4V}{x^3}; \quad \frac{\partial^2 A}{\partial y \partial x} = 2; \quad \frac{\partial^2 A}{\partial y^2} = \frac{4V}{y^3}$$

Dette gir

$$\frac{\partial^2 A}{\partial x^2}(\sqrt[3]{V}, \sqrt[3]{V}) = 4$$

$$\frac{\partial^2 A}{\partial y \partial x}(\sqrt[3]{V}, \sqrt[3]{V}) = 2$$

$$\frac{\partial^2 A}{\partial y^2}(\sqrt[3]{V}, \sqrt[3]{V}) = 4$$

og  $D = \begin{vmatrix} 4 & 2 \\ 2 & 4 \end{vmatrix} = 16 - 4 = 12$ . Følgelig er  $(\sqrt[3]{V}, \sqrt[3]{V})$  et lokalt minimum.  
Legg merke til at siden

$$z = \frac{V}{xy} = \frac{V}{\sqrt[3]{V} \cdot \sqrt[3]{V}} = \sqrt[3]{V}$$

svarer dette lokale minimumet til at vi lar kassen være en kube (alle sider like lange) med overflateareal

$$A = 6 \cdot V^{2/3}$$

La oss oppsummere våre resultater så langt: Vi har vist at funksjonen  $A$  bare har ett stasjonært punkt, og det er et lokalt minimum for  $x = \sqrt[3]{V}$  og  $y = \sqrt[3]{V}$ . Dersom det finnes et globalt minimum, må dette være i punktet  $(\sqrt[3]{V}, \sqrt[3]{V})$ . Mange vil nok slå seg til ro med at dette betyr at  $(\sqrt[3]{V}, \sqrt[3]{V})$  er et globalt minimum, men det finnes faktisk en annen mulighet – det kan tenkes at  $A$  nærmer seg en lavere “minimalverdi” enn  $6V^{2/3}$  uten noen gang å nå frem til den. Vi skal nå vise at dette ikke kan skje og at  $A_0 = 6V^{2/3}$  faktisk er den minste verdien arealet kan ha.

Vi tar utgangspunkt i at uttrykket for arealet

$$A(x, y) = 2xy + \frac{2V}{y} + \frac{2V}{x}$$

består av tre positive ledd  $2xy$ ,  $\frac{2V}{y}$  og  $\frac{2V}{x}$ . Skal arealet være mindre enn  $A_0$ , må i hvert fall hvert av disse leddene være mindre enn  $A_0$ . Begynner vi bakfra, ser vi at skal  $\frac{2V}{x}$  være mindre enn  $A_0$ , må

$$x > \frac{2V}{A_0} = \frac{1}{3}V^{1/3},$$

dvs. punktet  $(x, y)$  må ligge til høyre for den vertikale linjen  $x = \frac{1}{3}V^{1/3}$ . Tilsvarende ser vi at skal  $\frac{2V}{y}$  være mindre enn  $A_0$ , må

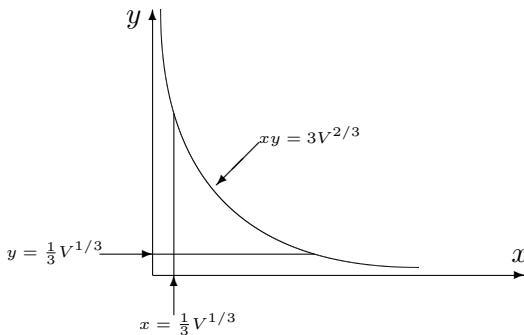
$$y > \frac{1}{3}V^{1/3}$$

dvs. punktet  $(x, y)$  må ligge over for den horisontale linjen  $y = \frac{1}{3}V^{1/3}$ . Til slutt ser vi at skal  $2xy$  være mindre enn  $A_0$ , må

$$xy \leq \frac{A_0}{2} = 3V^{2/3}$$

dvs. punktet  $(x, y)$  må ligge under hyperbelen  $xy = 3V^{2/3}$ .

Kombinerer vi disse kravene, ser vi at  $(x, y)$  må ligge i det avgrensede området på figur 6 dersom det skal være noe håp om at  $A(x, y) < A_0 = 6V^{2/3}$ . Vi legger også merke til at på randen av det avgrensede området vil  $A(x, y) > A_0$  – her er nemlig ett av de tre leddene  $2xy$ ,  $\frac{2V}{y}$  og  $\frac{2V}{x}$  lik  $A_0$  mens de to andre er positive.



Figur 6

Inkluderer vi randen, er det avgrensede området lukket og begrenset. Ifølge ekstremalverdisetningen i forrige seksjon har  $A$  et (globalt) minimumspunkt i dette området. Dette minimumspunktet kan ikke ligge på randen siden verdien på randen hele tiden er større enn verdien  $A_0$  i det indre punktet  $(\sqrt[3]{V}, \sqrt[3]{V})$ . Altså må minimumspunktet ligge i det indre, og ifølge setning 5.9.2 må det være et stasjonært punkt. Siden det eneste stasjonære punktet er  $(\sqrt[3]{V}, \sqrt[3]{V})$ , må dette være et globalt minimum for  $A$  på det avgrensede området. Siden funksjonsverdien utenfor dette området alltid er større enn  $A(\sqrt[3]{V}, \sqrt[3]{V})$ , må  $(\sqrt[3]{V}, \sqrt[3]{V})$  være det globale minimumet for  $A$  på hele definisjonsområdet. ♣

Det siste resonnementet i eksemplet ovenfor er ganske komplisert, og i praktiske oppgaver hender det ofte at man utelater argumenter av denne typen. Istedent argumenterer man for at det ut i fra oppgavens praktiske tolkning må finnes et maksimum eller minimum. I eksemplet ovenfor virker det imidlertid ikke så lett å gi et slikt argument.

### Oppgaver til seksjon 5.9

**1.** Finn de stasjonære punktene til funksjonen:

$$\begin{array}{ll} \text{a)} & f(x, y) = x^2 + 2y^2 - 4x + 4y \\ \text{b)} & f(x, y) = x^2 + y^2 - xy \\ \text{c)} & f(x, y) = x^2 + 2xy - 2y^2 + x + 7y \end{array} \quad \begin{array}{ll} \text{d)} & f(x, y) = xe^{y^2+x} \\ \text{e)} & f(x, y) = xy - \frac{1}{x} + \frac{2}{y} \end{array}$$

**2.** Finn de stasjonære punktene og avgjør om de er lokale maksimums-, minimums- eller sadelpunkter:

$$\begin{array}{l} \text{a)} \quad f(x, y) = x^2 + y^2 - 2x + 4y \\ \text{b)} \quad f(x, y) = x^2y^2 - 4xy + 6x - 6y \\ \text{c)} \quad f(x, y) = e^{x^2+3y^2} \\ \text{d)} \quad f(x, y) = \frac{1}{1-x+y+x^2+y^2} \\ \text{e)} \quad f(x, y) = x + \ln(x^2 + y^2) \end{array}$$

**3.** Finn det stasjonære punktet til funksjonen

$$f(x, y) = 3x^2 + 2xy + 2y^2 - 2x + 6y$$

og avgjør om det er et lokalt maksimumspunkt, minimimumspunkt eller sadelpunkt

**4.** Finn de stasjonære punktene til funksjonen

$$f(x, y) = x^3 + 5x^2 + 3y^2 - 6xy$$

og avgjør om de er lokale maksimumspunkter, minimimumspunkter eller sadelpunkter.

**5.** Finn de stasjonære punktene til funksjonen

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

og avgjør om de er lokale maksimumspunkter, minimimumspunkter eller sadelpunkter.

**6.** Finn de stasjonære punktene til funksjonen

$$f(x, y) = (x + y^2)e^x$$

og avgjør om de er lokale maksimumspunkter, minimimumspunkter eller sadelpunkter.

**7.** Finn de stasjonære punktene til funksjonen

$$f(x, y, z) = xyz - x^2 - y^2 - z^2$$

og avgjør om de er lokale maksimumspunkter, minimimumspunkter eller sadelpunkter.

**8.** (Eksamensoppgave fra MAT1110, 13/6, 2007)

- a) Finn de stasjonære punktene til  $f(x, y) = 2x^2y + 4xy - y^2$ .

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

**9.** (Eksamens i MAT1110, 13/8, 2007)

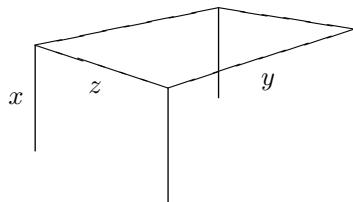
- a) Finn de stasjonære punktene til  $f(x, y) = (x^2 + y^2)e^x$ .  
 b) Avgjør om de stasjonære punktene er lokale maksimumspunkter, lokale minimumspunkter eller sadelpunkter.

**10.** La  $f(x, y) = (x^2 - y^2)e^{-\frac{x^2+y^2}{2}}$ .

- a) Finn de stasjonære punktene til  $f$  og avgjør om de er lokale maksimumspunkter, lokale minimumspunkter eller sadelpunkter.  
 b) Finn maksimum og minimum til  $f$  på området  $\{(x, y) \mid |x| \leq 1, |y| \leq 3\}$ . Bruk MATLAB eller en lommeregner til å tegne grafen.

**11.** Du skal lage en boks med volum  $V$ . Boksen skal ha bunn og fire sideflater, men ingen topp. Hvordan skal du lage boksen for at overflatearealet skal bli minst mulig?

**12.** Du skal lage en ramme av stålrør som skal brukes som reisverk til et telt. Rammen består av fire bein med lengde  $x$  festet til et rektangel med sider  $y$  og  $z$ , se figur under. Lengdene  $x, y$  og  $z$  måles i meter.



Volumet  $V = xyz$  av teltet skal være  $500 \text{ m}^3$ . Din oppgave er å lage teltet slik at den totale lengden  $L$  av stålrør som går med blir minst mulig.

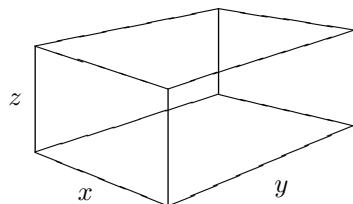
- a) Begrunn at lengden  $L$  kan skrives som

$$L(x, y) = 4x + 2y + \frac{1000}{xy}$$

og finn de partiellderiverte  $\frac{\partial L}{\partial x}$  og  $\frac{\partial L}{\partial y}$ .

- b) Bestem de dimensjonene av teltet som gjør totallengden av stålrør minst mulig.

**13.** Vi skal bygge en rettvinklet kasse uten lokk, se figur nedenfor. Kassen skal ha sidelengder  $x$  og  $y$  og høyde  $z$ . Selve "skjelettet" til kassen skal lages av 12 tynne rør (markert med streker på figuren). Den totale lengden rør vi skal benytte er 56 meter.



- a) Begrunn at arealet  $A$  av kassens utside (de fire veggene pluss bunnen) som funksjon av  $x$  og  $y$  kan skrives

$$A(x, y) = 28x + 28y - 2x^2 - 2y^2 - 3xy.$$

og finn de partiellederiverte av  $A$ . Bestem deretter eventuelle punkter  $(x, y)$  der begge de partiellederiverte er null, og avgjør om disse punktene er lokale minimumspunkter for  $A$ , lokale maksimumspunkter for  $A$  eller ingen av delene.

- b) Finn maksimumsverdien for  $A(x, y)$  på området i  $xy$ -planet gitt ved  $0 \leq x \leq 14$ ,  $0 \leq y \leq 14$ . Hvordan bør sidelengdene  $x$  og  $y$  velges for at arealet av kassens utside skal bli størst mulig? (Begrunn svaret.)

- 14.** (Eksamensoppgave i MAT 1100, 9/12, 2005) Oslo kommune planlegger å bygge et akvarium med volum  $5000 \text{ m}^3$ . Kostnadene er gitt ved:

Fronten – en glassplate: 1000 kr. per  $\text{m}^2$ .

Sidekantene – 3 stk. i stål: 300 kr. per  $\text{m}^2$

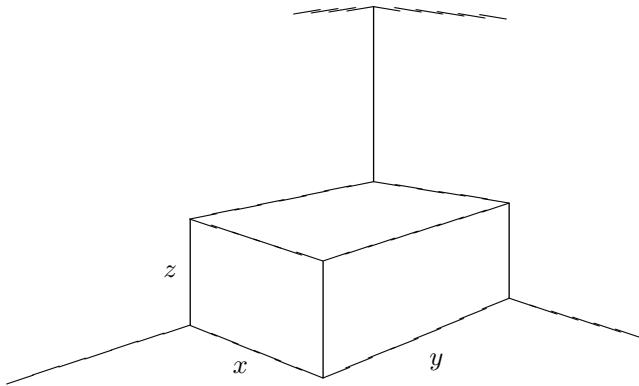
Bunnen – i sement: 500 kr. per  $\text{m}^2$

Anta glassplaten har lengde  $l$  og høyde  $h$ . Forklar hvorfor materialene koster

$$f(l, h) = 100 \left( 13lh + \frac{30000}{l} + \frac{25000}{h} \right)$$

Finn  $l$  og  $h$  som minimaliserer materialkostnadene.

- 15.** Et baderom har takhøyde 3 m og et kvadratisk gulv som er  $4 \text{ m}^2$ . I dette rommet skal vi plassere et badekar av lengde  $x$ , bredde  $y$  og høyde  $z$  (målt i meter) og med volum  $xyz = \frac{2}{3} \text{ m}^3$ . Badekaret skal plasseres i det ene hjørnet av baderommets romskjegge, se tegningen.



Figur 4

Ut fra dimensjonene på rommet og karet får vi:  $0 < x \leq 2$ ,  $0 < y \leq 2$ ,  $0 < z \leq 3$  (og derfor også  $xy = \frac{2}{3z} \geq \frac{2}{9} (\text{m})^2$ ). Vi skal flislegge de to veggene badekaret

berører, samt badegulvet, men vi flislegger bare de delene av veggene og gulvet som badekaret ikke dekker. Vi bruker to forskjellige typer fliser til veggene og gulvet. Prisen på veggflisene er 90 kr/m<sup>2</sup> og på gulvflisene 60 kr/m<sup>2</sup>. La  $P(x, y)$  betegne totalprisen på flisene som funksjon av  $x$  og  $y$  angitt i kr.

- a) Vis at vi får  $P(x, y) = 1320 - \frac{60}{x} - \frac{60}{y} - 60xy$ , og beregn  $\frac{\partial P}{\partial x}$  og  $\frac{\partial P}{\partial y}$ .
- b) For å få oversikt over utgiftene til flisleggingen ønsker vi å finne den verdien av  $(x, y)$  som gjør at  $P(x, y)$  blir størst mulig. Finn denne verdien av  $(x, y)$  og den tilsvarende verdien til  $P$ .

**16.** Det amerikanske postvesenet ekspederer bare pakker der summen av lengde, bredde og høyde er mindre enn 108 tommer. Hva er det største volumet en kasseformet pakke kan ha?

**17.** En fabrikk produserer to modeller av en vare. Det koster 400 kr. å lage standardmodellen og 600 kr. å lage luksusmodellen. Undersøkelser viser at når utsalgsprisene for standardmodellen og luksusmodellen er hhv.  $x$  og  $y$  kroner, så får fabrikken solgt  $5(y - x)$  eksemplarer av standardmodellen og  $450000 + 500(x - 2y)$  eksemplarer av luksusmodellen. Hvordan skal prisene settes for å maksimere fortjenesten?

**18.** To bedrifter konkurrerer om å selge nesten identiske varer i samme marked. En økning i produksjonen hos den ene bedriften fører derfor til svikt i inntektene hos den andre. Hvis bedrift  $A$  produserer  $x$  enheter per måned, og bedrift  $B$  produserer  $y$  enheter per måned, er de månedlige fortjenestene gitt ved

$$P = 12000x - \frac{x^2}{2} - \frac{y^2}{4} \quad \text{for bedrift } A$$

$$Q = 12000y - \frac{y^2}{2} - \frac{x^2}{6} \quad \text{for bedrift } B$$

- a) Dersom bedriftene ikke samarbeider om å fastsette produksjon, er det naturlig å anta at hver av bedriftene uavhengig av den andre fastsetter sin produksjon slik at egen fortjeneste blir så stor som mulig. Dessuten antar hver av bedriftene at den andre gjør det samme. Forklar hvorfor produksjonsnivået  $(x, y)$  er løsningen av ligningssystemet

$$\frac{\partial P}{\partial x} = 0 \quad \text{og} \quad \frac{\partial Q}{\partial y} = 0$$

Finn fortjenestene til  $A$  og  $B$  i dette tilfellet.

- b) Bedriftsledelsen i de to bedriftene tror at ved å samarbeide om produksjonsnivået kan de øke den totale fortjenesten til bedriftene. Forklar at det optimale produksjonsnivået nå er løsningen av ligningssystemet

$$\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial x} = 0 \quad \text{og} \quad \frac{\partial P}{\partial y} + \frac{\partial Q}{\partial y} = 0$$

Finn fortjenestene til  $A$  og  $B$  i dette tilfellet.

- c) Anta at bedriftene i hemmelighet har samarbeidet om å fastsette produksjonsnivået. Etter en stund oppdager bedrift  $B$ , som tidligere var mest lønnsom, at bedrift  $A$  nå er blitt markedsleder. Bedrift  $B$  bestemmer seg derfor for å bryte avtalen uten å si fra til  $A$ . Gitt at  $A$  fastholder sitt produksjonsnivå fra b), hvordan skal  $B$  velge sitt produksjonsnivå for å få størst mulig fortjeneste? Hva blir fortjenestene til  $A$  og  $B$  i dette tilfellet?

**19.** I denne oppgaven er  $f(x, y) = x^4 + y^4$ .

- a) Vis at  $(0,0)$  er et stasjonært punkt der Hesse-determinanten  $D$  er lik null.  
Vis at  $(0,0)$  er et minimumspunkt for funksjonen.
- b) Lag en funksjon  $g(x, y)$  der  $(0,0)$  er et maksimumspunkt, men hvor Hesse-determinanten i  $(0,0)$  er lik null.
- c) Lag en funksjon  $h(x, y)$  slik at  $(0,0)$  er et sadelpunkt, men hvor Hesse-determinanten i  $(0,0)$  er lik null.

**20.** I denne oppgaven skal vi se på noen viktige forskjeller mellom funksjoner av henholdsvis én og flere variable.

- a) La  $f: \mathbb{R} \rightarrow \mathbb{R}$  være en deriverbar funksjon av én variabel og anta at det eneste stasjonære punktet til  $f$  er et lokalt maksimum i  $a$ . Vis at da er  $a$  et globalt maksimum for  $f$ .

- b) La

$$f(x, y) = 1 - x^2 - (1 + x)^3 y^2$$

Vis at  $(0,0)$  er det eneste stasjonære punktet til  $f$ .

- c) Vis at  $(0,0)$  er et lokalt maksimum, men ikke et globalt maksimum for  $f$ . Bruk MATLAB til å tegne grafen til  $f$ , og tenk gjennom forskjellen på én og flere dimensjoner.

- d) La  $g: \mathbb{R} \rightarrow \mathbb{R}$  være en kontinuerlig funksjon av én variabel og anta at  $a$  og  $b$  er to lokale maksimumspunkter for  $f$ . Vis at det finnes et lokalt minimumspunkt mellom  $a$  og  $b$ .

- e) La

$$g(x, y) = 4x^2 e^y - 2x^4 - e^{4y}$$

Vis at de stasjonære punktene til  $g$  er  $(-1, 0)$  og  $(1, 0)$ .

- f) Vis at begge de to stasjonære punktene til  $g$  er lokale maksimumspunkter. Bruk MATLAB til å tegne grafen til  $g$ , og tenk gjennom forskjellen på én og flere dimensjoner.

## 5.10 Lagranges multiplikatormetode

I forrige seksjon så vi hvordan vi kan finne de lokale maksimums- og minimumspunktene til en funksjon  $f(x_1, x_2, \dots, x_m)$  av flere variable når  $x_1, x_2, \dots, x_m$  får lov til å ha alle verdiene i definisjonsområdet til  $f$ . Vi skal nå se hva som skjer når vi har begrensninger (såkalte *betingelser*) på variablene.

I de enkleste problemene av denne typen har vi to funksjoner  $f(x, y)$  og  $g(x, y)$ , og vi ønsker å finne den største og/eller minste verdien til  $f(x, y)$  blant de punktene som tilfredsstiller *bibetingelsen*  $g(x, y) = b$ , der  $b$  er en gitt konstant. La oss begynne med et enkelt (men langt!) eksempel.

**Eksempel 1:** Vi skal finne maksimums- og minimumsverdien til funksjonen  $f(x, y) = xy$  på sirkelen  $x^2 + y^2 = 1$ . Setter vi  $g(x, y) = x^2 + y^2$ , ser vi at dette er en bibetingelse  $g(x, y) = 1$  av typen vi beskrev ovenfor. Det er flere måter å løse dette problemet på. Den mest naturlige er kanskje å løse ligningen  $x^2 + y^2 = 1$  for  $y$  og sette inn (substituere) resultatet i  $f$ . Da får vi to optimeringsproblemer med én variabel,

$$h(x) = x\sqrt{1 - x^2}$$

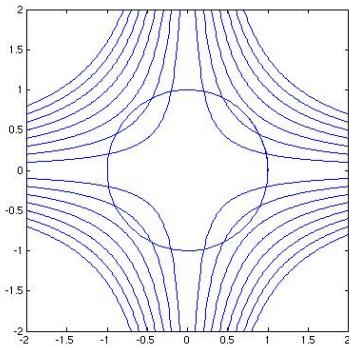
for øvre halvsirkel, og

$$k(x) = -x\sqrt{1 - x^2}$$

for nedre halvsirkel. Det er lett å finne maksimumspunktene til disse funksjonene ved vanlige metoder.

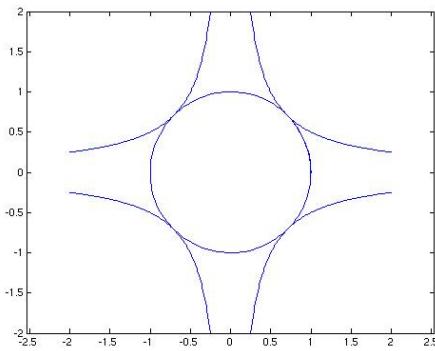
Selv om “substitusjonsmetoden” fungerer bra i dette eksemplet, har den en fundamental svakhet. Dersom bibetingelsen er mer komplisert enn  $x^2 + y^2 = 1$ , klarer vi ikke å løse ligningen for én av variablene, og hele forsøket vårt bryter sammen. Vi ønsker derfor å finne en metode som ikke er basert på at vi løser ligninger og substituerer.

I figur 1 har vi tegnet opp punktene som tilfredsstiller bibetingelsene (sirkelen) sammen med noen av nivåkurvene til funksjonen  $f(x, y) = xy$ . Nivåkurvene tilsvarer funksjonsverdiene fra -1.6 til 1.6 med trinn på 0.2. Absoluttverdien til funksjonen vokser med  $x$  og  $y$ , så det er de ytterste nivåkurvene som svarer til høye positive og negative funksjonsverdier (legg merke til at  $f$  har positive verdier i første og tredje kvadrant, og negative verdier i annen og fjerde kvadrant).



Figur 1: Nivåkurver og bibetingeskurve

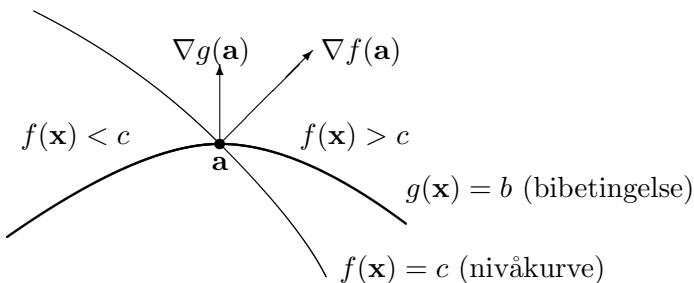
Vi ser at det er noen nivåkurver som ikke skjærer bibetingeskuren i det hele tatt — de tilsvarer verdier som funksjonen ikke kan ha så lenge vi innskrenker oss til punkter på sirkelen. Nivåkurver som skjærer sirkelen, tilsvarer verdier som funksjonen har på sirkelen. De største og minste verdiene får vi når nivåkurvene bare berører sirkelen og går ut igjen. Figur 2 viser denne situasjonen.



Figur 2: Optimale nivåkurver

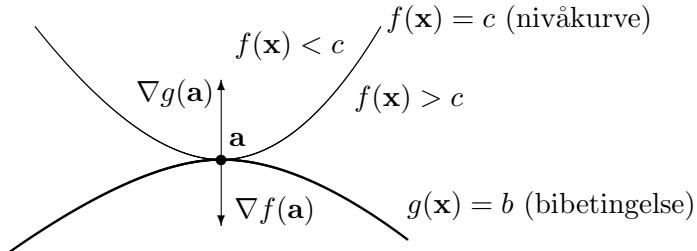
Nivåkurvene i figur 2 tilsvarer verdiene  $\frac{1}{2}$  (i første og tredje kvadrant) og  $-\frac{1}{2}$  (i annen og fjerde kvadrant), så maksimumsverdien til funksjonen på sirkelen er  $\frac{1}{2}$  og minimumsverdien er  $-\frac{1}{2}$ .

Legg merke til at nivåkurvene i figur 2 tangerer bibetingeskuren. Det er lett å innse at dette er et fenomen som ikke bare gjelder i dette eksemplet, men som gjelder generelt når man skal optimere en funksjon  $f(\mathbf{x}, \mathbf{y})$  under en bibetingelse  $g(\mathbf{x}, \mathbf{y}) = b$  — dersom nivåkurven *krysser* bibetingeskuren, vil vi normalt ha større verdier på den ene siden av skjæringspunktet og mindre på den andre (se figur 3).



Figur 3:  $\mathbf{a}$  er ikke et ekstremalpunkt under bibetingelsen  $g(\mathbf{x}) = b$

Dette betyr at når vi leter etter våre maksimums- og minimumspunkter, så må vi lete etter punkter der nivåkurven *tangerer* bibetingeskuren, eller — sagt med andre ord — der normalen til nivåkurven er parallel med normalen til bibetingeskuren (se figur 4).



Figur 4:  $\mathbf{a}$  er et minimumspunkt under bibetingelsen  $g(\mathbf{x}) = b$

Disse normalene er lette å finne siden vi vet at gradienter alltid står normalt på nivåkurver (husk setning 3.7.2), og bibetingelseskuren er en nivåkurve for funksjonen  $g(x, y)$ . Vi leter altså etter punkter der de to gradientene  $\nabla f(x, y)$  og  $\nabla g(x, y)$  er parallelle, dvs. punkter der det finnes et tall  $\lambda$  slik at  $\nabla f(x, y) = \lambda \nabla g(x, y)$ .

I eksemplet vårt er

$$\nabla f(x, y) = \begin{pmatrix} y \\ x \end{pmatrix} \quad \text{og} \quad \nabla g(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

og vi er på jakt etter punkter der disse er parallelle, dvs. punkter der det finnes et tall  $\lambda$  slik at

$$\begin{pmatrix} y \\ x \end{pmatrix} = \lambda \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

Skriver vi ut denne ligningen komponentvis, får vi

$$y = 2\lambda x$$

$$x = 2\lambda y$$

Dette gir oss to ligninger med tre ukjente,  $x$ ,  $y$  og  $\lambda$ . Den nye ukjente  $\lambda$  som har sneket seg inn i regnestykket, kalles en *Lagrangemultiplikator* og har gitt navn til hele metoden. I tillegg til ligningene ovenfor har vi en tredje ligning siden punktet vårt må tilfredsstille bibetingelsen:

$$x^2 + y^2 = 1$$

Dette ligningssystemet med tre ligninger og tre ukjente kan løses på flere måter. La oss først observere at ingen av de ukjente  $x$ ,  $y$  kan være 0, for hvis den ene er det, må den andre også være det, og da får vi ikke oppfylt ligningen  $x^2 + y^2 = 1$ . Dette medfører at heller ikke  $\lambda$  kan være 0. Dermed kan vi dele den første av ligningene våre på den andre, og få

$$\frac{y}{x} = \frac{x}{y}$$

som gir  $y^2 = x^2$ . Setter vi dette inn i den tredje ligningen, får vi  $2x^2 = 1$  som gir  $x = \pm\frac{\sqrt{2}}{2}$ . Siden  $y^2 = x^2$ , får vi også  $y = \pm\frac{\sqrt{2}}{2}$ . Dermed har vi fire punkter vi må se videre på:  $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ ,  $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ ,  $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$  og  $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ . Setter vi inn i funksjonen  $f(x, y) = xy$ , får vi

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

og

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

Dette stemmer svært godt med våre grafiske undersøkelser ovenfor, og det er derfor rimelig å tro at vi har en maksimumsverdi  $\frac{1}{2}$  som oppnås i punktene  $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$  og  $(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$  og en minimumsverdi  $-\frac{1}{2}$  som oppnås i punktene  $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$  og  $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2})$ . ♣

La oss oppsummere eksemplet ovenfor i litt mer generelle vendinger. Vi har en funksjon  $f(x, y)$  som vi ønsker å maksimere eller minimere under bibetingelsen  $g(x, y) = b$ , der  $b$  er en konstant. Da må vi lete etter punkter på bibetingelseskurven der  $\nabla f(x, y) = \lambda \nabla g(x, y)$ . Skriver vi ut denne ligningen komponentvis, får vi to ligninger med to ukjente

$$\frac{\partial f}{\partial x}(x, y) = \lambda \frac{\partial g}{\partial x}(x, y)$$

$$\frac{\partial f}{\partial y}(x, y) = \lambda \frac{\partial g}{\partial y}(x, y)$$

I tillegg har vi bibetingelsen

$$g(x, y) = b$$

slik at vi får tre ligninger med tre ukjente. Løser vi dette ligningssystemet, vil vi (under svært generelle betingelser) ha funnet alle potensielle maksimums- og minimumspunkter for problemet vårt.

Vi kan generalisere enda litt lenger. Anta at vi har en funksjon

$$f(x_1, x_2, \dots, x_m)$$

av  $m$  variable og en bibetingelse

$$g(x_1, x_2, \dots, x_m) = b$$

Setter vi opp den samme ligningen  $\nabla f(x_1, x_2, \dots, x_m) = \lambda \nabla g(x_1, x_2, \dots, x_m)$  som før og skriver den ut komponentvis, får vi  $m$  ligninger med  $m+1$  ukjente

$x_1, x_2, \dots, x_m, \lambda$ :

$$\begin{aligned}\frac{\partial f}{\partial x_1}(x_1, x_2, \dots, x_m) &= \lambda \frac{\partial g}{\partial x_1}(x_1, x_2, \dots, x_m) \\ \frac{\partial f}{\partial x_2}(x_1, x_2, \dots, x_m) &= \lambda \frac{\partial g}{\partial x_2}(x_1, x_2, \dots, x_m) \\ &\vdots \quad \vdots \quad \vdots \\ \frac{\partial f}{\partial x_m}(x_1, x_2, \dots, x_m) &= \lambda \frac{\partial g}{\partial x_m}(x_1, x_2, \dots, x_m)\end{aligned}$$

Legger vi til bibetingelsen

$$g(x_1, x_2, \dots, x_m) = b$$

har vi  $m + 1$  ligninger med  $m + 1$  ukjente. Igjen viser det seg at dette ligningssystemet gir oss alle mulige maksimums- og minimumspunkter.

**Bemerkning:** Ser du i litteraturen, vil du finne at bibetingelsene formuleres litt forskjellig — i noen bøker finner du alltid formen  $g(x_1, x_2, \dots, x_m) = 0$ , mens andre tillater  $g(x_1, x_2, \dots, x_m) = b$  for en vilkårlig  $b \in \mathbb{R}$ . Egentlig er det ikke noen forskjell på disse formene — har vi et krav av typen  $g(x_1, x_2, \dots, x_m) = b$ , innfører vi bare en ny funksjon

$$\tilde{g}(x_1, x_2, \dots, x_m) = g(x_1, x_2, \dots, x_m) - b,$$

og dermed har vi en bibetingelse av typen

$$\tilde{g}(x_1, x_2, \dots, x_m) = 0$$

Siden  $\nabla \tilde{g} = \nabla g$ , blir betingelsene for maksimums-/minimumspunkt uforandret. Vi skal derfor formulere resultatene våre for tilfellet  $g(x_1, x_2, \dots, x_m) = b$ , men i bevisene nøye oss med å se på tilfellet  $g(x_1, x_2, \dots, x_m) = 0$  som ofte er notasjonsmessig enklere.

**Teorem 5.10.1 (Lagranges multiplikatormetode med én bibetingelse)** Anta at  $U$  er en åpen delmengde av  $\mathbb{R}^m$ , og at  $f, g : U \rightarrow \mathbb{R}$  er to funksjoner med kontinuerlige partiellderiverte. La  $b$  være et reelt tall, og anta at  $\bar{x} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)$  er et lokalt maksimums- eller minimumspunkt for  $f$  på mengden

$$A = \{\mathbf{x} \in U \mid g(\mathbf{x}) = b\}$$

Da er enten  $\nabla g(\bar{x}) = 0$ , eller det finnes en konstant  $\lambda \in \mathbb{R}$  slik at

$$\nabla f(\bar{x}) = \lambda \nabla g(\bar{x})$$

Legg merke til at det i teoremet er kommet inn en ekstra mulighet som vi ikke har hatt med tidligere, nemlig at  $\nabla g(\bar{\mathbf{x}}) = 0$ . Dette er bare naturlig — dersom  $\nabla g(\bar{\mathbf{x}}) = 0$ , bryter vårt geometriske resonnement sammen, og hva som helst kan hende. Vi understreker også at teoremet bare hjelper oss å finne *potensielle* maksimums- og minimumspunkter — et punkt som tilfredsstiller betingelsene, behøver ikke å være noen av delene, men kan være et generalisert sadelpunkt.

Før vi beviser teoremet, tar vi med et eksempel på bruken.

**Eksempel 2:** Vi skal finne minimumsverdien til funksjonen

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

under bibetingelsen  $x^2 + 4y^2 - z = 0$ . Legg merke til at problemet har en geometrisk tolkning — vi ønsker å finne det punktet  $(x, y, z)$  på flaten  $z = x^2 + 4y^2$  som har kortest avstand til punktet  $(3, 0, 0)$ .

Lar vi  $g(x, y, z) = x^2 + 4y^2 - z = 0$ , ser vi at

$$\nabla g(x, y, z) = \begin{pmatrix} 2x \\ 8y \\ -1 \end{pmatrix}$$

Siden  $\nabla g$  aldri er null, slipper vi å bry oss om tilfellet  $\nabla g(\bar{\mathbf{x}}) = 0$ . Vi ser videre at

$$\nabla f(x, y, z) = \begin{pmatrix} 2x - 6 \\ 2y \\ 2z \end{pmatrix}$$

Skriver vi ligningen  $\nabla f(x, y, z) = \lambda \nabla g(x, y, z)$  på komponentform, får vi (etter å ha forkortet litt)

$$\begin{aligned} x - 3 &= \lambda x \\ y &= 4\lambda y \\ 2z &= -\lambda \end{aligned}$$

I tillegg har vi bibetingelsen

$$x^2 + 4y^2 - z = 0$$

En av utfordringene ved Lagranges multiplikatormetode er å løse ligningene vi kommer frem til. De kan være av forskjellig type, og det er ikke lett å gi generelle råd om hvordan det er lurt å gå frem. I dette tilfellet ser det ut til å være larest å starte med ligning nummer to,  $y = 4\lambda y$ . Her er det to muligheter. Dersom  $y \neq 0$ , må  $\lambda = \frac{1}{4}$ . Dersom  $y = 0$ , kan derimot  $\lambda$  være hva som helst. Vi ser på disse tilfellene hver for seg:

Tilfellet  $\lambda = \frac{1}{4}$ : Den øverste ligningen blir nå til  $x - 3 = \frac{1}{4}x$ , som gir  $x = 4$ , og den tredje ligningen gir  $z = -\frac{1}{8}$ . Setter vi dette inn i den nederste ligningen, får vi

$$16 + 4y^2 + \frac{1}{8} = 0$$

som åpenbart ikke har noen løsning. Tilfellet  $\lambda = \frac{1}{4}$  fører derfor ikke frem.

Tilfellet  $y = 0$ : Vi sitter nå igjen med tre ligninger for  $x$ ,  $z$  og  $\lambda$ , nemlig

$$\begin{aligned} x - 3 &= \lambda x \\ 2z &= -\lambda \\ x^2 - z &= 0 \end{aligned}$$

Eliminerer vi  $z$  fra de to siste, ser vi at  $\lambda = -2x^2$ , og setter vi dette inn i den øverste ligningen, sitter vi igjen med  $x - 3 = -2x^3$ , dvs.

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

Vi ser at  $x = 1$  er en løsning av denne ligningen. For å undersøke om det finnes flere løsninger, polynomdividerer vi  $2x^3 + x - 3$  med  $x - 1$ , og får  $2x^2 + 2x + 3$  som ikke har reelle røtter. Dermed har vi bare én løsning for  $x$ , nemlig  $x = 1$ . Siden  $x^2 - z = 0$ , følger det at  $z = 1$ , (det følger også at  $\lambda = -2$ , men  $\lambda$  er vi egentlig ikke interessert i).

Vi har dermed sett at den eneste løsningen av ligningssystemet er  $x = 1, y = 0, z = 1$ . Siden den geometriske tolkningen forteller oss at funksjonen må ha et minimumspunkt, er det dette vi har funnet. ♣

Vi er nå klar til å bevise Lagranges multiplikatoreteorem for én bibetingelse. Ideen i beviset er den samme som vi hadde i begynnelsen av eksempel 1: vi løser ligningen  $g(x_1, x_2, \dots, x_m) = b$  for én av de variable og substituerer inn i  $f$ . På denne måten får vi et “fritt ekstremalverdiproblem” (dvs. et ekstremalverdiproblem uten bibetingelser). I praksis er dette en ubrukelig metode fordi vi ikke greier å løse ligningen  $g(x_1, x_2, \dots, x_m) = 0$ , men teoretisk fungerer den fordi implisitt funksjonsteorem forteller oss at det finnes en løsning med de egenskapene vi trenger.

*Bevis for teorem 5.10.1:* Som påpekt ovenfor er det nok å vise teoremet for  $b = 0$ . Vi skal vise at dersom  $\mathbf{x}$  er et lokalt maksimums- eller minimumspunkt og  $\nabla g(\bar{\mathbf{x}}) \neq 0$ , så finnes det en konstant  $\lambda$  slik at  $\nabla f(\bar{\mathbf{x}}) = \lambda \nabla g(\bar{\mathbf{x}})$ . Siden vi antar at  $\nabla g(\bar{\mathbf{x}}) \neq 0$ , finnes det minst én variabel  $x_i$  slik at  $\frac{\partial g}{\partial x_i}(\bar{\mathbf{x}}) \neq 0$ . Ved eventuelt å bytte om på variablene, kan vi anta at dette er den siste variabelen  $x_m$ . Siden denne variabelen kommer til å spille en litt spesiell rolle i beviset, bytter vi navn på den og kaller den  $y$  istedenfor  $x_m$  (dette er bare for å gjøre beviset lettere å lese). Vi har dermed funksjoner  $f(x_1, \dots, x_{m-1}, y)$  og  $g(x_1, \dots, x_{m-1}, y)$ , der  $\frac{\partial g}{\partial y}(\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y}) \neq 0$ .

Siden  $\frac{\partial g}{\partial y}(\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y}) \neq 0$ , finnes det ifølge implisitt funksjonsteorem (teorem 5.7.3) en deriverbar funksjon  $\phi$  definert i en omegn om  $(\bar{x}_1, \dots, \bar{x}_{m-1})$  slik at  $\phi(\bar{x}_1, \dots, \bar{x}_{m-1}) = \bar{y}$  og

$$g(x_1, \dots, x_{m-1}, \phi(x_1, \dots, x_{m-1})) = 0$$

i denne omegnen. Dette betyr at funksjonen

$$h(x_1, \dots, x_{m-1}) = f((x_1, \dots, x_{m-1}, \phi(x_1, \dots, x_{m-1})))$$

har et vanlig ekstremalpunkt (uten bibetingelser) i punktet  $(\bar{x}_1, \dots, \bar{x}_{m-1})$ . Dermed er alle de partiellderiverte  $\frac{\partial h}{\partial x_i}$  lik null i dette punktet. Bruker vi kjerneregelen, er dermed

$$\begin{aligned} 0 = \frac{\partial h}{\partial x_i}(\bar{x}_1, \dots, \bar{x}_{m-1}) &= \frac{\partial f}{\partial x_i}(\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y}) + \\ &+ \frac{\partial f}{\partial y}(\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y}) \frac{\partial \phi}{\partial x_i}(\bar{x}_1, \dots, \bar{x}_{m-1}) \end{aligned}$$

Implisitt funksjonsteorem gir oss at

$$\frac{\partial \phi}{\partial x_i}(\bar{x}_1, \dots, \bar{x}_{m-1}) = -\frac{\frac{\partial g}{\partial x_i}(\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y})}{\frac{\partial g}{\partial y}(\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y})}$$

Setter vi dette inn i den foregående ligningen (husk at  $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_{m-1}, \bar{y})$ ), får vi etter litt omgruppering

$$\frac{\partial f}{\partial x_i}(\bar{\mathbf{x}}) = \frac{\frac{\partial f}{\partial y}(\bar{\mathbf{x}})}{\frac{\partial g}{\partial y}(\bar{\mathbf{x}})} \frac{\partial g}{\partial x_i}(\bar{\mathbf{x}})$$

Setter vi  $\lambda = \frac{\frac{\partial f}{\partial y}(\bar{\mathbf{x}})}{\frac{\partial g}{\partial y}(\bar{\mathbf{x}})}$ , har vi dermed

$$\frac{\partial f}{\partial x_i}(\bar{\mathbf{x}}) = \lambda \frac{\partial g}{\partial x_i}(\bar{\mathbf{x}})$$

for  $i = 1, \dots, m-1$ . Det gjenstår å vise at formelen også holder for den siste variabelen, dvs. at  $\frac{\partial f}{\partial y}(\bar{\mathbf{x}}) = \lambda \frac{\partial g}{\partial y}(\bar{\mathbf{x}})$ . Dette er bare en triviell utregning:

$$\lambda \frac{\partial g}{\partial y}(\bar{\mathbf{x}}) = \frac{\frac{\partial f}{\partial y}(\bar{\mathbf{x}})}{\frac{\partial g}{\partial y}(\bar{\mathbf{x}})} \frac{\partial g}{\partial y}(\bar{\mathbf{x}}) = \frac{\partial f}{\partial y}(\bar{\mathbf{x}})$$

Dermed har vi vist at  $\nabla f(\bar{\mathbf{x}}) = \lambda \nabla g(\bar{\mathbf{x}})$ , og beviset er fullført.  $\square$

Lagranges multiplikatormetode har mange anvendelser. Vi skal komme tilbake til noen av disse senere i seksjonen, men for øyeblikket skal vi nøye

oss med å følge opp en problemstilling fra forrige seksjon. I den seksjonen viste vi hvordan man kan finne lokale maksimums- og minimumspunkter i *det indre* av et område ved å lete etter punkter der alle de partiellderiverte er null. Dersom området vi er interessert i er lukket, er det også mulig at noen av maksimums- eller minimumspunktene ligger på randen (husk eksempel 5 i seksjon 5.9). Disse kan vi finne ved hjelp av Lagranges multiplikatormetode. Vi illustrerer fremgangsmåten med et enkelt eksempel.

**Eksempel 3:** Vi skal finne maksimums- og minimumsverdien til funksjonen

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

på området

$$A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

Siden området er lukket og funksjonen er kontinuerlig, vet vi fra ekstremalverdisetningen (5.8.2) at det finnes (globale) maksimums- og minimumspunkter. Hvis et slikt punkt ligger i det indre av området, vet vi at de partiellderiverte må være null i punktet. Siden

$$\frac{\partial f}{\partial x} = 2x \quad \text{og} \quad \frac{\partial f}{\partial y} = -3y^2,$$

ser vi at det eneste stasjonære punktet er  $(0, 0)$  og at  $f(0, 0) = 0$ . Dette er vår første kandidat til tittelen som maksimums- og minimumspunkt. De andre kandidatene må ligge på randen

$$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\},$$

så vi bruker Lagranges multiplikatormetode med  $f(x, y) = x^2 - y^3$  og  $g(x, y) = x^2 + y^2$ . Vi har

$$\nabla f(x, y) = \begin{pmatrix} 2x \\ -3y^2 \end{pmatrix} \quad \text{og} \quad \nabla g(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$

og skriver vi ligningen  $\nabla f(x, y) = \lambda \nabla g(x, y)$  på komponentform, får vi (etter litt forkorting) ligningssystemet

$$\begin{aligned} x &= \lambda x \\ -3y^2 &= 2\lambda y \\ x^2 + y^2 &= 1 \end{aligned}$$

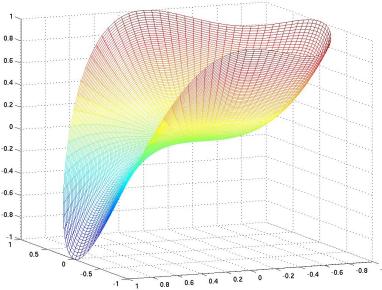
Den første ligningen kan oppfylles på to måter, enten er  $x = 0$  eller så er  $\lambda = 1$ . Vi ser på tilfellene hver for seg. Hvis  $x = 0$ , følger det fra den siste ligningen at  $y = \pm 1$ . Dette betyr at  $(0, \pm 1)$  er mulige ekstremalpunkter. Setter vi isteden  $\lambda = 1$ , får den andre ligningen i systemet formen  $-3y^2 = 2y$ . Denne ligningen har to løsninger,  $y = 0$  og  $y = -\frac{2}{3}$ . Setter vi disse løsningene

inn i den tredje ligningen, ser vi at  $y = 0$  gir  $x = \pm 1$  og at  $y = -\frac{2}{3}$  gir  $x = \pm \frac{\sqrt{5}}{3}$ .

Ialt har vi dermed sju kandidater:  $(0, 0)$ ,  $(0, \pm 1)$ ,  $(\pm 1, 0)$  og  $(\pm \frac{\sqrt{5}}{3}, -\frac{2}{3})$ . For å finne maksimum og minimum, regner vi ut alle funksjonsverdiene:

$$f(0, 0) = 0, f(0, \pm 1) = \mp 1, f(\pm 1, 0) = 1, f\left(\pm \frac{\sqrt{5}}{3}, -\frac{2}{3}\right) = \frac{22}{27}$$

Dette viser at maksimumsverdien 1 finner vi i punktene  $(0, -1)$ ,  $(\pm 1, 0)$ , mens minimumsverdien -1 finner vi i punktet  $(0, 1)$ .



Figur 5: Grafisk fremstilling av flaten  $f(x, y) = x^2 - y^3$

Figur 5 viser grafen. Du ser tydelig de tre maksimumspunktene og det ene minimumspunktet. Punktene  $(\pm \frac{\sqrt{5}}{3}, -\frac{2}{3})$  er lokale minimumspunkter når du går fra topp til topp langs randen. Det indre punktet  $(0, 0)$  er et sadelpunkt (men du kan ikke vise det ved annenderivertesten siden determinanten  $D$  er null).



**Bemerkning om Lagrangefunksjoner:** Vi skal straks gå over til å studere Lagranges metode for problemer med flere bibetingelser, men før vi gjør det, tar vi med en liten bemerkning om en annen måte å fremstille teorien ovenfor på. Selv om denne fremstillingen bare er en omforming av den vi hittil har brukt, ser de to beskrivelsene så forskjellige ut ved første øyekast at det kan være lurt å kjenne forbindelsen mellom dem.

Den alternative fremstillingen er basert på såkalte *Lagrangefunksjoner*. Dersom vi ønsker å optimere funksjonen  $f(x_1, \dots, x_m)$  under bibetingelsen  $g(x_1, \dots, x_m) = b$ , definerer vi *Lagrangefunksjonen*  $L$  ved

$$L(x_1, \dots, x_m, \lambda) = f(x_1, \dots, x_m) - \lambda(g(x_1, \dots, x_m) - b)$$

(legg merke til at  $\lambda$  nå er blitt en variabel på linje med  $x_1, \dots, x_m$ ). La oss glemme det opprinnelige problemet et øyeblikk, og heller finne de stasjonære

punktene til  $L$ . Regner vi ut de partiellderiverte, ser vi at:

$$\begin{aligned}\frac{\partial L}{\partial x_1} &= \frac{\partial f}{\partial x_i} - \lambda \frac{\partial g}{\partial x_1} \\ &\vdots \quad \vdots \quad \vdots \\ \frac{\partial L}{\partial x_m} &= \frac{\partial f}{\partial x_m} - \lambda \frac{\partial g}{\partial x_m} \\ \frac{\partial L}{\partial \lambda} &= -g(x_1, \dots, x_m) + b\end{aligned}$$

Setter vi disse uttrykkene lik 0, får vi akkurat de ligningene vi skal løse når vi bruker Lagranges multiplikatormetode — å bruke denne metoden er altså det samme som å finne de stasjonære punktene til Lagrangefunksjonen  $L$ . I en del bøker vil du derfor se at man løser optimeringsproblemer under bibetingelser ved å skrive opp en Lagrangefunksjon og så finne de stasjonære punktene til denne funksjonen.

Det er en ting til du bør være klar. Noen bøker bytter fortegn på  $\lambda$ -leddet i Lagrangefunksjonen og skriver den som

$$L(x_1, \dots, x_m, \lambda) = f(x_1, \dots, x_m) + \lambda(g(x_1, \dots, x_m) - b)$$

Dette spiller ingen rolle — maksimums- og minimumspunktene er de samme som før, det er bare  $\lambda$  som får omvendt fortegn av det den ellers ville ha fått.

### Lagranges multiplikatormetode med flere bibetingelser

Vi skal nå se på Lagranges multiplikatormetode når vi ønsker å optimere en funksjon

$$f(x_1, x_2, \dots, x_m)$$

under *flere* bibetingelser

$$\begin{aligned}g_1(x_1, x_2, \dots, x_m) &= b_1 \\ g_2(x_1, x_2, \dots, x_m) &= b_2 \\ &\vdots \quad \vdots \quad \vdots \\ g_k(x_1, x_2, \dots, x_m) &= b_k\end{aligned}$$

Normalt må vi ha  $k < m$  for å få et fornuftig ekstremalproblem, og vi skal derfor anta at dette alltid er tilfellet.

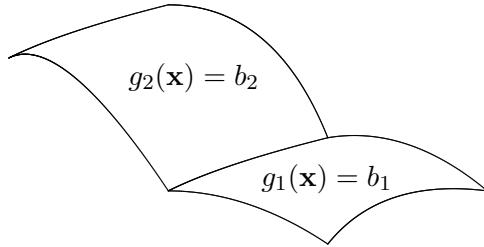
For å få en følelse for problemet ser vi først på tilfellet der vi ønsker å maksimere en funksjon

$$f(x, y, z)$$

av tre variable under to bibetingelser

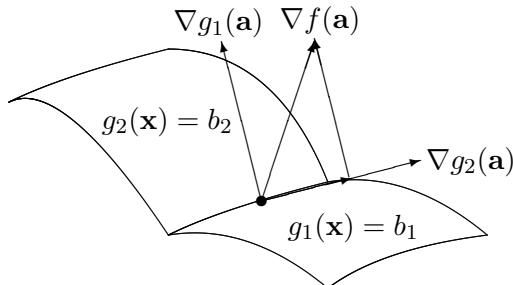
$$\begin{aligned} g_1(x, y, z) &= b_1 \\ g_2(x, y, z) &= b_2 \end{aligned}$$

De to ligningene  $g_1(x, y, z) = b_1$  og  $g_2(x, y, z) = b_2$  vil normalt definere to flater i rommet som skjærer hverandre langs en kurve (se figur 6). Problemet er altså å finne den største verdien til  $f$  langs denne kurven.



Figur 6: De to flatene  $g_1(\mathbf{x}) = b_1$  og  $g_2(\mathbf{x}) = b_2$  skjærer hverandre i en kurve

Husk at gradienten til  $f$  peker i den retningen hvor  $f$  vokser raskest. Dersom  $\nabla f$  ikke står normalt på kurven, er det rimelig å tro at funksjonen langs kurven stiger i den retningen hvor  $\nabla f$  peker. Skal vi derfor ha maksimum i et punkt, må  $\nabla f$  i dette punktet stå normalt på kurven, dvs. den må ligge i normalplanet til kurven. Dette normalplanet et utspekt av normalvektorene til flatene (prøv å forstå dette geometrisk!), og  $\nabla f$  må derfor være en lineærkombinasjon av normalvektorene  $\nabla g_1$  og  $\nabla g_2$  til de to flatene (se figur 7).



Figur 7:  $\nabla f(\mathbf{a})$  som lineærkombinasjon av  $\nabla g_1(\mathbf{a})$  og  $\nabla g_2(\mathbf{a})$

Vi venter derfor å finne maksimalverdien i et punkt  $(\bar{x}, \bar{y}, \bar{z})$  der det finnes konstanter  $\lambda_1$  og  $\lambda_2$  slik at

$$\nabla f(\bar{x}, \bar{y}, \bar{z}) = \lambda_1 \nabla g_1(\bar{x}, \bar{y}, \bar{z}) + \lambda_2 \nabla g_2(\bar{x}, \bar{y}, \bar{z})$$

Følgende teorem forteller oss at denne geometriske intuisjonen er riktig.

**Teorem 5.10.2 (Lagranges multiplikatormetode med flere bibetingelser)** *Anta at  $U$  er en åpen delmengde av  $\mathbb{R}^m$ , og at  $f, g_1, g_2, \dots, g_k : U \rightarrow \mathbb{R}$*

er funksjoner med kontinuerlige partiellderiverte. Dersom  $b_1, b_2, \dots, b_k$  er reelle tall og  $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)$  er et lokalt maksimums- eller minimumspunkt for  $f$  på mengden

$$A = \{\mathbf{x} \in U \mid g_1(\mathbf{x}) = b_1, g_2(\mathbf{x}) = b_2, \dots, g_k(\mathbf{x}) = b_k\}$$

så er enten  $\nabla g_1(\bar{\mathbf{x}}), \nabla g_2(\bar{\mathbf{x}}), \dots, \nabla g_k(\bar{\mathbf{x}})$  lineært avhengige, eller det finnes konstanter  $\lambda_1, \lambda_2, \dots, \lambda_k$  slik at

$$\nabla f(\bar{\mathbf{x}}) = \lambda_1 \nabla g_1(\bar{\mathbf{x}}) + \lambda_2 \nabla g_2(\bar{\mathbf{x}}) + \dots + \lambda_k \nabla g_k(\bar{\mathbf{x}})$$

Før vi ser på beviset, skal vi ta en nærmere kikk på hva teoremet sier. Legg merke til at vi nå har et ligningssystem med  $m+k$  ukjente  $x_1, x_2, \dots, x_m, \lambda_1, \lambda_2, \dots, \lambda_k$ , men at vi også har  $m+k$  ligninger: Skriver vi ut ligningen

$$\nabla f(\mathbf{x}) = \lambda_1 \nabla g_1(\mathbf{x}) + \lambda_2 \nabla g_2(\mathbf{x}) + \dots + \lambda_k \nabla g_k(\mathbf{x})$$

komponentvis, får vi  $m$  ligninger, og bibetingelsene

$$\begin{aligned} g_1(\mathbf{x}) &= b_1 \\ g_2(\mathbf{x}) &= b_2 \\ &\vdots && \vdots \\ g_k(\mathbf{x}) &= b_k \end{aligned}$$

gir oss de  $k$  siste. La oss se på et enkelt eksempel.

**Eksempel 4:** Vi skal minimere funksjonen  $f(x, y, z) = x^2 + y^2 + z^2$  under bibetingelsene

$$x + 2y - z = 2$$

$$-x + y + 2z = 1$$

(Dette er ekvivalent med å finne det punktet på skjæringslinjen mellom planene  $x + 2y - z = 2$  og  $-x + y + 2z = 1$  som ligger nærmest origo, så det er klart at problemet har en løsning). Vi regner ut gradientene til  $f$  og funksjonene  $g_1(x, y, z) = x + 2y - z$ ,  $g_2(x, y, z) = -x + y + 2z$ :

$$\nabla f(x, y, z) = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}, \quad \nabla g_1(x, y, z) = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \quad \nabla g_2(x, y, z) = \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$$

Ifølge teoremet ovenfor leter vi etter punkter der

$$\begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$$

Skriver vi ut ligningen komponentvis, får vi

$$2x = \lambda_1 - \lambda_2$$

$$2y = 2\lambda_1 + \lambda_2$$

$$2z = -\lambda_1 + 2\lambda_2$$

og i tillegg har vi bibetingelsene

$$x + 2y - z = 2$$

$$-x + y + 2z = 1,$$

altså fem ligninger med fem ukjente. Ligningssystemet er lineært og kan løses ved våre standardmetoder, men vi velger en snarvei. Fra de tre første ligningene, får vi uttrykkene  $x = \frac{\lambda_1}{2} - \frac{\lambda_2}{2}$ ,  $y = \lambda_1 + \frac{\lambda_2}{2}$ ,  $z = -\frac{\lambda_1}{2} + \lambda_2$ , som vi setter inn i de to siste ligningene. Resultatet er

$$3\lambda_1 - \frac{\lambda_2}{2} = 2$$

$$-\frac{\lambda_1}{2} + 3\lambda_2 = 1$$

Løser vi dette ligningssystemet, får vi  $\lambda_1 = \frac{26}{35}$  og  $\lambda_2 = \frac{16}{35}$ . Setter vi inn i uttrykkene for  $x$ ,  $y$  og  $z$ , får vi  $x = \frac{1}{7}$ ,  $y = \frac{34}{35}$ ,  $z = \frac{3}{35}$ . Siden det geometriske minimaliseringsproblemets vårt åpenbart har en løsning, og  $x = \frac{1}{7}$ ,  $y = \frac{34}{35}$ ,  $z = \frac{3}{35}$  er den eneste kandidaten, er problemet løst. ♣

Vi skal nå se på beviset for teorem 5.10.2 (Lagranges multiplikatoreteorem for flere bibetingelser). Ideen er akkurat den samme som for én bibetingelse, men utførelsen blir litt mer komplisert fordi matriser erstatter tall en del steder i argumentet. Vi får blant annet bruk for *rangteoremet* for matriser (teorem 4.7.9) som sier at vi alltid kan finne like mange lineært uavhengige søyler som lineært uavhengige rader i en matrise.

\**Bevis for teorem 5.10.2:* Det er nok å vise at hvis gradientene  $\nabla g_1(\bar{\mathbf{x}})$ ,  $\nabla g_2(\bar{\mathbf{x}})$ , ...,  $\nabla g_k(\bar{\mathbf{x}})$  er lineært uavhengige, så finnes det konstanter  $\lambda_1$ ,  $\lambda_2, \dots, \lambda_k$  slik at

$$\nabla f(\bar{\mathbf{x}}) = \lambda_1 \nabla g_1(\bar{\mathbf{x}}) + \lambda_2 \nabla g_2(\bar{\mathbf{x}}) + \cdots + \lambda_k \nabla g_k(\bar{\mathbf{x}})$$

Som i tilfellet med én bibetingelse er det også her nok å se på tilfellet der alle  $b_i$ -ene er 0. Lar vi  $\mathbf{G} : \mathbb{R}^m \rightarrow \mathbb{R}^k$  være funksjonen

$$\mathbf{G}(\mathbf{x}) = \begin{pmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_k(\mathbf{x}) \end{pmatrix},$$

kan vi dermed sammenfatte alle bibetingelsene i formelen  $\mathbf{G}(\mathbf{x}) = \mathbf{0}$ . Deriverer vi, ser vi at radene i Jacobi-matrisen  $\mathbf{G}'(\mathbf{x})$  rett og slett er gradientene  $\nabla g_1(\mathbf{x}), \nabla g_2(\mathbf{x}), \dots, \nabla g_k(\mathbf{x})$ . Siden disse gradientene er lineært uavhengige i punktet  $\bar{\mathbf{x}}$ , vet vi fra *rangteoremet 4.7.9* for matriser at Jacobi-matrisen  $\mathbf{G}(\bar{\mathbf{x}})$  har  $k$  lineært uavhengige søyler. Ved eventuelt å bytte om på rekkefølgen til variablene kan vi anta at dette er de  $k$  siste søylene. Siden de  $m - k$  første variablene og de  $k$  siste kommer til å spille ulike roller i resten av beviset, bytter vi navn på dem for å gjøre forskjellen tydeligere. Vi setter

$$\mathbf{z} = (x_1, x_2, \dots, x_n)$$

der  $n = m - k$ , og

$$\mathbf{y} = (x_{n+1}, x_{n+2}, \dots, x_m)$$

Den "partielle" Jacobi-matrisen  $\frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}})$  består av de siste  $k$  søyrene i den "fulle" Jacobi-matrisen  $\mathbf{G}'(\bar{\mathbf{x}})$ , og siden disse søyrene er lineært uavhengige, er  $\frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}})$  inverterbar. Ifølge teorem 5.7.4 (den vektorvaluerte versjonen av implisitt funksjonsteorem) finnes det da en deriverbar funksjon  $\Phi$  definert i en omegn om  $\bar{\mathbf{z}} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$  slik at

$$\mathbf{G}(\mathbf{z}, \Phi(\mathbf{z})) = \mathbf{0}$$

og  $\Phi(\bar{\mathbf{z}}) = \bar{\mathbf{y}} = (\bar{x}_{n+1}, \bar{x}_{n+2}, \dots, \bar{x}_m)$ . Vi vet også at

$$\Phi'(\bar{\mathbf{z}}) = - \left( \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) \right)^{-1} \frac{\partial \mathbf{G}}{\partial \mathbf{z}}(\bar{\mathbf{x}}) \quad (5.10.1)$$

Fra konstruksjonen ovenfor og betingelsene i teoremet følger det at funksjonen

$$h(\mathbf{z}) = f(\mathbf{z}, \Phi(\mathbf{z}))$$

har et lokalt ekstremalpunkt i  $\bar{\mathbf{z}}$ , og følgelig er  $\nabla h(\bar{\mathbf{z}}) = \mathbf{0}$ . Vi skal nå bruke kjerneregelen til å regne ut  $\nabla h$ . Det er da viktig å huske på at Jacobi-matrisen til et skalarfelt er det samme som gradienten *forutsatt at gradienten oppfattes som en radvektor*. Alle graderenter i dette beviset må derfor oppfattes som radvektorer. Vi får også bruk for at "kjernefunksjonen"

$\Psi(\mathbf{z}) = \begin{pmatrix} \mathbf{z} \\ \Phi(\mathbf{z}) \end{pmatrix}$  har Jacobi-matrice

$$\Psi'(\mathbf{z}) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \\ \frac{\partial \phi_1}{\partial z_1} & \frac{\partial \phi_1}{\partial z_2} & \dots & \frac{\partial \phi_1}{\partial z_n} \\ \frac{\partial \phi_2}{\partial z_1} & \frac{\partial \phi_2}{\partial z_2} & \dots & \frac{\partial \phi_2}{\partial z_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial \phi_k}{\partial z_1} & \frac{\partial \phi_k}{\partial z_2} & \dots & \frac{\partial \phi_k}{\partial z_n} \end{pmatrix}$$

Legg merke til at den nedre delen av denne matrisen er Jacobi-matrisen  $\Phi'(\mathbf{z})$ . Ifølge kjerneregelen har vi

$$\nabla h(\mathbf{z}) = \nabla f(\mathbf{z}, \Phi(\mathbf{z})) \Phi'(\mathbf{z})$$

Dersom vi lar

$$\nabla_{\mathbf{z}} f(\mathbf{z}, \mathbf{y}) = \left( \frac{\partial f}{\partial z_1}(\mathbf{z}, \mathbf{y}), \dots, \frac{\partial f}{\partial z_n}(\mathbf{z}, \mathbf{y}) \right)$$

være den gradienten vi får dersom vi bare tenker på  $f$  som en funksjon av  $z$ -variablene, og

$$\nabla_{\mathbf{y}} f(\mathbf{z}, \mathbf{y}) = \left( \frac{\partial f}{\partial y_1}(\mathbf{z}, \mathbf{y}), \dots, \frac{\partial f}{\partial y_k}(\mathbf{z}, \mathbf{y}) \right)$$

være den gradienten vi får dersom vi bare tenker på  $f$  som en funksjon av  $y$ -variablene, så kan ligningen ovenfor skrives:

$$\nabla h(\mathbf{z}) = \nabla_{\mathbf{z}} f(\mathbf{z}, \Phi(\mathbf{z})) + \nabla_{\mathbf{y}} f(\mathbf{z}, \Phi(\mathbf{z})) \Phi'(\mathbf{z})$$

Siden  $\nabla h(\bar{\mathbf{z}}) = \mathbf{0}$ , har vi dermed

$$\mathbf{0} = \nabla_{\mathbf{z}} f(\bar{\mathbf{z}}, \Phi(\bar{\mathbf{z}})) + \nabla_{\mathbf{y}} f(\bar{\mathbf{z}}, \Phi(\bar{\mathbf{z}})) \Phi'(\bar{\mathbf{z}})$$

Fra ligning (5.10.1) ovenfor vet vi at

$$\Phi'(\bar{\mathbf{z}}) = - \left( \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) \right)^{-1} \frac{\partial \mathbf{G}}{\partial \mathbf{z}}(\bar{\mathbf{x}})$$

og setter vi dette inn i ligningen ovenfor, får vi (husk at  $\bar{\mathbf{x}} = (\bar{\mathbf{z}}, \Phi(\bar{\mathbf{z}}))$ ):

$$\mathbf{0} = \nabla_{\mathbf{z}} f(\bar{\mathbf{x}}) - \nabla_{\mathbf{y}} f(\bar{\mathbf{x}}) \left( \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) \right)^{-1} \frac{\partial \mathbf{G}}{\partial \mathbf{z}}(\bar{\mathbf{x}})$$

Dette ser komplisert ut, men er egentlig ikke så ille. Observer at  $\nabla_{\mathbf{y}} f(\bar{\mathbf{x}})$  er en  $1 \times k$ -matrise og at  $\frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}})^{-1}$  er en  $k \times k$ -matrise. Ganger vi sammen disse, får vi en  $1 \times k$ -matrise  $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$ . Dermed kan ligningen ovenfor skrives

$$\nabla_{\mathbf{z}} f(\bar{\mathbf{x}}) = \Lambda \frac{\partial \mathbf{G}}{\partial \mathbf{z}}(\bar{\mathbf{z}})$$

Ganger vi ut høyresiden og bruker at radene i  $\frac{\partial \mathbf{G}}{\partial \mathbf{z}}(\bar{\mathbf{z}})$  er gradientene  $\nabla_{\mathbf{z}} g_i(\bar{\mathbf{z}})$ , får vi

$$\nabla_{\mathbf{z}} f(\bar{\mathbf{x}}) = \lambda_1 \nabla_{\mathbf{z}} g_1(\bar{\mathbf{x}}) + \lambda_2 \nabla_{\mathbf{z}} g_2(\bar{\mathbf{x}}) + \cdots + \lambda_k \nabla_{\mathbf{z}} g_k(\bar{\mathbf{x}})$$

Dette er nesten det vi skulle vise, det eneste problemet er at vi har de *begrensede* gradientene  $\nabla_{\mathbf{z}} f$ ,  $\nabla_{\mathbf{z}} g_1 \dots, \nabla_{\mathbf{z}} g_k$  og ikke de *fulle* gradientene  $\nabla f$ ,  $\nabla g_1 \dots \nabla g_k$ . Det gjenstår derfor å vise at

$$\nabla_{\mathbf{y}} f(\bar{\mathbf{x}}) = \lambda_1 \nabla_{\mathbf{y}} g_1(\bar{\mathbf{x}}) + \lambda_2 \nabla_{\mathbf{y}} g_2(\bar{\mathbf{x}}) + \cdots + \lambda_k \nabla_{\mathbf{y}} g_k(\bar{\mathbf{x}})$$

Dette er ekvivalent med å vise at

$$\nabla_{\mathbf{y}} f(\bar{\mathbf{x}}) = \Lambda \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}),$$

og siden  $\Lambda = \nabla_{\mathbf{y}} f(\bar{\mathbf{x}}) \left( \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) \right)^{-1}$ , er dette en enkel utregning:

$$\Lambda \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) = \nabla_{\mathbf{y}} f(\bar{\mathbf{x}}) \left( \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) \right)^{-1} \frac{\partial \mathbf{G}}{\partial \mathbf{y}}(\bar{\mathbf{x}}) = \nabla_{\mathbf{y}} f(\bar{\mathbf{x}}) \quad \square$$

**Bemerkning:** Vi skal ikke komme nærmere inn på det her, men nevner i forbifarten at det også finnes annenderivertester for ekstremalverdiproblemer med bibetingelser.

**Bemerkning om Lagrangefunksjoner:** Også når vi skal optimere en funksjon  $f(x_1, \dots, x_m)$  under flere bibetingelser  $g_1(x_1, \dots, x_m) = b_1, \dots, g_k(x_1, \dots, x_m) = b_k$  er det mulig å formulere problemstillingen ved hjelp av en Lagrangefunksjon  $L$ . I dette tilfellet får  $L$  formen

$$\begin{aligned} L(x_1, \dots, x_m, \lambda_1, \dots, \lambda_k) &= f(x_1, \dots, x_m) \\ &- \lambda_1(g_1(x_1, \dots, x_m) - b_1) - \dots - \lambda_k(g_k(x_1, \dots, x_m) - b_k) \end{aligned}$$

Regner vi ut de partielle deriverte til  $L$ , får vi:

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= \frac{\partial f}{\partial x_1} - \lambda_1 \frac{\partial g_1}{\partial x_1} - \dots - \lambda_k \frac{\partial g_k}{\partial x_1} \\ &\vdots && \vdots \\ \frac{\partial L}{\partial x_m} &= \frac{\partial f}{\partial x_m} - \lambda_1 \frac{\partial g_1}{\partial x_m} - \dots - \lambda_k \frac{\partial g_k}{\partial x_m} \\ \frac{\partial L}{\partial \lambda_1} &= -g_1(x_1, \dots, x_m) + b_1 \\ &\vdots && \vdots \\ \frac{\partial L}{\partial \lambda_k} &= -g_k(x_1, \dots, x_m) + b_k \end{aligned}$$

Setter vi disse uttrykkene lik 0, får vi de ligningene vi må løse når vi bruker Lagranges multiplikatormetode. Akkurat som for problemer med én bibetingelse kan vi altså løse optimeringsproblemer med flere bibetingelser ved å finne de stasjonære punktene til Lagrangefunksjonen.

Vær oppmerksom på at noen bøker bytter fortegn på  $\lambda$ -leddene i Lagrangefunksjonen slik at den blir seende slik ut:

$$\begin{aligned} L(x_1, \dots, x_m, \lambda_1, \dots, \lambda_k) &= f(x_1, \dots, x_m) \\ &+ \lambda_1(g_1(x_1, \dots, x_m) - b_1) + \dots + \lambda_k(g_k(x_1, \dots, x_m) - b_k) \end{aligned}$$

Dette spiller ingen rolle — maksimums- og minimumspunktene er de samme som før, det er bare  $\lambda$ -ene som får omvendt fortegn av det de ellers ville ha fått.

## Økonomisk tolkning av Lagrangemultiplikatorer

Lagranges multiplikatormetode brukes mye i økonomiske fag. Det er ikke så vanskelig å forstå hvorfor — i økonomi er man opptatt av maksimums- og minimumsproblemer (man ønsker f.eks. å maksimere inntektene og minimere utgiftene), men samtidig har man naturlige bibetingelser — man kan f.eks. ha en begrenset sum å kjøpe råvarer for, eller man har et begrenset antall arbeidstimer å fordele på ulike oppgaver.

I de eksemplene vi har sett på hittil, har Lagrangemultiplikatorene spilt en underordnet rolle; de har vært hjelpestørrelser vi har trengt for å løse problemet vårt, men de har ikke hatt noen selvstendig betydning. I en del økonomiproblemer spiller imidlertid Lagrangemultiplikatorene en viktig rolle.

La oss tenke oss av vi ønsker å maksimere en inntektsfunksjon  $f(\mathbf{x})$  under bibetingelsene  $g_1(\mathbf{x}) = b_1, g_2(\mathbf{x}) = b_2, \dots, g_k(\mathbf{x}) = b_k$ . Dersom vi endrer verdiene  $b_1, b_2, \dots, b_k$ , må vi selvfølgelig regne med at både maksimalpunktet  $\bar{\mathbf{x}}$  og maksimalverdien  $\bar{y} = f(\bar{\mathbf{x}})$  endrer seg. Vi kan derfor tenke på disse som funksjoner av  $\mathbf{b} = (b_1, b_2, \dots, b_k)$ , altså  $\bar{\mathbf{x}}(\mathbf{b}), \bar{y}(\mathbf{b}) = f(\bar{\mathbf{x}}(\mathbf{b}))$ . Dersom  $\lambda_1, \lambda_2, \dots, \lambda_k$  er Lagrangemultiplikatorene som tilhører maksimumspunktet  $\bar{\mathbf{x}}(\mathbf{b})$ , må vi regne med at disse også vil variere med  $\mathbf{b}$ , og vi antar for enkelhets skyld at de er veldefinerte funksjoner av  $\mathbf{b}$ , altså  $\lambda_1(\mathbf{b}), \lambda_2(\mathbf{b}), \dots, \lambda_k(\mathbf{b})$ . Det er lurt å tenke på  $b_1, b_2, \dots, b_k$  som *innsatsfaktorer* i produksjonen —  $b_1$  er kanskje det totale beløpet vi er villige til å kjøpe råvarer for,  $b_2$  er det totale antall arbeidstimer vi er villige til å bruke i produksjonen,  $b_3$  beløpet vi bruker på å videreutvikle produktene osv.

Et naturlig spørsmål er hvordan en endring i innsatsfaktorene vil påvirke inntektene — hvor mye vil vi f.eks. tjene på å øke arbeidsinnsatsen med 10%? Disse endringene måles av de partiellderiverte

$$\frac{\partial \bar{y}}{\partial b_i}(\mathbf{b}) = \frac{\partial f(\bar{\mathbf{x}}(\mathbf{b}))}{\partial b_i}$$

Som vi snart skal se, er

$$\frac{\partial \bar{y}}{\partial b_i}(\mathbf{b}) = \lambda_i(\mathbf{b})$$

Dette betyr at dersom vi gir innsatsfaktoren  $b_i$  en liten økning  $\Delta b_i$ , så øker inntektene med  $\lambda_i(\mathbf{b})\Delta b_i$ . Dersom kostnadene ved å øke  $b_i$  én enhet er mindre enn  $\lambda_i(\mathbf{b})$ , så lønner det seg altså å øke innsatsfaktoren  $b_i$ , men dersom kostnadene er større enn  $\lambda_i(\mathbf{b})$ , lønner det seg å redusere  $b_i$ . Av denne grunn kalles  $\lambda_i(\mathbf{b})$  *likevektsprisen* til innsatsfaktor  $b_i$  (den kalles også *skyggeprisen* for å understreke at den ikke nødvendigvis har noe med den virkelige prisen å gjøre).

La oss nå vise at

$$\frac{\partial \bar{y}}{\partial b_i}(\mathbf{b}) = \lambda_i(\mathbf{b})$$

Vi skal ikke gjennomføre et fullstendig matematisk resonnement, men vise at denne formelen følger dersom vi antar at de involverte funksjonene er deriverbare (det går an å vise at dette er tilfellet under svært rimelige betingelser). La oss begynne med å se på bibetingelsene. Siden de alltid er oppfylt, har vi

$$g_j(\bar{\mathbf{x}}(\mathbf{b})) = b_j$$

Deriverer vi dette uttrykket mhp.  $b_i$ , får vi (husk kjerneregelen på venstresiden!):

$$\sum_{n=1}^m \frac{\partial g_j}{\partial x_n}(\bar{\mathbf{x}}(\mathbf{b})) \frac{\partial \bar{x}_n}{\partial b_i}(\mathbf{b}) = \begin{cases} 1 & \text{hvis } i = j \\ 0 & \text{ellers} \end{cases} \quad (5.10.2)$$

Deriverer vi inntektsfunksjonen mhp.  $b_i$ , får vi tilsvarende

$$\frac{\partial \bar{y}}{\partial b_i}(\mathbf{b}) = \frac{\partial}{\partial b_i} f(\bar{\mathbf{x}}(\mathbf{b})) = \sum_{n=1}^m \frac{\partial f}{\partial x_n}(\bar{\mathbf{x}}(\mathbf{b})) \frac{\partial \bar{x}_n}{\partial b_i}(\mathbf{b})$$

Ifølge Lagrangebetingelsene er

$$\frac{\partial f}{\partial x_n}(\bar{\mathbf{x}}(\mathbf{b})) = \sum_{j=1}^k \lambda_j(\mathbf{b}) \frac{\partial g_j}{\partial x_n}(\bar{\mathbf{x}}(\mathbf{b}))$$

og setter vi dette inn i uttrykket ovenfor, får vi

$$\begin{aligned} \frac{\partial \bar{y}}{\partial b_i}(\mathbf{b}) &= \sum_{n=1}^m \sum_{j=1}^k \lambda_j(\mathbf{b}) \frac{\partial g_j}{\partial x_n}(\bar{\mathbf{x}}(\mathbf{b})) \frac{\partial \bar{x}_n}{\partial b_i}(\mathbf{b}) = \\ &= \sum_{j=1}^k \lambda_j(\mathbf{b}) \sum_{n=1}^m \frac{\partial g_j}{\partial x_n}(\bar{\mathbf{x}}(\mathbf{b})) \frac{\partial \bar{x}_n}{\partial b_i}(\mathbf{b}) = \lambda_i(\mathbf{b}) \end{aligned}$$

der vi i siste overgang har brukt formel (5.10.2).

## Oppgaver til seksjon 5.10

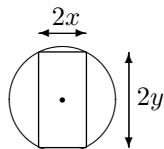
1. Finn maksimums- og minimumspunktene (hvis de finnes) til funksjonen  $f$  under bibetingelsen(e).

- a)  $f(x, y) = 4x - 3y$  når  $x^2 + y^2 = 1$ .
- b)  $f(x, y) = xy$  når  $9x^2 + y^2 = 18$
- c)  $f(x, y, z) = x^2 + y^2 + z^2$  når  $2x - 3y + 2z = 17$
- d)  $f(x, y, z) = x + y + z$  når  $x^2 + y^2 = 1$  og  $2x + z = 1$
- e)  $f(x, y, z) = 2x + 3y$  når  $3x^2 + 2y^2 = 3$
- f)  $f(x, y, z) = x^2 - 2x + 2y^2 + z^2 + z$  når  $x + y + z = 1$  og  $2x - y - z = 5$

- 2.** Finn punktene på flaten  $z^2 - xy = 1$  som ligger nærmest origo.
- 3.** Finn punktene på skjæringkurven mellom flatene  $x^2 + y^2 = 1$  og  $x^2 - xy + y^2 - z^2 = 1$  som ligger nærmest origo.
- 4.** Løs oppgave 11 i forrige seksjon ved hjelp av Lagranges multiplikatorer.
- 5.** Løs oppgave 12 i forrige seksjon ved hjelp av Lagranges multiplikatorer.
- 6.** Løs oppgave 13 i forrige seksjon ved hjelp av Lagranges multiplikatorer.
- 7.** Løs oppgave 16 i forrige seksjon ved hjelp av Lagranges multiplikatorer.
- 8.** La  $f(x, y) = \ln(x^2 + y^2 + 1) - \frac{x^2}{2} + y^2$ .
- Finn de stasjonære punktene til  $f$  og avgjør om de er lokale maksimumspunkter, lokale minimumspunkter eller sadelpunkter.
  - Finn maksimum og minimum til  $f$  på området  $\{(x, y) \mid x^2 + y^2 \leq 1\}$ . (Hint: Det kan lønne seg å bytte til polarkoordinater.)
- 9.** (Eksamensoppgave fra MAT 1110, 13/6, 2008) Forklar at funksjonen  $f(x, y) = 2x + 4y$  har maksimums- og minimumspunkter under bibetingelsen  $\{(x, y) \in \mathbf{R}^2 \mid x^2 + y^2 = 4\}$ . Finn disse maksimums- og minimumspunktene.
- 10.** Kjeglesnittet  $K$  består av alle punkter  $(x, y)$  som oppfyller ligningen  $9x^2 + 4y^2 - 18x + 16y = 11$
- Hva slags kjeglesnitt er  $K$ ? Lag en tegning av  $K$  der alle viktige størrelser er tegnet inn (f.eks. sentrum, brennpunkter, halvakser, asymptoter etter hva som er aktuelt).
  - Forklar at funksjonen  $f(x, y) = 2x + y$  har en største og minste verdi når den begrenses til mengden  $K$ . Finn disse maksimum- og minimumsverdiene.
- 11.** En rektangulær boks med kanter parallelle med koordinataksene skal plasseres inni ellipsoiden  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ . Hva er det største volumet denne boksen kan ha?
- 12.** Av en sylinderformet stokk med radius  $r$  skal det skjæres ut en bjelke med bredde  $2x$  og høyde  $2y$ . (Se figur). Bæreevnen til bjelken er proporsjonal med  $x$  og med kvadratet av  $y$ , dvs. den er gitt ved funksjonen

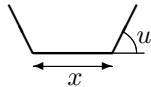
$$f(x, y) = kxy^2$$

der  $k$  er en konstant.



Finn de verdier av  $x$  og  $y$  som gir størst verdi for  $f(x, y)$ .

- 13.** En renne skal lages ved at en  $b$  cm bred metallplate brettes opp symmetrisk på begge sider. Figuren viser et tverrsnit av rennen. Hvordan må vi velge bredden  $x$  og vinkelen  $u$  for at tverrsnittet skal få størst mulig areal?



- 14.** Herons formel sier at arealet til en trekant med sider  $x, y, z$  er

$$A = \sqrt{s(s-x)(s-y)(s-z)}$$

der  $s = \frac{x+y+z}{2}$  er halve omkretsen. Bruk Lagranges multiplikatormetode til å vise at blant alle trekanner med samme omkrets, er det den likesidede som har størst areal.

- 15.** La  $S$  være ellipsoideflaten

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

- a) Finn ligningen for tangentplanet til  $S$  i punktet  $(x_0, y_0, z_0)$ .
- b) Anta at  $x_0, y_0, z_0 > 0$ . Finn volumet til pyramiden avgrenset av koordinatplanene og tangentplanet til  $S$  i  $(x_0, y_0, z_0)$ .
- c) Finn det punktet  $(x, y, z)$  på  $S$  med  $x, y, z \geq 0$  som gjør produktet  $xyz$  størst mulig.
- d) Hva er det minste volumet pyramiden i b) kan ha?

- 16.** I økonomi regner man ofte at profitten kan modelleres som en *Cobb-Douglas-funksjon*. Det betyr at profitten er gitt ved  $P(x, y) = Kx^\alpha y^\beta$  der  $K, \alpha, \beta$  er positive konstanter, og  $x$  og  $y$  står for det beløpet man investerer i forskjellige "innsatsfaktorer", f.eks. kan  $x$  være investeringen i råvarer og  $y$  investeringen i arbeidskraft. Anta at den totale investeringen er gitt, dvs. at  $x + y = S$ , der  $S$  er en konstant.

- a) Vis at profitten er størst når  $x = \frac{\alpha S}{\alpha+\beta}$ ,  $y = \frac{\beta S}{\alpha+\beta}$ . Hva er den maksimale profitten?
- b) Generaliser resultatet ovenfor til flere innsatsfaktorer. Anta at

$$P(x_1, x_2, \dots, x_n) = Kx_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$$

(der  $K, \alpha_1, \alpha_2, \dots, \alpha_n$  er positive konstanter) og finn maksimumsverdien til  $P$  under bibetingelsen  $x_1 + x_2 + \dots + x_n = S$ .

- 17.** Anta at en forbruker har valget mellom to varetyper. De to vareslagene koster hhv.  $p$  og  $q$  kroner per enhet. I økonomiske modeller regner man ofte med at nytten forbrukeren har av å kjøpe  $x$  enheter av den ene varetypen og  $y$  enheter av den andre, er gitt ved en *nyttefunksjon* av typen

$$U(x, y) = a \ln x + b \ln y$$

der  $a$  og  $b$  er positive konstanter. Dersom forbrukeren har  $S$  kroner til rådighet, ønsker hun å maksimere nyttefunksjon under bibetingelsen  $px + qy = S$ . Vist at

hun får maksimalt utbytte ved å velge  $x = \frac{aS}{p(a+b)}$  og  $y = \frac{bS}{q(a+b)}$ .

**18.** Et firma produserer to vareslag. Det har et samlet produksjonsbudsjett på  $S$  kroner i året og et utviklingsbudsjett på  $T$  kroner i året. Firmaet regner at hvis det bruker  $x$  kroner på produksjon av vareslag 1 og samtidig bruker  $y$  kroner på å videreutvikle produktet, vil overskuddet fra vareslag 1 være

$$U(x, y) = Ax^\alpha y^{1-\alpha}$$

der  $A$  og  $\alpha$  er konstanter,  $0 \leq \alpha < 1$ . Hvis firmaet på samme måte bruker  $z$  kroner på produksjon av vareslag 2 og samtidig bruker  $u$  kroner på videreutvikling, regner det med at overskuddet fra vareslag 2 vil være

$$V(z, u) = Bz^\beta u^{1-\beta}$$

der  $B$  og  $\beta$  er konstanter,  $0 \leq \beta < 1$ .

- a) Forklar hvorfor det er naturlig for firmaet å optimere størrelsen  $U(x, y) + V(z, u)$  under bibetingelsene  $x + z = S$ ,  $y + u = T$ .
- b) Vis at Lagranges multiplikatormetode leder til ligningene

$$\begin{aligned} \alpha A \left(\frac{x}{y}\right)^{\alpha-1} &= \lambda \\ (1-\alpha)A \left(\frac{x}{y}\right)^\alpha &= \mu \\ \beta B \left(\frac{z}{u}\right)^{\beta-1} &= \lambda \\ (1-\beta)B \left(\frac{z}{u}\right)^\beta &= \mu \\ x + z &= S \\ y + u &= T \end{aligned}$$

- c) Vis ved å kombinere de to første ligningene ovenfor at  $\frac{x}{y} = \frac{\alpha}{1-\alpha} \frac{\mu}{\lambda}$ . Vis også at  $\frac{z}{u} = \frac{\beta}{1-\beta} \frac{\mu}{\lambda}$ .
- d) Vis ved å kombinere den første og tredje ligningen ovenfor at

$$\alpha A \left(\frac{x}{y}\right)^{\alpha-1} = \beta B \left(\frac{z}{u}\right)^{\beta-1}$$

Sett inn uttrykkene for  $\frac{x}{y}$  og  $\frac{z}{u}$  fra c) og vis at dette leder til formelen

$$\frac{\mu}{\lambda} = \left[ \frac{B\beta^\beta(1-\beta)^{1-\beta}}{A\alpha^\alpha(1-\alpha)^{1-\alpha}} \right]^{\frac{1}{\alpha-\beta}}$$

- e) For enkelthets skyld kaller vi uttrykket på høyre side av uttrykket ovenfor for  $K$ . Vi har med andre ord  $\frac{\mu}{\lambda} = K$ . Ifølge c) har vi dermed  $\frac{x}{y} = \frac{\alpha}{1-\alpha} K$  og  $\frac{z}{u} = \frac{\beta}{1-\beta} K$ . Sett disse uttrykkene inn i de to siste ligningen i ligningssystemet i b), og finn  $x, y, z$  og  $u$ .

**19.** I denne oppgaven skal vi se på en anvendelse av Lagranges multiplikatormetode i statistisk fysikk. La oss begynne med en kort skisse av den fysiske problemstillingen (selv om det egentlig ikke er nødvendig å skjonne fysikken for å løse oppgaven): Vi har et system med  $N$  partikler ( $N$  er et svært stort tall) som kan fordelle seg på  $n$  energinivåer  $E_1, E_2, \dots, E_n$ . Den totale energien til systemet er  $U$ , og sannsynligheten for å finne en tilfeldig partikkell på energinivå  $E_i$  er  $p_i$ . Målet er å finne den mest sannsynlige fordelingen av partiklene på energinivåene. Denne fordelingen er gitt ved vektoren  $(x_1, x_2, \dots, x_n)$  der  $x_i$  er antall partikler på nivå  $E_i$ .

Etter noen innledende betrakninger (som er ren sannsynlighetsregning) kommer man frem til at man ønsker å finne maksimum til funksjonen

$$f(x_1, x_2, \dots, x_n) = N - \sum_{i=1}^n x_i \ln \left( \frac{x_i}{p_i} \right)$$

under bibetingelsene

$$\sum_{i=1}^n x_i = N \quad (\text{det totale antall partikler er } N)$$

$$\sum_{i=1}^n x_i E_i = U \quad (\text{den totale energien er } U)$$

- a) La  $g$  og  $h$  være funksjonene  $g(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i$ ,  $h(x_1, x_2, \dots, x_n) = \sum_{i=1}^n x_i E_i$ . Vis at

$$\frac{\partial f}{\partial x_i} = -\ln \left( \frac{x_i}{p_i} \right) + 1, \quad \frac{\partial g}{\partial x_i} = 1 \quad \text{og} \quad \frac{\partial h}{\partial x_i} = E_i$$

- b) Vis at Lagranges multiplikatormetode gir ligningene

$$-\ln \left( \frac{x_i}{p_i} \right) + 1 = \lambda + \mu E_i \quad \text{for } i = 1, 2, \dots, n$$

i tillegg til bibetingelsene  $\sum_{i=1}^n x_i = N$  og  $\sum_{i=1}^n x_i E_i = U$ .

- c) Vis at  $x_i = p_i e^{-\lambda - \mu E_i + 1}$ .

- d) Vi innfører nå *partisjonsfunksjonen*  $Z = \sum_{i=1}^N p_i e^{-\mu E_i}$ . Bruk den første bibetingelsen til å vise at  $x_i = \frac{N}{Z} p_i e^{-\mu E_i}$ . Dermed har vi kvittet oss med den første multiplikatoren  $\lambda$ .

- e) Vis at den gjenværende multiplikatoren  $\mu$  er bestemt av ligningen  $U = \frac{N}{Z} \sum_{i=1}^n p_i E_i e^{-\mu E_i}$ .

Dette viser at  $\mu$  er knyttet til den totale energien til systemet. Faktisk viser det seg at  $\mu = \frac{1}{kT}$  der  $T$  er temperaturen til systemet (målt i grader Kelvin) og  $k$  er en konstant (*Boltzmanns konstant*). Vi ender dermed opp med den fundamentale sammenhengen  $x_i = \frac{N}{Z} p_i e^{-\frac{E_i}{kT}}$  som kalles *Maxwell-Boltzmann-fordelingen*.

- 20.** (Eksamens i MAT 1110 15/8, 2008, noe utvidet) I denne oppgaven er

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{23} \\ \vdots & \vdots & \dots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix}$$

en symmetrisk  $n \times n$ -matrise, og  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$  er funksjonen  $f(\mathbf{x}) = (\mathbf{Ax}) \cdot \mathbf{x}$  (der  $(\mathbf{Ax}) \cdot \mathbf{x}$  betegner skalarproduktet mellom vektorene  $\mathbf{Ax}$  og  $\mathbf{x}$ ).

- a) Vis at dersom  $\mathbf{x}$  er en egenvektor for  $A$  med egenverdi  $\lambda$ , så er  $f(\mathbf{x}) = \lambda|\mathbf{x}|^2$ .
- b) Vis at for alle vektorer  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  er

$$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^n a_{ii}x_i^2 + 2 \sum_{i \neq j} a_{ij}x_i x_j$$

der den siste summen er over alle par av ulike indekser  $1 \leq i, j \leq n$ .

- c) La  $S = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| = 1\}$  være det  $n$ -dimensionale kuleskallet om origo med radius 1. Forklar at når vi innskrenker  $f$  til  $S$ , så har funksjonen maksimums- og minimumspunkter. Bruk Lagranges multiplikatormetode til å vise at disse maksimums- og minimumspunktene er egenvektorer til  $A$ . Vis til slutt at maksimumsverdien til  $f$  på  $S$  er den største egenverdien til  $A$ , mens minimumsverdien er den minste egenverdien til  $A$ .

## 5.11 Gradientmetoden

I de foregående seksjonene har vi studert optimeringsproblemer for funksjoner av flere variable. Vi har sett at både problemer *med* bibetingelser og problemer *uten* bibetingelser leder til ligningssystemer som skal løses. Disse ligningssystemene blir fort så kompliserte at de ikke kan løses for hånd, og man må derfor bruke datamaskiner til å finne tilnærmede løsninger. Newtons metode er et utmerket redskap for numerisk løsning av ligningssystemer, og det er selvfølgelig utviklet metoder der man kobler optimeringsproblemer direkte til Newtons metode. Men det finnes også andre metoder der man går direkte løse på optimeringsproblemene uten å gå veien om stasjonære punkter. I denne seksjonen skal vi se kort på en slik metode — “gradientmetoden” eller “den bratteste nedstigningsmetoden” (“method of steepest descent”).

Grunnideen i denne metoden er enkel. Vi vet at gradienten til en funksjon peker i den retningen hvor funksjonen stiger raskest, og ønsker vi å finne et minimumspunktet for funksjonen (metoden presenteres gjerne som en metode for å finne minimumspunkter), er det naturlig å gå i *motsatt* retning av gradienten. Etter å ha gått i denne retningen et stykke, stopper vi opp, regner ut gradienten på nytt, og fortsetter i motsatt retning av den nye gradienten osv. Denne metoden bringer oss stadig lengre ned, og med litt flaks burde den lede oss til et (lokalt) minimumspunkt.

Et viktig spørsmål er hvor langt vi skal gå i én retning før vi stopper opp og regner ut en ny gradient. Går vi for kort, blir metoden ineffektiv fordi vi må regne ut nye graderenter oftere enn nødvendig, og går vi for langt, risikerer vi å passere minimumspunktet og gå langt ut på den andre siden. Et naturlig valg er å fortsette i den retningen vi har begynt så lenge det går

nedover, og først beregne en ny gradient når vi kommer til et sted der det begynner å gå oppover.

La oss se hvordan dette ser ut matematisk. Anta at ønsker å finne et (lokalt) minimumspunkt for funksjonen  $f(\mathbf{x})$ , og at vi har et startpunkt  $\mathbf{x}_0$  som vi tror ikke ligger altfor langt unna det minimumspunktet vi er på jakt etter. Gradienten  $\nabla f(\mathbf{x}_0)$  gir oss den retningen hvor funksjonen vokser brattest, og vi ønsker å gå i motsatt retning, altså langs linjen

$$\mathbf{r}(t) = \mathbf{x}_0 - \nabla f(\mathbf{x}_0)t$$

Setter vi dette uttrykket inn i  $f$ , får vi en funksjon  $g$  av én variabel

$$g(t) = f(\mathbf{r}(t))$$

Vi ønsker å finne minimumspunktet til  $g$ , så vi regner ut den deriverte ved hjelp av kjerneregelen

$$g'(t) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t) = \nabla f(\mathbf{x}_0 - \nabla f(\mathbf{x}_0)t) \cdot \nabla f(\mathbf{x}_0)$$

Vi er på jakt etter en  $t > 0$  slik at  $g(t) = 0$ , dvs.

$$\nabla f(\mathbf{x}_0 - \nabla f(\mathbf{x}_0)t) \cdot \nabla f(\mathbf{x}_0) = 0$$

(finnes det flere slike  $t$ 'er, velger vi den første). Selv om dette bare er én ligning med én ukjent, er det slett ikke sikkert vi kan løse den for hånd, og da må vi bruke f.eks. Newtons metode for å finne en tilnærmet verdi. Når vi har funnet en løsning  $t_0$ , setter vi

$$\mathbf{x}_1 = \mathbf{x}_0 - \nabla f(\mathbf{x}_0)t_0$$

og gjentar hele prosedyren med  $\mathbf{x}_1$  som utgangspunkt. På denne måten får vi en følge  $\{\mathbf{x}_n\}$  som (forhåpentligvis) konvergerer mot et lokalt minimum.

Det går an å analysere gradientmetoden teoretisk (omtrent som vi tidligere har analysert Newtons metode) og komme frem til sikre kriterier for konvergens og gode estimater for konvergenshastighet. Vi skal ikke gjøre dette her, men nøyde oss med et enkelt eksempel som viser metoden i praksis.

**Eksempel 1:** Vi skal bruke gradientmetoden på funksjonen

$$f(x, y) = x^2 + 4y^2$$

Det er lett å se at  $f$  har ett eneste minimumspunkt, nemlig  $(0, 0)$ , så poenget med eksemplet er ikke å finne minimumspunktet, men å studere hvordan gradientmetoden virker.

Vi trenger åpenbart gradienten til  $f$ , så la oss regne den ut med en gang:

$$\nabla f(x, y) = (2x, 8y)$$

Anta at vi starter iterasjonen i et punkt  $\mathbf{x}_0 = (x_0, y_0)$ . Ifølge teorien ovenfor, leter vi etter en  $t > 0$  som løser ligningen

$$\nabla f(\mathbf{x}_0 - \nabla f(\mathbf{x}_0)t) \cdot \nabla f(\mathbf{x}_0)$$

Bruker vi at  $\nabla f(x, y) = (2x, 8y)$  og  $\mathbf{x}_0 - \nabla f(\mathbf{x}_0)t = (x_0 - 2x_0t, y_0 - 8y_0t)$ , får vi ligningen

$$(2(x_0 - 2x_0t), 8(y_0 - 8y_0t)) \cdot (2x_0, 8y_0) = 0$$

Ganger vi ut og forkorter, gir dette ligningen

$$x_0^2 - 2x_0t + 16y_0^2 - 128y_0^2t = 0$$

som har løsningen

$$t_0 = \frac{x_0^2 + 16y_0^2}{2x_0^2 + 128y_0^2}$$

Dermed får vi

$$\mathbf{x}_1 = \mathbf{x}_0 - \nabla f(\mathbf{x}_0)t_0 = (x_0, y_0) - (2x_0, 8y_0) \frac{x_0^2 + 16y_0^2}{2x_0^2 + 128y_0^2}$$

dvs.

$$x_1 = x_0 \left( 1 - \frac{x_0^2 + 16y_0^2}{x_0^2 + 64y_0^2} \right) = \frac{48x_0y_0^2}{x_0^2 + 64y_0^2}$$

$$y_1 = y_0 \left( 1 - \frac{4x_0^2 + 64y_0^2}{x_0^2 + 64y_0^2} \right) = -\frac{3x_0^2y_0}{x_0^2 + 64y_0^2}$$

Tilsvarende formler gjelder selvfølgelig senere i iterasjonen; vi har

$$x_{n+1} = \frac{48x_ny_n^2}{x_n^2 + 64y_n^2}$$

$$y_{n+1} = -\frac{3x_n^2y_n}{x_n^2 + 64y_n^2}$$

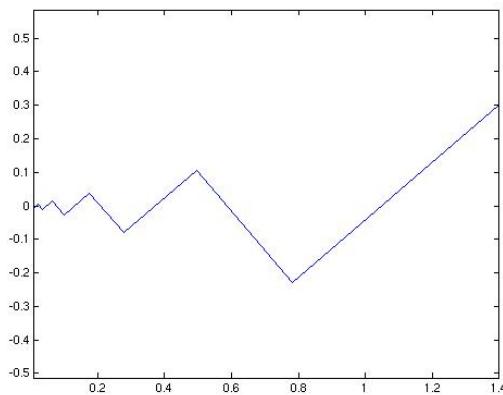
La oss skrive et lite MATLAB-program for å se hvordan iterasjonen forløper:

```
function [x,y]=gradient(a,b,N)
x=zeros(1,N);
y=zeros(1,N);
x(1)=a;
y(1)=b;
for n=1:N
    x(n+1)=48*x(n)*y(n)^2/(x(n)^2+64*y(n)^2));
    y(n+1)=-3*x(n)^2*y(n)/(x(n)^2+64*y(n)^2));
end
```

Gir vi nå kommandoene

```
>> [x,y]=gradient(1.4,.3,10);
>> plot(x,y)
>> axis 'equal'
```

får vi figuren nedenfor som viser hvordan gradientmetoden gir en følge som nærmer seg nullpunktet  $(0,0)$  når vi starter i punktet  $(1.4, 0.3)$ .



Figur 1: Gradientmetoden



Legg merke til at i figuren ovenfor ser gradientmetoden ut til å “overskyte” ved å gå litt for langt i hvert skritt. Dette er ganske vanlig og gjelder ikke bare eksemplet vi nå har studert. Gradientmetoden har også andre svakheter, og det finnes derfor en rekke videreutviklinger av metoden, men disse skal vi ikke komme inn på her.

### Oppgave til seksjon 5.11

1. Skriv et MATLAB-program som bruker gradientmetoden til å finne minimumspunktet til  $f(x, y) = x^2 + 2xy + 2y^2 + x - y$ . Sjekk resultatet ved å finne minimumspunktet ved regning.



## Kapittel 6

# Multippel integrasjon

I dette kapitlet skal vi se på integrasjon av funksjoner av flere variable. Slike *multiple integraler* defineres på samme måte som integraler av funksjoner av én variabel, og de regnes ut ved at man beregner flere “vanlige” integraler etter hverandre. Disse delene av teorien er derfor ikke spesielt vanskelige, men det er selvfølgelig en fordel å være flink i tradisjonell integralregning.

Det som er vanskeligst med de nye integralene, er å stille dem opp. Istedentfor å integrere over intervaller på tallinjen, skal vi nå integrerer over områder i  $\mathbb{R}^n$ , og for å stille opp integralene, må vi beskrive disse områdene ved hjelp av formler. Ofte blir beskrivelsene enklere når man bruker andre koordinatsystemer enn det vanlige (kartesiske), og vi kommer nå virkelig til å få bruk for de alternativene koordinatsystemene vi innførte i seksjon 3.7: polarkoordinater, sylinderkoordinater og kulekoordinater. I noen tilfeller strekker ingen av disse “standardsystemene” til, og da må vi selv lage et passende koordinatsystem. Å bytte koordinatsystem på denne måten svarer til å skifte variabel i et vanlig integral, men den flervariable teorien er mer geometrisk — vi bytter som regel (men ikke alltid) koordinatssystem for å få et enklere område å integrere over og ikke for å få en enklere integrand.

Multiple integraler har viktige anvendelser. Mange av disse er naturlige generaliseringer av anvendelser du allerede kjenner for vanlige integraler — vi kan f.eks. bruke multiple integraler til å regne ut arealer, volumer og masser. I tillegg finnes det mange “nye” anvendelser i f.eks. statistikk og fysikk. Det viktigste i dette kapitlet er imidlertid ikke å lære om flest mulig slike anvendelser; det viktigste er å lære hvordan man selv går frem når man skal bruke multiple integraler til å løse problemer i andre fag eller i andre deler av matematikken.

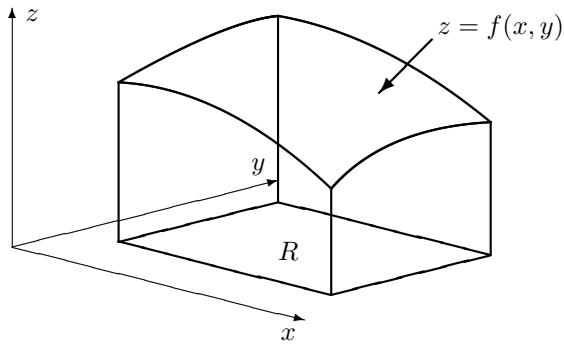
I praksis må man ofte bruke numeriske metoder for å regne ut multiple integraler. Det krever mye regnekapasitet, og som regel er det verdt å bruke litt tid på å skrive integralene på enklest mulig form før man prøver seg på en numerisk integrasjon. I dette kapitlet skal vi nøyne oss med å se litt på hvordan MATLAB håndterer multiple integraler (men det er nok til å gi deg erfaring med hvor tunge slike integralberegninger er!).

For at ikke geometrien skal bli for komplisert og notasjonen for uoversiktlig, skal vi hovedsakelig koncentrere oss om funksjoner av to og tre variable. Den generelle teorien for funksjoner av  $n$ -variable er imidlertid en gansk grei generalisering av den to- og tre-dimensjonale teorien, og du vil neppe få store problemer om du skulle støte på et fire- eller fem-dimensjonalt integral i andre kurs. For fullstendighetens

skyld har vi likevel tatt med en kommentar om slike integraler helt til slutt i seksjon 6.10.

## 6.1 Dobbeltintegraler over rektangler

Vi skal først se hvordan vi kan integrere en funksjon  $f(x, y)$  av to variable over et rektangel  $R$  i  $xy$ -planet. Slike *dobbeltintegraler* kan brukes til så mangt, men i utgangspunktet er det lurt å tenke seg at  $f$  er en positiv funksjon, og at vi ønsker å regne ut volumet under funksjonsgrafen og over rektangelet  $R$  (se figur 1).



Figur 1: Volumet under grafen  $z = f(x, y)$

La oss begynne med litt notasjon: Dersom  $R$  er rektangelet

$$R = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, c \leq y \leq d\},$$

skriver vi ofte

$$R = [a, b] \times [c, d]$$

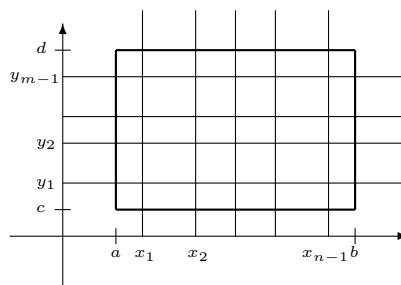
En *partisjon*  $\Pi$  av  $R$  består av en partisjon

$$a = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = b$$

av  $[a, b]$ , og en partisjon

$$c = y_1 < y_2 < \dots < y_{m-1} < y_m = d$$

av  $[c, d]$ . Vi tenker oss at denne partisjonen deler opp  $R$  i et rutenett som vist på figur 2.



Figur 2: En partisjon av  $R$

Vi lar

$$R_{ij} = [x_{i-1}, x_i] \times [y_{j-1}, y_j]$$

være den  $ij$ -te ruten i denne oppdelingen (dvs. den  $i$ -te ruten når vi teller i  $x$ -retning, og den  $j$ -te når vi teller i  $y$ -retning).

La oss nå tenke oss at  $f : R \rightarrow \mathbb{R}$  er en begrenset funksjon (den kan godt ha negative verdier, men vi skriver som om den var positiv). For å få en tilnærming til volumet under grafen til  $f$ , lar vi

$$m_{ij} = \inf\{f(x, y) \mid (x, y) \in R_{ij}\}$$

og

$$M_{ij} = \sup\{f(x, y) \mid (x, y) \in R_{ij}\}$$

være henholdsvis infimum og supremum til  $f$  over  $R_{ij}$ . Lar vi

$$|R_{ij}| = (x_i - x_{i-1})(y_j - y_{j-1})$$

være arealet til ruten  $R_{ij}$ , ser vi at den *nedre trappesummen*

$$N(\Pi) = \sum_{i=1}^n \sum_{j=1}^m m_{ij} |R_{ij}|$$

og den *øvre trappesummen*

$$\mathcal{O}(\Pi) = \sum_{i=1}^n \sum_{j=1}^m M_{ij} |R_{ij}|$$

er henholdsvis en nedre og en øvre tilnærming til volumet under grafen til  $f$  (dersom dette finnes!). Dersom vi kan få disse verdiene så nær hverandre vi måtte ønske ved å lage partisjonen fin nok, er det naturlig å definere den felles grenseverdien til å være integralet av  $f$  over rektanglet  $R$ .

**Definisjon 6.1.1** Anta at  $R = [a, b] \times [c, d]$  er et rektangel i  $\mathbb{R}^2$  og at  $f : R \rightarrow \mathbb{R}$  er en begrenset funksjon. Da definerer vi *øvreintegralet til  $f$  over  $R$*  som

$$\overline{\iint_R f(x, y) dx dy} = \inf\{\mathcal{O}(\Pi) \mid \Pi \text{ er en partisjon av } R\}$$

og *nedreintegralet til  $f$  over  $R$*  som

$$\underline{\iint_R f(x, y) dx dy} = \sup\{N(\Pi) \mid \Pi \text{ er en partisjon av } R\}$$

Dersom  $\overline{\iint_R f(x, y) dx dy} = \underline{\iint_R f(x, y) dx dy}$ , sier vi at  $f$  er integrerbar over  $R$ , og definerer (dobbelt)integralet til  $f$  over  $R$  til å være

$$\iint_R f(x, y) dx dy = \overline{\iint_R f(x, y) dx dy} = \underline{\iint_R f(x, y) dx dy}$$

**Bemerkning:** Vi skal skrive to integraltegn  $\iint$  i dobbeltintegraler for å understreke at vi integrerer med hensyn på to variable, men siden denne skrivemåten blir slitsom etterhvert, er det mange videregående bøker som bare bruker ett integraltegn. På samme måte skal vi veksle litt mellom betegnelsene “integral” og “dobbeltintegral” — vi bruker “dobbeltintegral” når vi sammenligner dobbeltintegraler med andre integraltyper, men nøyser oss ofte med “integral” ellers.

Legg merke til at definisjonene ovenfor er helt analoge til dem vi har for vanlige integraler (se *Kalkulus*, seksjon 8.2). Også en del fundamentale regneregler er som før:

**Setning 6.1.2** *Anta  $R = [a, b] \times [c, d]$  er et rektangel i  $\mathbb{R}^2$ . Anta at  $f, g : R \rightarrow \mathbb{R}$  er integrerbare funksjoner og at  $k$  er en konstant. Da er*

- (i)  *$kf$  integrerbar og  $\iint_R kf(x, y) dx dy = k \iint_R f(x, y) dx dy$ .*
- (ii)  *$f+g$  integrerbar og  $\iint_R (f(x, y) + g(x, y)) dx dy = \iint_R f(x, y) dx dy + \iint_R g(x, y) dx dy$ .*
- (iii) *Hvis  $f(x, y) \leq g(x, y)$  for alle  $(x, y) \in R$ , er  $\iint_R f(x, y) dx dy \leq \iint_R g(x, y) dx dy$ .*

*Bevis:* Vi overlater beviset til leserne. Se oppgavene til seksjonen for hint.  $\square$

Vi har nå definert dobbeltintegraler over rektangler og sett på deres enkleste egenskaper. Våre neste oppgaver er

- (i) å vise at alle kontinuerlige funksjoner er integrerbare
- (ii) å finne metoder for å regne ut dobbeltintegraler

Vi begynner med den første oppgaven.

### Integrasjon av kontinuerlige funksjoner

For å vise at kontinuerlige funksjoner er integrerbare, trenger vi et begrep og et resultat fra seksjon 5.3 (har du ikke lest denne seksjonen ennå, er det sannsynlig at like greit å ta teorem 6.1.5 nedenfor til etterretning og heller vende tilbake til seksjon 5.3 ved en senere anledning). Vi begynner med å minne om definisjon 5.3.1

**Definisjon 6.1.3** *Anta at  $f : A \rightarrow \mathbb{R}$  er en funksjon av  $n$  variable. Vi sier at  $f$  er uniformt kontinuerlig på en mengde  $B \subset A$  dersom det til enhver  $\epsilon > 0$  finnes en  $\delta > 0$ , slik at hvis  $\mathbf{u}, \mathbf{v} \in B$  og  $|\mathbf{u} - \mathbf{v}| < \delta$ , så er  $|f(\mathbf{u}) - f(\mathbf{v})| < \epsilon$ .*

Nøkkelresultatet er teorem 5.3.2 som vi skriver opp på nytt her:

**Teorem 6.1.4** *Anta at  $K$  er en lukket, begrenset delmengde av  $\mathbb{R}^n$ . Enhver funksjon  $f$  som er kontinuerlig på  $K$ , er også uniformt kontinuerlig på  $K$ .*

Vi har nå det redskapet vi trenger for å vise at alle kontinuerlige funksjoner er integrerbare.

**Teorem 6.1.5** *Anta at  $R = [a, b] \times [c, d]$  er et rektangel i  $\mathbb{R}^2$ , og at  $f : R \rightarrow \mathbb{R}$  er kontinuerlig. Da er  $f$  integrerbar over  $R$ .*

*Bevis:* Siden enhver øvre trappesum er større enn eller lik enhver nedre trappesum (se oppgave 3 hvis du ønsker hjelp til å vise dette), så er det nok å vise at for enhver  $\epsilon > 0$ , finnes det et partisjon  $\Pi$  slik at  $\mathcal{O}(\Pi) - N(\Pi) \leq \epsilon$ . Fra setningen ovenfor vet vi at  $f$  er uniformt kontinuerlig på  $R$ , så det finnes en  $\delta > 0$  slik at hvis  $|\mathbf{u} - \mathbf{v}| < \delta$ , så er  $|f(\mathbf{u}) - f(\mathbf{v})| < \frac{\epsilon}{|R|}$ , der  $|R| = (b-a)(d-c)$  er arealet til  $R$ . Velger vi partisjonen så fin at avstanden mellom to punkter i samme delrekktangel  $R_{ij}$  alltid er mindre enn  $\delta$ , må  $M_{ij} - m_{ij} \leq \frac{\epsilon}{|R|}$ . Dermed er

$$\mathcal{O}(\Pi) - N(\Pi) = \sum_{i=1}^n \sum_{j=1}^m (M_{ij} - m_{ij}) |R_{ij}| \leq \frac{\epsilon}{|R|} \sum_{i=1}^n \sum_{j=1}^m |R_{ij}| = \epsilon$$

der vi har brukt at  $\sum_{i=1}^n \sum_{j=1}^m |R_{ij}| = |R|$  siden summen av arealene til alle smårekktanglene er lik arealet til det store rektanglet.  $\square$

Ovenfor har vi definert dobbeltintegralet ved hjelp av øvre og nedre trappesummer. Som du sikkert husker fra teorien for vanlige integraler, lønner det seg ofte å tenke på slike integraler som grensen for Riemann-summer. Vi har et tilsvarende begrep for dobbeltintegraler. Dersom vi har en partisjon  $\Pi$  gitt ved

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

$$c = y_0 < y_1 < \dots < y_{m-1} < y_m = d,$$

kan vi lage et *utplukk*  $U$  ved å velge en  $\mathbf{c}_{ij}$  i hvert delrekktangel  $R_{ij}$ . Den tilhørende *Riemann-summen* er

$$R(\Pi, U) = \sum_{i=1}^n \sum_{j=1}^m f(\mathbf{c}_{ij}) |R_{ij}|$$

Legg merke til at siden  $m_{ij} \leq f(\mathbf{c}_{ij}) \leq M_{ij}$ , så er

$$N(\Pi) \leq R(\Pi, U) \leq \mathcal{O}(\Pi)$$

Vi ønsker å vise at Riemann-summene nærmer seg integralet når partisjonen  $\Pi$  blir finere og finere. Som et mål på finheten, bruker vi *maskevidden* som er lik den lengste diagonalen i rekktanglene  $R_{ij}$  — med andre ord

$$|\Pi| = \max \left\{ \sqrt{(x_i - x_{i-1})^2 + (y_j - y_{j-1})^2} : 1 \leq i \leq n, 1 \leq j \leq m \right\}$$

**Setning 6.1.6** *Anta at  $\{\Pi_n\}$  er en følge av partisjoner av rekktanglet  $R = [a, b] \times [c, d]$  slik at maskevidden  $|\Pi_n|$  går mot null, og la  $U_n$  være et utplukk for  $\Pi_n$ . For alle kontinuerlige funksjoner  $f : R \rightarrow \mathbb{R}$  er da*

$$\iint_R f(x, y) dx dy = \lim_{n \rightarrow \infty} R(\Pi_n, U_n)$$

*Bevis:* Anta at  $\epsilon > 0$  er gitt. Siden  $f$  er uniformt kontinuerlig og maskevidden til partisjonene går mot null, finnes det en  $N \in \mathbb{N}$  slik at hvis  $n \geq N$ , så er  $|f(\mathbf{z}) - f(\mathbf{w})| < \frac{\epsilon}{|R|}$  for alle  $\mathbf{z}, \mathbf{w}$  som hører til i samme delrekktangel  $R_{ij}$ . På samme måte som i forrige bevis medfører dette at  $\mathcal{O}(\Pi_n) - N(\Pi_n) < \epsilon$ . Siden både  $\iint_R f(x, y) dx dy$  og  $R(\Pi_n, U_n)$  ligger mellom  $\mathcal{O}(\Pi_n)$  og  $N(\Pi_n)$ , betyr dette at

$$|\iint_R f(x, y) dx dy - R(\Pi_n, U_n)| < \epsilon$$

for alle  $n \geq N$ , og setningen er dermed bevist.  $\square$

**Bemerkning:** Setningen ovenfor gjelder faktisk for alle *integrerbare* funksjoner  $f$  og ikke bare for de kontinuerlige. For å bevise dette må vi vise ulikheten  $\mathcal{O}(\Pi_n) - N(\Pi_n) < \epsilon$  uten å bruke uniform kontinuitet. Den følger fra et generelt resultat som sier at dersom  $f$  er integrerbar, så finnes det for enhver  $\epsilon > 0$  en  $\delta > 0$  slik at hvis  $|\Pi| < \delta$ , så er  $\mathcal{O}(\Pi) - N(\Pi) < \epsilon$ . Det krever litt fingerferdighet å bevise dette, men har du lyst til å prøve deg, kan du bruke beviset for lemma 8.5.6 i *Kalkulus* som modell.

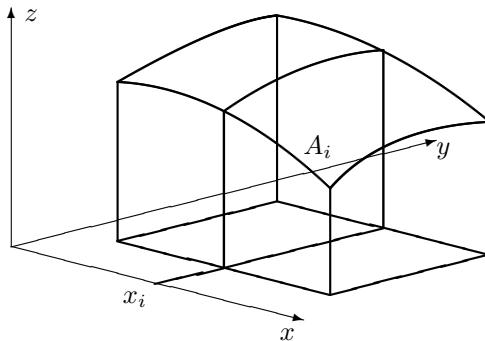
### Itererte integraler

Vi har nå definert integralet av en funksjon av to variable over et rektangel og vist at alle kontinuerlige funksjoner er integrerbare, men vi har foreløpig ingen effektive metoder for å regne ut dobbeltintegraler. For å finne frem til slike metoder skal vi ta utgangspunkt i en annen idé for å regne ut volumet under en funksjonsgraf.

Som tidligere deler vi intervallet  $[a, b]$  i  $n$  deler ved punktene

$$a = x_0 < x_1 < x_2 < \dots < x_n = b.$$

For hver  $x_i$  tar vi et tverrsnitt av legemet normalt på  $x$ -aksen (se figur 3).



Figur 3: Tverrsnittet  $A_i$

Arealet av dette tverrsnittet er

$$A_i = \int_c^d f(x_i, y) dy.$$

Tar vi dette arealet  $A_i$  og ganger med lengden  $x_i - x_{i-1}$  av det  $i$ -te intervallet, får vi et volum som er svært nær volumet til den delen av legemet som ligger mellom  $x_{i-1}$  og  $x_i$ . Ved å summere ser vi at

$$\sum_{i=1}^n \int_c^d f(x_i, y) dy (x_i - x_{i-1})$$

er en god tilnærming til volumet og at denne tilnærmingen bør bli bedre dess finere partisjonen er. Vi bør altså ha

$$V \approx \sum_{i=1}^n \left[ \int_c^d f(x_i, y) dy \right] (x_i - x_{i-1})$$

Definerer vi en funksjon  $F$  ved

$$F(x) = \int_c^d f(x, y) dy$$

ser vi at

$$\sum_{i=1}^n \left[ \int_c^d f(x_i, y) dy \right] \Delta x = \sum_{i=1}^n F(x_i) \Delta x$$

er en Riemann-sum for  $F$ . Følgelig vil

$$\sum_{i=1}^n F(x_i) \Delta x \longrightarrow \int_a^b F(x) dx = \int_a^b \left[ \int_c^d f(x, y) dy \right] dx$$

når oppdelingen blir finere. Kombinerer vi de formlene vi nå har, ser vi at vi burde ha

$$V = \int_a^b \left[ \int_c^d f(x, y) dy \right] dx.$$

I dette uttrykket kalles

$$\int_a^b \left[ \int_c^d f(x, y) dy \right] dx$$

et *iterert integral* og det har følgende tolkning: Først integrerer vi  $f(x, y)$  som en funksjon av  $y$  mens vi later som  $x$  er en konstant. Deretter integrerer vi resultatet som en funksjon av  $x$  ( $y$  har nå forsvunnet fra uttrykket siden vi har satt inn grensene  $y = c$  og  $y = d$ ).

I argumentet ovenfor kuttet vi opp volumet i skiver normalt på  $x$ -aksen. Ved isteden å kutte opp volumet i skiver normalt på  $y$ -aksen, kan vi på akkurat samme måte komme frem til formelen

$$V = \int_c^d \left[ \int_a^b f(x, y) dx \right] dy$$

Sammenligner vi de tre uttrykkene vi nå har for volumet  $V$ , ser vi at

$$\iint_R f(x, y) dx dy = \int_a^b \left[ \int_c^d f(x, y) dy \right] dx = \int_c^d \left[ \int_a^b f(x, y) dx \right] dy$$

Fordelen ved de to siste uttrykkene er at de bare involverer ting vi kan, nemlig vanlig integrasjon av funksjoner av én variabel. Vi har altså redusert integrasjon i to variable til to ganger integrasjon i én variabel.

Før vi gjennomfører et fullverdig matematisk bevis for formlene ovenfor (argumentet vi hittil har gitt, inneholder litt for mye ønsketenking om hva forskjellige uttrykk ”bør” konvergere mot), skal vi se hvordan de brukes i praksis.

**Eksempel 1:** Beregn  $\iint_R x^2y \, dxdy$  der  $R = [0, 1] \times [1, 3]$ . Ifølge formelen ovenfor er

$$I = \iint_R x^2y \, dxdy = \int_1^3 \left[ \int_0^1 x^2y \, dx \right] dy$$

Integratorer vi  $\int_0^1 x^2y \, dx$  som om  $y$  var en konstant, får vi

$$\int_0^1 x^2y \, dx = \left[ \frac{1}{3}x^3y \right]_{x=0}^{x=1} = \frac{1}{3} \cdot 1^3 \cdot y - \frac{1}{3} \cdot 0^3 \cdot y = \frac{1}{3}y$$

Dermed er

$$I = \int_1^3 \frac{1}{3}y \, dy = \left[ \frac{y^2}{6} \right]_1^3 = \frac{3^2}{6} - \frac{1^2}{6} = \frac{8}{6} = \frac{4}{3}$$

La oss for treningens skyld også regne ut integralene i motsatt rekkefølge:

$$I = \iint_R x^2y \, dxdy = \int_0^1 \left[ \int_1^3 x^2y \, dy \right] dx$$

Integratorer vi  $\int_1^3 x^2y \, dy$  som om  $x$  var en konstant, får vi

$$\int_1^3 x^2y \, dy = \left[ \frac{1}{2}x^2y^2 \right]_{y=1}^{y=3} = \frac{1}{2}x^2 \cdot 3^2 - \frac{1}{2}x^2 \cdot 1^2 = 4x^2$$

Altså er

$$I = \int_0^1 4x^2 \, dx = \left[ \frac{4}{3}x^3 \right]_0^1 = \frac{4}{3}$$



La oss se på et eksempel til.

**Eksempel 2:** Beregn  $I = \iint_R y^3 e^{xy^2} \, dxdy$  der  $R = [0, 1] \times [0, 2]$ .

Vi får

$$\begin{aligned} I &= \int_0^2 \left[ \int_0^1 y^3 e^{xy^2} dx \right] dy = \int_0^2 \left[ ye^{xy^2} \right]_{x=0}^{x=1} dy \\ &= \int_0^2 (ye^{y^2} - y) dy = \left[ \frac{1}{2}e^{y^2} - \frac{1}{2}y^2 \right]_0^2 = \\ &= \frac{1}{2}e^4 - \frac{1}{2} \cdot 2^2 - \frac{1}{2}e^0 + \frac{1}{2} \cdot 0^2 = \frac{1}{2}(e^4 - 5) \end{aligned}$$

Hvis du forsøker å regne ut dette integralet ved å integrere med hensyn på  $y$  først, får du atskillig større problemer. Det er faktisk ikke uvanlig at den ene integrasjonsrekkefølgen gir mye enklere regninger enn den andre, så det kan lønne seg å se skikkelig på integralet før man begynner. ♣

Vi skal nå bevise at vi virkelig har lov til å regne ut dobbeltintegraler ved hjelp av iterert integrasjon. For å dekke fremtidige anvendelser er vi nødt til å formulere resultatet for funksjoner som ikke nødvendigvis er kontinuerlige. Dette gjør at selve teoremet blir litt langt og kronglete, men beviset er heldigvis ikke så vanskelig.

**Teorem 6.1.7** *Anta at  $R = [a, b] \times [c, d]$  er et rektangel i  $\mathbb{R}^2$  og at  $f : R \rightarrow \mathbb{R}$  er integrerbar. Dersom funksjonen*

$$y \mapsto f(x, y)$$

*er integrerbar over  $[c, d]$  for alle  $x \in [a, b]$ , så er funksjonen  $F(x) = \int_c^d f(x, y) dy$  integrerbar over  $[a, b]$  og*

$$\iint_R f(x, y) dxdy = \int_a^b \left[ \int_c^d f(x, y) dy \right] dx$$

*Tilsvarende gjelder om vi bytter om variablene: Dersom funksjonen*

$$x \mapsto f(x, y)$$

*er integrerbar over  $[a, b]$  for alle  $y \in [c, d]$ , så er funksjonen  $G(y) = \int_a^b f(x, y) dx$  integrerbar over  $[c, d]$  og*

$$\iint_R f(x, y) dxdy = \int_c^d \left[ \int_a^b f(x, y) dx \right] dy$$

Før vi beviser teoremet, formulerer vi et resultat for kontinuerlige funksjoner som er lettere å få oversikt over (funksjonene som inngår er automatisk integrerbare).

**Korollar 6.1.8** *Anta at  $R = [a, b] \times [c, d]$  er et rektangel i  $\mathbb{R}^2$  og at  $f : R \rightarrow \mathbb{R}$  er kontinuerlig. Da er*

$$\iint_R f(x, y) dx dy = \int_a^b \left[ \int_c^d f(x, y) dy \right] dx = \int_c^d \left[ \int_a^b f(x, y) dx \right] dy$$

*Bevis:* Alle kontinuerlige funksjoner (både i én og to variable) er integrerbare, og betingelsene i teoremet er derfor oppfylt.  $\square$

*Bevis for teorem 6.1.7:* Vi nøyer oss med å vise den første halvdelen av teoremet — den andre vises på akkurat samme måte med  $x$ -er og  $y$ -er byttet om.

Vi starter med en partisjon  $\Pi$  av rektanglet  $R$  gitt ved

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

$$c = y_0 < y_1 < \dots < y_{m-1} < y_m = d$$

og lar  $P$  være partisjonen av intervallet  $[a, b]$  gitt ved

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

(vi bruker altså de samme  $x$ -delepunktene som i  $\Pi$ ). Vi lar  $N(\Pi)$  og  $\mathcal{O}(\Pi)$  være henholdsvis nedre og øvre trappesum for  $f$  med hensyn på  $\Pi$ , og vi lar  $n(P)$  og  $\phi(P)$  være henholdsvis nedre og øvre trappesum for funksjonen  $F(x) = \int_c^d f(x, y) dy$  med hensyn på  $P$ . Observer at dersom vi kan vise at

$$N(\Pi) \leq n(P) \quad \text{og} \quad \phi(P) \leq \mathcal{O}(\Pi),$$

så vil det følge at  $F$  er integrerbar på  $[a, b]$  og at

$$\int_a^b F(x) dx = \iint_R f(x, y) dx dy$$

(dette er fordi  $f$  er integrerbar over  $R$ , og vi derfor kan få både  $N(\Pi)$  og  $\mathcal{O}(\Pi)$  så nær  $\iint_R f(x, y) dx dy$  vi måtte ønske).

La oss vise at  $\phi(P) \leq \mathcal{O}(\Pi)$  — den andre ulikheten  $N(\Pi) \leq n(P)$  bevises på akkurat samme måte og overlates til leserne. Per definisjon er

$$\phi(P) = \sum_{i=1}^n N_i(x_i - x_{i-1})$$

der

$$N_i = \sup\{F(x) \mid x \in [x_{i-1}, x_i]\}$$

For alle  $x \in [x_i, x_{i-1}]$  er

$$F(x) = \int_c^d f(x, y) dy = \sum_{j=1}^m \int_{y_{j-1}}^{y_j} f(x, y) dy \leq \sum_{j=1}^m M_{ij}(y_j - y_{j-1})$$

der  $M_{ij}$  som vanlig er supremum til  $f(x, y)$  over rektanglet  $R_{ij}$ . Dermed er også

$$N_i \leq \sum_{j=1}^m M_{ij}(y_j - y_{j-1})$$

og vi får

$$\phi(P) = \sum_{i=1}^n N_i(x_i - x_{i-1}) \leq \sum_{i=1}^n \sum_{j=1}^m M_{ij}(x_i - x_{i-1})(y_j - y_{j-1}) = \mathcal{O}(\Pi)$$

og teoremet er bevist.  $\square$

## Dobbelintegraler i MATLAB

Det er lett å bruke MATLAB til å regne ut dobbeltintegraler over rektangler, men kommandonavnene kan virke litt mystiske første gang du ser dem. Den grunnleggende kommandoen er `>> dblquad`. Her står `dbl` for “double”, mens `quad` står for “quadrature”, et gammelt ord for integrasjon. I tillegg må du fortelle MATLAB hvilken funksjon du vil integrere, og da kan du f.eks. bruke funksjonstilordningskomandoen `@`. For å integrere funksjonen  $f(x,y) = x^2y$  over rektanglet  $R = [0, 1] \times [-1, 2]$ , skriver vi

```
>> dblquad(@(x,y)x.^2.*y,0,1,-1,2)
```

og får svaret

```
ans =
0.5000
```

Legg merke til syntaksen i kommandoen: Først skriver vi `@(x,y)x.^2.*y` for å definere funksjonen, deretter kommer grensene i  $x$ -retning, 0 og 1, og til slutt grensene i  $y$ -retning, -1 og 2.

Skal du arbeide mye med et litt komplisert funksjonsuttrykk, kan det lønne seg å lagre funksjonsuttrykket på en m-fil og bare kalle på filnavnet. Vil du gjøre dette med funksjonene ovenfor, lager du en m-fil `integrand.m` med dette innholdet

```
function z=integrand(x,y);
z=x.^2.*y;
```

og utfører beregningene ved å skrive

```
>> dblquad(@integrand,0,1,-1,2)
```

## Oppgaver til seksjon 6.1

1. Regn ut dobbeltintegralene

- a)  $\iint_R xy \, dx dy$  der  $R = [1, 2] \times [2, 4]$
- b)  $\iint_R (x + \sin y) \, dx dy$  der  $R = [0, 1] \times [0, \pi]$
- c)  $\iint_R x^2 e^y \, dx dy$  der  $R = [-1, 1] \times [0, 1]$
- d)  $\iint_R x \cos(xy) \, dx dy$  der  $R = [1, 2] \times [\pi, 2\pi]$

e)  $\iint_R xye^{x^2y} dxdy \quad \text{der } R = [0, 2] \times [1, 2]$

f)  $\iint_R \ln(xy) dxdy \quad \text{der } R = [1, e] \times [1, e]$

g)  $\iint_R \frac{1}{1+x^2y} dxdy \quad \text{der } R = [1, \sqrt{3}] \times [0, 1]$

2. Bruk MATLAB til å regne ut integralene i oppgave 1.

3. Vis at enhver nedre trappesum til en funksjon  $f$  over et rektangel  $R$  er mindre enn eller lik enhver øvre trappesum, dvs. at hvis  $\Pi_1$  og  $\Pi_2$  er to partisjoner av  $R$ , så er  $N(\Pi_1) \leq \mathcal{O}(\Pi_2)$ . Hint: La  $\Pi$  være en tredje partisjon som inneholder alle delepunkter i både  $\Pi_1$  og  $\Pi_2$ , og vis at  $N(\Pi_1) \leq N(\Pi) \leq \mathcal{O}(\Pi) \leq \mathcal{O}(\Pi_2)$

4. Vis at  $f$  er integrerbar over  $R$  hvis og bare hvis det for hver  $\epsilon > 0$  finnes en partisjon  $\Pi$  slik at  $\mathcal{O}(\Pi) - N(\Pi) < \epsilon$ . (Du kan få bruk for forrige oppgave)

5. Bevis setning 6.1.2 (du kan få bruk for de to foregående oppgavene).

6. Vis at funksjonen  $f(x) = x^2$  ikke er uniformt kontinuerlig på  $\mathbb{R}$ .

7. Anta at  $f : R \rightarrow \mathbb{R}$  er en kontinuerlig funksjon på et rektangel  $R = [a, b] \times [c, d]$ . Vis at det finnes et punkt  $(\bar{x}, \bar{y})$  i  $R$  slik at

$$\frac{\iint_R f(x, y) dxdy}{|R|} = f(\bar{x}, \bar{y})$$

der  $|R|$  er arealet til  $R$ . Dette kalles ofte *middeleverdisetningen for dobbeltintegraler*.

## 6.2 Dobbeltintegraler over begrensede områder

Hittil har vi bare definert dobbeltintegraler  $\iint_R f(x, y) dxdy$  når  $R$  er et rektangel, men i praksis må vi ofte integrere over mer kompliserte områder. Vi skal nå se hvordan dette kan gjøres.

For å slippe å begynne helt på nytt skal vi bruke et knep som knytter de nye integralene til dem vi allerede har definert. Anta at vi ønsker å integrere en funksjon  $f$  over et begrenset område  $A$ . Vi plukker først ut et rektangel  $R$  som er stort nok til å inneholde  $A$ . Deretter definerer vi funksjonen  $f_A : R \rightarrow \mathbb{R}$  ved

$$f_A(x, y) = \begin{cases} f(x, y) & \text{hvis } (x, y) \in A \\ 0 & \text{hvis } (x, y) \notin A \end{cases}$$

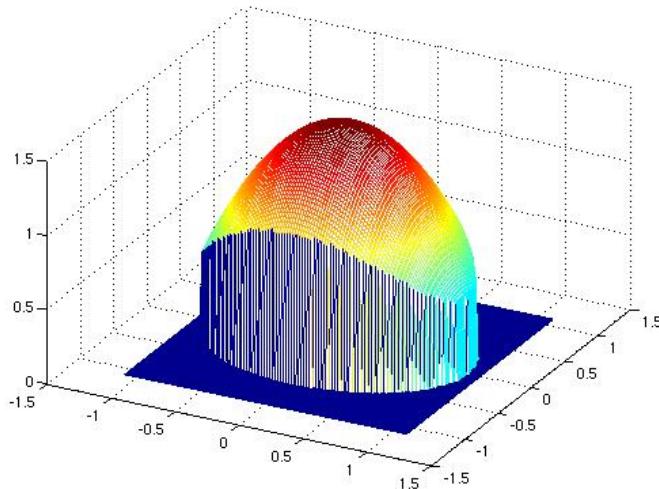
Vi sier at  $f$  er *integrerbar* over  $A$  hvis  $f_A$  er integrerbar over  $R$ , og i så fall definerer vi

$$\iint_A f(x, y) dxdy = \iint_R f_A(x, y) dxdy$$

**Bemerkning:** Strengt tatt burde vi vise at definisjonen ovenfor ikke avhenger av hvilket rektangel  $R$  vi bruker (så lenge det inneholder  $A$ ). Vi overlater dette til de

ivrigste leserne.

Figur 1 viser et eksempel på en funksjon  $f_A$ . Denne funksjonen er “tvunget” til å være 0 utenfor en ellipse i planet.



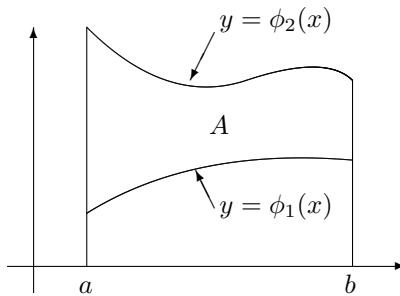
Figur 1: En funksjon  $f_A$ .

Vi skal foreløpig ikke gå nærmere inn på de teoretiske aspektene ved dette integralbegrepet, men nevner bare at siden vi nå også må ta hensyn til hvor komplisert mengden  $A$  er, er denne teorien mer innviklet enn den vi så på ovenfor. La oss først konsentrere oss om de problemene som oftest dukker opp i praksis.

Anta at  $\phi_1, \phi_2: [a, b] \rightarrow \mathbb{R}$  er to kontinuerlige funksjoner av én variabel slik at  $\phi_1(x) \leq \phi_2(x)$ . Vi får da definert et område  $A$  i planet ved

$$A = \{(x, y) : a \leq x \leq b, \phi_1(x) \leq y \leq \phi_2(x)\}$$

(se figur 2).



Figur 2: Område av type I

Vi skal si at et slikt område er av *type I*. Følgende resultat er intuitivt rimelig, og vi skal foreløpig bruke det med bare “et halvt bevis”.

**Setning 6.2.1** *Anta at  $A$  er av type I, og at  $f: A \rightarrow \mathbb{R}$  er kontinuerlig. Da er  $f$*

integrerbar over  $A$  og

$$\iint_A f(x, y) dx dy = \int_a^b \left[ \int_{\phi_1(x)}^{\phi_2(x)} f(x, y) dy \right] dx$$

*Bevis:* Vi putter  $A$  inn i rektanglet  $R = [a, b] \times [c, d]$ . Det er lett å sjekke at funksjonen  $y \mapsto f_A(x, y)$  er integrerbar over  $[c, d]$ , så setningen ville ha fulgt fra teorem 6.1.7 dersom vi bare visste at  $f_A$  var integrerbar over  $R = [a, b] \times [c, d]$ . Dette er lettere å vise når vi har utviklet litt mer maskineri, så vi utsetter denne delen av beviset til seksjon 6.6.  $\square$

**Eksempel 1:** Regn ut  $\iint_A (x + y^2) dx dy$  der

$$A = \{(x, y) : 0 \leq x \leq 1 \quad \text{og} \quad x \leq y \leq e^x\}$$

Ifølge teoremet ovenfor er

$$\iint_A (x + y^2) dx dy = \int_0^1 \left[ \int_x^{e^x} (x + y^2) dy \right] dx$$

Vi regner ut det innerste integralet:

$$\int_x^{e^x} (x + y^2) dy = \left[ xy + \frac{y^3}{3} \right]_{y=x}^{y=e^x} = xe^x + \frac{e^{3x}}{3} - x^2 - \frac{x^3}{3}$$

Setter vi dette inn i uttrykket ovenfor, får vi

$$\iint_A (x + y^2) dx dy = \int_0^1 \left( xe^x + \frac{e^{3x}}{3} - x^2 - \frac{x^3}{3} \right) dx$$

De tre siste integralene kan regnes ut direkte, men i det første må vi bruke delvis integrasjon. Vi setter  $u = x$ ,  $v' = e^x$  og får  $u' = 1$ ,  $v = e^x$ . Dette gir

$$\begin{aligned} \int_0^1 xe^x dx &= [x \cdot e^x]_0^1 - \int_0^1 1 \cdot e^x dx = \left[ xe^x - e^x \right]_0^1 = \\ &= 1 \cdot e^1 - e^1 - 0 \cdot e^0 + e^0 = 1 \end{aligned}$$

Dermed er

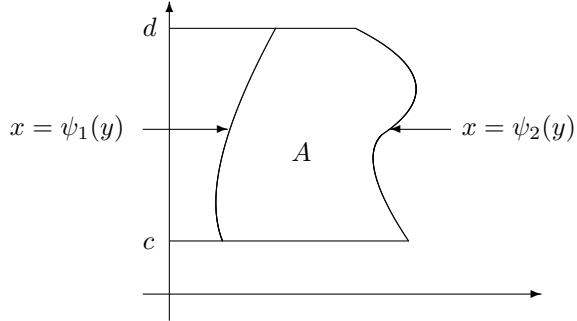
$$\begin{aligned} \iint_A xe^x dx &= \int_0^1 \left( xe^x + \frac{e^{3x}}{3} - x^2 - \frac{x^3}{3} \right) dx = \\ &= 1 + \left[ \frac{e^{3x}}{9} - \frac{x^3}{3} - \frac{x^4}{3 \cdot 4} \right]_0^1 = 1 + \frac{e^3}{9} - \frac{1}{3} - \frac{1}{12} - \frac{1}{9} = \frac{17}{36} + \frac{e^3}{9} \end{aligned}$$



Vi skal nå se på en annen situasjon der  $x$ -aksen og  $y$ -aksen har byttet roller. Vi sier at en mengde  $A \subset \mathbb{R}^2$  er av *type II* dersom det finnes tall  $c, d$ , der  $c < d$ , og to kontinuerlige funksjoner  $\psi_1, \psi_2: [c, d] \rightarrow \mathbb{R}$ , der  $\psi_1(y) \leq \psi_2(y)$  for alle  $y \in [c, d]$ , slik at

$$A = \{(x, y) : c \leq y \leq d \quad \text{og} \quad \psi_1(y) \leq x \leq \psi_2(y)\}$$

(se figur 3).



Figur 3: Område av type II

Vi har følgende parallel til setning 6.2.1.

**Setning 6.2.2** Anta at  $f: A \rightarrow \mathbb{R}$  er kontinuerlig, og at

$$A = \{(x, y) : c \leq y \leq d \quad \text{or} \quad \psi_1(y) \leq x \leq \psi_2(y)\}$$

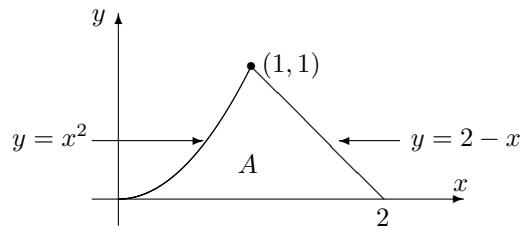
er av type II. Da er  $f$  integrererbar over  $A$  og

$$\iint_A f(x, y) \, dxdy = \int_c^d \left[ \int_{\psi_1(y)}^{\psi_2(y)} f(x, y) \, dx \right] dy$$

*Bevis:* Identisk med beviset for 6.2.1 med  $x$  og  $y$  byttet om.

□

**Eksempel 2:** Regn ut  $\iint_A xy^2 \, dxdy$  der  $A$  er området avgrenset av  $x$ -aksen og de to grafene  $y = x^2$  og  $y = 2 - x$  for  $x \geq 0$ . Figur 4 viser området  $A$ .



Figur 4: Området i eksempel 2

Dette området er av type II siden det ligger klemt mellom to funksjonsgrøper og ytterpunktene  $y = 0$  og  $y = 1$ . For å finne funksjonsuttrykkene  $\psi_1(y)$  og  $\psi_2(y)$

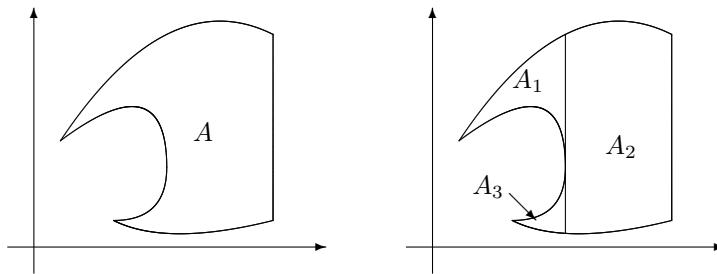
må vi løse ligningene  $y = x^2$  og  $y = 2 - x$  med hensyn på  $x$ . Vi får  $\psi_1(y) = \sqrt{y}$ ,  $\psi_2(y) = 2 - y$ . Altså er

$$\begin{aligned}
 \iint_A xy^2 \, dx \, dy &= \int_0^1 \left[ \int_{\sqrt{y}}^{2-y} xy^2 \, dx \right] dy = \\
 &= \int_0^1 \left[ \frac{1}{2}x^2y^2 \right]_{x=\sqrt{y}}^{x=2-y} dy = \\
 &= \int_0^1 \left( \frac{1}{2}(2-y)^2y^2 - \frac{1}{2}(\sqrt{y})^2y^2 \right) dy = \\
 &= \int_0^1 \left( \frac{y^4}{2} - \frac{5}{2}y^3 + 2y^2 \right) dy = \left[ \frac{y^5}{10} - \frac{5}{8}y^4 + \frac{2}{3}y^3 \right]_0^1 \\
 &= \left[ \frac{1}{10} - \frac{5}{8} + \frac{2}{3} \right] = \frac{17}{120} \quad \clubsuit
 \end{aligned}$$

Ved å bryte et komplisert område opp i enklere deler, kan vi integrere over områder som er atskillig verre enn type I og type II. Figur 5a) viser et eksempel – området  $A$  er hverken av type I eller II, men deler vi det opp som vist i figur 5b), får vi tre deler som alle er av type I. Vi kan da regne ut  $\iint_A f(x, y) \, dx \, dy$  ved hjelp av formelen

$$\iint_A f(x, y) \, dx \, dy = \iint_{A_1} f(x, y) \, dx \, dy + \iint_{A_2} f(x, y) \, dx \, dy + \iint_{A_3} f(x, y) \, dx \, dy$$

og teknikkene ovenfor.

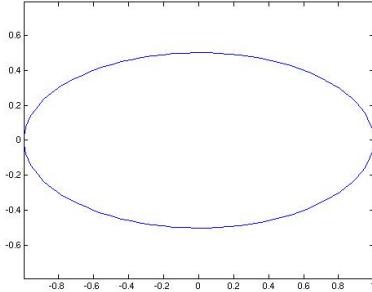


Figur 5: Oppdeling av integrasjonsområdet

La oss avslutte denne seksjonen med et eksempel som viser behovet for å integrere over andre områder enn rektangler. Eksemplet viser også de lange og ganske kompliserte utregningene vi ofte får.

**Eksempel 3:** Vi skal regne ut volumet som ligger over  $xy$ -planet og under grafen til  $f(x, y) = 1 - x^2 - 4y^2$ . Denne funksjonen er en paraboloid med snuten oppover,

og den er positiv når  $x^2 + 4y^2 \leq 1$ . Legg merke til at  $x^2 + 4y^2 = 1$  er en ellipse med store halvaksen  $a = 1$  langs  $x$ -aksen og lille halvaksen  $b = \frac{1}{2}$  langs  $y$ -aksen, og vi ønsker altså å integrere over innsiden av denne ellipsen (se figur 6).



Figur 6: Integrasjonsområdet.

Integratorer vi i  $y$ -retning først, må vi integrere fra den nedre “halvellipsen”  $y = -\frac{1}{2}\sqrt{1-x^2}$  til den øvre  $y = \frac{1}{2}\sqrt{1-x^2}$ , mens den ytre integrasjonen i  $x$ -retning går fra  $-1$  til  $1$ . Volumet  $V$  er altså gitt av integralet

$$V = \int_{-1}^1 \left[ \int_{-\frac{1}{2}\sqrt{1-x^2}}^{\frac{1}{2}\sqrt{1-x^2}} (1-x^2-4y^2) dy \right] dx$$

Integratorer vi med hensyn på  $y$ , får vi

$$\begin{aligned} V &= \int_{-1}^1 \left[ \left( y - x^2 y - \frac{4}{3} y^3 \right) \Big|_{y=-\frac{1}{2}\sqrt{1-x^2}}^{y=\frac{1}{2}\sqrt{1-x^2}} \right] dx = \\ &= \int_{-1}^1 \left[ \left( \frac{1}{2}\sqrt{1-x^2} - \frac{1}{2}x^2\sqrt{1-x^2} - \frac{4}{3} \left( \frac{1}{2}\sqrt{1-x^2} \right)^3 \right) - \right. \\ &\quad \left. - \left( -\frac{1}{2}\sqrt{1-x^2} + \frac{1}{2}x^2\sqrt{1-x^2} + \frac{4}{3} \left( \frac{1}{2}\sqrt{1-x^2} \right)^3 \right) \right] dx = \\ &= \int_{-1}^1 \left( \sqrt{1-x^2} - x^2\sqrt{1-x^2} - \frac{1}{3}(1-x^2)^{3/2} \right) dx = \\ &= \int_{-1}^1 \left( (1-x^2)\sqrt{1-x^2} - \frac{1}{3}(1-x^2)^{3/2} \right) dx = \int_{-1}^1 \frac{2}{3}(1-x^2)^{3/2} dx \end{aligned}$$

For å løse dette integralet, setter vi  $u = \arcsin x$ . Da er  $x = \sin u$ ,  $dx = \cos u du$  og de nye grensene er  $\arcsin(-1) = -\frac{\pi}{2}$  og  $\arcsin(1) = \frac{\pi}{2}$ . Dermed har vi

$$V = \int_{-1}^1 \frac{2}{3}(1-x^2)^{3/2} dx = \frac{2}{3} \int_{-\pi/2}^{\pi/2} (1-\sin^2 u)^{3/2} \cos u du = \frac{2}{3} \int_{-\pi/2}^{\pi/2} \cos^4 u du$$

Trigonometriske integraler av denne typen dukker ofte opp når vi løser dobbelt-integraler, og det kan være lurt å ta en titt på seksjon 9.4 i *Kalkulus* som (blant annet) gir en innføring i hvordan man løser slike integraler. I integralet ovenfor er det lurt å bruke formelen for cosinus til den dobbelte vinkelen baklengs: Vi har  $\cos 2u = 2\cos^2 u - 1$  som gir  $\cos^2 u = \frac{1+\cos 2u}{2}$ . Setter vi dette inn i integralet ovenfor, får vi

$$\begin{aligned} V &= \frac{2}{3} \int_{-\pi/2}^{\pi/2} \cos^4 u \, du = \frac{2}{3} \int_{-\pi/2}^{\pi/2} \left( \frac{1+\cos 2u}{2} \right)^2 \, du = \\ &= \frac{1}{6} \int_{-\pi/2}^{\pi/2} (1 + 2\cos 2u + \cos^2 2u) \, du = \\ &= \frac{1}{6} \left[ u + \sin 2u \right]_{-\pi/2}^{\pi/2} + \frac{1}{6} \int_{-\pi/2}^{\pi/2} \cos^2 2u \, du = \frac{\pi}{6} + \frac{1}{6} \int_{-\pi/2}^{\pi/2} \cos^2 2u \, du \end{aligned}$$

For å løse det siste integralet, bruker vi igjen formelen for cosinus til den dobbelte vinkelen baklengs, og får  $\cos^2 2u = \frac{1+\cos 4u}{2}$ . Dermed er

$$\begin{aligned} V &= \frac{\pi}{6} + \frac{1}{6} \int_{-\pi/2}^{\pi/2} \cos^2 2u \, du = \frac{\pi}{6} + \frac{1}{12} \int_{-\pi/2}^{\pi/2} (1 + \cos 4u) \, du = \\ &= \frac{\pi}{6} + \frac{1}{12} \left[ u - \frac{1}{4} \sin 4u \right]_{-\pi/2}^{\pi/2} = \frac{\pi}{6} + \frac{\pi}{12} = \frac{\pi}{4} \end{aligned}$$

♣

Regnestykket ovenfor er langt og komplisert, men problemene vi støter på er ganske typiske for dobbeltintegraler. Med litt trening lærer du å beherske dem.

### Mer om dobbeltintegraler i MATLAB

MATLAB kan også brukes til å regne ut dobbeltintegraler over mer generelle områder. Teknikken vi bruker, ligger tett opp til definisjonen. Hovedtrikset er at MATLAB-kommandoer av typen `>>(f(x,y)<=g(x,y))` (legg merke til de ytterste parentesene) definerer funksjonen

$$F(x, y) = \begin{cases} 1 & \text{hvis } f(x, y) \leq g(x, y) \\ 0 & \text{ellers} \end{cases}$$

Skriver vi f.eks. `>>(x.^2+y.^2<=1)` (legg igjen merke til de ytterste parentesene!) har vi definert en funksjon

$$F(x, y) = \begin{cases} 1 & \text{hvis } x^2 + y^2 \leq 1 \\ 0 & \text{hvis } x^2 + y^2 > 1 \end{cases}$$

Skriver vi nå  $\gg (x.^3-y.^2) .*(x.^2+y.^2<=1)$  har vi dermed definert funksjonen

$$G(x, y) = \begin{cases} x^3 - y^2 & \text{hvis } x^2 + y^2 \leq 1 \\ 0 & \text{hvis } x^2 + y^2 > 1 \end{cases}$$

Det er akkurat slike funksjoner vi trenger for å regne ut integraler over generelle områder.

La oss bruke MATLAB til å regne ut integralet i eksempel 1 ovenfor, dvs.  $\iint_A (x + y^2) dx dy$  der

$$A = \{(x, y) : 0 \leq x \leq 1 \text{ og } x \leq y \leq e^x\}$$

Vi må først finne et rektangel  $R$  som omfatter mengden  $A$ . Siden den minste verdien  $y$  kan ha i området vårt er 0 og den største er  $e^1 = e < 3$ , kan vi f.eks. bruke  $R = [0, 1] \times [0, 3]$ . For å fange opp betingelsen  $x \leq y$ , innfører vi funksjonen  $(x \leq y)$ , og for å fange opp betingelsen  $y \leq e^x$ , innfører vi funksjonen  $(y \leq \exp(x))$ . Uttrykket

$$(x+y.^2) .*(x \leq y) .*(y \leq \exp(x))$$

vil dermed fange opp den funksjonen vi er interessert i. For å utføre integrasjonene, skriver vi

```
>> dblquad(@(x,y)(x+y.^2).*(x<=y).*(y<=exp(x)),0,1,0,3)
```

```
ans =
2.7040
```

La oss også prøve oss på integrasjon i eksempel 3. Her skal vi integrere  $f(x, y) = 1 - x^2 - 4y^2$  over området  $x^2 + 4y^2 \leq 1$ . Vi vet at ellipsen har sentrum i origo og halvakser  $a = 1$ ,  $b = \frac{1}{2}$ , så hele området ligger inni rektanglet  $R = [-1, 1] \times [-\frac{1}{2}, \frac{1}{2}]$ . For å regne ut integralet, gir vi kommandoen

```
>> dblquad(@(x,y)(1-x.^2-4*y.^2).*(x.^2+4*y.^2<=1),-1,1,-0.5,0.5)
```

```
ans =
0.7854
```

## Oppgaver til seksjon 6.2

1. Regn ut dobbeltintegralene

- a)  $\iint_R x^2 y \, dx dy$  der  $R = \{(x, y) : 0 \leq x \leq 2 \text{ og } 0 \leq y \leq x\}$
- b)  $\iint_R (x + 2xy) \, dx dy$  der  $R = \{(x, y) : 0 \leq x \leq 3 \text{ og } x \leq y \leq 2x + 1\}$
- c)  $\iint_R y \, dx dy$  der  $R = \{(x, y) : 1 \leq y \leq 2 \text{ og } y \leq x \leq y^2\}$
- d)  $\iint_R x \cos y \, dx dy$  der  $R = \{(x, y) : 0 \leq y \leq \frac{\pi}{2} \text{ og } 0 \leq x \leq \sin y\}$

- e)  $\iint_R e^{x^2} dx dy$  der  $R$  er området i første kvadrant avgrenset av  $x$ -aksen og linjene  $x = 1$  og  $y = x$ .
- f)  $\iint_R x^2 y dx dy$  der  $R$  er området avgrenset av kurvene  $y = x^2$  og  $y = \sqrt{x}$ .
- g)  $\iint_R x \cos(x + y) dx dy$  der  $R$  er trekanten med hjørner  $(0, 0), (\pi, 0), (\pi, \pi)$ .
- h)  $\iint_R \frac{1}{\sqrt{1-y^2}} dx dy$  der  $R$  er området gitt ved  $0 \leq y \leq \sin x$  og  $0 \leq x \leq \frac{\pi}{2}$ .
- i)  $\iint_R x dx dy$  der  $R$  er området mellom kurvene  $y = \ln x$  og  $y = \frac{x-1}{e-1}$ .

**2.** Bruk MATLAB til å regne ut integralene i oppgave 1.

**3.** Noen integraler er enklere å regne ut hvis vi bytter integrasjonsrekkefølgen. Løs disse integralene ved å utføre integrasjonene i motsatt rekkefølge. *Hint:* Lag en skisse over integrasjonsområdet før du prøver å bytte integrasjonsrekkefølgen.

$$\begin{aligned} a) & \int_0^1 \left[ \int_y^1 e^{x^2} dx \right] dy \\ b) & \int_0^{\frac{\pi}{2}} \left[ \int_x^{\frac{\pi}{2}} \frac{\sin y}{y} dy \right] dx \\ c) & \int_0^1 \left[ \int_{\sqrt{x}}^1 e^{\frac{x}{y^2}} dy \right] dx \end{aligned}$$

**4.** Vis at verdien til  $\iint_A f(x, y) dx dy$  ikke avhenger av hvilket rektangel  $R$  vi bruker i definisjonen.

**5.** Anta at  $R_1$  og  $R_2$  er to disjunkte mengder (dvs. at  $R_1 \cap R_2 = \emptyset$ ), og at  $f$  er integrerbar over både  $R_1$  og  $R_2$ . Vis at

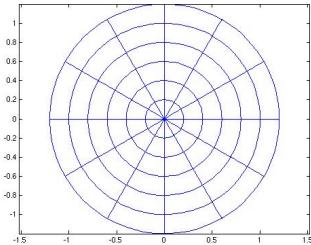
$$\iint_{R_1 \cup R_2} f(x, y) dx dy = \iint_{R_1} f(x, y) dx dy + \iint_{R_2} f(x, y) dx dy$$

*Hint:* Du kan bruke setning 6.1.2(ii).

### 6.3 Dobbeltintegraler i polarkoordinater

I en del dobbeltintegraler er det enklere å uttrykke integranden eller integrasjonsområdet (eller begge deler) ved hjelp av polarkoordinater, og da er det som regel også enklere å utføre integrasjonen i polarkoordinater. Når vi arbeider i polarkoordinater, er det naturlig å bruke en annen type partisjoner enn når vi arbeider i vanlige (kartesiske) koordinater. Istedentfor å bruke et rutenett definert av konstante  $x$ - og  $y$ -verdier, er det naturlig å bruke ett

med konstante  $r$ - og  $\theta$ -verdier. Figur 1 viser et slikt rutenett der sirklene svarer til faste  $r$ -verdier og strålene ut fra origo svarer til faste  $\theta$ -verdier.

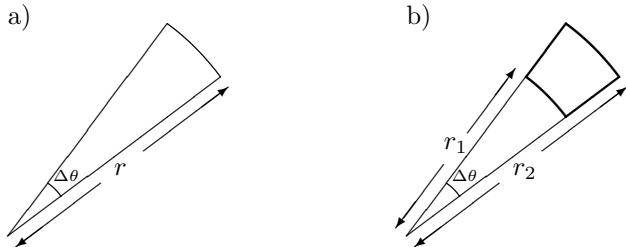


Figur 1. Partisjon i polarkoordinater

Før vi går videre, lønner det seg å finne en formel for arealet til rutene i nettet. Vi minner først om at arealet til sirkelsektoren i figur 2a) er  $\frac{1}{2}r^2\Delta\theta$  (forutsatt at vinkelen måles i radianer). Husker du ikke dette, er formelen lett å utlede: Arealet til hele sirkelen er  $\pi r^2$ , og den delen vi ser på, utgjør  $\frac{\Delta\theta}{2\pi}$  av det hele, og har følgelig areal  $\pi r^2 \cdot \frac{\Delta\theta}{2\pi} = \frac{1}{2}r^2\Delta\theta$ . Arealet til "ruten" i figur 2b) er differensen mellom to sirkelsegmenter som begge har samme vinkel  $\Delta\theta$ , men der den største har radius  $r_2$  og den minste radius  $r_1$ . Arealet er dermed

$$A = \frac{1}{2}r_2^2\Delta\theta - \frac{1}{2}r_1^2\Delta\theta = \frac{1}{2}(r_2 + r_1)(r_2 - r_1)\Delta\theta = r^*\Delta r\Delta\theta$$

der  $r^* = \frac{r_2+r_1}{2}$  er den midlere radien og  $\Delta r = r_2 - r_1$  er differensen mellom radiene.



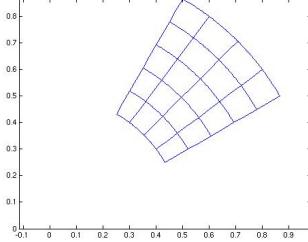
Figur 2: Et arealelement i polarkoordinater

Anta nå at vi har et "rekktangel"  $S$  i polarkoordinater, dvs. et område i planet avgrenset av sirkelbuene  $r = a$ ,  $r = b$  og strålene  $\theta = \alpha$ ,  $\theta = \beta$ . En partisjon

$$a = r_0 < r_1 < \dots < r_{n-1} < r_n = b$$

$$\alpha = \theta_0 < \theta_1 < \dots < \theta_{m-1} < \theta_m = \beta$$

av  $S$  gir opphav til en oppdeling som vist på figur 3.



Figur 3: Rutenett i polarkoordinater

Hvis  $S_{ij}$  er den  $ij$ -te ruten i denne oppdelingen (den som inneholder punkter  $(r, \theta)$  der  $r_{i-1} \leq r \leq r_i$  og  $\theta_{j-1} \leq \theta \leq \theta_j$ ), så vet vi fra utregningen ovenfor at arealet til  $S_{ij}$  er

$$|S_{ij}| = r_{ij}^*(r_i - r_{i-1})(\theta_j - \theta_{j-1})$$

der  $r_{ij}^* = \frac{r_i + r_{i-1}}{2}$  er den midlere radien. Det naturlige midtpunktet i  $S_{ij}$  er punktet med polarkoordinater  $(r_{ij}^*, \theta_{ij}^*)$  der  $\theta_{ij}^* = \frac{\theta_{j-1} + \theta_j}{2}$ . Dette punktet har vanlige koordinater  $(x_{ij}^*, y_{ij}^*) = (r_{ij}^* \cos \theta_{ij}^*, r_{ij}^* \sin \theta_{ij}^*)$ , og har vi en (positiv) funksjon  $f$  definert på  $S$ , ser vi at volumet under funksjonsgrafen er tilnærmet gitt ved

$$\begin{aligned} V &\approx \sum_{i=1}^n \sum_{j=1}^m f(x_{ij}^*, y_{ij}^*) |S_{ij}| = \\ &= \sum_{i=1}^n \sum_{j=1}^m f(r_{ij}^* \cos \theta_{ij}^*, r_{ij}^* \sin \theta_{ij}^*) r_{ij}^* (r_i - r_{i-1})(\theta_j - \theta_{j-1}) \end{aligned}$$

Dette er en Riemann-sum for funksjonen  $f(r \cos \theta, r \sin \theta)r$  over rektangelet  $R = [a, b] \times [\alpha, \beta]$ , og når oppdelingen blir finere og finere, vil den nærme seg integralet

$$\iint_R f(r \cos \theta, r \sin \theta)r dr d\theta$$

(legg merke til faktoren  $r$  som har kommet inn i tillegg til funksjonen  $f$ !) Dermed burde

$$V = \iint_R f(r \cos \theta, r \sin \theta)r dr d\theta$$

På den annen side — dersom vi ikke hadde byttet til polarkoordinater, men holdt oss til  $x$  og  $y$ , ville volumet være gitt som

$$V = \iint_S f(x, y) dx dy$$

og vi har dermed

$$\iint_S f(x, y) dx dy = \iint_R f(r \cos \theta, r \sin \theta)r dr d\theta$$

der  $S$  er integrasjonsområdet beskrevet i vanlige koordinater, og  $R$  er integrasjonsområdet beskrevet i polarkoordinater.

Argumentet ovenfor er ikke et matematisk bevis siden det baserer seg på “ikke-matematisk” intuisjon om hva volum er, men det kan være liten tvil om at konklusjonen er riktig. Vi formulerer den som en setning.

**Setning 6.3.1** Anta at  $S$  er et område i  $xy$ -planet som i polarkoordinater kan beskrives ved ulikheterne  $a \leq r \leq b$ ,  $\alpha \leq \theta \leq \beta$ . For enhver kontinuerlig funksjon  $f : S \rightarrow \mathbb{R}$  er da

$$\iint_S f(x, y) dx dy = \iint_R f(r \cos \theta, r \sin \theta) r dr d\theta$$

der  $R = [a, b] \times [\alpha, \beta]$ . Med andre ord

$$\begin{aligned} \iint_S f(x, y) dx dy &= \int_a^b \left[ \int_{\alpha}^{\beta} f(r \cos \theta, r \sin \theta) r d\theta \right] dr = \\ &= \int_{\alpha}^{\beta} \left[ \int_a^b f(r \cos \theta, r \sin \theta) r dr \right] d\theta \end{aligned}$$

Vi skal ikke gi et formelt bevis for denne setningen nå siden den vil følge fra et mer generelt resultat om skifte av variabel som vi skal bevise i seksjon 6.7. Istedet ser vi på et eksempel.

**Eksempel 1:** Vi skal regne ut  $\iint_S \ln(x^2 + y^2) dx dy$  der  $S$  er området i første kvadrant avgrenset av  $x$ -aksen, linjen  $y = x$  og de to sirklene  $x^2 + y^2 = 4$ ,  $x^2 + y^2 = 16$  (lag en figur). I polarkoordinater kan  $A$  beskrives ved

$$2 \leq r \leq 4 \quad \text{og} \quad 0 \leq \theta \leq \frac{\pi}{4}$$

Setningen ovenfor forteller oss at

$$\iint_A \ln(x^2 + y^2) dx dy = \int_2^4 \left[ \int_0^{\frac{\pi}{4}} \ln(r^2) \cdot r d\theta \right] dr = \frac{\pi}{4} \int_2^4 r \ln r^2 dr$$

Dette integralet kan vi løse ved delvis integrasjon. Vi setter  $u = \ln r^2$ ,  $v' = r$  og får  $u' = \frac{2}{r}$ ,  $v = \frac{1}{2}r^2$ . Dermed er

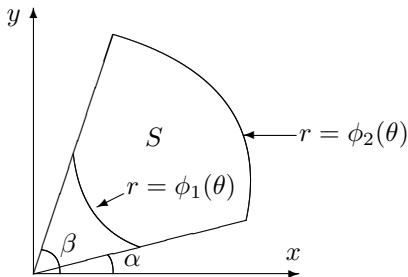
$$\int_2^4 r \ln r^2 dr = \left[ \frac{1}{2}r^2 \ln r^2 \right]_2^4 - \int_2^4 r dr = 8 \cdot \ln 16 - 2 \ln 4 - \left[ \frac{r^2}{2} \right]_2^4 = 28 \ln 2 - 6$$

og vi får

$$\iint_S \ln(x^2 + y^2) dx dy = \frac{\pi}{4} \int_2^4 r \ln r^2 dr = 7\pi \ln 2 - \frac{3\pi}{2}$$



Vi kan også bruke polarkoordinater på mer generelle områder enn de vi har beskrevet ovenfor. En type som ofte dukker opp, er den vi ser i figur 4 — her ligger  $\theta$  mellom to faste grenser  $\alpha$  og  $\beta$ , mens  $r$  er begrenset av to funksjoner  $\phi_1(\theta)$  og  $\phi_2(\theta)$ .



Figur 4: Et område i polarkoordinater

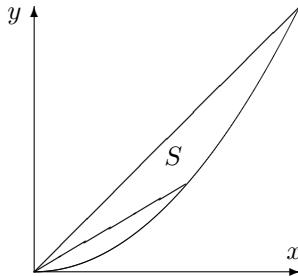
**Setning 6.3.2** La \$S\$ være et område i \$xy\$-planet som i polarkoordinater kan beskrives ved at \$\alpha \leq \theta \leq \beta\$ og \$\phi\_1(\theta) \leq r \leq \phi\_2(\theta)\$, der \$\phi\_1, \phi\_2 : [\alpha, \beta] \rightarrow \mathbb{R}\$ er to kontinuerlige, ikke-negative funksjoner slik at \$\phi\_1(\theta) \leq \phi\_2(\theta)\$ for alle \$\theta \in [\alpha, \beta]\$. Da er

$$\iint_S f(x, y) dx dy = \int_{\alpha}^{\beta} \left[ \int_{\phi_1(\theta)}^{\phi_2(\theta)} f(r \cos \theta, r \sin \theta) r dr \right] d\theta$$

Også dette resultatet vil følge fra det generelle resultatet om skifte av variable, så vi tar ikke med beviset her. Istedent ser vi på et eksempel.

**Eksempel 2:** Vi skal regne ut \$\iint\_S \frac{y}{\sqrt{x^2+y^2}} dx dy\$ der \$S\$ er området i første kvadrant avgrenset av parabelen \$y = x^2\$ og linjene \$y = \frac{\sqrt{3}}{3}x\$ og \$y = x\$ (se figur 5). Vi ser at integrasjonsområdet et avgrenset av verdiene \$\theta = \frac{\pi}{6}\$ og \$\theta = \frac{\pi}{4}\$ (tilsvarende de to linjene \$y = \frac{\sqrt{3}}{3}x\$ og \$y = x\$). Holder vi \$\theta\$ fast og lar \$r\$ variere, ser vi at vi kommer inn i integrasjonsområdet ved \$r = 0\$ og går ut av det når vi treffer parablen \$y = x^2\$. I polarkoordinater blir denne ligningen til \$r \sin \theta = r^2 \cos^2 \theta\$, som gir \$r = \frac{\sin \theta}{\cos^2 \theta}\$. Integrasjonsområdet er altså gitt ved

$$\frac{\pi}{6} \leq \theta \leq \frac{\pi}{4} \quad \text{og} \quad 0 \leq r \leq \frac{\sin \theta}{\cos^2 \theta}$$



Figur 5: Integrasjonsområdet i eksempel 2

Integralet blir dermed (husk  $r$ -faktoren!)

$$I = \iint_S \frac{y}{\sqrt{x^2 + y^2}} dx dy = \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \left[ \int_0^{\frac{\sin \theta}{\cos^2 \theta}} \frac{r \sin \theta}{r} r dr \right] d\theta = \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \left[ \int_0^{\frac{\sin \theta}{\cos^2 \theta}} r \sin \theta dr \right] d\theta$$

Integrasjon med hensyn på  $r$  gir

$$I = \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \left[ \frac{r^2}{2} \sin \theta \right]_{r=0}^{r=\frac{\sin \theta}{\cos^2 \theta}} d\theta = \frac{1}{2} \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \frac{\sin^3 \theta}{\cos^4 \theta} d\theta$$

Skriver vi om det siste integralet slik:

$$I = \frac{1}{2} \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \frac{\sin^3 \theta}{\cos^4 \theta} d\theta = \frac{1}{2} \int_{\frac{\pi}{6}}^{\frac{\pi}{4}} \frac{(1 - \cos^2 \theta) \sin \theta}{\cos^4 \theta} d\theta,$$

ser vi at  $u = \cos \theta$  er en naturlig substitusjon. Den gir

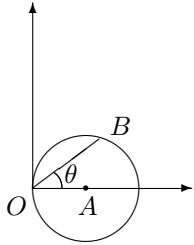
$$I = \frac{1}{2} \int_{\frac{\sqrt{3}}{2}}^{\frac{\sqrt{2}}{2}} \frac{(u^2 - 1)}{u^4} du$$

Resten er enkel integrasjon:

$$\begin{aligned} I &= \frac{1}{2} \int_{\frac{\sqrt{3}}{2}}^{\frac{\sqrt{2}}{2}} \frac{(u^2 - 1)}{u^4} du = \frac{1}{2} \int_{\frac{\sqrt{3}}{2}}^{\frac{\sqrt{2}}{2}} \left( \frac{1}{u^2} - \frac{1}{u^4} \right) du = \\ &= \frac{1}{2} \left[ -\frac{1}{u} + \frac{1}{3} \frac{1}{u^3} \right]_{\frac{\sqrt{3}}{2}}^{\frac{\sqrt{2}}{2}} = \frac{1}{2} \left( -\sqrt{2} + \frac{1}{3} 2\sqrt{2} + \frac{2\sqrt{3}}{3} - \frac{1}{3} \frac{8\sqrt{3}}{9} \right) = \frac{5\sqrt{3}}{27} - \frac{\sqrt{2}}{6} \quad \clubsuit \end{aligned}$$

Vi tar også med et eksempel der vi må arbeide litt for å finne grensene.

**Eksempel 3:** Vi skal integrere funksjonen  $f(x, y) = \sqrt{x^2 + y^2}$  over området  $S$  gitt ved  $x^2 - 2x + y^2 \leq 0$ . Vi har lyst til å skifte til polarkoordinater fordi integranden da blir så mye penere. Spørsmålet er hvordan området  $S$  kan beskrives i polarkoordinater. Vi ser først at dersom vi fullfører kvadratet, kan ulikheten  $x^2 - 2x + y^2 \leq 0$  skrives  $(x - 1)^2 + y^2 \leq 1$  som er en sirkelskive med sentrum i  $(1, 0)$  og radius 1. På figur 6 har vi tegnet inn denne sirkelen og en vinkel  $\theta$ . Vi er interessert i hvor langt vi kan gå i  $\theta$ -retning og fortsatt holde oss innenfor området, dvs. vi er interessert i lengden på linjestykket  $OB$ .



Figur 6: Integrasjonsområdet i eksempel 3

Vi må bruke litt geometri. Trekk normalen fra  $A(1,0)$  ned på linjestykket  $OB$ . Den treffer  $OB$  i et punkt  $C$  som ligger midt mellom  $O$  og  $B$  (hvorfor?), og ved definisjonen av cosinus er  $|OC| = |OA| \cos \theta = \cos \theta$ . Siden  $OB$  er dobbelt så lang, har vi  $|OB| = 2 \cos \theta$ . Det betyr at starter vi i  $O$  og går i en fast  $\theta$ -retning, går vi inn i integrasjonsområdet  $S$  når  $r = 0$  og ut når  $r = 2 \cos \theta$ . Legg også merke til at de  $\theta$ -verdiene som gir punkter i området  $S$ , ligger mellom  $-\frac{\pi}{2}$  og  $\frac{\pi}{2}$ . Vi får dermed dette integralet

$$\begin{aligned} I &= \iint_S f(x, y) \, dx \, dy = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left[ \int_0^{2 \cos \theta} f(r \sin \theta, r \cos \theta) r \, dr \right] d\theta = \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left[ \int_0^{2 \cos \theta} r^2 \, dr \right] d\theta \end{aligned}$$

der vi i siste skritt har brukt at  $\sqrt{x^2 + y^2} = r$ .

Det gjenstår bare å regne ut integralet. Vi får først

$$I = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left[ \int_0^{2 \cos \theta} r^2 \, dr \right] d\theta = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left[ \frac{r^3}{3} \right]_{r=0}^{r=2 \cos \theta} d\theta = \frac{8}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^3 \theta \, d\theta$$

Dette integralet løser vi ved å skrive  $\cos^3 \theta = (1 - \sin^2 \theta) \cos \theta$  og så substituere  $u = \sin \theta$ ,  $du = \cos \theta \, d\theta$ . Dette gir

$$\begin{aligned} I &= \frac{8}{3} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (1 - \sin^2 \theta) \cos \theta \, d\theta = \frac{8}{3} \int_{-1}^1 (1 - u^2) \, du = \\ &= \frac{8}{3} \left[ u - \frac{u^3}{3} \right]_{-1}^1 = \frac{8}{3} \left[ \left( 1 - \frac{1^3}{3} \right) - \left( (-1) - \frac{(-1)^3}{3} \right) \right] = \frac{32}{9} \end{aligned}$$

♣

Trigonometriske integraler av den typen vi har sett i de to siste eksemplene, er ganske vanlige når vi bruker polarkoordinater til å løse dobbeltintegraler. Synes

du slike integraler er vanskelige, finner du noen tips i seksjon 9.4 i *Kalkulus*. Se spesielt på de eksemplene som omhandler integraler på formene

$$\int \sin^n x \, dx, \quad \int \cos^n x \, dx \quad \text{og} \quad \int \sin^n x \cos^m x \, dx$$

### Oppgaver til seksjon 6.3

**1.** Løs integralet ved å bruke polarkoordinater:

- a)  $\iint_R xy^2 \, dxdy$  der  $R$  er området i første kvadrant som ligger innenfor sirkelen  $x^2 + y^2 = 9$
- b)  $\iint_R (x^2 + y^2) \, dxdy$  der  $R$  er området i første kvadrant som ligger innenfor sirkelen  $x^2 + y^2 = 25$  og mellom linjene  $y = 0$  og  $y = x$ .
- c)  $\iint_R e^{x^2+y^2} \, dxdy$  der  $R$  er området mellom sirklene om origo med radier lik 1 og 4.
- d)  $\iint_R xy \, dxdy$  der  $R$  er området i første kvadrant avgrenset av  $x$ -aksen, linjen  $y = x$  og sirkelen  $x^2 + y^2 = 1$
- e)  $\iint_R (x^2 - y^2) \, dxdy$  der  $R$  er området i tredje kvadrant som ligger mellom linjene  $y = \sqrt{3}x$ ,  $y = \frac{\sqrt{3}}{3}x$  og innenfor sirkelen  $x^2 + y^2 = 1$ .
- f)  $\iint_R \sqrt{2 - x^2 - y^2} \, dxdy$  der  $R$  er den delen av sirkelen  $x^2 + y^2 \leq 1$  som ligger i første kvadrant.
- g)  $\iint_R (x^2 + y^2)^{\frac{3}{2}} \, dxdy$  der  $R$  er sirkelen  $(x - 1)^2 + y^2 \leq 1$

**2.** Bruk MATLAB til å regne ut integralene i oppgave 1.

**3.** La  $R$  være sirkelskiven med sentrum i  $(0,1)$  og radius 1.

- a) Vis at dersom  $f$  er en kontinuerlig funksjon definert på hele  $R$ , så er

$$\iint_R f(x, y) \, dxdy = \int_0^\pi \left[ \int_0^{2\sin\theta} f(r \cos\theta, r \sin\theta) r \, dr \right] d\theta$$

- b) Regn ut  $\iint_R \sqrt{x^2 + y^2} \, dxdy$ .

**4.** Vi har en positiv kontinuerlig funksjon  $r : [\alpha, \beta] \rightarrow \mathbb{R}$ . Et område  $A$  består av de punktene i planet som har polarkoordinater  $(r, \theta)$  slik at  $\alpha \leq \theta \leq \beta$ ,  $0 \leq r \leq r(\theta)$ . Vis at arealet til  $A$  er

$$|A| = \frac{1}{2} \int_\alpha^\beta r(\theta)^2 \, d\theta$$

Bruk denne formelen til å finne arealet til området avgrenset av kurven

$$r(\theta) = \sin(2\theta) \quad \theta \in [0, \frac{\pi}{2}]$$

Lag en skisse av området.

## 6.4 Anvendelser av dobbeltintegraler

Vi har allerede påpekt at dobbeltintegraler kan brukes til å regne ut volumer — dersom  $f$  er en positiv funksjon på området  $A$ , er  $V = \iint_A f(x, y) dx dy$  volumet mellom funksjonsgrafen og området  $A$ . I denne seksjonen skal vi se på noen andre anvendelser av dobbeltintegralet. Det er ikke sikkert at det er de anvendelsene vi har tatt med nedenfor, du kommer til å ha mest bruk for, så det viktigste med dette avsnittet er ikke å pugge formlene utenat, men å lære seg å stille opp dobbeltintegraler i praktiske situasjoner. Avsnittet gir også flere eksempler på hvordan man regner ut dobbeltintegraler.

### Arealberegninger i planet

La oss begynne med arealer. Dersom funksjonen  $f$  er konstant lik 1, er volumet under funksjonen lik arealet til det området vi integrerer over (volum er lik grunnflate ganger høyde). Arealet til et område  $A$  er derfor gitt ved

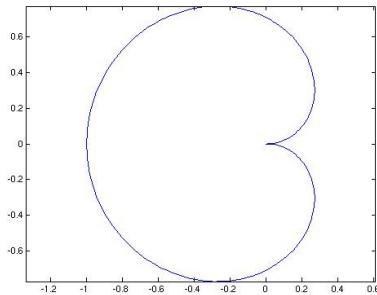
$$\text{areal}(A) = \iint_A 1 dx dy$$

(forutsatt at integralet eksisterer, men dette er spørsmål vi skal se nærmere på i seksjon 6.6). La oss se på et eksempel.

**Eksempel 1.** En kurve er gitt i polarkoordinater ved  $r = \sin \frac{\theta}{2}$ , der  $0 \leq \theta \leq 2\pi$ . Vi kan tegne kurven i MATLAB ved hjelp av kommandoene

```
>> t=linspace(0,2*pi,200);
>> x=sin(.5*t).*cos(t);
>> y=sin(.5*t).*sin(t);
>> plot(x,y)
>> axis'equal'
```

og resultatet blir som i figur 1. Vår oppgave er å regne ut arealet til området  $S$  avgrenset av kurven.



Figur 1: Kurven i eksempel 1.

Vi vet at arealet er gitt ved

$$\text{areal}(S) = \iint_S 1 \, dx dy = \iint_R r \, dr d\theta$$

der  $R$  er området beskrevet i polarkoordinater. Dette gir

$$\text{areal}(S) = \iint_R r \, dr d\theta = \int_0^{2\pi} \left[ \int_0^{\sin \frac{\theta}{2}} r \, dr \right] d\theta = \frac{1}{2} \int_0^{2\pi} \sin^2 \frac{\theta}{2} \, d\theta$$

Formelen for cosinus til det dobbelte av en vinkel sier at  $\cos 2u = 1 - 2 \sin^2 u$ , og bruker vi denne med  $v = \frac{\theta}{2}$ , ser vi at  $\sin^2 \frac{\theta}{2} = \frac{1 - \cos \theta}{2}$ . Dermed er

$$\text{areal}(S) = \frac{1}{4} \int_0^{2\pi} (1 - \cos \theta) \, d\theta = \frac{1}{4} \left[ \theta - \sin \theta \right]_0^{2\pi} = \frac{\pi}{2} \quad \clubsuit$$

### Massemiddelpunkt

Den neste anvendelsen vi skal se på, er massemiddelpunkt. La oss først tenke oss at vi har  $n$  partikler med masse  $m_1, m_2, \dots, m_n$  plassert i punktene  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$  i planet. *Massemiddelpunktet* (eller *massesenteret*) til disse partiklene er da definert til å være punktet

$$\bar{\mathbf{r}} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2 + \cdots + m_n \mathbf{r}_n}{m_1 + m_2 + \cdots + m_n} = \frac{\sum_{i=1}^n m_i \mathbf{r}_i}{\sum_{i=1}^n m_i}$$

Skriver vi ut definisjonen komponentvis, ser vi at koordinatene til  $\mathbf{r}$  er gitt ved

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2 + \cdots + m_n x_n}{m_1 + m_2 + \cdots + m_n} = \frac{\sum_{i=1}^n m_i x_i}{\sum_{i=1}^n m_i}$$

og

$$\bar{y} = \frac{m_1 y_1 + m_2 y_2 + \cdots + m_n y_n}{m_1 + m_2 + \cdots + m_n} = \frac{\sum_{i=1}^n m_i y_i}{\sum_{i=1}^n m_i}$$

der  $\mathbf{r}_i = (x_i, y_i)$ . Massemiddelpunktet er viktig i fysikk og mekanikk. Det er et “balansepunkt” for systemet i følgende forstand: Dersom vi tenker oss at partiklene våre hviler på en tynn glassplate som selv ikke veier noe, vil denne glassplaten balansere perfekt på en blyantspiss plassert i massemiddelpunktet.

Vi ønsker nå å generalisere begrepet massemiddelpunkt til sammenhengende legemer. Anta at vi har en tynn plate der massetettheten kan variere. La oss si at massetettheten i punktet  $\mathbf{r} = (x, y)$  er  $f(\mathbf{r}) = f(x, y)$ . Dette betyr at dersom vi kutter ut en liten bit av platen med areal  $A$  rundt punktet  $\mathbf{r}$ , vil massen til denne biten være tilnærmet lik  $f(\mathbf{r})|A|$ , der  $A$  er arealet til  $A$ .

Anta nå at vi kutter opp  $A$  i et kvadratisk rutenett på den vanlige måten, og at  $A_{ij}$  er den  $ij$ -te ruten. Plukker vi ut et punkt  $\mathbf{r}_{ij}^* = (x_{ij}^*, y_{ij}^*)$  i  $A_{ij}$ , vil systemet vårt mekanisk oppføre seg omtrent som en partikkelsamling der punkter med masse  $f(x_{ij}^*, y_{ij}^*)|A_{ij}|$  befinner seg i posisjon  $(x_{ij}^*, y_{ij}^*)$ . Dette tilnærmede systemet har massemiddelpunkt med koordinater

$$\bar{x} = \frac{\sum_{i=1}^n \sum_{j=1}^m x_{ij}^* f(x_{ij}^*, y_{ij}^*) |A_{ij}|}{\sum_{i=1}^n \sum_{j=1}^m f(x_{ij}^*, y_{ij}^*) |A_{ij}|}$$

$$\bar{y} = \frac{\sum_{i=1}^n \sum_{j=1}^m y_{ij}^* f(x_{ij}^*, y_{ij}^*) |A_{ij}|}{\sum_{i=1}^n \sum_{j=1}^m f(x_{ij}^*, y_{ij}^*) |A_{ij}|}$$

Legg merke til at summene i disse uttrykkene er Riemann-summer for integralene  $\iint_A xf(x, y) dx dy$ ,  $\iint_A yf(x, y) dx dy$  og  $\iint_A f(x, y) dx dy$ . Lar vi oppdelingen bli finere og finere, konvergerer dermed uttrykkene ovenfor mot

$$\bar{x} = \frac{\iint_A xf(x, y) dx dy}{\iint_A f(x, y) dx dy}$$

og

$$\bar{y} = \frac{\iint_A yf(x, y) dx dy}{\iint_A f(x, y) dx dy}$$

Det er naturlig å definere  $\bar{\mathbf{r}} = (\bar{x}, \bar{y})$  til å være masse middelpunktet til platen vår.

**Eksempel 2:** Finn masse middelpunktet til kvadratet  $A = [0, 1] \times [0, 1]$  når tettheten er gitt ved  $f(x, y) = xy^2$ .

La oss først regne ut nevneren i uttrykkene ovenfor, nemlig:

$$\begin{aligned} \iint_A f(x, y) dx dy &= \int_0^1 \left[ \int_0^1 xy^2 dx \right] dy = \int_0^1 \left[ \frac{1}{2}x^2 y^2 \right]_{x=0}^{x=1} dy = \\ &= \int_0^1 \frac{1}{2}y^2 dy = \left[ \frac{1}{6}y^3 \right]_0^1 = \frac{1}{6} \end{aligned}$$

Neste trinn er å regne ut telleren i uttrykket for  $\bar{x}$ :

$$\begin{aligned}\iint_A xf(x, y) \, dxdy &= \int_0^1 \left[ \int_0^1 x^2 y^2 \, dx \right] dy = \int_0^1 \left[ \frac{1}{3} x^3 y^2 \right]_{x=0}^{x=1} dy = \\ &= \int_0^1 \frac{1}{3} y^2 \, dy = \left[ \frac{1}{9} y^3 \right]_0^1 = \frac{1}{9}\end{aligned}$$

Dette gir

$$\bar{x} = \frac{\iint_A xf(x, y) \, dxdy}{\iint_A f(x, y) \, dxdy} = \frac{\frac{1}{9}}{\frac{1}{6}} = \frac{2}{3}$$

Så regner vi ut telleren i uttrykket for  $\bar{y}$ :

$$\begin{aligned}\iint_A yf(x, y) \, dxdy &= \int_0^1 \left[ \int_0^1 xy^3 \, dx \right] dy = \int_0^1 \left[ \frac{1}{2} x^2 y^3 \right]_{x=0}^{x=1} dy = \\ &= \int_0^1 \frac{1}{2} y^3 \, dy = \left[ \frac{1}{8} y^4 \right]_0^1 = \frac{1}{8}\end{aligned}$$

Dermed er

$$\bar{y} = \frac{\iint_A yf(x, y) \, dxdy}{\iint_A f(x, y) \, dxdy} = \frac{\frac{1}{8}}{\frac{1}{6}} = \frac{3}{4}$$

Massemiddelpunktet er altså  $\bar{\mathbf{r}} = (\bar{x}, \bar{y}) = \left(\frac{2}{3}, \frac{3}{4}\right)$ . ♣

Et viktig spesialtilfelle av teorien ovenfor får vi når tetthetsfunksjonen  $f$  er konstant lik 1. Da er

$$\bar{x} = \frac{\iint_A x \, dxdy}{\iint_A 1 \, dxdy} = \frac{\iint_A x \, dxdy}{\text{Areal}(A)}$$

og

$$\bar{y} = \frac{\iint_A y \, dxdy}{\iint_A 1 \, dxdy} = \frac{\iint_A y \, dxdy}{\text{Areal}(A)}$$

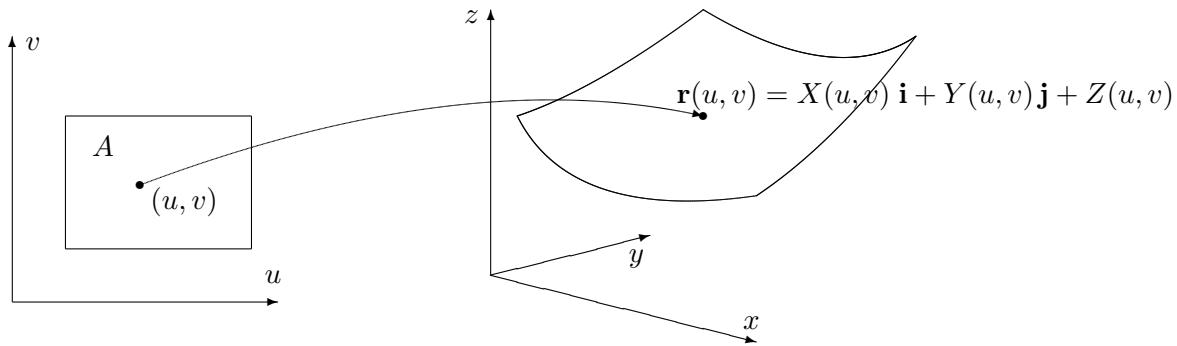
Punktet  $\bar{\mathbf{r}} = (\bar{x}, \bar{y})$  kalles da *sentroiden* til området  $A$ .

### Arealet til flater

I seksjon 3.9 studerte vi parametriserte flater, og vi begynner med en rask repetisjon. En parametrisert flate består av et område  $A$  i  $\mathbb{R}^2$  og en kontinuerlig funksjon  $\mathbf{r} : A \rightarrow \mathbb{R}^3$  som vi ofte skriver på denne måten

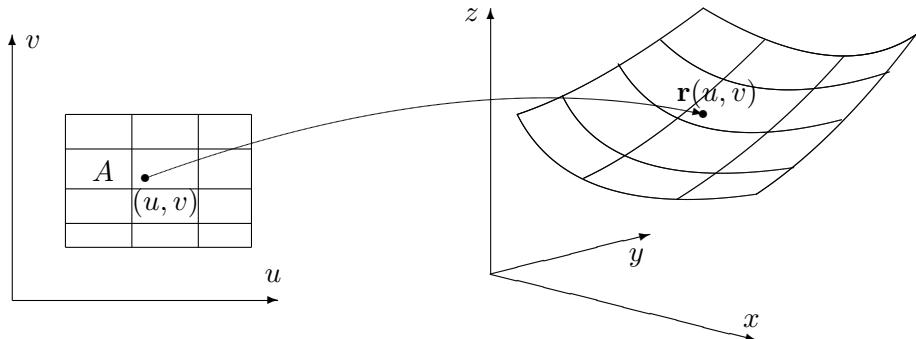
$$\mathbf{r}(u, v) = X(u, v)\mathbf{i} + Y(u, v)\mathbf{j} + Z(u, v)\mathbf{k}, \quad \text{der } (u, v) \in A$$

Figur 2 viser hvordan en slik funksjon “løfter” området  $A$  opp på en flate. Vi kaller  $\mathbf{r}$  en *parametrisering av flaten. Vi skal nå se hvordan vi kan bruke dobbeltintegraler over  $A$  til å beregne arealet til denne flaten.*



Figur 2: Parametrisert flate

Som vanlig tar vi utgangspunkt i en oppdeling av  $A$  i et rutenett. Funksjonen  $\mathbf{r}$  flytter dette rutenettet opp på flaten (se figur 3).



Figur 3: Oppdeling av flate

Vi skal først prøve å beregne et tilnærmet areal for hver av de små rutene opp på flaten. Vi starter med den  $ij$ -te ruten  $A_{ij}$  i oppdelingen av  $A$ . Den har sitt nedre venstre hjørne i punktet  $(u_{i-1}, v_{j-1})$ , sitt nedre høyre hjørne i punktet  $(u_i, v_{j-1})$  og sitt øvre venstre hjørne i punktet  $(u_{i-1}, v_j)$ . Disse punktene avbildes opp på punktene  $\mathbf{r}(u_{i-1}, v_{j-1})$ ,  $\mathbf{r}(u_i, v_{j-1})$ ,  $\mathbf{r}(u_{i-1}, v_j)$  på flaten. Dersom oppdelingen er fin nok, vil parallellogrammet utspent av vektorene  $\mathbf{r}(u_i, v_{j-1}) - \mathbf{r}(u_{i-1}, v_{j-1})$  og  $\mathbf{r}(u_{i-1}, v_j) - \mathbf{r}(u_{i-1}, v_{j-1})$  omtrent falle sammen med bildet  $\mathbf{r}(A_{ij})$  av  $A_{ij}$  oppå flaten (parallellogrammet vil være en slags flat ”takstein” som dekker det lett krummede flateelementet  $\mathbf{r}(A_{ij})$ ).

Arealet til flateelementet  $\mathbf{r}(A_{ij})$  er derfor tilnærmet lik arealet til paral-

lellogrammet, som er gitt ved lengden til kryssproduktet

$$\left| (\mathbf{r}(u_i, v_{j-1}) - \mathbf{r}(u_{i-1}, v_{j-1})) \times (\mathbf{r}(u_{i-1}, v_j) - \mathbf{r}(u_{i-1}, v_{j-1})) \right|$$

Bruker vi at

$$\mathbf{r}(u_i, v_{j-1}) - \mathbf{r}(u_{i-1}, v_{j-1}) \approx \frac{\partial \mathbf{r}}{\partial u}(u_{i-1}, v_{j-1})(u_i - u_{i-1})$$

og

$$\mathbf{r}(u_{i-1}, v_j) - \mathbf{r}(u_{i-1}, v_{j-1}) \approx \frac{\partial \mathbf{r}}{\partial v}(u_{i-1}, v_{j-1})(v_j - v_{j-1}),$$

får vi derfor at arealet av flateelementet  $\mathbf{r}(A_{ij})$  er tilnærmet lik

$$\left| \frac{\partial \mathbf{r}}{\partial u}(u_{i-1}, v_{j-1}) \times \frac{\partial \mathbf{r}}{\partial v}(u_{i-1}, v_{j-1}) \right| (u_i - u_{i-1})(v_j - v_{j-1})$$

med bedre og bedre tilnærming dess finere oppdelingen er. Summerer vi, ser vi at arealet til hele flaten er tilnærmet gitt ved

$$\sum_{i=1}^n \sum_{j=1}^m \left| \frac{\partial \mathbf{r}}{\partial u}(u_{i-1}, v_{j-1}) \times \frac{\partial \mathbf{r}}{\partial v}(u_{i-1}, v_{j-1}) \right| (u_i - u_{i-1})(v_j - v_{j-1})$$

Dette er en Riemann-sum for funksjonen  $|\frac{\partial \mathbf{r}}{\partial u}(u, v) \times \frac{\partial \mathbf{r}}{\partial v}(u, v)|$ , og lar vi oppdelingen bli finere og finere, nærmer Riemann-summene seg integralet

$$\iint_A \left| \frac{\partial \mathbf{r}}{\partial u}(u, v) \times \frac{\partial \mathbf{r}}{\partial v}(u, v) \right| dudv$$

Vi har dermed kommet frem til denne formelen for arealet til en parametrisert flate:

$$\text{Flateareal} = \iint_A \left| \left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right)(u, v) \right| dudv$$

der vi har skrevet  $(\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v})(u, v)$  istedenfor  $\frac{\partial \mathbf{r}}{\partial u}(u, v) \times \frac{\partial \mathbf{r}}{\partial v}(u, v)$  for å spare litt plass.

**Bemerkning:** Du må være litt forsiktig når du bruker denne formelen og passe på at flaten ikke overlapper med seg selv. Er det overlapp, vil formelen regne arealet til den overlappende biten to ganger.

**Eksempel 3:** I seksjon 3.9 så vi at en torus (smultring) har parametriseringen

$$\mathbf{r}(u, v) = (R + r \cos u) \cos v \mathbf{i} + (R + r \cos u) \sin v \mathbf{j} + r \sin u \mathbf{k},$$

der  $u, v \in [0, 2\pi]$ . La oss regne ut overflatearealet til denne torusen. Vi ser først at

$$\frac{\partial \mathbf{r}}{\partial u}(u, v) = -r \sin u \cos v \mathbf{i} - r \sin u \sin v \mathbf{j} + r \cos u \mathbf{k}$$

og

$$\frac{\partial \mathbf{r}}{\partial v}(u, v) = -(R + r \cos u) \sin v \mathbf{i} + (R + r \cos u) \cos v \mathbf{j}$$

Dermed blir

$$\left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right)(u, v) = -r(R + r \cos u) (\cos u \cos v \mathbf{i} + \cos u \sin v \mathbf{j} + \sin u \mathbf{k})$$

og

$$\left| \left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right)(u, v) \right| = r(R + r \cos u)$$

der vi har brukt formelen  $\sin^2 x + \cos^2 x = 1$  en rekke ganger både for  $x = u$  og  $x = v$ . Nå er integrasjonen grei:

$$\text{Arealet til torusen} = \iint_A \left| \left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right)(u, v) \right| dudv =$$

$$= \int_0^{2\pi} \left[ \int_0^{2\pi} r(R + r \cos u) dv \right] du = 2\pi \int_0^{2\pi} r(R + r \cos u) du = 4\pi^2 Rr \quad \clubsuit$$

Et viktig spesialtilfelle av teorien ovenfor er grafer  $z = f(x, y)$  til funksjoner av to variable. Som vi så i seksjon 3.9, kan de oppfattes som parametriserte flater på formen

$$\mathbf{r}(x, y) = x \mathbf{i} + y \mathbf{j} + f(x, y) \mathbf{k} \quad x, y \in A$$

Vi ser at

$$\frac{\partial \mathbf{r}}{\partial x}(x, y) = \mathbf{i} + \frac{\partial f}{\partial x}(x, y) \mathbf{k}$$

og

$$\frac{\partial \mathbf{r}}{\partial y}(x, y) = \mathbf{j} + \frac{\partial f}{\partial y}(x, y) \mathbf{k}$$

som gir

$$\left( \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} \right)(x, y) = -\frac{\partial f}{\partial x}(x, y) \mathbf{i} - \frac{\partial f}{\partial y}(x, y) \mathbf{j} + \mathbf{k}$$

og

$$\left| \left( \frac{\partial \mathbf{r}}{\partial x} \times \frac{\partial \mathbf{r}}{\partial y} \right)(x, y) \right| = \sqrt{1 + \left( \frac{\partial f}{\partial x}(x, y) \right)^2 + \left( \frac{\partial f}{\partial y}(x, y) \right)^2}$$

Arealet til funksjonsgrafen  $z = f(x, y)$  over området  $A$  er dermed

$$\text{Areal} = \iint_A \sqrt{1 + \left( \frac{\partial f}{\partial x}(x, y) \right)^2 + \left( \frac{\partial f}{\partial y}(x, y) \right)^2} dx dy$$

**Eksempel 4:** Vi skal regne ut arealet til den delen av paraboloiden  $f(x, y) = x^2 + y^2$  som ligger over sirkelskiven  $A$  gitt ved  $x^2 + y^2 \leq 1$ . Siden

$$\frac{\partial f}{\partial x} = 2x \quad \text{og} \quad \frac{\partial f}{\partial y} = 2y,$$

får vi

$$\text{Areal} = \iint_A \sqrt{1 + 4x^2 + 4y^2} \, dx \, dy$$

Går vi over til polarkoordinater, ser vi at

$$\text{Areal} = \int_0^1 \left[ \int_0^{2\pi} r \sqrt{1 + 4r^2} \, d\theta \right] dr = 2\pi \int_0^1 r \sqrt{1 + 4r^2} \, dr$$

Vi innfører nå en ny variabel  $u = 1 + 4r^2$ , som gir  $du = 8r \, dr$  og

$$\text{Areal} = \frac{\pi}{4} \int_1^5 \sqrt{u} \, du = \frac{\pi}{4} \left[ \frac{2}{3} u^{3/2} \right]_1^5 = \frac{\pi}{6} (5\sqrt{5} - 1)$$



### Flateintegraler av skalarfelt

Vi arbeider fortsatt med en parametrisert flate  $\mathbf{r}(u, v)$ ,  $(u, v) \in A$ , som ovenfor, men vi tenker oss at flaten er et fysisk objekt med varierende massetettet — la oss si at massetettheten i punktet  $(x, y, z)$  på flaten er  $f(x, y, z)$ . Gjentar vi argumentene ovenfor, kommer vi raskt frem til at massen til hele flaten må være gitt av

$$\iint_A f(\mathbf{r}(u, v)) \left| \left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right) (u, v) \right| \, du \, dv$$

Uttrykk av denne typen dukker opp så ofte at det lønner seg å ha et eget navn og en egen notasjon for dem:

**Definisjon 6.4.1** Anta at  $T$  er flaten gitt ved parametriseringen  $\mathbf{r}(u, v)$ ,  $(u, v) \in A$ . Anta at  $f$  er en kontinuerlig funksjon på  $T$ . Da er flateintegralet til  $f$  over  $T$  definert ved

$$\iint_T f \, dS = \iint_A f(\mathbf{r}(u, v)) \left| \left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right) (u, v) \right| \, du \, dv$$

(forutsatt at dobbeltintegralet er definert).

Det er en del å holde styr på når man skal regne ut et flateintegral, så vi tar med et eksempel som viser håndgrepene.

**Eksempel 5:** Vi skal regne ut flateintegralet  $\iint_T f \, dS$  der  $f(x, y, z) = xyz^2$  og der  $T$  er flaten gitt ved

$$\mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + u^2 \mathbf{k} \quad \text{for } u \in [0, 2], v \in [0, 2\pi]$$

(dette er en del av paraboloiden  $z = x^2 + y^2$ ). Først regner vi ut

$$\frac{\partial \mathbf{r}}{\partial u} = \cos v \mathbf{i} + \sin v \mathbf{j} + 2u \mathbf{k}$$

og

$$\frac{\partial \mathbf{r}}{\partial v} = -u \sin v \mathbf{i} + u \cos v \mathbf{j}$$

Tar vi vektorproduktet, får vi

$$\frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} = -2u^2 \cos v \mathbf{i} - 2u^2 \sin v \mathbf{j} + u \mathbf{k}$$

og

$$\left| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right| = u \sqrt{4u^2 + 1}$$

Vi ser også at

$$f(\mathbf{r}(u, v)) = f(u \cos v, u \sin v, u^2) = (u \cos v)(u \sin v)(u^2)^2 = u^6 \sin v \cos v$$

Nå kan vi sette inn i integralet:

$$\begin{aligned} \iint_T f \, dS &= \iint_A f(\mathbf{r}(u, v)) \left| \left( \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right) (u, v) \right| \, dudv = \\ &= \iint_{[0,2] \times [0,2\pi]} \left( u^6 \sin v \cos v \right) \left( u \sqrt{4u^2 + 1} \right) \, dudv = \\ &= \int_0^2 \left[ \int_0^{2\pi} u^7 \sin v \cos v \sqrt{4u^2 + 1} \, dv \right] du = \\ &= \int_0^2 u^7 \sqrt{4u^2 + 1} \left[ \int_0^{2\pi} \sin v \cos v \, dv \right] du \end{aligned}$$

(vi kan ta  $u$ -faktorene utenfor siden de regnes som konstanter når vi integrerer med hensyn på  $v$ ). Siden  $\int \sin v \cos v \, dv = \frac{1}{2} \sin^2 v + C$  (sett  $z = \sin v$  hvis du ikke ser dette direkte), får vi

$$\iint_T f \, dS = \int_0^2 u^7 \sqrt{4u^2 + 1} \left[ \frac{1}{2} \sin^2 v \right]_0^{2\pi} \, du = 0$$



Vi kan nå utvide våre formler for massemiddelpunkt til massefordelinger på flater. Dersom  $f$  er massetettheten på flaten  $T$ , er  $x$ -,  $y$ - og  $z$ -komponentene til massemiddelpunktet  $\bar{\mathbf{r}} = (\bar{x}, \bar{y}, \bar{z})$  gitt ved

$$\begin{aligned}\bar{x} &= \frac{\iint_T xf(x, y, z) dS}{\iint_T f(x, y, z) dS} \\ \bar{y} &= \frac{\iint_T yf(x, y, z) dS}{\iint_T f(x, y, z) dS} \\ \bar{z} &= \frac{\iint_T zf(x, y, z) dS}{\iint_T f(x, y, z) dS}\end{aligned}$$

Vær oppmerksom på at massemiddelpunktet som regel ligger utenfor flaten  $T$ .

**Bemerkning:** I kapittel 3 definerte vi kurveintegraler både av skalarfelt (integraler av typen  $\int_C f ds$ ) og av vektorfelt (integraler av typen  $\int_C \mathbf{F} \cdot d\mathbf{r}$ ). Flateintegralene vi definerte ovenfor, korresponderer til den første typen linjeintegraler. Det finnes også flateintegraler som korresponderer til den andre typen linjeintegraler. De betegnes gjerne med  $\iint_T \mathbf{F} \cdot \mathbf{n} dS$  og brukes mye i fysikk og mekanikk. Vi skal ikke komme nærmere inn på dem her.

### Oppgaver til seksjon 6.4

1. Beregn volumet til området  $E$  når

- a)  $E = \{(x, y, z) : 0 \leq x \leq 2, 0 \leq y \leq 1, 0 \leq z \leq x + y^2\}$
- b)  $E = \{(x, y, z) : 0 \leq x \leq 1, 0 \leq y \leq x, 0 \leq z \leq x^2\}$
- c)  $E = \{(x, y, z) : -1 \leq x \leq 1, 0 \leq y \leq 2, -xy \leq z \leq 3 - xy\}$
- d)  $E$  er området over  $xy$ -planet og under grafen  $z = \sqrt{32 - 2x^2 - 2y^2}$
- e)  $E$  er området som ligger under grafen  $z = x^2 - y^2$  og over sirkelskiven  $x^2 + y^2 \leq 1$
- f)  $E$  er området som ligger over  $xy$ -planet og under grafen  $z = 4 - (x - 2)^2 - (y + 1)^2$

2. En trekantet plate har hjørner i  $(0, 0)$ ,  $(1, 0)$  og  $(1, 1)$  og tetthet  $f(x, y) = x$ . Finn massemiddelpunktet.

3. En plate dekker området  $R = \{(x, y) \mid 0 \leq x \leq 1, x^2 \leq y \leq 1\}$  og har tetthet  $f(x, y) = xy$ . Finn massemiddelpunktet.

4. Regn ut overflatearealet til en kule med radius  $R$ .

- 5.** Finn arealet til flaten  $z = x^2 - y^2$ ,  $x^2 + y^2 \leq 4$ .
- 6.** Finn arealet av den delen av kjegleflaten  $z^2 = x^2 + y^2$  som ligger mellom  $z = 0$  og  $z = 1$ .
- 7.** Finn arealet av den delen av kuleflaten  $x^2 + y^2 + z^2 = 1$  som ligger over sirkelen  $(x - \frac{1}{2})^2 + y^2 \leq \frac{1}{4}$ .

**8.** Vis at arealet til en flate

$$\mathbf{r}(r, \theta) = r \cos \theta \mathbf{i} + r \sin \theta \mathbf{j} + f(r, \theta) \mathbf{k} \quad \text{der } (r, \theta) \in A$$

er (under passende betingelser) gitt ved

$$\iint_A \sqrt{1 + \left(\frac{\partial f}{\partial r}\right)^2 + \frac{1}{r^2} \left(\frac{\partial f}{\partial \theta}\right)^2} r \, dr \, d\theta$$

**9.** Regn ut flatearealet  $\iint_T x^2 \, dS$  når  $T$  er flaten gitt ved  $z = x^2 + y^2$  der  $x^2 + y^2 \leq 1$ .

**10.** Regn ut flateintegralet  $\iint_T xyz^2 \, dS$  der  $T$  er den delen av cylinderflaten  $x^2 + y^2 = 4$  der  $x \geq 0$ ,  $y \geq 0$  og  $0 \leq z \leq 1$ .

**11.** Regn ut flateintegralet  $\iint_T z^2 \, dS$  når  $T$  er torusen

$$\mathbf{r}(u, v) = (5 + 3 \cos u) \cos v \mathbf{i} + (5 + 3 \cos u) \sin v \mathbf{j} + 3 \sin u \mathbf{k} \quad u, v \in [0, 2\pi]$$

Du kan få bruk for noen av regningene i eksempel 3.

**12.** En cylinderflate  $T$  har parametriseringen

$$\mathbf{r}(u, v) = u \mathbf{i} + 5 \cos v \mathbf{j} + 5 \sin v \mathbf{k}, \quad u \in [0, 2], v \in [0, 2\pi]$$

Tegn en skisse av flaten og regn ut flateintegralet  $\iint_T x \, dS$ .

**13.** Forklar at en kule om origo med radius  $R$  har parametriseringen

$$\mathbf{r}(\theta, \phi) = R \cos \theta \sin \phi \mathbf{i} + R \sin \theta \sin \phi \mathbf{j} + R \cos \phi \mathbf{k}, \quad \theta \in [0, 2\pi], \phi \in [0, \pi]$$

Regn ut  $\left| \frac{\partial \mathbf{r}}{\partial \theta} \times \frac{\partial \mathbf{r}}{\partial \phi} \right|$ . Regn også ut flateintegralet  $\iint_T xy \, dS$  når  $T$  er den delen av kuleflaten som ligger i første oktant (dvs. området der  $x, y, z \geq 0$ ).

**14.** En avkortet kjegle  $T$  har parametriseringen

$$\mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + u \mathbf{k} \quad u \in [1, 2], v \in [0, 2\pi]$$

Regn ut  $\left| \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v} \right|$ . Regn også ut flateintegralet  $\iint_T x^2 z \, dS$ .

**15.** Finn massemiddelpunktet til halvkulen  $x^2 + y^2 + z^2 = 1$ ,  $z \geq 0$ , når massetetheten er 1.

**16.** (Prøveeksamen i MAT1110, 2004)

- (a) La  $T$  være området som ligger inni både sylinderen  $x^2 - 2x + y^2 = 0$  ogkulen  $x^2 + y^2 + z^2 = 4$ . Finn volumet til  $T$ .
- (b) Finn arealet av den delen av overflaten til  $T$  som ligger på kuleflaten  $x^2 + y^2 + z^2 = 4$ .

**17.** (Eksamens i MAT1110 13/8, 2004) La  $D$  være det begrensede området i  $\mathbb{R}^3$  som ligger over  $xy$ -planet og inni både paraboloiden  $z = 4 - x^2 - y^2$  og sylinderen  $x^2 + y^2 = 1$ .

- a) Finn volumet til  $D$ .
- b) Finn arealet av den delen av randflaten til  $D$  som ligger på paraboloiden  $z = 4 - x^2 - y^2$ .

**18.** (Eksamens i MAT1110 13/6, 2007)

- a) Vis at volumet til området avgrenset av planet  $2x + 4y - z = -4$  og paraboloiden  $z = x^2 + y^2$  er gitt ved

$$V = \iint_D (2x + 4y - x^2 - y^2 + 4) \, dx \, dy$$

der  $D$  er sirkelen med sentrum i  $(1, 2)$  og radius 3.

- b) Regn ut  $V$ .

**19.** (Eksamens i MAT1110 14/6, 2006)  $R$  er området i  $\mathbb{R}^3$  avgrenset av paraboloiden  $z = x^2 + y^2$  og planet  $z = 2x + 6y - 6$ .

- a) Forklar at volumet til  $R$  er

$$V = \iint_A (2x + 6y - 6 - x^2 - y^2) \, dx \, dy$$

der  $A = \{(x, y) \in \mathbb{R}^2 \mid (x - 1)^2 + (y - 3)^2 \leq 4\}$ .

- b) Regn ut  $V$ .

- c) Vis at vektorfeltet

$$\mathbf{F}(x, y, z) = y^2 z \mathbf{i} + 2xyz \mathbf{j} + xy^2 \mathbf{k}$$

er konserватivt. Regn ut  $\oint_C \mathbf{G} \cdot d\mathbf{r}$  der

$$\mathbf{G}(x, y, z) = (y^2 z + z) \mathbf{i} + 2xyz \mathbf{j} + xy^2 \mathbf{k}$$

og der  $C$  er skjæringskurven til flatene  $z = x^2 + y^2$  og  $z = 2x + 6y - 6$ . Kurven er orientert mot klokken når du ser den ovenfra.

## 6.5 Greens teorem

I denne seksjonen skal vi se på en interessant sammenheng mellom linjeintegraler og dobbeltintegraler. Husk fra kapittel 3 at dersom

$$\mathbf{F}(x, y) = P(x, y) \mathbf{i} + Q(x, y) \mathbf{j}$$

er et vektorfelt og  $\mathcal{C}$  er en kurve i planet parametrisert ved

$$\mathbf{r}(t) = x(t) \mathbf{i} + y(t) \mathbf{j} \quad \text{der } t \in [a, b],$$

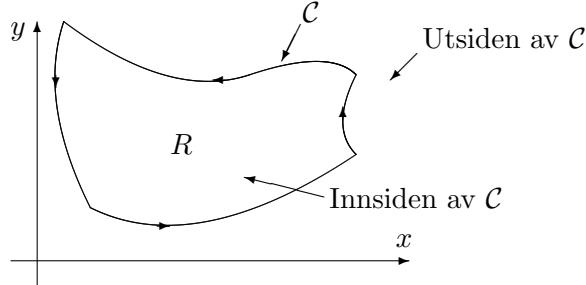
så er linjeintegralet  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  definert ved

$$\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_a^b \mathbf{F}(\mathbf{r}(t)) \cdot \mathbf{r}'(t) dt = \int_a^b \left( P(x(t), y(t))x'(t) + Q(x(t), y(t))y'(t) \right) dt$$

Inspirert av det siste uttrykket skriver man ofte

$$\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} = \int_{\mathcal{C}} P dx + Q dy$$

der  $dx$  er en forkortelse for  $x'(t) dt$  og  $dy$  er en forkortelse for  $y'(t) dt$ .



Figur 1: En enkel, lukket kurve  $\mathcal{C}$  og området  $R$  den avgrenser

For å formulere Greens teorem trenger vi noen begreper. Husk at en parametrisert kurve  $\mathbf{r} : [a, b] \rightarrow \mathbb{R}^2$  er *lukket* dersom den starter og ender i samme punkt, dvs. dersom  $\mathbf{r}(a) = \mathbf{r}(b)$ . Vi sier at kurven er *enkel* dersom den ikke skjærer eller berører seg selv underveis, dvs. dersom  $\mathbf{r}(s) \neq \mathbf{r}(t)$  for alle  $s, t \in [a, b]$ ,  $s \neq t$ . En enkel, lukket kurve deler planet i to deler, en *utsiden* og en *innsiden* (se figur 1). Dette kan virke opplagt, men er forbausende vanskelig å bevise. Med *området avgrenset av kurven* skal vi mene samlingen av alle punkter som ligger enten på eller på innsiden av kurven. Vi kan nå formulere teoremet:

**Teorem 6.5.1 (Greens teorem)** *Anta at  $\mathcal{C}$  er en enkel, lukket kurve med en stykkevis glatt parametrisering  $\mathbf{r}$ , og la  $R$  være området avgrenset av  $\mathcal{C}$ .*

Dersom de partiellderiverte til  $P$  og  $Q$  er kontinuerlige i et åpent område som inneholder  $R$ , så er

$$\int_{\mathcal{C}} P \, dx + Q \, dy = \iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy$$

der  $\mathcal{C}$  er orientert mot klokken

**Bemerkning:** Greens teorem er en to-dimensjonal versjon av analysens fundamentalteorem. Fundamentalteoremet sier at oppførselen til én funksjon  $f$  over det indre av et intervall  $[a, b]$ , oppfanges av oppførselen til en annen funksjon  $F$  på randen til intervallet — vi har

$$F(b) - F(a) = \int_a^b f(x) \, dx$$

Tilsvarende sier Greens teorem at oppførselen til funksjonen  $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}$  over det indre av området  $R$ , oppfanges av oppførselen til en annen funksjon  $\mathbf{F} = P \mathbf{i} + Q \mathbf{j}$  over randen  $\mathcal{C}$  til området.

Så generelt som det er formulert ovenfor, er Greens teorem vanskelig å bevise, og vi skal derfor nøye oss med å bevise noen spesialtilfeller som er tilstrekkelig for de fleste regnetekniske formål. Men før vi gjør det, skal vi se på noen eksempler.

Greens teorem kan brukes begge veier — både til å omdanne linjeintegraler til dobbeltintegraler og omvendt. Det enkleste er å omdanne linjeintegraler til dobbeltintegraler.

**Eksempel 1:** Vi skal regne ut linjeintegralet  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  der  $\mathcal{C}$  er omkretsen til kvadratet  $R = [0, 1] \times [2, 4]$  orientert mot klokken, og  $\mathbf{F}$  er gitt ved

$$\mathbf{F}(x, y) = \left( x + \frac{y^3}{3} \right) \mathbf{i} + (2x + y^2) \mathbf{j}$$

Vi ser at siden  $P(x, y) = x + \frac{y^3}{3}$  og  $Q(x, y) = 2x + y^2$ , så er

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 2 - y^2$$

Dermed har vi

$$\begin{aligned} \int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r} &= \int_{\mathcal{C}} P \, dx + Q \, dy = \iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \\ &= \int_2^4 \left[ \int_0^1 (2 - y^2) \, dx \right] dy = \int_2^4 (2 - y^2) \left[ \int_0^1 1 \, dx \right] dy = \int_2^4 (2 - y^2) \, dy = \end{aligned}$$

$$= \left[ 2y - \frac{y^3}{3} \right]_2^4 = -\frac{44}{3}$$

♣

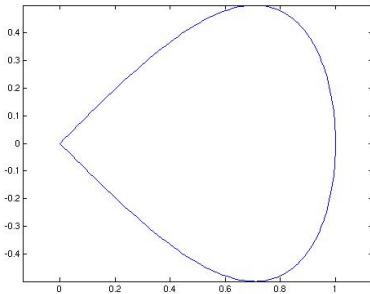
Vi ser at vi har to fordeler av å bruke Greens teorem i eksemplet over — for det første er det enklere å dobbeltintegrasjonen over et rektangel enn å parametrisere alle de fire kantene, og for det andre blir integranden enklere når den deriveres.

La oss nå se på et eksempel der vi bruker Greens teorem i motsatt retning — fra dobbeltintegral til linjeintegral.

**Eksempel 2:** Vi skal regne ut dobbeltintegralet  $\iint_R xy \, dx dy$  der  $R$  er området omsluttet av kurven  $\mathcal{C}$  gitt ved

$$\mathbf{r}(t) = \sin t \mathbf{i} + \sin t \cos t \mathbf{j} \quad t \in [0, \pi]$$

Bruker vi MATLAB til å tegne opp kurven, får vi denne figuren



Figur 2: Området i eksempel 2

Siden det er kurven som er gitt, kan det være lurt å bruke Greens teorem til å regne ut dobbeltintegralet. Vi trenger da et vektorfelt

$$\mathbf{F}(x, y) = P(x, y) \mathbf{i} + Q(x, y) \mathbf{j}$$

der  $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = xy$ . Det er mange valgmuligheter, men siden  $y$ -komponenten er den “styggeste” delen av parametriseringen, kan det lønne seg å velge et vektorfelt der  $Q = 0$ . Dermed må vi ha  $-\frac{\partial P}{\partial y} = xy$ , og  $P = -\frac{1}{2}xy^2$  er et naturlig valg. Vi har dermed

$$\begin{aligned} I &= \iint_R xy \, dx dy = \iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dx dy = \\ &= \int_C P \, dx + Q \, dy = \int_C -\frac{1}{2}xy^2 \, dx + 0 \, dy \end{aligned}$$

Fra parametriseringen ser vi at  $dx = \cos t dt$ , så vi får

$$\begin{aligned} I &= \int_C -\frac{1}{2}xy^2 dx = -\frac{1}{2} \int_0^\pi \sin t (\sin t \cos t)^2 \cos t dt = \\ &= -\frac{1}{2} \int_0^\pi \sin^3 t \cos^3 t dt = -\frac{1}{16} \int_0^\pi \sin^3 2t dt \end{aligned}$$

der vi i siste overgang har brukt formelen  $\sin 2t = 2 \sin t \cos t$ . Skriver vi  $\sin^3 2t = (1 - \cos^2 2t) \sin 2t$  og innfører en ny variabel  $u = \cos 2t$ , får vi

$$I = -\frac{1}{16} \int_0^\pi (1 - \cos^2 2t) \sin 2t dt = \frac{1}{32} \int_1^1 (1 - u^2) du = 0$$

(hadde vi vært smarte, kunne vi på et tidlig tidspunkt ha brukt symmetri til å se at integralet må være 0, men knepene i utregningen av integralet er også greie å kjenne til). ♣

Av og til er det nyttig å bruke Greens teorem til å regne ut arealer.

**Korollar 6.5.2** *Anta at  $C$  er en enkel, lukket kurve med en stykkevis glatt parametrisering  $\mathbf{r}$ , og la være  $R$  området avgrenset av  $C$ . Da er arealet til  $R$  gitt ved*

$$\text{Areal}(R) = \int_C x dy = - \int_C y dx = \frac{1}{2} \int_C -y dx + x dy$$

der kurveintegralene er orientert mot klokken.

*Bevis:* Disse formlene er direkte konsekvenser av Greens teorem. For å vise den første, bruker vi Greens teorem på vektorfeltet  $\mathbf{F}(x, y) = 0 \mathbf{i} + x \mathbf{j}$ . Da er

$$\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 1$$

og vi får

$$\int_C x dy = \iint_R 1 dx dy = \text{Areal}(R)$$

De to andre formlene utledes på tilsvarende måte. □

**Eksempel 3:** La oss regne ut arealet til ellipsen

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Denne ellipsen har parametriseringen

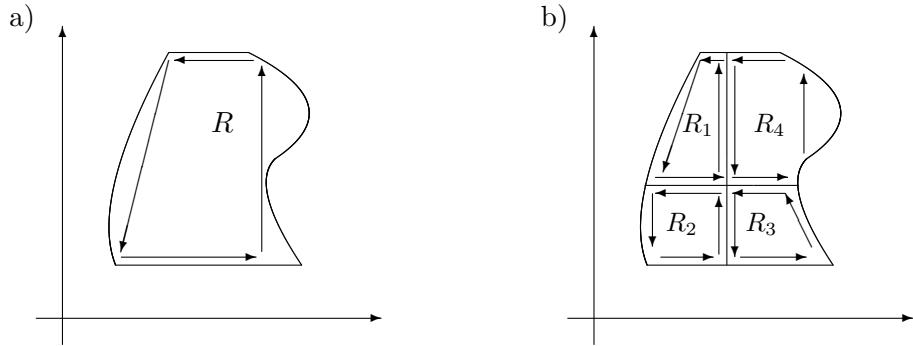
$$\mathbf{r}(t) = a \cos t \mathbf{i} + b \sin t \mathbf{j} \quad t \in [0, 2\pi]$$

som gir  $dx = -a \sin t \, dt$  og  $dy = b \cos t \, dt$ . Setter vi dette inn i den siste av de tre formlene for arealet i korollar 6.5.2, får vi

$$\begin{aligned} \text{Arealet av ellipse} &= \frac{1}{2} \int_C -y \, dx + x \, dy = \\ &= \frac{1}{2} \int_0^{2\pi} [ - (b \sin t)(-a \sin t) \, dt + (a \cos t)(b \cos t) \, dt ] = \int_0^{2\pi} \frac{ab}{2} = \pi ab \end{aligned}$$

♣

Vi skal bevise Greens teorem for “enkle” områder  $R$ , men før vi gjør det, skal vi ta med oss en liten observasjon som gjør det mulig å utvide resultatet fra enkle til mer kompliserte områder. Anta at figur 3a) viser et område  $R$  vi ønsker å bruke Greens teorem på, men at vi er litt usikre på om formelen gjelder for områder av denne typen. Anta videre at vi kan dele  $R$  opp i fire regioner  $R_1, R_2, R_3$  og  $R_4$  som vi vet Greens teorem kan brukes på. Figur 3b) viser oppdelingen med piler som indikerer retningen de forskjellige randkurvene gjennomløpes i.



Figur 3: Oppdeling i enklere områder

Bruker vi Greens teorem på hvert av disse områdene, får vi

$$\iint_{R_1} \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dxdy = \int_{C_1} P \, dx + Q \, dy$$

$$\iint_{R_2} \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dxdy = \int_{C_2} P \, dx + Q \, dy$$

$$\iint_{R_3} \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dxdy = \int_{C_3} P \, dx + Q \, dy$$

$$\iint_{R_4} \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) \, dxdy = \int_{C_4} P \, dx + Q \, dy$$

Legger vi sammen disse ligningene, får vi

$$\begin{aligned} \iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy &= \int_{C_1} (P dx + Q dy) + \\ &+ \int_{C_2} (P dx + Q dy) + \int_{C_3} (P dx + Q dy) + \int_{C_4} (P dx + Q dy) \end{aligned}$$

siden vi åpenbart kan legge sammen de fire dobbeltintegralene. Dersom vi på tilvarende måte kunne legge sammen de fire linjeintegralene, hadde vi hatt Greens teorem for området  $R$ . Det er slett ikke opplagt at vi kan gjøre dette — kurvene  $C_1, C_2, C_3$  og  $C_4$  inneholder jo deler som ikke er med i randkurven  $C$  til det opprinnelige området  $R$  (på figuren er dette de loddrette og vannrette skillelinjene mellom områdene  $R_1, R_2, R_3$  og  $R_4$ ). Legg merke til at disse “ekstra” linjestykkene gjennomløpes to ganger, én i hver retning. Siden motsatt rettede linjeintegraler har samme tallverdi, men motsatte fortegn (husk setning 3.4.4), vil disse bidragene kansellere hverandre, og vi står igjen med

$$\begin{aligned} \int_{C_1} (P dx + Q dy) + \int_{C_2} (P dx + Q dy) + \int_{C_3} (P dx + Q dy) + \\ + \int_{C_4} (P dx + Q dy) = \int_C P dx + Q dy \end{aligned}$$

Kombinerer vi ligningene våre, får vi

$$\iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_C P dx + Q dy$$

som er Greens teorem for området  $R$ .

Det viser seg at det vi nettopp har vist, er et generelt fenomen — når vi deler et komplisert område opp i enklere bestanddeler, vil integralene over de “indre” kurvebitene alltid kansellere (dette forutsetter at området ikke har hull; vi skal komme tilbake til dette nedenfor). For praktisk bruk av Greens teorem holder det for oss å bevise teoremet for enkle områder og overlate til brukerne å vise at de kompliserte områdene de kommer borti, kan deles opp i enklere biter på en sånn måte at de “indre” linjeintegralene kansellerer.

Før vi begynner på beviset, lønner det seg også å observere at Greens teorem består av to deler — den ene sier at

$$\int_C P dx = - \iint_R \frac{\partial P}{\partial y} dx dy,$$

den andre at

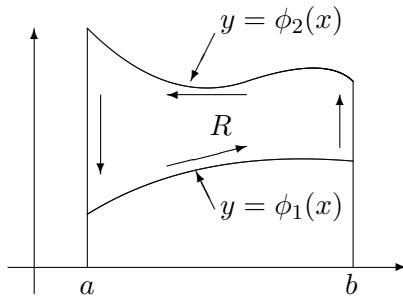
$$\int_C Q dy = \iint_R \frac{\partial Q}{\partial x} dx dy,$$

Vi skal vise disse likhetene hver for seg og for litt forskjellig type område (at områdene er forskjellig, spiller ingen rolle så lenge vi kan bruke sammenslåingsteknikken ovenfor til å lappe sammen delene til mer generelle helheter).

Vi skal først vise  $\int_C P dx = - \iint_R \frac{\partial P}{\partial y} dxdy$  for områder av type I. Husk at et slikt område er på formen

$$R = \{(x, y) : a \leq x \leq b, \phi_1(x) \leq y \leq \phi_2(x)\}$$

Figur 4 viser et område av type I med omløpsretningen fra Greens teorem.



Figur 4: Område av type I

**Lemma 6.5.3** *Anta at  $\phi_1, \phi_2 : [a, b] \rightarrow \mathbb{R}$  er to deriverbare funksjoner slik at  $\phi_1(x) < \phi_2(x)$  for alle  $x \in (a, b)$ . La*

$$R = \{(x, y) : a \leq x \leq b, \phi_1(x) \leq y \leq \phi_2(x)\}$$

*og la  $C$  være randen til  $R$  orientert mot klokken. Anta at  $P$  er en funksjon av to variable med kontinuerlige partiellderiverte i  $R$ . Da er*

$$\int_C P dx = - \iint_R \frac{\partial P}{\partial y} dxdy$$

*Bevis:* Vi begynner med å regne ut linjeintegralet. Randkurven  $C$  består av fire deler  $C_1, C_2, C_3$  og  $C_4$ , der  $C_1$  er den nedre funksjonsgrafen,  $C_2$  er det loddrette linjestykket til høyre,  $C_3$  er den øvre funksjonsgrafen og  $C_4$  er det loddrette linjestykket til venstre. Vi har dermed

$$\int_C P dx = \int_{C_1} P dx + \int_{C_2} P dx + \int_{C_3} P dx + \int_{C_4} P dx$$

Det andre og det fjerde av disse integralene (de to langs de loddrette linjestikkene) er null fordi  $dx = x'(t) dt$  er null siden  $x(t)$  er konstant når vi beveger oss loddrett.

For å regne ut  $\int_{C_1} P dx$  parametriserer vi den nedre funksjonsgrafen ved

$$\mathbf{r}_1(t) = t \mathbf{i} + \phi_1(t) \mathbf{j}; \quad t \in [a, b]$$

Da er  $dx = x'(t) dt = dt$ , og vi får

$$\int_{C_1} P dx = \int_a^b P(t, \phi_1(t)) dt$$

For å regne ut  $\int_{C_3} P dx$  må vi parametrisere den øvre funksjonsgrafen. Den naturlige parametriseringen er

$$\mathbf{r}_2(t) = t \mathbf{i} + \phi_2(t) \mathbf{j}; \quad t \in [a, b],$$

men siden den parametriserer kurven i gal retning, må vi passe på å putte et minus foran integraltegnet. Dermed er

$$\int_{C_3} P dx = - \int_a^b P(t, \phi_2(t)) dt$$

og vi får

$$\int_C P dx = \int_a^b P(t, \phi_1(t)) dt - \int_a^b P(t, \phi_2(t)) dt$$

Til sammenligning regner vi nå ut dobbeltintegralet  $\iint_R \frac{\partial P}{\partial y} dxdy$ . Ved hjelp av analysens fundamentalteorem får vi

$$\begin{aligned} \iint_R \frac{\partial P}{\partial y} dxdy &= \int_a^b \left[ \int_{\phi_1(x)}^{\phi_2(x)} \frac{\partial P}{\partial y}(x, y) dy \right] dx = \\ &= \int_a^b [P(x, \phi_2(x)) - P(x, \phi_1(x))] dx \end{aligned}$$

Sammenligner vi de resultatene vi nå har, ser vi at

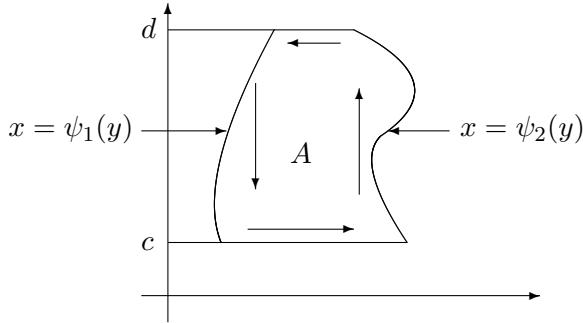
$$\int_C P dx = - \iint_R \frac{\partial P}{\partial y} dxdy$$

□

Vi skal nå vise  $\int_C Q dy = \iint_R \frac{\partial Q}{\partial x} dxdy$  for områder av type II. Vi minner om at slike områder er på formen

$$R = \{(x, y) : c \leq y \leq d, \psi_1(y) \leq x \leq \psi_2(y)\}$$

Figur 5 viser et område av type II med omløpsretningen i Greens teorem inntegnet.



Figur 5: Område av type II

**Lemma 6.5.4** Anta at  $\psi_1, \psi_2 : [c, d] \rightarrow \mathbb{R}$  er to deriverbare funksjoner slik at  $\psi_1(y) < \psi_2(y)$  for alle  $y \in (c, d)$ . La

$$R = \{(x, y) : c \leq y \leq d, \psi_1(y) \leq x \leq \psi_2(y)\}$$

og la  $\mathcal{C}$  være randen til  $R$  orientert mot klokken. Anta at  $Q$  er en funksjon av to variable med kontinuerlige partiellderiverte i  $R$ . Da er

$$\int_{\mathcal{C}} Q dy = \iint_R \frac{\partial Q}{\partial x} dx dy$$

*Bevis:* Beviset er en kopi av det foregående med rollene til  $x$ - og  $y$ -aksen byttet om. Vi begynner med å regne ut linjeintegralet. Randkurven  $\mathcal{C}$  består av fire deler  $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$  og  $\mathcal{C}_4$ , der  $\mathcal{C}_1$  er den venstre funksjonsgrafen,  $\mathcal{C}_2$  er det nederste, vannrette linjestykket,  $\mathcal{C}_3$  er den høye funksjonsgrafen og  $\mathcal{C}_4$  er det øverste, vannrette linjestykket. Vi har dermed

$$\int_{\mathcal{C}} Q dy = \int_{\mathcal{C}_1} Q dy + \int_{\mathcal{C}_2} Q dy + \int_{\mathcal{C}_3} Q dy + \int_{\mathcal{C}_4} Q dy$$

Det andre og det fjerde av disse integralene (de to langs de vannrette linjestykrene) er null fordi  $dy = y'(t) dt$  er null siden  $y(t)$  er konstant når vi beveger oss vannrett.

For å regne ut  $\int_{\mathcal{C}_1} Q dy$  parametriserer vi den venstre funksjonsgrafen ved

$$\mathbf{r}_1(t) = \psi_1(t) \mathbf{i} + t \mathbf{j}; \quad t \in [c, d]$$

Siden denne parametriseringen går i feil retning, må vi kompensere ved å sette et minus foran linjeintegralet. Siden  $dy = y'(t) dt = dt$ , får vi dermed

$$\int_{\mathcal{C}_1} Q dy = - \int_c^d Q(\psi_1(t), t) dt$$

For å regne ut  $\int_{\mathcal{C}_3} Q dy$  parametriserer vi den høye funksjonsgrafen ved

$$\mathbf{r}_2(t) = \psi_2(t) \mathbf{i} + t \mathbf{j}; \quad t \in [c, d],$$

og får

$$\int_{C_3} Q \, dy = \int_c^d Q(\psi_2(t), t) \, dt$$

Dermed har vi

$$\int_C Q \, dy = - \int_c^d Q(\psi_1(t), t) \, dt + \int_c^d Q(\psi_2(t), t) \, dt$$

Til slutt regner vi ut dobbeltintegralet  $\iint_R \frac{\partial Q}{\partial x} \, dxdy$ . Ved hjelp av analysens fundamentalteorem får vi

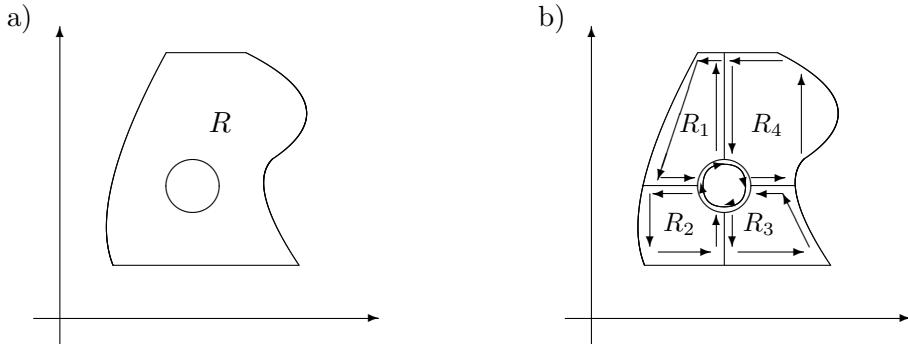
$$\begin{aligned} \iint_R \frac{\partial Q}{\partial x} \, dxdy &= \int_c^d \left[ \int_{\psi_1(y)}^{\phi_2(y)} \frac{\partial Q}{\partial x}(x, y) \, dx \right] dy = \\ &= \int_c^d [Q(\psi_2(y), y) - Q(\psi_1(y), y)] \, dy \end{aligned}$$

Sammenligner vi de resultatene vi nå har, ser vi at

$$\int_C Q \, dy = \iint_R \frac{\partial Q}{\partial x} \, dxdy$$

□

Med de to lemmaene ovenfor til disposisjon kan selv den mest nidkjære med god samvittighet bruke Greens teorem på alle de sammenhengende, begrensete mengdene uten hull som dukker opp i praksis. Vi kan også bruke lemmaene på områder *med* hull, men da dukker det opp en ny effekt.



Figur 6: Et område med hull

Figur 6a) viser et område  $R$  med hull. I b) har vi delt opp området i fire delområder  $R_1, R_2, R_3, R_4$  som vi kan bruke Greens teorem på. Dette gir oss

$$\iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy = \int_{C_1} P \, dx + Q \, dy + \int_{C_2} P \, dx + Q \, dy +$$

$$+ \int_{C_3} P \, dx + Q \, dy + \int_{C_4} P \, dx + Q \, dy$$

der  $C_i$  er randen til  $R_i$ . Tar vi en kikk på figur 6b) ser vi at alle “indre” bidrag til linjeintegralene kansellerer med fire unntak — linjeintegralene langs hullet forekommer bare én gang, og de er orientert i *negativ* omløpsretning (altså med klokken). Vi sitter altså igjen med

$$\iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_C P \, dx + Q \, dy - \int_{C_h} P \, dx + Q \, dy$$

der  $C$  er den “ytre” randen til  $R$ , og  $C_h$  er den indre randen (mot hullet), begge med positiv orientering. Vi får en tilsvarende formel for områder med flere hull.

Det neste eksemplet viser at man må være litt forsiktig når man bruker Greens teorem.

**Eksempel 4:** Vi lar  $C$  være enhetssirkelen orientert mot klokken og setter

$$\mathbf{F}(x, y) = P(x, y) \mathbf{i} + Q(x, y) \mathbf{j} = -\frac{y}{x^2 + y^2} \mathbf{i} + \frac{x}{x^2 + y^2} \mathbf{j}$$

La oss først regne ut linjeintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  på vanlig måte. Siden enhets-sirkelen har parametriseringen

$$\mathbf{r}(t) = \cos t \mathbf{i} + \sin t \mathbf{j}, \quad t \in [0, 2\pi]$$

får vi

$$\begin{aligned} \int_C \mathbf{F} \cdot d\mathbf{r} &= \int_0^{2\pi} \left( -\frac{\sin t}{\cos^2 t + \sin^2 t} \mathbf{i} + \frac{\cos t}{\cos^2 t + \sin^2 t} \mathbf{j} \right) \cdot (-\sin t \mathbf{i} + \cos t \mathbf{j}) \, dt = \\ &= \int_0^{2\pi} 1 \, dt = 2\pi \end{aligned}$$

La oss så forsøke å bruke Greens teorem isteden. Vi observerer at

$$\frac{\partial Q}{\partial x} = \frac{1 \cdot (x^2 + y^2) - x \cdot 2x}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}$$

$$\frac{\partial P}{\partial y} = -\frac{1 \cdot (x^2 + y^2) - y \cdot 2y}{(x^2 + y^2)^2} = \frac{y^2 - x^2}{(x^2 + y^2)^2}$$

Dermed er  $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = 0$ , og ifølge Greens teorem burde vi vel få

$$\int_C \mathbf{F} \cdot d\mathbf{r} = \int_C P \, dx + Q \, dy = \iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \iint_R 0 \, dx dy = 0$$

selv om dette ikke stemmer med utregningene ovenfor?

Hva er galt? Ser vi nøyere etter, ser vi at  $\mathbf{F}$  ikke er definert i origo, og at vi derfor strengt tatt ikke kan bruke Greens teorem. Vi kan imidlertid prøve å unngå problemet på følgende måte. Vi lar  $C_r$  være en liten sirkel om origo med radius  $r$ , og bruker Greens teorem på området mellom de to sirklene. Vi burde da få

$$0 = \iint_R \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy = \int_C P dx + Q dy - \int_{C_r} P dx + Q dy \quad (6.5.1)$$

Lar vi  $r$  gå mot null, får vi en kortere og kortere kurve å integrere over, og vi ville tro at

$$\int_{C_r} P dx + Q dy \rightarrow 0 \quad \text{når } r \rightarrow 0$$

I så fall sitter vi nok en gang igjen med

$$\int_C P dx + Q dy = 0$$

Men la oss se nærmere på integralet  $\int_{C_r} P dx + Q dy$ . Dersom vi parametriserer  $C_r$  med

$$\mathbf{s}(t) = r \cos t \mathbf{i} + r \sin t \mathbf{j}, \quad t \in [0, 2\pi]$$

får vi

$$\begin{aligned} & \int_{C_r} P dx + Q dy = \\ &= \int_0^{2\pi} \left( -\frac{r \sin t}{r^2 \cos^2 t + r^2 \sin^2 t} \mathbf{i} + \frac{r \cos t}{r^2 \cos^2 t + r^2 \sin^2 t} \mathbf{j} \right) \cdot (-r \sin t \mathbf{i} + r \cos t \mathbf{j}) dt \\ &= \int_0^{2\pi} 1 dt = 2\pi \end{aligned}$$

Nå stemmer alt! Går vi tilbake til ligning (6.5.1) ovenfor, ser vi at

$$\int_C P dx + Q dy = \int_{C_r} P dx + Q dy = 2\pi$$

akkurat som vi skulle ha.

Dette eksemplet har en dobbelt moral (men ikke en dobbeltmoral!). Det ene poenget er at vi må være forsiktig når vi bruker Greens teorem og sjekke at funksjonene vi bruker, virkelig er definert i alle punkter vi integrerer over. Det andre er at det ikke er sikkert at integralet over stadig kortere kurver går mot null dersom integranden samtidig går mot uendelig. Vær oppmerksom på at dette ikke er matematiske spissfindigheter — grunnleggende fysiske krefter som gravitasjon og elektromagnetisme har en form som gjør at slike problemstillinger stadig dukker opp i praksis.



Greens teorem er et eksempel på det som ofte kalles *vektoranalyse*. I tre (og høyere) dimensjoner finnes det enda viktigere eksempler på slike teoremer — *Stokes' teorem* og *Gauss' teorem* (også kjent som *divergensteoremet*). Det første av disse binder sammen et integral over en flate med et integral langs randkurven til flaten, mens det andre binder sammen et integral over et romlegeme med flateintegralet over randen til dette legemet. Vi ser at temaet hele tiden er det samme — et integral over et område sammenlignes med et annet integral over randen til området.

### Oppgaver til seksjon 6.5

**1.** Bruk Greens teorem til å regne ut linjeintegralene. I alle tilfeller er kurven  $\mathcal{C}$  positivt orientert.

- a)  $\int_{\mathcal{C}} (x^2 + y) \, dx + x^2y \, dy$  der  $\mathcal{C}$  er omkretsen til kvadratet med hjørner i  $(0, 0)$ ,  $(2, 0)$ ,  $(2, 2)$  og  $(0, 2)$ .
- b)  $\int_{\mathcal{C}} x^2y^3 \, dx + x^3y^2 \, dy$  der  $\mathcal{C}$  er omkretsen til trekanten med hjørner i  $(0, 0)$ ,  $(3, 0)$ ,  $(3, 1)$
- c)  $\int_{\mathcal{C}} (x^2y + y) \, dx + (xy + x) \, dy$  der  $\mathcal{C}$  er omkretsen til trapeset med hjørner i  $(0, 0)$ ,  $(1, 0)$ ,  $(1, 2)$  og  $(0, 1)$
- d)  $\int_{\mathcal{C}} (x^2y + xe^x) \, dx + (xy^3 + e^{\sin y}) \, dy$  der  $\mathcal{C}$  er omkretsen til området avgrenset av parabelen  $y = x^2$  og linjestykke med endepunkter  $(-1, 1)$  og  $(2, 4)$ .

**2.** Kurven  $\mathcal{C}$  er gitt ved

$$\mathbf{r}(t) = t \sin(t) \mathbf{i} + (2\pi t - t^2) \mathbf{j}, \quad t \in [0, 2\pi]$$

Skisser kurven (f.eks. ved å bruke MATLAB) og regn ut arealet til området den avgrenser.

**3.** Kurven  $\mathcal{C}$  er gitt ved

$$\mathbf{r}(t) = \sin 2t \mathbf{i} + t \cos t \mathbf{j}, \quad t \in [0, \frac{\pi}{2}]$$

Skisser kurven (f.eks. ved å bruke MATLAB) og regn ut arealet til området avgrenset av kurven.

**4.** Regn ut arealet avgrenset av kurven

$$\mathbf{r}(t) = a \cos^3 t \mathbf{i} + b \sin^3 t \mathbf{j}, \quad t \in [0, 2\pi]$$

der  $a$  og  $b$  er to positive tall.

**5.** Regn ut  $\iint_R x \, dxdy$  der  $R$  er området avgrenset av kurven

$$\mathbf{r}(t) = (t - t^2) \mathbf{i} + (t - t^3) \mathbf{j}, \quad t \in [0, 1]$$

**6.** Regn ut  $\iint_R y \, dxdy$  der  $R$  er området avgrenset av kurven

$$\mathbf{r}(t) = \sin t \mathbf{i} + t^2 \mathbf{j}, \quad t \in [-\pi, \pi]$$

**7.** (Eksamens i MAT1110 13/6, 2005) La  $D$  være området i  $\mathbb{R}^2$  som består av punkter  $(x, y)$  som oppfyller ulikhettene  $x^2 + y^2 \leq 1$  og  $y \geq 0$ . La  $\mathcal{C}$  være randen til  $D$  orientert mot urviseren. Finn verdien av kurveintegralet

$$\int_{\mathcal{C}} (xy + \ln(x^2 + 1)) \, dx + (4x + e^{y^2} + 3 \arctan y) \, dy$$

**8.** (Eksamens i MAT1110 13/8, 2004) La  $D$  være det begrensede området i  $\mathbb{R}^2$  som er avgrenset av parabelen  $y = 1 - x^2$  og  $x$ -aksen. La  $\mathcal{C}$  være den lukkede randkurven til  $D$ . Orienter  $\mathcal{C}$  mot urviseren.

- a) Regn ut kurveintegralet  $I = \int_{\mathcal{C}} -y \, dx + x^2 \, dy$  ved direkte utregning av kurveintegralet.
- b) Regn ut  $I$  ved å beregne et dobbeltintegral  $\iint_D f(x, y) \, dxdy$  av et passelig skalarfelt  $f(x, y)$ .

**9.** (Prøveeksamen i MAT1110, 2004) La  $D$  være området i  $\mathbb{R}^2$  bestemt av ulikhettene  $x \geq 0$ ,  $y \geq 0$  og  $x \leq y \leq 2 - x^2$ .

- a) Sett opp et dobbeltintegral som har verdi lik arealet til  $D$  og regn ut verdien av dette dobbeltintegralet.
- b) Beregn arealet av  $D$  ved å beregne linjeintegralet av et passelig vektorfelt langs randen til  $D$ .

**10.** (Eksamens i MAT1110 14/6, 2004).

- a) La  $D$  være området av punkter  $(x, y)$  i  $\mathbb{R}^2$  som oppfyller ulikhettene:  $x^2 + y^2 \leq 1$ ,  $x \geq 0$ ,  $y \geq 0$  og  $0 \leq y \leq x$ . Lag en skisse av området og beregn dobbeltintegralet  $I = \iint_D (x + y^2) \, dxdy$  ved å innføre polarkoordinater.
- b) Beregn  $I$  ved å regne ut direkte et kurveintegral  $\int_{\mathcal{C}} P \, dx + Q \, dy$  av et passelig vektorfelt  $F = P \mathbf{i} + Q \mathbf{j}$  langs den stykkevis glatte kurven  $\mathcal{C}$  som utgjør randen til  $D$ .

**11.** (Eksamens i MAT1110 13/6, 2007)  $R$  er rektanglet med hjørner i  $(1, 1)$ ,  $(3, 1)$ ,  $(3, 2)$ ,  $(1, 2)$ , og  $\mathcal{C}$  er omkretsen til  $R$  orientert mot klokken. Finn  $\int_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  der

$$\mathbf{F}(x, y) = (xy^2 - y) \mathbf{i} + (x^2y + x) \mathbf{j}$$

**12.** (Kontinuasjonseksemplen i MAT1110, 2006)

- a) En ellipse har ligningen

$$9x^2 + 4y^2 - 18x + 16y = 11$$

Finn sentrum og halvaksene til ellipsen, og lag en skisse av ellipsen i koordinatsystemet.

- b) Vis at

$$\mathbf{r}(t) = (1 + 2 \cos t) \mathbf{i} + (-2 + 3 \sin t) \mathbf{j}, \quad t \in [0, 2\pi)$$

er en parametrisering av ellipsen i a). Regn ut  $\oint_{\mathcal{C}} \mathbf{F} \cdot d\mathbf{r}$  der

$$\mathbf{F}(x, y) = y^2 \mathbf{i} + x \mathbf{j}$$

og der  $\mathcal{C}$  er ellipsen med positiv orientering.

c) Regn ut

$$\iint_R (1 - 2y) \, dx \, dy$$

der  $R$  er området avgrenset av ellipsen.

**13.** Det er en nær sammenheng mellom Greens teorem og teorien for konservative vektorfelt i seksjon 3.5. Bruk Greens teorem til å vise at dersom  $\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$  er et konservativt felt, så er  $\int_C \mathbf{F} \cdot d\mathbf{r} = 0$  for alle enkle, lukkede, stykkevis glatte kurver  $C$ .

## 6.6 Jordan-målbare mengder

I seksjon 6.4 så vi hvordan vi kunne regne ut arealet til en mengde  $A$  ved hjelp av dobbeltintegraler:

$$\text{Areal}(A) = \iint_A 1 \, dx \, dy$$

Dette forutsetter selvfølgelig at integralet  $\iint_A 1 \, dx \, dy$  er definert, og det er det samme som at funksjonen

$$1_A(x) = \begin{cases} 1 & \text{hvis } x \in A \\ 0 & \text{hvis } x \notin A \end{cases}$$

er integrerbar. I denne seksjonen skal vi finne betingelser som sikrer at  $1_A$  er integrerbar, og vi skal bruke disse betingelsene til å finne kriterier som garanterer at generelle integraler av typen  $\iint_A f(x, y) \, dx \, dy$  eksisterer. Dette vil blant annet sette oss i stand til å bevise at dobbeltintegraler av kontinuerlige funksjoner over områder av type I og type II virkelig eksisterer (husk seksjon 6.2).

Sammenlignet med de foregående er denne seksjonen ganske teoretisk, og det er ikke sikkert at alle har lyst til å lese alle bevisene. Det er imidlertid viktig å få med seg definisjonen av begrepene "Jordan-målbar" og "innhold null" som vi skal få bruk for i senere seksjoner.

**Definisjon 6.6.1** Vi sier at en begrenset mengde  $A \subset \mathbb{R}^2$  er Jordan-målbar dersom  $1_A$  er integrerbar.

**Bemerkning:** Det finnes et sterkere begrep *Lebesgue-målbar* (sterkere i den forstand at flere mengder er Lebesgue-målbare enn Jordan-målbare) som du vil støte på i videregående kurs. Når folk bare sier "målbar", mener de som regel "Lebesgue-målbar", og vi skal derfor dra med oss "Jordan"-forstavelsen i dette heftet.

Det viser seg at det eneste som kan hindre en mengde i å være Jordan-målbar, er at randen ikke er “liten” nok. Den neste definisjonen presiserer hva “liten” betyr i denne sammenhengen.

**Definisjon 6.6.2** En begrenset mengde  $B \subset \mathbb{R}^2$  har innhold 0 dersom det for hver  $\epsilon > 0$  finnes endelig mange rektangler

$$R_1 = [a_1, b_1] \times [c_1, d_1], R_2 = [a_2, b_2] \times [c_2, d_2], \dots, R_n = [a_n, b_n] \times [c_n, d_n]$$

slik at

$$B \subset R_1 \cup R_2 \cup \dots \cup R_n$$

og summen av arealene til  $R_1, R_2, \dots, R_n$  er mindre enn  $\epsilon$ .

En begrenset mengde har altså innhold null dersom den kan dekkes av endelige familier av rektangler med vilkårlig lite samlet areal. Mengder med innhold null er “små” sett med to-dimensjonale øyne. Man vil f.eks. vente at en-dimensjonale objekter som en kontinuerlig funksjonsgraf har innhold null (vi skal vise dette i beviset for setning 6.6.5).

Vi kan nå formulere det første hovedresultatet vårt:

**Teorem 6.6.3** En begrenset mengde  $A \subset \mathbb{R}^2$  er Jordan-målbar hvis og bare hvis randen  $\partial A$  til  $A$  har innhold 0.

Vi trenger noen forberedelser før vi kan bevise teoremet.

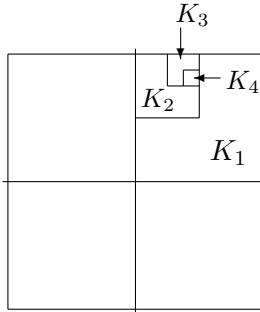
**Lemma 6.6.4** Anta at  $A \subset \mathbb{R}^2$  er en begrenset mengde, og at  $R$  er et rektangel som inneholder  $A$  i sitt indre. Anta at  $\Pi$  er en partisjon som deler  $R$  inn i delrektangler  $R_{ij} = [a_{i-1}, a_i] \times [b_{j-1}, b_j]$ .

- (i) Dersom  $\mathbf{a} \in \partial A$ , så er  $\mathbf{a}$  med i et delrektangel  $R_{ij}$  som både inneholder punkter som er med i  $A$ , og punkter som ikke er det.
- (ii) Dersom et delrektangel  $R_{ij}$  både inneholder punkter som er med i  $A$ , og punkter som ikke er det, så inneholder  $R_{ij}$  et randpunkt  $\mathbf{a} \in \partial A$ .

*Bevis:* (i) Siden  $\mathbf{a}$  ligger i  $R$ , må det høre til minst ett delrektangel  $R_{ij}$  (siden naborektangler har felles rand, kan  $\mathbf{a}$  godt høre til flere rektangler). Dersom  $\mathbf{a}$  ligger i det indre av  $R_{ij}$ , må  $R_{ij}$  åpenbart inneholde både punkter som er med i  $A$  og punkter som ikke er det (husk at  $\mathbf{a} \in \partial A$ ). Dersom  $\mathbf{a}$  ligger på randen til  $R_{ij}$  (enten på en kant eller i et hjørne), kan det hende at  $R_{ij}$  enten bare inneholder punkter som er med i  $A$  eller bare punkter som ikke er med i  $A$ , men da er det lett å overbevise seg om at ett av de andre smårektanglene som  $\mathbf{a}$  hører til, må inneholde både punkter som er med i  $A$ , og punkter som ikke er det (lag en tegning).

(ii) Vi må vise at dersom  $R_{ij}$  både inneholder punkter som er med i  $A$ , og punkter som ikke er det, så må  $R_{ij}$  inneholde et punkt fra  $\partial A$ . Vi skal

bruke samme teknikk som i beviset for Bolzano-Weierstrass' teorem 5.2.3. Vi deler først  $R_{ij}$  opp i fire mindre, lukkede rektangler, og observerer at minst ett av dem (kall det  $K_1$ ) må inneholde både punkter som er fra  $A$ , og punkter som ikke er det (tenk gjennom dette — det er ikke helt opplagt). Deler vi på samme måte  $K_1$  inn i fire mindre rektangler, må minst ett av dem både inneholde punkter som er med i  $A$ , og punkter som ikke er det. Ved å fortsette på denne måten får vi en følge  $\{K_n\}$  av stadig mindre rektangler som alle inneholder både punkter som er med i  $A$ , og punkter som ikke er med i  $A$  (se figur 1).



Figur 1: Rektanglene  $K_n$

Lar vi  $\mathbf{a}_n$  være nedre, venstre hjørne i  $K_n$ , vil følgen  $\{\mathbf{a}_n\}$  konvergere mot et punkt  $\mathbf{a}$  i  $R_{ij}$ . Siden hver  $K_n$  både inneholder punkter som er med i  $A$ , og punkter som ikke er det, finnes det punkter av begge typer vilkårlig nær  $\mathbf{a}$ , og  $\mathbf{a}$  må derfor ligge på randen  $\partial A$  til  $A$ .  $\square$

Før vi går løs på beviset for teorem 6.6.3 gjør vi en liten observasjon til. Dersom  $\Pi$  er en partisjon av rektanglet  $R$ , og  $A$  er en delmengde av  $R$ , vil hvert delrektangel  $R_{ij}$  høre til én av følgende tre kategorier:

- (i) Kategori 1: Alle punktene i  $R_{ij}$  er med i  $A$
- (ii) Kategori 2: Ingen av punktene i  $R_{ij}$  er med i  $A$
- (iii) Kategori 3: Noen (men ikke alle) punktene i  $R_{ij}$  er med i  $A$

Bevis for setning 6.6.3: Anta først at  $A$  er Jordan-målbar. Vi må vise at randen  $\partial A$  til  $A$  har innhold 0.

La  $R$  være et rektangel som inneholder  $A$  i sitt indre. At  $A$  er Jordan-målbar, betyr at funksjonen  $\mathbf{1}_A$  er integrerbar over  $R$ . Gitt  $\epsilon > 0$ , finnes det da en partisjon  $\Pi$  av  $R$  slik at  $\mathcal{O}(\Pi) - N(\Pi) < \epsilon$ , der  $\mathcal{O}(\Pi)$  og  $N(\Pi)$  er øvre og nedre trappesum til funksjonen  $\mathbf{1}_A$ . Dersom delrektanglet  $R_{ij}$  er av kategori 1 eller 2, er supremum  $M_{ij}$  og infimum  $m_{ij}$  over rektanglet  $R_{ij}$  like (henholdsvis 1 for den første kategorien og 0 for den andre), mens for den

tredje kategorien er  $M_{ij} = 1$  og  $m_{ij} = 0$ . Dette betyr at

$$\epsilon > \emptyset(\Pi) - N(\Pi) = \sum_{R_{ij} \text{ av tredje kategori}} |R_{ij}|$$

der  $|R_{ij}|$  er arealet til rektanglet  $R_{ij}$ . Fra lemmaet ovenfor vet vi at ethvert randpunkt  $\mathbf{a}$  er med i et rektangel  $R_{ij}$  av kategori 3, og dermed utgjør rektanglene av kategori 3 en overdekning av randen  $\partial A$ . Dermed har vi funnet en overdekning av  $\partial A$  med totalt areal mindre enn  $\epsilon$ , og følgelig har  $\partial A$  innhold 0.

Anta så at randen  $\partial A$  til  $A$  har innhold 0. Gitt en  $\epsilon > 0$ , må vi finne en partisjon  $\Pi$  slik at  $\emptyset(\Pi) - N(\Pi) < \epsilon$ . Dette vil medføre at  $1_A$  er integrerbar, og dermed bevise teoremet.

Siden  $\partial A$  har innhold 0, finnes det en samling rektangler

$$S_1 = [a_1, b_1] \times [c_1, d_1], S_2 = [a_2, b_2] \times [c_2, d_2], \dots, S_n = [a_n, b_n] \times [c_n, d_n]$$

som dekker  $\partial A$ , og har samlet areal mindre enn  $\epsilon$ . Lag en partisjon  $\Pi$  av  $R$  ved å bruke punktene  $a_1, a_2, \dots, a_n$  og  $b_1, b_2, \dots, b_n$  som delepunkter langs  $x$ -aksen og  $c_1, c_2, \dots, c_n$  og  $d_1, d_2, \dots, d_n$  som delepunkter langs  $y$ -aksen. Observer at hvert av smårektanglene  $R_{ij}$  til denne partisjonen ligger inni et av de opprinnelige rektanglene  $S_1, S_2, \dots, S_n$  (lag en figur!). De av rektanglene  $R_{ij}$  som inneholder punkter fra  $\partial A$  må derfor ha samlet areal mindre enn  $\epsilon$ .

Akkurat som ovenfor vil rektanglene  $R_{ij}$  være av tre forskjellige kategorier, og som ovenfor har vi

$$\emptyset(\Pi) - N(\Pi) = \sum_{R_{ij} \text{ av tredje kategori}} |R_{ij}|$$

Fra lemmaet vet vi at ethvert rektangel av kategori 3 inneholder et punkt fra  $\partial A$ , og dermed har vi

$$\emptyset(\Pi) - N(\Pi) = \sum_{R_{ij} \text{ av tredje kategori}} |R_{ij}| \leq \epsilon$$

Beviset er fullført. □

Vi kan nå vise at ethvert område av type I eller II (se seksjon 6.2) er Jordan-målbart.

**Setning 6.6.5** *Ethvert område av type I eller II er Jordan-målbart.*

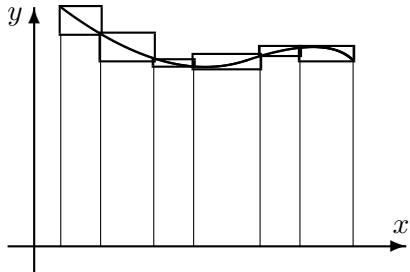
*Bevis:* Vi nøyer oss med å bevise setningen for et område  $A$  av type I. Ifølge teoremet ovenfor er det nok å vise at randen til  $A$  har innhold 0. Siden randen til  $A$  består av fire deler (to funksjonsgrafer og to loddrette linjestykker), er

det nok å vise at hver av disse delene har innhold 0 (sjekk dette!). Vi tar for oss funksjonsgrafene og overlater linjestykkene til leserne.

Anta altså at  $\phi : [a, b] \rightarrow \mathbb{R}$  er en kontinuerlig funksjon. Vi må vise at funksjonsgrafen

$$G = \{(x, \phi(x)) \mid x \in [a, b]\}$$

har innhold null. Figur 2 viser ideen i beviset som består i å overdekke grafen med rektanglene som skiller en øvre trappesum fra en nedre.



Figur 2: Overdekning av en graf

Her er detaljene: Siden alle kontinuerlige funksjoner er integrerbare, finnes det for enhver  $\epsilon > 0$  en partisjon  $\Pi$  av intervallet  $[a, b]$  slik at  $\mathcal{O}(\Pi) - N(\Pi) < \epsilon$ . Ser vi på den tilsvarende overdekningen i figur 2, ser vi at det totale arealet til de overdekkende rektanglene er

$$\sum_{i=1}^n (M_i - m_i)(t_i - t_{i-1}) < \mathcal{O}(\Pi) - N(\Pi) \leq \epsilon$$

Dermed har vi funnet en overdekning av grafen  $G$  med totalt areal mindre enn  $\epsilon$ , og følgelig har grafen innhold 0.  $\square$

Vi kan nå bevise et kriterium for integrerbarhet av kontinuerlige funksjoner.

**Teorem 6.6.6** *Anta at  $A \subset \mathbb{R}^2$  er en lukket, begrenset, Jordan-målbar mengde. Da er enhver kontinuerlig funksjon  $f : A \rightarrow \mathbb{R}$  integrerbar over  $A$  (dvs. at integralet  $\iint_A f(x, y) dx dy$  eksisterer). Spesielt er alle kontinuerlige funksjoner integrerbare over områder av type I og II.*

*Bevis:* La  $R$  være et rektangel som inneholder  $A$ . Vi må vise at funksjonen

$$f_A(x, y) = \begin{cases} f(x, y) & \text{hvis } (x, y) \in A \\ 0 & \text{ellers} \end{cases}$$

er integrerbar over  $R$ . Det er nok å vise at  $f$  er begrenset og at det for hver  $\epsilon > 0$ , finnes en partisjon  $\Pi$  av  $R$  slik at  $\mathcal{O}(\Pi) - N(\Pi) < \epsilon$ .

Ifølge ekstremalverdisetningen 5.7.2 er  $f$  begrenset, og det finnes derfor et tall  $M$  slik at  $|f(\mathbf{x})| \leq M$  for alle  $\mathbf{x} \in A$ . Dette betyr at dersom  $R_{ij}$  er et delrekktangel i en partisjon av  $R$ , så er alltid  $M_{ij} - m_{ij} \leq 2M$ .

Fra setning 6.6.3 vet vi at det finnes en partisjon  $\hat{\Pi}$  slik at hvis  $\phi(\hat{\Pi})$  og  $n(\hat{\Pi})$  er henholdsvis øvre og nedre trappesum til funksjonen  $1_A$  med hensyn på  $\hat{\Pi}$ , så er  $\phi(\hat{\Pi}) - n(\hat{\Pi}) < \frac{\epsilon}{4M}$ . Ifølge setning 6.1.4 er  $f : A \rightarrow \mathbb{R}$  uniformt kontinuerlig, og det finnes derfor en  $\delta > 0$  slik at hvis  $\mathbf{a}$  og  $\mathbf{b}$  er to punkter i  $A$  med  $|\mathbf{a} - \mathbf{b}| < \delta$ , så er  $|f(\mathbf{a}) - f(\mathbf{b})| < \frac{\epsilon}{2|R|}$ , der  $|R|$  som vanlig er arealet til  $R$ . Ved å legge inn enda flere delepunkter hvis nødvendig, kan vi forfine  $\hat{\Pi}$  til en partisjon  $\Pi$  slik at alle punkter som hører til samme delrekktangel  $R_{ij}$ , har innbyrdes avstand mindre enn  $\delta$ . Vi har

$$\begin{aligned}\mathcal{O}(\Pi) - N(\Pi) &= \sum_{\text{alle rektangler } R_{ij}} (M_{ij} - m_{ij}) |R_{ij}| = \\ &= \sum_{\substack{\text{rektangler av} \\ \text{kategori 1}}} (M_{ij} - m_{ij}) |R_{ij}| + \sum_{\substack{\text{rektangler av} \\ \text{kategori 2}}} (M_{ij} - m_{ij}) |R_{ij}| + \\ &\quad + \sum_{\substack{\text{rektangler av} \\ \text{kategori 3}}} (M_{ij} - m_{ij}) |R_{ij}|\end{aligned}$$

I det siste uttrykket er

$$\sum_{\substack{\text{rektangler av} \\ \text{kategori 1}}} (M_{ij} - m_{ij}) |R_{ij}| \leq \sum_{\substack{\text{rektangler av} \\ \text{kategori 1}}} \frac{\epsilon}{2|R|} |R_{ij}| < \frac{\epsilon}{2}$$

på grunn av den uniforme kontinuiteten. Videre er

$$\sum_{\substack{\text{rektangler av} \\ \text{kategori 2}}} (M_{ij} - m_{ij}) |R_{ij}| = 0$$

siden  $M_{ij} = m_{ij} = 0$  i dette tilfellet. Endelig er

$$\begin{aligned}\sum_{\substack{\text{rektangler av} \\ \text{kategori 3}}} (M_{ij} - m_{ij}) |R_{ij}| &\leq 2M \sum_{\substack{\text{rektangler av} \\ \text{kategori 3}}} |R_{ij}| \leq \\ &\leq 2M (\phi(\Pi) - n(\Pi)) \leq \frac{\epsilon}{2}\end{aligned}$$

(husk at  $\phi(\Pi)$  og  $n(\Pi)$  er øvre og nedre trappesum til funksjonen  $1_A$ , og at vi har laget partisjonen slik at  $\phi(\Pi) - n(\Pi) < \frac{\epsilon}{4M}$ ). Alt i alt er da

$$\mathcal{O}(\Pi) - N(\Pi) < \frac{\epsilon}{2} + 0 + \frac{\epsilon}{2} = \epsilon$$

som viser at  $f$  er integrerbar. Utsagnene om integraler over områder av type I og II følger nå fra setning 6.6.5.  $\square$

**Bemerkning:** Hvis man arbeider litt hardere, kan man beskrive nøyaktig hvilke funksjoner  $f$  som er integrerbare over et rektangel  $R$  — det viser seg at  $f$  er integrerbar hvis og bare hvis mengden av punkter der  $f$  er diskontinuerlig, har *mål* null (dette er en mer generell betingelse enn å ha *innhold* null). Resultatet ovenfor holder imidlertid for våre behov i dette kapitlet — spesielt er det tilstrekkelig for beviset av formelen for skifte av variabel.

### Oppgaver til seksjon 6.6

1. Vis at dersom  $A_1, A_2, \dots, A_m$  er delmengder av  $\mathbb{R}^n$  med innhold null, så har også  $A_1 \cup A_2 \cup \dots \cup A_m$  innhold null.
2. I denne oppgaven er  $A$  og  $B$  to delmengder av  $\mathbb{R}^2$ .
  - a) Vis at  $\partial(A \cup B) \subset \partial A \cup \partial B$  og  $\partial(A \cap B) \subset \partial A \cap \partial B$ .
  - b) Vis at dersom  $A$  og  $B$  er Jordan-målbare, så er også  $A \cup B$  og  $A \cap B$  Jordan-målbare (du kan få bruk for resultatet i den forrige oppgaven).
3. Anta at  $r : [a, b] \rightarrow \mathbb{R}$  er kontinuerlig funksjon. Vis at mengden av alle punkter med polarkoordinater  $(\theta, r(\theta))$  der  $\theta \in [a, b]$  har innhold 0.

## 6.7 Skifte av variable i dobbeltintegral

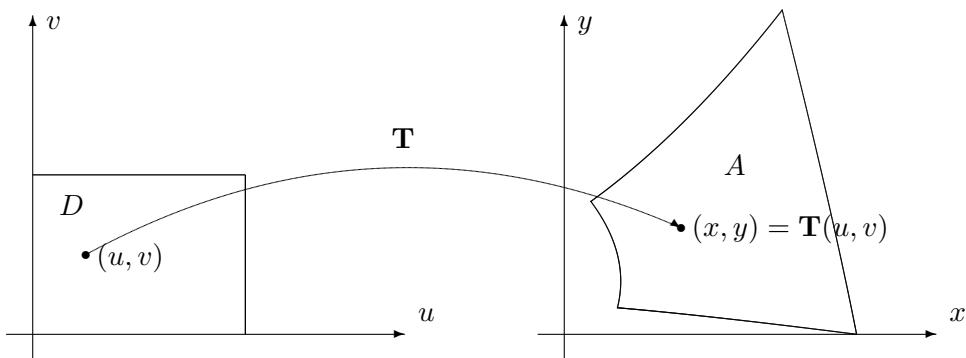
Husk hvordan vi skifter variabel i et vanlig integral  $\int_a^b f(x) dx$ : Dersom vi ønsker å innføre en ny variabel  $u = g(x)$ , regner vi ut  $x = h(u)$  og  $dx = h'(u) du$ , og setter inn (husk å bytte integrasjonsgrenser):

$$\int_a^b f(x) dx = \int_{g(a)}^{g(b)} f(h(u))h'(u) du$$

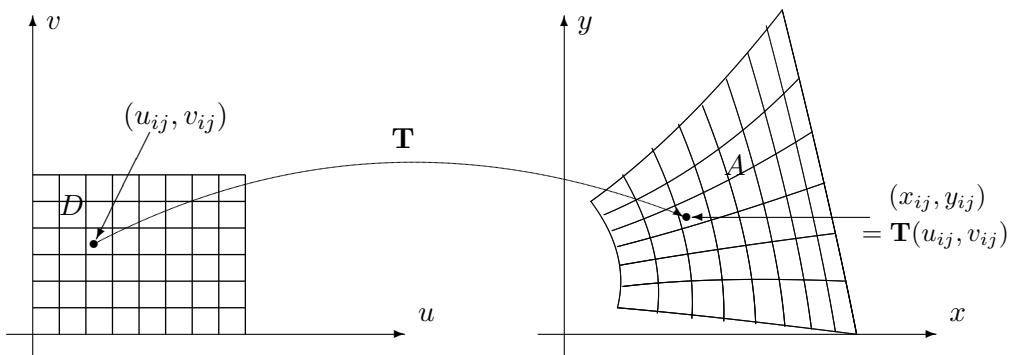
Vi ser at vi får to nye ingredienser i integralet etter variabelskiftet — et nytt intervall  $[g(a), g(b)]$  å integrere over og en ny faktor  $h'(u)$  i integranden. I denne seksjonen skal vi se at akkurat det samme skjer for dobbeltintegraler — vi kan skifte variabel i disse integralene også, men da får vi et nytt område å integrere over og en ny faktor i integranden.

I vanlige integraler skifter vi variabel for å få en enklere integrand. I dobbeltintegraler skifter vi vel så ofte variable for å få et enklere område å integrere over. Vi skal ta denne situasjonen som utgangspunkt for vår diskusjon. Vi tenker oss at vi skal regne ut integralet til en funksjon  $f$  over området  $A$  til *høyre* på figur 1, altså integralet  $\iint_A f(x, y) dx dy$ . Området  $A$  er ganske uregelmessig, og vi må bryte dobbeltintegralet opp i flere biter

for å få regnet det ut på vanlig måte. La oss nå tenke oss at det finnes en funksjon  $\mathbf{T} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  som avbilder et atskillig penere område  $D$  på  $A$  slik figur 1 viser. Kanskje vi kan gjøre om dobbeltintegralet  $\iint_A f(x, y) dx dy$  til et dobbeltintegral over det enklere området  $D$ ? Kanskje dette integralet har noe å gjøre med funksjonen  $g(u, v) = f(\mathbf{T}(u, v))$ ? (Funksjonen  $g$  er altså definert slik at verdien til  $g$  i punktet  $(u, v) \in D$  er lik verdien til  $f$  i det "tilsvarende" punktet  $(x, y) = \mathbf{T}(u, v)$  i  $A$ .)

Figur 1:  $\mathbf{T}$  avbilder  $D$  på  $A$ 

For å undersøke dette spørsmålet nærmere tenker vi oss at vi har delt opp  $D$  i et rutenett, og at vi har bruktt  $\mathbf{T}$  til å flytte dette rutenettet bort til  $A$  slik figur 2 viser.

Figur 2:  $\mathbf{T}$  flytter rutenettet fra  $D$  til  $A$ 

Vi plukker ut et punkt  $(u_{ij}, v_{ij})$  i den  $ij$ -te ruten  $D_{ij}$  i oppdelingen av  $D$ , og lar  $(x_{ij}, y_{ij}) = \mathbf{T}(u_{ij}, v_{ij})$  være det tilsvarende punktet i  $ij$ -te ruten  $A_{ij}$  i oppdelingen av  $A$ . Er oppdelingen fin nok, burde

$$\sum_{ij} f(x_{ij}, y_{ij}) |A_{ij}|$$

(der  $|A_{ij}|$  er arealet til  $A_{ij}$ ) være en god tilnærming til dobbeltintegralet

$\iint_A f(x, y) dx dy$ . Tilsvarende burde

$$\sum_{ij} g(u_{ij}, v_{ij}) |D_{ij}|$$

være en god tilnærming til dobbeltintegralet  $\iint_D g(x, y) dudv$ .

Hva er sammenhengen mellom disse to uttrykkene? Per definisjon av  $g$  er  $g(u_{ij}, v_{ij}) = f(x_{ij}, y_{ij})$ , så funksjonsverdiene i de to summene er de samme, men hva med arealene  $|A_{ij}|$  og  $|D_{ij}|$ ? I seksjon 3.8 observerte vi at tallverdien til Jacobi-determinanten

$$\det \mathbf{T}'(u, v) = \begin{vmatrix} \frac{\partial T_1}{\partial u}(u, v) & \frac{\partial T_1}{\partial v}(u, v) \\ \frac{\partial T_2}{\partial u}(u, v) & \frac{\partial T_2}{\partial v}(u, v) \end{vmatrix}$$

er en forstørrelsesfaktor for avbildningen  $\mathbf{T}$ , og derfor er

$$|A_{ij}| \approx |\det \mathbf{T}'(u_{ij}, v_{ij})| |D_{ij}|$$

Vi burde derfor ha

$$\sum_{ij} f(x_{ij}, y_{ij}) |A_{ij}| \approx \sum_{ij} g(u_{ij}, v_{ij}) |\det \mathbf{T}'(u_{ij}, v_{ij})| |D_{ij}|$$

med bedre og bedre tilnærming dess finere oppdelingen er. Summene i dette uttrykket er begge en slags Riemann-summer som burde nærme seg hvert sitt integral, og dermed sitter vi igjen med

$$\begin{aligned} \iint_A f(x, y) dx dy &= \iint_D g(u, v) |\det \mathbf{T}'(u, v)| dudv = \\ &= \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv \end{aligned}$$

der vi i siste overgang har brukt at  $g(u, v) = f(\mathbf{T}(u, v))$ .

Argumentet ovenfor er langt fra et matematisk bevis, men det er så pass overbevisende at resultatet burde være riktig under passende forutsetninger. Vi skal nå formulere et presist resultat som vi skal bevise mot slutten av seksjonen. Det er slett ikke det mest generelle resultatet man kan bevise, men det holder for våre anvendelser og er mulig å bevise med de redskapene vi har til disposisjon. Husk at en funksjon  $\mathbf{T}$  er *injektiv* dersom  $\mathbf{a} \neq \mathbf{b}$  medfører  $\mathbf{T}(\mathbf{a}) \neq \mathbf{T}(\mathbf{b})$ .

**Teorem 6.7.1 (Skifte av variable i dobbeltintegral)** La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^2$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at  $\det \mathbf{T}' \neq 0$  på hele  $U$ . Hvis  $D \subset U$  er en lukket, Jordan-målbar mengde, og  $f : \mathbf{T}(D) \rightarrow \mathbb{R}$  er en kontinuerlig funksjon, så er

$$\iint_A f(x, y) dx dy = \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv$$

der  $A = \mathbf{T}(D)$ .

Før vi ser på noen eksempler, skal vi innføre litt hendig notasjon. Dersom vi kaller komponentene til  $\mathbf{T}(u, v)$  for  $x(u, v)$  og  $y(u, v)$  slik at

$$\mathbf{T}(u, v) = (x(u, v), y(u, v)),$$

blir Jacobi-determinanten

$$\det \mathbf{T}'(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u}(u, v) & \frac{\partial x}{\partial v}(u, v) \\ \frac{\partial y}{\partial u}(u, v) & \frac{\partial y}{\partial v}(u, v) \end{vmatrix}$$

Det er vanlig å bruke notasjonen

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u}(u, v) & \frac{\partial x}{\partial v}(u, v) \\ \frac{\partial y}{\partial u}(u, v) & \frac{\partial y}{\partial v}(u, v) \end{vmatrix}$$

Formelen for skifte av variabel blir da

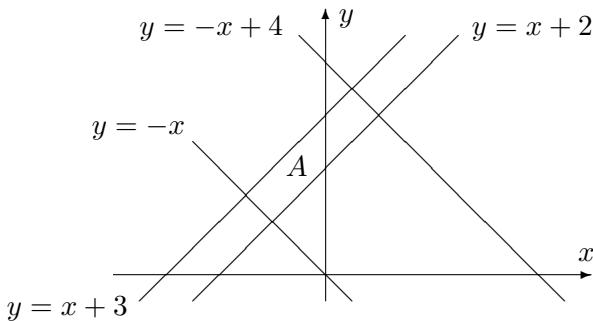
$$\iint_A f(x, y) dx dy = \iint_D f(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv$$

Det er instruktivt å sammenligne denne formelen med den formelen for skifte av variabel i vanlige integraler som vi startet denne seksjonen med, altså

$$\int_a^b f(x) dx = \int_{g(a)}^{g(b)} f(h(u)) h'(u) du$$

(likheten blir enda mer slående om vi skriver  $h'(u) = \frac{dx}{du}$ ).

**Eksempel 1:** Beregn integralet  $\iint_A xy dx dy$  der  $A$  består av området avgrenset av linjene  $y = -x$ ,  $y = -x + 4$ ,  $y = x + 2$ ,  $y = x + 3$ . Figur 3 viser området  $A$ .



Figur 3: Området  $A$

Legg merke til at  $A$  består av de punktene  $(x, y)$  som oppfyller ulikhettene

$$2 \leq y - x \leq 3 \quad \text{og} \quad 0 \leq y + x \leq 4$$

Ideen er å innføre nye variable  $u = y - x$ ,  $v = y + x$  slik at rendene til området blir parallelle med koordinataksene. Løser vi disse ligningene med hensyn på  $y$  og  $x$ , får vi

$$x = \frac{-u+v}{2}, \quad y = \frac{u+v}{2}$$

Avbildningene våre blir dermed  $x(u, v) = \frac{-u+v}{2}$  og  $y(u, v) = \frac{u+v}{2}$ . Lar vi  $D$  være rektanglet

$$D = \{(u, v) : 2 \leq u \leq 3 \text{ og } 0 \leq v \leq 4\}$$

ser vi at  $T(D) = A$ . Dermed har vi

$$\iint_A xy \, dx dy = \iint_D \left( \frac{-u+v}{2} \right) \left( \frac{u+v}{2} \right) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv$$

Vi regner ut Jacobi-determinanten:

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}$$

Dette gir oss

$$\begin{aligned} \iint_A xy \, dx dy &= \int_2^3 \left[ \int_0^4 \left( \frac{-u+v}{2} \right) \left( \frac{u+v}{2} \right) \cdot \frac{1}{2} dv \right] du \\ &= \frac{1}{8} \int_2^3 \left[ \int_0^4 (v^2 - u^2) dv \right] du = \frac{1}{8} \int_2^3 \left[ \frac{v^3}{3} - u^2 v \right]_{v=0}^{v=4} du \\ &= \frac{1}{8} \int_2^3 \left( \frac{64}{3} - 4u^2 \right) du = \frac{1}{8} \left[ \frac{64}{3}u - \frac{4}{3}u^3 \right]_2^3 \\ &= \frac{1}{8} \left[ \frac{64}{3} \cdot 3 - \frac{4}{3} \cdot 3^3 - \frac{64}{3} \cdot 2 + \frac{4}{3} \cdot 2^3 \right] = -\frac{1}{2} \end{aligned}$$



Før vi går løs på det neste eksemplet, viser vi frem et triks som av og til er nyttig. Siden funksjonen

$$(x, y) = \mathbf{T}(u, v)$$

er injektiv, har den en omvendt funksjon  $\mathbf{T}^{-1}$  som vi kan finne (i hvert fall i prinsippet!) ved å løse ligningen ovenfor for  $x$  og  $y$ :

$$(u, v) = \mathbf{T}^{-1}(x, y)$$

Jacobi-determinanten til denne funksjonen er

$$\det((\mathbf{T}^{-1})'(x, y)) = \frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}$$

Fra omvendt funksjonsteorem 5.6.2 vet vi at Jacobi-matrisen til  $\mathbf{T}^{-1}$  er den inverse til Jacobi-matrisen til  $\mathbf{T}$ . Det samme må gjelde for Jacobi-determinantene (siden  $\det(A^{-1}) = (\det A)^{-1}$  ifølge korollar 4.9.15), og dermed har vi

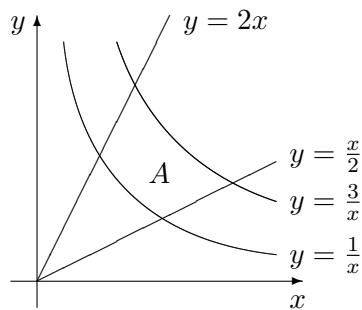
$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix}^{-1} = \left( \frac{\partial(u, v)}{\partial(x, y)} \right)^{-1}$$

Formelen for skifte av variable kan derfor også skrives:

$$\iint_A f(x, y) dx dy = \iint_D f(x(u, v), y(u, v)) \left| \frac{\partial(u, v)}{\partial(x, y)} \right|^{-1} du dv$$

I noen tilfeller er denne formen enklere å bruke enn den opprinnelige slik det neste eksemplet illustrerer.

**Eksempel 2:** Vi skal regne ut integralet  $\iint_A \frac{x}{y} dx dy$  der  $A$  er området i første kvadrant avgrenset av kurvene  $y = \frac{1}{x}$ ,  $y = \frac{3}{x}$  og linjene  $y = \frac{x}{2}$ ,  $y = 2x$ . Figur 4 viser området.



Figur 4: Området  $A$

For treningens skyld skal vi regne ut integralet på to måter; først skal vi bruke "standardmetoden", og så skal vi bruke metoden vi nettopp presenterte med å regne ut  $\frac{\partial(u, v)}{\partial(x, y)}$  istedenfor  $\frac{\partial(x, y)}{\partial(u, v)}$ . I begge tilfeller skal vi bruke det samme variabelskiftet, så la oss finne det først. Omformer vi litt på ligningene som beskriver området  $A$ , får vi betingelsene

$$1 \leq xy \leq 3 \quad \text{og} \quad \frac{1}{2} \leq \frac{y}{x} \leq 2$$

Vi innfører derfor nye variable

$$u = xy \quad \text{og} \quad v = \frac{y}{x}$$

og ser at betingelsene ovenfor nå blir til  $1 \leq u \leq 3$ ,  $\frac{1}{2} \leq v \leq 2$ . Det nye integrasjonsområdet vi skal integrere  $u$  og  $v$  over, er derfor rektanglet

$$D = [1, 3] \times [\frac{1}{2}, 2]$$

Metode 1: For å komme videre, løser vi ligningene  $u = xy$  og  $v = \frac{y}{x}$  for  $x$  og  $y$ . Fra den andre ligningen får vi  $y = vx$ , og setter vi dette inn i den første, ser vi at  $u = x(vx)$  som gir  $x = \sqrt{\frac{u}{v}}$  (husk at vi er i første kvadrant og derfor bare er interessert i positive  $x$ -verdier). Setter vi dette inn i ligningen  $y = vx$ , får vi  $y = \sqrt{uv}$ . Vi har altså

$$x = \sqrt{\frac{u}{v}} \quad \text{og} \quad y = \sqrt{uv}$$

Vi kan nå regne ut Jacobi-determinanten

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{1}{2\sqrt{uv}} & -\frac{\sqrt{u}}{2\sqrt{v^3}} \\ \frac{\sqrt{v}}{2\sqrt{u}} & \frac{\sqrt{u}}{2\sqrt{v}} \end{vmatrix} = \frac{1}{4v} + \frac{1}{4v} = \frac{1}{2v}$$

Dermed er vi klare til å skrive opp det nye integralet:

$$\iint_A \frac{x}{y} dx dy = \iint_D \frac{\sqrt{\frac{u}{v}}}{\sqrt{uv}} \frac{1}{2v} dudv = \iint_D \frac{1}{2v^2} dudv$$

Resten er enkelt:

$$\begin{aligned} \iint_D \frac{1}{2v^2} dudv &= \int_1^3 \left[ \int_{\frac{1}{2}}^2 \frac{1}{2v^2} du \right] dv = \int_1^3 \frac{3}{2} \cdot \frac{1}{2v^2} dv = \\ &= \int_1^3 \frac{3}{4v^2} dv = \left[ -\frac{3}{4v} \right]_1^3 = -\frac{1}{4} + \frac{3}{4} = \frac{1}{2} \end{aligned}$$

Metode 2: Når vi bruker denne metoden, skal regne ut Jacobi-determinanten  $\frac{\partial(u, v)}{\partial(x, y)}$ , så vi kan derivere med en gang uten å løse for  $x$  og  $y$  først:

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} y & x \\ -\frac{y}{x^2} & \frac{1}{x} \end{vmatrix} = \frac{y}{x} + \frac{y}{x} = \frac{2y}{x} = 2v$$

Dette gir  $\left| \frac{\partial(u, v)}{\partial(x, y)} \right|^{-1} = \frac{1}{2v}$ . Observerer vi at  $\frac{x}{y} = \frac{1}{v}$ , kan vi nå sette rett inn i integralet:

$$\iint_A \frac{x}{y} dx dy = \iint_D \frac{1}{v} \frac{1}{2v} dudv = \iint_D \frac{1}{2v^2} dudv$$

Resten av regningen er som ovenfor.

Vi ser at i dette eksemplet gir metode 2 kortere og enklere utregninger enn metode 1. Det er flere grunner til dette. For det første er det lettere å derivere  $u$  og  $v$  med hensyn på  $x$  og  $y$  enn omvendt, og for det andre er det enkelt å uttrykke integranden og Jacobi-determinanten  $\frac{\partial(u,v)}{\partial(x,y)}$  ved hjelp av  $u$  og  $v$  uten først å måtte finne  $x$  og  $y$  uttrykt ved  $u$  og  $v$ . Er ikke disse betingelsene oppfylt, er det ofte vel så raskt å løse integralet ved hjelp av metode 1. ♣

Du har kanskje lurt på sammenhengen mellom integrasjon i polarkoordinater, som vi behandlet i seksjon 6.3, og teorien ovenfor. Den er svært tett — vi kan tenke på integrasjon i polarkoordinater som resultatet av et variabelskifte der vi innfører nye variable  $r$  og  $\theta$  slik at

$$x = r \cos \theta \quad \text{og} \quad y = r \sin \theta$$

Dette gir oss Jacobi-determinanten

$$\frac{\partial(x, y)}{\partial(r, \theta)} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cos^2 \theta + r \sin^2 \theta = r$$

Den ekstra faktoren  $r$  i polarintegrasjonen er altså Jacobi-determinanten til variabelskiftet. Formelen for skifte av variable i dobbeltintegraler blir nå til den vanlige formelen for integrasjon i polarkoordinater:

$$\iint_A f(x, y) dx dy = \iint_D f(r \cos \theta, r \sin \theta) r dr d\theta$$

der  $D$  er området  $A$  beskrevet i polarkoordinater.

**Bemerkning:** Ser du nøye etter, vil du se at vi egentlig bare har lov til å bruke teorem 6.7.1 på polarkoordinater dersom origo *ikke* er med i integrasjonsområdet — betingelsen om at det  $\mathbf{T}' \neq 0$  er nemlig ikke oppfylt i origo. Det er imidlertid ingen problemer med å bruke polarkoordinater på områder som inneholder 0. Vil man begrunne dette, kan man først fjerne en liten sirkelskive  $B(\mathbf{0}, \epsilon)$  fra integrasjonsområdet og så se på grenseverdien når  $\epsilon$  går mot 0.

Vi tar med et eksempel der det er naturlig å bruke polarkoordinater.

**Eksempel 3:** Vi skal finne volumet til en kule med radius  $a$ . Ligningen for en kuleflate med radius  $a$  og sentrum i origo er

$$x^2 + y^2 + z^2 = a^2$$

Løser vi for  $z$ , får vi

$$z = \pm \sqrt{a^2 - x^2 - y^2}.$$

Halve volumet ligger dermed under funksjonsgrafen  $f(x, y) = \sqrt{a^2 - x^2 - y^2}$  og over sirkelen

$$S\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq a^2\}$$

Dermed er

$$V = 2 \iint_S \sqrt{a^2 - x^2 - y^2} dx dy.$$

Skifter vi til polarkoordinater, får vi

$$V = 2 \int_0^a \left[ \int_0^{2\pi} \sqrt{a^2 - r^2} r d\theta \right] dr$$

(husk Jacobi-determinanten er  $r$ ). Dette gir

$$V = 4\pi \int_0^a \sqrt{a^2 - r^2} r dr$$

For å løse dette integralet innfører vi en ny variabel  $u = a^2 - r^2$ . Da er  $du = -2r dr$ , og vi får

$$\int \sqrt{a^2 - r^2} r dr = -\frac{1}{2} \int \sqrt{u} du = -\frac{1}{3} u^{3/2} + c = \frac{1}{3} (a^2 - r^2)^{3/2} + c$$

Dermed er

$$V = 4\pi \left[ -\frac{1}{3} (a^2 - r^2)^{3/2} \right]_0^a = \frac{4}{3} \pi a^3$$



Bruk av polarkoordinater er så vanlig at du fritt kan bruke formelen  $\frac{\partial(x,y)}{\partial(r,\theta)} = r$  uten å måtte begrunne den. Vær forvrig oppmerksom på at det ofte kan være lurt å bruke polarkoordinater med et annet sentrum enn origo. Ønsker du f.eks. å integrere over en sirkel med sentrum i  $(a, b)$ , er det som regel lurt å innføre polarkoordinater med dette punktet som sentrum. Det betyr at du setter

$$x = a + r \cos \theta \quad \text{og} \quad y = b + r \sin \theta$$

Også i dette tilfellet blir Jacobi-determinanten  $\frac{\partial(x,y)}{\partial(r,\theta)} = r$ .

### \*Bevis for skifte av variabel i dobbeltintegraler

Vi skal nå bevise teorem 6.7.1. Beviset er langt og vanskelig, og vi skal stykke det opp i flere etapper. Først ser vi på hvordan affinavbildninger transformerer arealer. (Dersom du synes det er noe kjent ved dette, skyldes det nok at lemmaet nedenfor bare er en presisering av setning 2.9.3.)

**Lemma 6.7.2** *Anta at  $D \subset \mathbb{R}^2$  er en begrenset, Jordan-målbar mengde med areal  $|D|$ . Dersom  $\mathbf{F}(\mathbf{x}) = B\mathbf{x} + \mathbf{b}$  er en affinavbildning fra  $\mathbb{R}^2$  til  $\mathbb{R}^2$  med matrise  $B$ , så er bildet*

$$A = \mathbf{F}(D) = \{\mathbf{F}(\mathbf{x}) \mid \mathbf{x} \in D\}$$

*Jordan-målbart med areal  $|A| = |\det(B)| |D|$ . Tallverdien til determinanten er altså forstørrelsesfaktoren til affinavbildningen.*

*Bevis:* Fra seksjon 2.9 vet vi at formelen gjelder når  $D$  er et rektangel (strengt tatt viste vi den bare når  $D$  er et kvadrat, men beviset for rektangler er akkurat det samme). For å vise at den gjelder for en generell, begrenset, Jordan-målbar mengde  $D$ , må vi tilnærme  $D$  ved hjelp av rektangler. Her er detaljene:

Siden  $D$  er Jordan-målbar, er funksjonen  $\mathbf{1}_D$  integrerbar, og vi kan for enhver  $\epsilon > 0$  finne en partisjon  $\Pi$  av et rektangel  $R$  som innholder  $D$  slik at

$$\mathcal{O}(\Pi) - |D| < \frac{\epsilon}{2|\det B|}$$

og

$$|D| - N(\Pi) < \frac{\epsilon}{2|\det B|}$$

der  $\mathcal{O}(\Pi)$  er den øvre trappesummen til funksjonen  $\mathbf{1}_D$ , og  $N(\Pi)$  er den nedre trappesummen. Siden

$$\mathcal{O}(\Pi) = \sum_{\substack{\text{rektanglene } R_{ij} \\ \text{slik at } R_{ij} \cap D \neq \emptyset}} |R_{ij}|$$

og

$$N(\Pi) = \sum_{\substack{\text{rektanglene } R_{ij} \\ \text{slik at } R_{ij} \subset D}} |R_{ij}|,$$

vil det samlede bildet av de første rektanglene (de som inngår i summen til  $\mathcal{O}(\Pi)$ ) under  $\mathbf{F}$  gi en overdekning av  $\mathbf{F}(D)$  med samlet areal  $|\det(B)|\mathcal{O}(\Pi) < |\det B||D| + \frac{\epsilon}{2}$ , mens det samlede bildet av rektanglene i den andre summen, vil gi en mengde som ligger inni  $\mathbf{F}(D)$  med areal  $|\det(B)|N(\Pi) > |\det B||D| - \frac{\epsilon}{2}$ . Bildet  $\mathbf{F}(D)$  ligger dermed klemt mellom to mengder der den største har areal mindre enn  $|\det B||D| + \frac{\epsilon}{2}$ , og den minste har areal større enn  $|\det B||D| - \frac{\epsilon}{2}$ . Det følger at  $A = \mathbf{F}(D)$  er Jordan-målbar med areal  $|A| = |\det B||D|$ .  $\square$

**Bemerkning:** Er man virkelig kritisk, vil man si at vi har jukset litt i beviset ovenfor — vi har nemlig brukt to arealbegreper om hverandre uten å sjekke at de faktisk er sammenfallende. Det ene arealbegrepet er det uformelle som vi har med oss fra skolen, og som forteller oss hvordan vi kan

regne ut arealet til trekantene og parallelogrammer og lignende, mens det andre arealbegrepet er det vi har innført i dette kapitlet, og som bygger på dobbeltintegraler. Det er ikke vanskelig å vise at de to begrepene stemmer overens, men det tar litt tid, og vi overlater utfordringen til leserne!

Før vi for alvor går løs på beviset for teoremet, trenger vi noen tekniske hjelpeemidler. Den neste setningen er et generelt resultat om lukkede mengder som også er nyttig i mange sammenhenger.

**Setning 6.7.3** *Anta at  $B$  og  $C$  er to disjunkte (dvs. at  $B \cap C = \emptyset$ ), lukkede mengder i  $\mathbb{R}^n$ , og at minst én av dem er begrenset. Da finnes det en  $\delta > 0$  slik at  $|\mathbf{b} - \mathbf{c}| \geq \delta$  for alle  $\mathbf{b} \in B$  og  $\mathbf{c} \in C$ .*

*Bevis:* La oss si at det er  $B$  som er begrenset. Dersom resultatet ikke holder, kan vi for hver  $n \in \mathbb{N}$  finne et par av elementer  $\mathbf{b}_n \in B$  og  $\mathbf{c}_n \in C$  slik at  $|\mathbf{b}_n - \mathbf{c}_n| < \frac{1}{n}$ . Siden  $B$  er lukket og begrenset, har  $\{\mathbf{b}_n\}$  en delfølge  $\{\mathbf{b}_{n_k}\}$  som konvergerer mot et punkt  $\mathbf{b} \in B$ . Siden  $|\mathbf{b}_{n_k} - \mathbf{c}_{n_k}| \rightarrow 0$ , vil også  $\mathbf{c}_{n_k} \rightarrow \mathbf{b}$ . Siden  $C$  er lukket, må da  $\mathbf{b} \in C$ , og det er umulig siden  $B \cap C = \emptyset$ .  $\square$

I argumentene nedenfor er det ofte lurt å måle størrelsen til en vektor på en litt annen måte enn vanlig. Dersom  $\mathbf{x} = (x_1, x_2)$  er en vektor i  $\mathbb{R}^2$ , lar vi  $\|\mathbf{x}\| = \max\{|x_1|, |x_2|\}$ . Et kvadrat med sentrum i  $\mathbf{x}$  og sider  $\epsilon$  er nå gitt ved

$$K(\mathbf{x}, \epsilon) = \{\mathbf{y} \in \mathbb{R}^2 \mid \|\mathbf{y} - \mathbf{x}\| < \frac{\epsilon}{2}\}$$

Siden vi mäter vektorer på en litt uvanlig måte, er det naturlig också å mäle storrelsen til matriser på en annen måte enn vanlig. Dersom  $A$  er matrisen  $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ , lar vi

$$\|A\| = \max\{|a_{11}| + |a_{12}|, |a_{21}| + |a_{22}|\}$$

(OBS: Selv om vi bruker den samme notasjonen, er dette ikke den samme matrise-normen som vi brukte i setning 1.6.3). Her er et enkelt lemma om sammenhengen mellom disse normene:

**Lemma 6.7.4** *For alle  $\mathbf{x} \in \mathbb{R}^2$  og alle  $2 \times 2$ -matriser  $A$  har vi*

$$\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\|$$

*Bevis:* Siden

$$A\mathbf{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix},$$

er

$$\|A\mathbf{x}\| = \max\{|a_{11}x_1 + a_{12}x_2|, |a_{21}x_1 + a_{22}x_2|\}$$

Videre er

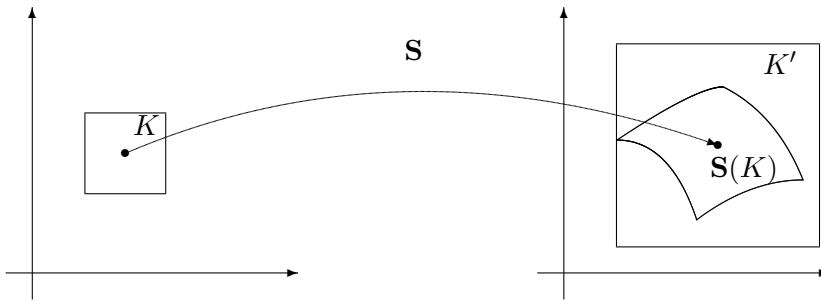
$$|a_{11}x_1 + a_{12}x_2| \leq |a_{11}| |x_1| + |a_{12}| |x_2| \leq (|a_{11}| + |a_{12}|) \max\{|x_1|, |x_2|\} \leq \|A\| \|\mathbf{x}\|$$

og

$$|a_{21}x_1 + a_{22}x_2| \leq |a_{21}| |x_1| + |a_{22}| |x_2| \leq (|a_{21}| + |a_{22}|) \max\{|x_1|, |x_2|\} \leq \|A\| \|\mathbf{x}\|,$$

og følgelig er  $\|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|$   $\square$

Grunnen til at vi innfører de nye normene, er at de egner seg utmerket til å beskrive hva som skjer med kvadrater når de transformeres ved hjelp av en funksjon. Figur 5 viser problemstillingen — vi er interessert i størrelsen til det minste kvadratet  $K'$  (med sider parallelle med aksene) vi kan putte bildet av det opprinnelige kvadratet  $K$  inn i.



Figur 5: Avbildning av kvadrater

**Setning 6.7.5** La  $K \subset \mathbb{R}^2$  være et kvadrat med areal  $|K|$  og sentrum  $\mathbf{c}$ . Anta at  $\mathbf{S} : K \rightarrow \mathbb{R}^2$  er en deriverbar avbildning slik at  $\|\mathbf{S}'(x, y)\| \leq C$  for alle punkter  $(x, y)$  i det indre av  $K$ . Da er  $\mathbf{S}(K)$  inneholdt i et kvadrat med sentrum i  $\mathbf{S}(\mathbf{c})$  og areal  $C^2|K|$

*Beweis:* La  $\mathbf{c} = (x_0, y_0)$  og la  $(x, y)$  være et annet punkt i  $K$ . Ved middelverdisetningen for funksjoner av flere variable (se setning 3.2.3) finnes det et punkt  $\mathbf{d}$  i det indre av  $K$  slik at

$$|\mathbf{S}_1(x, y) - \mathbf{S}_1(x_0, y_0)| = |\nabla \mathbf{S}_1(\mathbf{d}) \cdot (x - x_0, y - y_0)|$$

Hvis  $s$  er halvparten av sidekanten i det opprinnelige integralet, er  $|x - x_0| \leq s$  og  $|y - y_0| \leq s$ . Dermed er

$$|\mathbf{S}_1(x, y) - \mathbf{S}_1(x_0, y_0)| = \left( \left| \frac{\partial \mathbf{S}_1}{\partial x}(\mathbf{d}) \right| + \left| \frac{\partial \mathbf{S}_1}{\partial y}(\mathbf{d}) \right| \right) s \leq Cs$$

Tilsvarende resonnement for  $S_2$  gir  $|\mathbf{S}_2(x, y) - \mathbf{S}_2(x_0, y_0)| \leq Cs$ , og resultatet følger.  $\square$

Vi er nå ferdige med de første forberedelsene. De neste to resultatene gir oss viktig informasjon om målbarhet. Husk at tillukningen  $\overline{N}$  til en mengde  $N \subset \mathbb{R}^2$  består av  $N$  pluss randen til  $N$  — med andre ord  $\overline{N} = N \cup \partial N$ .

**Setning 6.7.6** *La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^2$  og anta at  $N$  er en mengde med innhold 0 slik at tillukningen til  $N$  er inneholdt i  $U$ . Dersom  $\mathbf{S} : U \rightarrow \mathbb{R}^2$  er en funksjon med kontinuerlige partiellderiverte, så har  $\mathbf{S}(N)$  innhold null.*

*Bevis:* Siden  $U$  er åpen, er mengden

$$U^c = \{\mathbf{a} \in \mathbb{R}^2 \mid \mathbf{a} \notin U\}$$

(den kalles *komplementet* til  $U$ ) lukket. Ifølge setning 6.7.3 finnes det dermed en  $\delta > 0$  slik at avstanden fra ethvert punkt i  $\overline{N}$  til det nærmeste punktet i  $U^c$  er minst  $\delta$ . La  $K$  være tillukningen til mengden

$$\{\mathbf{x} \in \mathbb{R}^2 : \text{det finnes } \mathbf{a} \in N \text{ slik at } |\mathbf{x} - \mathbf{a}| \leq \delta/2\}$$

Da er  $K$  en lukket og begrenset delmengde av  $U$ , og ifølge ekstremalverdi-setningen finnes et tall  $M$  slik at tallverdien til de partiellderiverte til  $\mathbf{S}$  er mindre enn  $\frac{M}{2}$  på mengden  $K$ . Dermed er  $\|\mathbf{S}'(\mathbf{x})\| \leq M$  for alle  $\mathbf{x} \in K$ .

Gitt en  $\epsilon > 0$ , kan vi finne en overdekning av  $N$  med samlet areal mindre enn  $\frac{\epsilon}{M^2}$  slik at alle rektanglene i overdekningen ligger inni  $K$ . Vi kan anta at alle rektanglene i denne overdekningen er kvadrater. (For å se dette velger du først en overdekning som består av rektangler, så utvider du hvert rektangel ørlite slik at alle hjørner har rasjonale koordinater, men det samlede arealet fortsatt er mindre enn  $\frac{\epsilon}{M^2}$ , og så observerer du at alle rektangler med rasjonale hjørner kan deles opp i et endelig antall kvadrater — det siste har noe med fellesnevner å gjøre!). Ser vi på bildene av alle kvadratene i overdekningen, forteller forrige setning oss at hvert av dem ligger inni et nytt kvadrat med areal maksimalt  $M^2$  ganger arealet til det opprinnelige kvadratet. De nye kvadratene overdekker dermed  $\mathbf{S}(N)$  og har et samlet areal mindre enn  $M^2 \frac{\epsilon^2}{M^2} = \epsilon$ . Følgelig har  $\mathbf{S}(N)$  innhold 0, og setningen er bevist.  $\square$

Det neste resultatet viser at injektive funksjoner med ikke-null Jacobideterminant bevarer målbarhet.

**Setning 6.7.7** *La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^2$ , og anta at  $D$  er en Jordan-målbar mengde med tillukning inneholdt i  $U$ . Anta at  $\mathbf{S} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at det  $\mathbf{S}'(\mathbf{x}) \neq 0$  for alle  $\mathbf{x} \in U$ . Da er  $\mathbf{S}(D)$  Jordan-målbar.*

*Bevis:* Ifølge teorem 6.6.3 er en mengde Jordan-målbar hvis og bare hvis randen har innhold 0. Dette betyr at randen  $N$  til  $D$  har innhold 0, og

ifølge forrige setning har dermed  $\mathbf{S}(N)$  innhold 0. Dersom vi kan vise at  $\partial\mathbf{S}(D) \subset \mathbf{S}(N)$ , så må  $\mathbf{S}(D)$  være målbar og teoremet følger (de to mengdene  $\partial\mathbf{S}(D)$  og  $\mathbf{S}(N)$  er faktisk like, men det har vi ikke bruk for).

Anta  $\mathbf{y} \in \partial\mathbf{S}(D)$ . Siden det  $\mathbf{S} \neq 0$  på  $U$ , forteller omvendt funksjons-teorem 5.6.2 oss at  $\mathbf{S}$  har en kontinuerlig omvendt funksjon  $\mathbf{S}^{-1} : \mathbf{S}(U) \rightarrow U$  som er definert i en omegn om  $\mathbf{y}$ . Siden  $\mathbf{y} \in \partial\mathbf{S}(D)$ , kan vi nå finne to følger fra  $\mathbf{S}(U)$  som konvergerer mot  $\mathbf{x}$  — den ene,  $\{\mathbf{y}_n\}$ , består av punkter som ligger i  $\mathbf{S}(D)$ , den andre,  $\{\mathbf{z}_n\}$ , består av punkter som ikke er med i  $\mathbf{S}(D)$ . Siden  $\mathbf{S}^{-1}$  er kontinuerlig, vil  $\mathbf{S}^{-1}(\mathbf{y}_n)$  og  $\mathbf{S}^{-1}(\mathbf{z}_n)$  begge konvergere mot  $\mathbf{x} = \mathbf{S}^{-1}(\mathbf{y})$ . Den første av disse følgene består av punkter som er med i  $D$ , den andre av punkter som ikke er med i  $D$ . Følgelig er  $\mathbf{x} \in N = \partial D$ . Siden  $\mathbf{y} = \mathbf{S}(\mathbf{x})$ , er dermed  $\mathbf{y} \in \mathbf{S}(N)$ , og setningen er bevist.  $\square$

Vi kan nå for alvor tenke på å bevise teoremet. Hovedideen i beviset er at hvis  $K$  er et lite kvadrat med sentrum  $\mathbf{a}$ , så vil bildet  $\mathbf{T}(K)$  ha areal tilnærmet lik  $|\det \mathbf{T}'(\mathbf{a})| |K|$ . For å gjennomføre beviset trenger vi å vite hvor god tilnærmelsen er. Dette skal vi gjøre på følgende indirekte måte. Vi bruker først den inverse lineæravbildningen  $B = \mathbf{T}'(\mathbf{a})^{-1}$  til å avbilde  $\mathbf{T}(K)$  tilbake til utgangspunktet — dvs. vi ser på mengden  $B\mathbf{T}(K)$ . Siden  $B$  og  $\mathbf{T}$  nesten er inverser til hverandre i det området vi ser på, er  $B\mathbf{T}(K)$  og  $K$  svært like, og vi skal bruke setning 6.7.6 til å vise at  $B\mathbf{T}(K)$  kan passes inn i et kvadrat  $K'$  som bare har litt større areal enn  $K$ . Dermed er  $\mathbf{T}'(\mathbf{a})(K')$  en mengde som inneholder  $\mathbf{T}(K)$ , og som har et areal som bare er ørlite grann større enn  $|\det \mathbf{T}'(\mathbf{a})| |K|$ . Det neste lemmaet oppsummerer de regnetekniske delene av dette resonnementet:

**Lemma 6.7.8** *La  $U$  være en åpen mengde i  $\mathbb{R}^2$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at  $\det \mathbf{T}' \neq 0$  på hele  $U$ . Dersom  $K$  er et lukket kvadrat inni  $U$ , og  $B : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  er en inverterbar matrise, så er*

$$|\mathbf{T}(K)| \leq |\det B|^{-1} \left( \sup_{\mathbf{u} \in K} \|B\mathbf{T}'(\mathbf{u})\| \right)^2 |K|$$

*Bevis:* La  $\mathbf{S}$  være den sammensatte funksjonen  $\mathbf{S}(\mathbf{u}) = B\mathbf{T}(\mathbf{u})$ . Da er  $\mathbf{S}'(\mathbf{u}) = B\mathbf{T}'(\mathbf{u})$  og

$$C = \sup_{\mathbf{u} \in K} \|\mathbf{S}'(\mathbf{u})\| = \sup_{\mathbf{u} \in K} \|B\mathbf{T}'(\mathbf{u})\|$$

eksisterer siden  $K$  er lukket og begrenset, og de partiellderiverte til  $\mathbf{T}$  er kontinuerlige. Ifølge setning 6.7.5 er  $\mathbf{S}(K)$  inneholdt i et kvadrat med areal mindre enn  $C^2 |K|$  og må derfor selv ha areal mindre enn  $C^2 |K|$  (vi vet fra forrige setning at  $\mathbf{S}(K)$  er Jordan-målbar og dermed har et areal). Siden  $\mathbf{T}(K) = B^{-1}(\mathbf{S}(K))$ , følger det fra lemma 6.7.2 at

$$|\mathbf{T}(K)| = |\det B|^{-1} |\mathbf{S}(K)| \leq |\det B|^{-1} C^2 |K|$$

og dermed er lemmaet bevist.  $\square$

Vi trenger også et lemma som viser at vi kan få konstanten

$$C = \sup_{\mathbf{u} \in K} \|B\mathbf{T}'(\mathbf{u})\|$$

ovenfor så nær 1 vi måtte ønske dersom  $K$  er et tilstrekkelig lite kvadrat, og  $B$  er  $\mathbf{T}'(\mathbf{a})^{-1}$  der  $\mathbf{a}$  er sentrum i  $K$ :

**Lemma 6.7.9** *La  $U$  være en åpen mengde i  $\mathbb{R}^2$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at det  $\mathbf{T}' \neq 0$  på hele  $U$ . Anta at  $R$  er en lukket, begrenset delmengde av  $U$ . For enhver  $\epsilon > 0$  finnes det da en  $\delta > 0$  slik at*

$$|\|\mathbf{T}'(\mathbf{v})^{-1}\mathbf{T}'(\mathbf{u})\| - 1| < \epsilon$$

og

$$|\det \mathbf{T}'(\mathbf{v})|^{-1} |\det \mathbf{T}'(\mathbf{u})| - 1| < \epsilon$$

for alle  $\mathbf{u}, \mathbf{v} \in R$  slik at  $|\mathbf{u} - \mathbf{v}| \leq \delta$ .

*Bevis:* Siden  $\mathbf{T}'$  er kontinuerlig på  $R$ , er  $(\mathbf{T}')^{-1}$  det også, og det følger at funksjonene

$$f(\mathbf{u}, \mathbf{v}) = \|\mathbf{T}'(\mathbf{v})^{-1}\mathbf{T}'(\mathbf{u})\|$$

og

$$g(\mathbf{u}, \mathbf{v}) = |\det \mathbf{T}'(\mathbf{v})|^{-1} |\det \mathbf{T}'(\mathbf{u})|$$

er kontinuerlige funksjoner på mengden

$$R \times R = \{(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^4 \mid \mathbf{u}, \mathbf{v} \in R\}$$

Siden  $R \times R$  er lukket og begrenset, er  $f$  og  $g$  uniformt kontinuerlige ifølge setning 6.1.4 (alias 5.3.2). Gitt  $\epsilon > 0$ , finnes det derfor en  $\delta > 0$  slik at

$$|f(\mathbf{u}, \mathbf{v}) - f(\mathbf{u}', \mathbf{v}')| < \epsilon \quad \text{og} \quad |g(\mathbf{u}, \mathbf{v}) - g(\mathbf{u}', \mathbf{v}')| < \epsilon$$

når  $|(\mathbf{u}, \mathbf{v}) - (\mathbf{u}', \mathbf{v}')| < \delta$ . Velger vi  $\mathbf{u}' = \mathbf{v}' = \mathbf{u}$ , følger lemmaet.  $\square$

Vi har nå kommet til selve kjernen i argumentet vårt:

**Lemma 6.7.10** *La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^2$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at det  $\mathbf{T}' \neq 0$  på hele  $U$ . Hvis  $D \subset U$  er en lukket, Jordan-målbar mengde, så er*

$$|\mathbf{T}(D)| \leq \iint_D |\det \mathbf{T}'(u, v)| \, du \, dv$$

(de to uttrykkene er faktisk like, men vi nøyer oss med ulikheten foreløpig).

*Bevis:* Velg et kvadrat  $R$  som inneholder tillukningen til  $D$ . For ethvert tall  $n$ , lar vi  $\Pi^{(n)}$  være den partisjonen av  $R$  som vi får ved å dele sidene til  $R$  inn i  $2^n$  like store deler. Velg  $N_0$  så stor at de delkvadratene  $R_{ij}$  i  $\Pi^{(N_0)}$  som overlapper med  $D$ , ligger inni  $U$  (dette er mulig ifølge setning 6.7.3). La  $K_n$  være unionen av alle (de lukkede) delkvadratene i  $\Pi^n$  som overlapper med  $D$ . Da er  $K_n$  en lukket, begrenset mengde, og dersom  $n \geq N_0$ , så ligger  $K_n$  inni  $K_{N_0}$ . Observer at siden  $D$  er Jordan-målbar, er  $\lim_{n \rightarrow \infty} |K_n| = |D|$  (tenk gjennom dette — det er ikke helt opplagt!).

Gitt en  $\epsilon > 0$ , velger vi  $n$  så stor at  $|K_n| \leq |D| + \epsilon$ . Ved eventuelt å velge  $n$  enda større, kan vi ifølge forrige lemma anta at hvis  $\mathbf{u}$  og  $\mathbf{v}$  ligger i samme delkvadrat i  $K_n$ , så er

$$|\|\mathbf{T}'(\mathbf{v})^{-1}\mathbf{T}'(\mathbf{u})\| - 1| < \epsilon$$

og

$$|\det \mathbf{T}'(\mathbf{v})|^{-1} |\det \mathbf{T}'(\mathbf{u})| - 1| < \epsilon$$

Anta at  $S_1, S_2, \dots, S_m$  er delkvadratene som utgjør  $K_n$  og la  $\mathbf{a}_i$  være sentrum i  $S_i$ . Bruker vi lemma 6.7.8 med  $B = \mathbf{T}(\mathbf{a}_i)^{-1}$ , får vi

$$|\mathbf{T}(S_i)| \leq |\det \mathbf{T}'(\mathbf{a}_i)|(1 + \epsilon)^2 |S_i|$$

Summerer vi opp, har vi

$$|\mathbf{T}(D)| \leq \sum_{i=1}^m |\mathbf{T}(S_i)| \leq (1 + \epsilon)^2 \sum_{i=1}^m |\det \mathbf{T}'(\mathbf{a}_i)| |S_i|$$

Fra lemma 6.7.9 ser vi også at for alle  $\mathbf{u} \in S_i$ , er

$$|\det \mathbf{T}'(\mathbf{u})| \geq (1 - \epsilon) |\det \mathbf{T}'(\mathbf{a}_i)|$$

Dermed er

$$|\det \mathbf{T}'(\mathbf{a}_i)| |S_i| = \iint_{S_i} |\det \mathbf{T}'(\mathbf{a}_i)| \, dudv \leq \frac{1}{1 - \epsilon} \iint_{S_i} |\det \mathbf{T}'(u, v)| \, dudv$$

og kombinerer vi dette med resultatet ovenfor, får vi

$$|\mathbf{T}(D)| \leq \frac{(1 + \epsilon)^2}{1 - \epsilon} \iint_{K_n} |\det \mathbf{T}'(u, v)| \, dudv$$

(husk at  $K_n = S_1 \cup S_2 \cup \dots \cup S_m$ ). Dersom  $C$  er supremum til  $|\det \mathbf{T}'(u, v)|$  på  $K_{N_0}$  (dette er en kontinuerlig funksjon over en lukket, begrenset mengde, så et supremum finnes), er

$$\iint_{K_n} |\det \mathbf{T}'(u, v)| \, dudv \leq \iint_D |\det \mathbf{T}'(u, v)| \, dudv + C\epsilon$$

(husk at vi har valgt  $n$  så stor at  $|K_n| \leq |D| + \epsilon$ ). Totalt har vi dermed

$$|\mathbf{T}(D)| \leq \frac{(1+\epsilon)^2}{1-\epsilon} \iint_D |\det \mathbf{T}'(u, v)| dudv + \frac{C\epsilon(1+\epsilon)^2}{1-\epsilon}$$

Siden  $\epsilon > 0$  er vilkårlig, kan vi få uttrykket på høyre side av ulikheten så nær  $\iint_D |\det \mathbf{T}'(u, v)| dudv$  vi måtte ønske ved å velge  $\epsilon$  liten nok. Det betyr at

$$|\mathbf{T}(D)| \leq \iint_D |\det \mathbf{T}'(u, v)| dudv$$

og beviset er fullført.  $\square$

Vi er nå gjennom det verste, og resten av beviset for teoremet består bare i å få bitene på plass. Først utvider vi lemmaet ovenfor fra mengder til positive funksjoner.

**Lemma 6.7.11** *La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^2$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at  $\det \mathbf{T}' \neq 0$  på hele  $U$ . Hvis  $D \subset U$  er en lukket, Jordan-målbar mengde, og  $f : T(D) \rightarrow \mathbb{R}$  er en ikke-negativ, kontinuerlig funksjon, så er*

$$\iint_{\mathbf{T}(D)} f(x, y) dx dy \leq \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv$$

*Bevis:* Fra teorem 6.6.3, teorem 6.6.6 og setning 6.7.6 vet vi at de to funksjonene er integrerbare over henholdsvis  $\mathbf{T}(D)$  og  $D$ . La  $\Pi$  være en partisjon av et rektangel som inneholder  $T(D)$ , og la

$$N(\Pi) = \sum_{ij} m_{ij} |R_{ij}|$$

være den nedre trappesummen til  $f_{T(D)}$ . Legg merke til at det bare er de kvadratene som ligger inni  $D$  som bidrar til summen. Lar vi

$$S_{ij} = \mathbf{T}^{-1}(R_{ij}),$$

vet vi fra forrige lemma at  $|R_{ij}| \leq \iint_{S_{ij}} |\det \mathbf{T}'(u, v)| dudv$ . Dermed er

$$\begin{aligned} N(\Pi) &\leq \sum_{R_{ij} \subset D} m_{ij} \iint_{S_{ij}} |\det \mathbf{T}'(u, v)| dudv \leq \\ &\leq \iint_{\bigcup S_{ij}} f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv \leq \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv \end{aligned}$$

Siden vi kan få  $N(\Pi)$  så nær  $\iint_{\mathbf{T}(D)} f(x, y) dx dy$  vi måtte ønske, så er

$$\iint_{\mathbf{T}(D)} f(x, y) dx dy \leq \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv$$

og lemmaet er bevist.  $\square$

Foreløpig har vi bare bevist teoremet vårt for positive funksjoner og med ulikhet istedenfor likhet. Det kan se ut som vi fortsatt har en lang vei å gå, men det viser seg at vi bare er et triks fra målet. La oss skrive opp teoremet på nytt før vi beviser det:

**Teorem 6.7.12 (Skifte av variable i dobbeltintegral)** *La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^2$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^2$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at det  $\mathbf{T}' \neq 0$  på hele  $U$ . Hvis  $D \subset U$  er en lukket, Jordan-målbar mengde, og  $f : \mathbf{T}(D) \rightarrow \mathbb{R}$  er en kontinuerlig funksjon, så er*

$$\iint_{\mathbf{T}(D)} f(x, y) \, dx dy = \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| \, du dv$$

*Bevis:* Anta først at  $f$  er ikke-negativ. Fra lemmaet ovenfor vet vi at

$$\iint_{\mathbf{T}(D)} f(x, y) \, dx dy \leq \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| \, du dv$$

For å få den omvendte ulikheten bruker vi trikset vårt. La  $E = \mathbf{T}(D)$ ,  $g(u, v) = f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)|$  og  $S = T^{-1}$ . Da er forutsetningene i lemmaet ovenfor oppfylt når vi erstatter  $D$ ,  $f$  og  $T$  med henholdsvis  $E$ ,  $g$  og  $S$  (sjekk dette!). Følgelig er

$$\iint_{\mathbf{S}(E)} g(u, v) \, du dv \leq \iint_E g(\mathbf{S}(x, y)) |\det \mathbf{S}'(x, y)| \, dx dy$$

Setter vi inn de opprinnelige uttrykkene og bruker at

$$\det \mathbf{T}'(\mathbf{S}(x, y)) \det \mathbf{S}'(x, y) = 1$$

(husk omvendt funksjonsteorem) og  $S(E) = D$ , omdannes ulikheten ovenfor til

$$\iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| \, du dv \leq \iint_{\mathbf{T}(D)} f(x, y) \, dx dy$$

som er den omvendte ulikheten av den vi har fra før. Dermed er teoremet bevist for ikke-negative funksjoner. For å utvide til generelle funksjoner observerer vi først at dersom  $f$  er en funksjon som oppfyller betingelsene i teoremet, så er  $f$  begrenset. Velger vi en tilstrekkelig stor, positiv konstant  $C$  er da  $f(x, y) + C$  en ikke-negativ funksjon, og dermed er

$$\iint_D (f(\mathbf{T}(u, v)) + C) |\det \mathbf{T}'(u, v)| \, du dv = \iint_{\mathbf{T}(D)} (f(x, y) + C) \, dx dy$$

Etter det vi allerede har vist ( $C$  er en positiv funksjon!), er

$$\iint_D C |\det \mathbf{T}'(u, v)| dudv = \iint_{\mathbf{T}(D)} C dx dy$$

og trekker vi dette fra ligningen ovenfor, får vi

$$\iint_{\mathbf{T}(D)} f(x, y) dx dy = \iint_D f(\mathbf{T}(u, v)) |\det \mathbf{T}'(u, v)| dudv$$

og teoremet er omsider bevist!  $\square$

### Oppgaver til seksjon 6.7

**1.** Løs dobbeltintegralene ved å bruke den angitte substitusjonen.

- a)  $\iint_A x^2 dx dy$  der  $A$  er området avgrenset av linjene  $y = x$ ,  $y = x + 1$ ,  $y = -x$ ,  $y = -x + 2$ . Sett  $u = y - x$ ,  $v = y + x$ .
- b)  $\iint_A x dx dy$  der  $A$  er parallellogrammet med hjørner  $(0, 0)$ ,  $(3, 0)$ ,  $(1, 1)$ ,  $(4, 1)$ . Sett  $u = x - y$ ,  $v = y$ .
- c)  $\iint_A xy dx dy$  der  $A$  er området begrenset av linjene  $y = \frac{x}{2}$ ,  $y = \frac{x}{2} + 2$ ,  $y = 2x$  og  $y = 2x - 2$ . Sett  $u = y - \frac{x}{2}$ ,  $v = y - 2x$ .

**2.** Bruk MATLAB til å regne ut integralene i oppgave 1 uten å skifte variabel.

**3.** Løs dobbeltintegralet ved å bruke den angitte substitusjonen:

- a)  $\iint_R xy dx dy$  der  $R$  er området avgrenset av linjene  $x + 2y = -1$ ,  $x + 2y = 3$ ,  $x = y + 1$ ,  $x = y + 4$ . Bruk substitusjonen  $u = x + 2y$ ,  $v = x - y$ .
- b)  $\iint_R (x^2 - y^2) e^{x+y} dx dy$  der  $R$  er kvadratet med hjørner  $(0, 0)$ ,  $(-1, 1)$ ,  $(1, 1)$ ,  $(0, 2)$ . Sett  $u = x + y$ ,  $v = x - y$ .
- c)  $\iint_R (y^2 - yx) dx dy$  der  $R$  er området avgrenset av linjene  $y = x$ ,  $y = 2x$  og kurvene  $y = \frac{2}{x}$ ,  $y = \frac{1}{x}$ . Sett  $u = yx$ ,  $v = \frac{y}{x}$ .

**4.** Bruk MATLAB til å regne ut integralene i oppgave 3 uten å skifte variabel.

**5.** Regn ut dobbeltintegralene.

- a)  $\iint_A \frac{e^{x-y}}{x+y} dx dy$  der  $A$  er området avgrenset av linjene  $y = x$ ,  $y = x + 5$ ,  $y = -x + 2$  og  $y = -x + 4$ .
- b)  $\iint_A xy dx dy$  der  $A$  er området avgrenset av linjene  $y = x$ ,  $y = 2x$  og kurvene  $y = \frac{1}{x}$ ,  $y = \frac{3}{x}$ .
- c)  $\iint_A y dx dy$  der  $A$  er området avgrenset av linjene  $y = \frac{x}{2}$ ,  $y = 2x$  og kurvene  $y = \frac{1}{x^2}$ ,  $y = \frac{2}{x^2}$ .

- d)  $\iint_A (3x - 2y) \, dx dy$  der  $A$  er parallellogrammet utspent av vektorene  $(2, 1)$  og  $(1, 3)$ .
- e)  $\iint_A x \, dx dy$  der  $A$  er området avgrenset av parablene  $y = x^2$ ,  $y = x^2 + 4$ ,  $y = (x - 1)^2$ ,  $y = (x - 1)^2 + 4$ .

**6.** Bruk MATLAB til å regne ut integralene i oppgave 5 uten å skifte variabel.

**7.** La  $A$  være parallellogrammet utspent av to vektorer  $\begin{pmatrix} a \\ b \end{pmatrix}$  og  $\begin{pmatrix} c \\ d \end{pmatrix}$  som ikke er parallelle, og la  $M$  være matrisen  $M = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$

- a) Vis at avbildningen  $\begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{T} \begin{pmatrix} u \\ v \end{pmatrix} = M \begin{pmatrix} u \\ v \end{pmatrix}$  avbilder enhetskvadratet  $K$  utspent av  $\mathbf{e}_1$  og  $\mathbf{e}_2$  på  $A$ .

- b) Vis at for alle kontinuerlige funksjoner  $f$  er

$$\iint_A f(x, y) \, dx dy = \det M \int_0^1 \int_0^1 f(au + cv, bu + dv) \, du \, dv$$

- c) Regn ut  $\iint_A e^{2x-3y} \, dx dy$  der  $A$  er parallellogrammet utspent av  $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$  og  $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ .

**8.** (Eksamensoppgave fra MAT1110 13/6, 2005)

- a) Gitt koordinatskiftet  $x = u \cos v$  og  $y = 2u \sin v$ . Beskriv linjen  $y = 2x$  i koordinatene  $u$  og  $v$ . La  $R$  være området i første kvadrant av  $xy$ -planet som er begrenset av  $x$ -aksen, linjen  $y = 2x$  og ellipsen  $x^2 + \frac{y^2}{4} = 1$ . Finn arealet av  $R$ .
- b) Finn arealet av flaten  $z = x^2 + \frac{y^2}{2}$ ,  $(x, y) \in R$  (der  $R$  er området beskrevet i punkt a)).

**9.** (Eksamensoppgave fra MAT1110 16/8, 2007)  $R$  er området i planet avgrenset av linjene  $y = x$ ,  $y = 2x$ ,  $y = -x + 1$ ,  $y = -x + 3$ . Lag en skisse av  $R$  og regn ut dobbeltintegralet  $\iint_R \frac{x+y}{x^2} \, dx dy$ .

**10.** I denne oppgaven skal vi se nærmere på en påstand fremsatt i beviset for setning 6.7.6. Anta at  $\Pi$  er en partisjon av rektanglet  $R = [a, b] \times [c, d]$ :

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

$$c = y_0 < y_1 < \dots < y_{m-1} < y_m = d$$

der alle  $x$ 'ene og  $y$ 'ene er rasjonale tall. Vis at det finnes en partisjon  $\hat{\Pi}$  som inneholder alle delepunktene  $x_i$  og  $y_j$  i  $\Pi$ , men der alle delrekktanglene  $\hat{R}_{ij}$  er kvadrater av samme størrelse. (Hint: Del opp  $[a, b]$  og  $[c, d]$  i delintervaller med lengde  $\frac{1}{N}$  der  $N$  er fellesnevneren til alle brøkene  $x_0, x_1, \dots, x_n, y_0, y_1, \dots, y_m$ .)

## 6.8 Uegentlige integraler i planet

Hittil har vi alltid integrert over begrensede områder i planet, men i mange anvendelser er det viktig å kunne integrere over ubegrensede områder, og vi skal nå se hvordan dette kan gjøres. Vi skal først definere slike integraler for ikke-negative funksjoner, og så utvide til mer generelle funksjoner senere. Litt notasjon før vi begynner: I denne seksjonen er

$$K_n = \{(x, y) \in \mathbb{R}^2 \mid |x|, |y| \leq n\}$$

det lukkede kvadratet med sentrum i origo og sidekant  $2n$ .

**Definisjon 6.8.1** La  $A$  være en delmengde av  $\mathbb{R}^2$  slik at  $A \cap K_n$  er Jordan-målbar for alle  $n \in \mathbb{N}$ . Hvis  $f : A \rightarrow \mathbb{R}$  er en positiv, kontinuerlig funksjon, definerer vi

$$\iint_A f(x, y) \, dx \, dy = \lim_{n \rightarrow \infty} \iint_{A \cap K_n} f(x, y) \, dx \, dy$$

dersom denne grenseverdien eksisterer. I så fall sier vi at det uegentlige integralet  $\iint_A f(x, y) \, dx \, dy$  konvergerer, i motsatt fall sier vi at det divergerer.

La oss se på et eksempel:

**Eksempel 1:** Vi skal integrere  $f(x, y) = \frac{y^2}{1+x^2}$  over området

$$A = \mathbb{R} \times [0, 2] = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq y \leq 2\}$$

$A$  er altså en uendelig lang stripe med bredde 2. Legg merke til at når  $n \geq 2$ , så er  $A \cap K_n = [-n, n] \times [0, 2]$ , og vi får

$$\iint_A f(x, y) \, dx \, dy = \lim_{n \rightarrow \infty} \int_{-n}^n \left[ \int_0^2 f(x, y) \, dy \right] dx$$

Setter vi inn funksjonsuttrykket og regner litt, ser vi at

$$\begin{aligned} \int_{-n}^n \left[ \int_0^2 \frac{y^2}{1+x^2} \, dy \right] dx &= \int_{-n}^n \left[ \frac{\frac{y^3}{3}}{1+x^2} \right]_{y=0}^{y=2} dx = \frac{8}{3} \int_{-n}^n \frac{1}{1+x^2} \, dx = \\ &= \frac{8}{3} \left[ \arctan x \right]_{-n}^n = \frac{8}{3} (\arctan n - \arctan(-n)) \longrightarrow \frac{8\pi}{3} \end{aligned}$$

der vi i siste overgang bruker at  $\lim_{n \rightarrow \infty} \arctan n = \frac{\pi}{2}$  og  $\lim_{n \rightarrow \infty} \arctan(-n) = -\frac{\pi}{2}$ . Dermed har vi vist at

$$\iint_{\mathbb{R} \times [0, 2]} \frac{y^2}{1+x^2} \, dx \, dy = \frac{8\pi}{3}$$



I definisjonen ovenfor brukte vi kvadratene  $K_n$  til å definere dobbeltintegraler over uendelige områder. Ofte er det mer behagelig å bruke områder av en annen type, for eksempel sirklene  $B(\mathbf{0}, n)$

**Setning 6.8.2** *Anta at  $A$  er en delmengde av  $\mathbb{R}^2$  slik at  $A \cap B(\mathbf{0}, n)$  er Jordan-målbar for alle  $n$ . Hvis  $f : A \rightarrow \mathbb{R}$  er en kontinuerlig, ikke-negativ funksjon, så er*

$$\int_A f(x, y) dx dy = \lim_{n \rightarrow \infty} \iint_{A \cap B(\mathbf{0}, n)} f(x, y) dx dy$$

(hvis integralet til venstre divergerer, er grenseverdien til høyre lik  $\infty$ )

*Bevis:* Siden rette linjer og sirkelbuer har innhold null, vil mengdene  $A \cap K_n$  være Jordan-målbare hvis og bare hvis mengdene  $A \cap B(\mathbf{0}, n)$  er det, så det er ingen problemer med målbarheten. Siden

$$K_{n/\sqrt{2}} \subset B(\mathbf{0}, n) \subset K_n$$

(lag en tegning!) og  $f$  er ikke-negativ, har vi

$$\iint_{A \cap K_{n/\sqrt{2}}} f(x, y) dx dy \leq \iint_{A \cap B(\mathbf{0}, n)} f(x, y) dx dy \leq \iint_{A \cap K_n} f(x, y) dx dy$$

Dersom integralet konvergerer, går både det venstre og det høyre uttrykket mot  $\iint_A f(x, y) dx dy$ , og dermed må uttrykket i midten også gjøre det. Dersom integralet divergerer, går uttrykkene til venstre og høyre mot  $\infty$ , og da må uttrykket i midten også gjøre det.  $\square$

Det neste eksemplet viser en overraskende bruk av dobbeltintegraler. Husk at vi hittil ikke er vært istand til å regne ut integralet  $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$  (som via normalfordelingen spiller en sentral rolle i sannsynlighetsregning og statistikk) fordi funksjonen  $e^{-\frac{x^2}{2}}$  har en antiderivert som ikke kan uttrykkes ved hjelp av de funksjonene vi kjenner.

**Eksempel 2:** Vi skal regne ut dobbeltintegralet  $\iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy$  på to forskjellige måter. Lar vi

$$I_n = \int_{-n}^n e^{-\frac{x^2}{2}} dx,$$

ser vi først at

$$\iint_{K_n} e^{-\frac{x^2+y^2}{2}} dx dy = \int_{-n}^n \left[ \int_{-n}^n e^{-\frac{x^2+y^2}{2}} dx \right] dy =$$

$$= \int_{-n}^n e^{-\frac{y^2}{2}} \left[ \int_{-n}^n e^{-\frac{x^2}{2}} dx \right] dy = I_n \int_{-n}^n e^{-\frac{y^2}{2}} dy = I_n^2$$

Lar vi  $n$  gå mot uendelig, ser vi at

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{\iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy}$$

(forutsatt at integralene konvergerer, men det skal vi vise om et øyeblikk).

La oss nå se hva som skjer når vi bruker sirklene  $B(\mathbf{0}, n)$  til å regne ut dobbeltintegralet. Da er det naturlig å bytte til polarkoordinater, og vi får

$$\iint_{B(\mathbf{0}, n)} e^{-\frac{x^2+y^2}{2}} dx dy = \int_0^n \left[ \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta \right] dr = 2\pi \int_0^n r e^{-\frac{r^2}{2}} dr$$

Nå har det skjedd et lite under — det nye integralet er lett å regne ut ved substitusjonen  $u = \frac{r^2}{2}$ . Vi får

$$\iint_{B(\mathbf{0}, n)} e^{-\frac{x^2+y^2}{2}} dx dy = 2\pi \int_0^n r e^{-\frac{r^2}{2}} dr = 2\pi \int_0^{\frac{n^2}{2}} e^{-u} du = 2\pi(1 - e^{-\frac{n^2}{2}})$$

Dette uttrykket går mot  $2\pi$  når  $n$  går mot uendelig, og følgelig er

$$\iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy = 2\pi$$

Kombinerer vi dette med utregningene ovenfor, ser vi at

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

som også kan skrives

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 1$$

Dette forklarer hvorfor  $\sqrt{2\pi}$  dukker opp i normalfordelingen. ♣

Hittil har vi bare sett på uegentlige integraler av ikke-negative funksjoner. For å utvide til generelle funksjoner, observerer vi først at enhver funksjon  $f$  kan skrives som en differanse mellom to positive funksjoner; vi har

$$f(x) = f_+(x) - f_-(x)$$

der

$$f_+(x) = \begin{cases} f(x) & \text{hvis } f(x) > 0 \\ 0 & \text{ellers} \end{cases}$$

$$f_-(x) = \begin{cases} -f(x) & \text{hvis } f(x) < 0 \\ 0 & \text{ellers} \end{cases}$$

Legg også merke til at  $|f(x)| = f_+(x) + f_-(x)$ .

**Definisjon 6.8.3** La  $A$  være en delmengde av  $\mathbb{R}^2$  slik at  $A \cap K_n$  er Jordan-målbar for alle  $n \in \mathbb{N}$ , og anta at  $f : A \rightarrow \mathbb{R}$  er en begrenset, kontinuerlig funksjon. Vi sier at integralet  $\iint_A f(x, y) dx dy$  konvergerer dersom begge integralene  $\iint_A f_+(x, y) dx dy$  og  $\iint_A f_-(x, y) dx dy$  konvergerer, og i så fall definerer vi

$$\iint_A f(x, y) dx dy = \iint_A f_+(x, y) dx dy - \iint_A f_-(x, y) dx dy$$

**Bemerkning:** Definisjonen ovenfor kan se litt tungvinn ut — hvorfor sier vi ikke bare at integralet  $\iint_A f(x, y) dx dy$  konvergerer dersom grenseverdien  $\lim_{n \rightarrow \infty} \iint_{A \cap K_n} f(x, y) dx dy$  eksisterer slik vi gjorde for ikke-negative funksjoner? Grunnen er rett og slett at vi da risikerer å få integraler med ganske rare egenskaper som er vanskelig å holde styr på. Bruker vi definisjonen ovenfor, gjelder de vanlige regnereglene for integraler fortsatt.

Legg forøvrig merke til at siden  $|f(x)| = f_+(x) + f_-(x)$  så kan de to kravene om at både  $\iint_A f_+(x, y) dx dy$  og  $\iint_A f_-(x, y) dx dy$  skal konvergere, reduseres til ett krav, nemlig at  $\iint_A |f(x, y)| dx dy$  konvergerer. Legg også merke til at hvis vi vet at integralet  $\iint_A f(x, y) dx dy$  konvergerer, så kan vi regne det ut ved

$$\iint_A f(x, y) dx dy = \lim_{n \rightarrow \infty} \iint_{A \cap K_n} f(x, y) dx dy$$

I så fall har vi nemlig

$$\begin{aligned} \iint_A f(x, y) dx dy &= \iint_A f_+(x, y) dx dy - \iint_A f_-(x, y) dx dy = \\ &= \lim_{n \rightarrow \infty} \iint_{A \cap K_n} f_+(x, y) dx dy - \lim_{n \rightarrow \infty} \iint_{A \cap K_n} f_-(x, y) dx dy = \\ &= \lim_{n \rightarrow \infty} \iint_{A \cap K_n} (f_+(x, y) - f_-(x, y)) dx dy = \lim_{n \rightarrow \infty} \iint_{A \cap K_n} f(x, y) dx dy \end{aligned}$$

Fra teorien for funksjoner av én variabel vet vi at det også finnes en annen type uegentlige integraler — de hvor integranden går mot uendelig. Tilsvarende tilfeller finnes også for dobbeltintegraler, men vi skal ikke grave oss ned i denne teorien, men nøye oss med å se på et eksempel.

**Eksempel 3:** Vi har lyst til å integrere funksjonen  $f(x, y) = \frac{1}{(x^2+y^2)^p}$ ,  $p > 0$ , over sirkelen  $A = B(\mathbf{0}, 1)$ . Siden  $f$  ikke er definert i origo, og  $f(x, y)$  nærmer seg uendelig når  $(x, y) \rightarrow \mathbf{0}$ , er det ikke opplagt hvordan vi skal gjøre dette. En naturlig idé er å fjerne en liten sirkel rundt origo og først integrere over

området  $A_\epsilon = \{\mathbf{x} \in \mathbb{R}^2 : \epsilon \leq |\mathbf{x}| \leq 1\}$ . Bytter vi til polarkoordinater, får vi da (anta først at  $p \neq 1$ ):

$$\begin{aligned}\iint_{A_\epsilon} \frac{1}{(x^2 + y^2)^p} dx dy &= \int_\epsilon^1 \left[ \int_0^{2\pi} \frac{1}{r^{2p}} r d\theta \right] dr = 2\pi \int_\epsilon^1 \frac{1}{r^{2p-1}} dr = \\ &= \frac{\pi}{1-p} \left[ \frac{1}{r^{2p-2}} \right]_\epsilon^1 = \frac{\pi}{1-p} (1 - \epsilon^{2-2p})\end{aligned}$$

Lar vi  $\epsilon$  gå mot 0, ser vi at det siste uttrykket går mor uendelig hvis  $p > 1$  og mot  $\frac{\pi}{1-p}$  hvis  $p < 1$ . I dette siste tilfellet er det rimelig å sette

$$\iint_A \frac{1}{(x^2 + y^2)^p} dx dy = \frac{\pi}{1-p},$$

mens vi i det første sier at integralet divergerer. Går vi tilbake til det utelatte tilfellet  $p = 1$ , får vi

$$\iint_{A_\epsilon} \frac{1}{(x^2 + y^2)} dx dy = 2\pi \int_\epsilon^1 \frac{1}{r} dr = 2\pi \left[ \ln r \right]_\epsilon^1 = -2\pi \ln \epsilon \rightarrow \infty$$

som viser at integralet divergerer også i dette tilfellet. ♣

### Oppgaver til seksjon 6.8

1. Regn ut  $\iint_A e^{-x^2-y^2} dx dy$  der  $A$  er området i første kavadrant mellom x-aksen og linjen  $y = x$ .
2. Avgør om integralet  $\iint_{\mathbb{R}^2} \frac{1}{1+x^2+y^2} dx dy$  konvergerer eller divergerer.
3. Avgør om integralet  $\iint_A x dx dy$  konvergerer når  $A$  er området i fjerde kvadrant mellom  $y$ -aksen funksjonsgrafen  $y = \ln x$
4. Avgør om integralet  $\iint_A xy dx dy$  konvergerer når  $A$  er området i første kvadrant under funksjonsgrafen  $y = \frac{1}{x}$
5. Regn ut integralet  $\iint_A \frac{x}{1+y^4} dx dy$  der  $A$  er området i første kvadrant som ligger over funksjonsgrafen  $y = x^2$ .
6. La  $A = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \geq 1\}$ . Avgjør for hvilke verdier av  $p$  integralet  $\iint_A \frac{1}{(x^2+y^2)^p} dx dy$  konvegerer, og regne ut verdien i disse tilfellene.

### 6.9 Trippelintegraler

Vi skal nå se hvordan vi kan integrere funksjoner av tre variable over områder i rommet. Teoretisk er ikke dette så veldig forskjellig fra å integrere funksjoner av to variable over områder i planet, men i praksis blir det ofte litt

mer komplisert fordi geometrien er mindre oversiktlig. Vi skal derfor legge hovedvekten på praktisk utregning av integraler og hoppe over utledninger som er nesten identiske med de to-dimensjonale.

La oss først se hvordan vi kan integrere en funksjon  $f(x, y, z)$  over en rektangulær boks

$$\begin{aligned} R &= [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2] = \\ &= \{(x, y, z) \in \mathbb{R}^3 \mid a_1 \leq x \leq a_2, b_1 \leq y \leq b_2, c_1 \leq z \leq c_2\} \end{aligned}$$

Vi begynner med å lage en partisjon  $\Pi$  av  $R$ . En slik partisjon består rett og slett av en partisjon av hvert av intervallene  $[a_1, a_2], [b_1, b_2], [c_1, c_2]$ :

$$a_1 = x_0 < x_1 < x_2 < \dots < x_{n-1} < x_n = a_2$$

$$b_1 = y_0 < y_1 < y_2 < \dots < y_{m-1} < y_m = b_2$$

$$c_1 = z_0 < z_1 < z_2 < \dots < z_{l-1} < z_l = c_2$$

Partisjonen  $\Pi$  deler  $R$  opp i mange mindre “delbokser”. Vi lar

$$R_{ijk} = [x_{i-1}, x_i] \times [y_{j-1}, y_j] \times [z_{k-1}, z_k]$$

være den  $ijk$ -te delboksen, og lar

$$m_{ijk} = \inf\{f(x, y, z) \mid (x, y, z) \in R_{ijk}\}$$

$$M_{ijk} = \sup\{f(x, y, z) \mid (x, y, z) \in R_{ijk}\}$$

være henholdsvis infimum og supremum til funksjonen  $f$  over denne delboksen. På vanlig måte definerer vi nå den nedre og den øvre trappesummen til  $f$  med hensyn på partisjonen  $\Pi$ :

$$N(\Pi) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l m_{ijk} |R_{ijk}|$$

og

$$\mathcal{O}(\Pi) = \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l M_{ijk} |R_{ijk}|$$

der

$$|R_{ijk}| = (x_i - x_{i-1})(y_j - y_{j-1})(z_k - z_{k-1})$$

er volumet til delboksen  $R_{ijk}$ .

Nedre- og øvreintegralet til  $f$  kan vi også definere på vanlig måte. Nedreintegralet er gitt ved

$$\iiint_R f(x, y, z) dx dy dz = \sup\{N(\Pi) \mid \Pi \text{ er en partisjon av } R\}$$

og øvreintegralet ved

$$\overline{\iiint_R} f(x, y, z) \, dx dy dz = \inf \{ \mathcal{O}(\Pi) \mid \Pi \text{ er en partisjon av } R \}$$

Dermed har vi kommet frem til definisjonen av trippelintegraler:

**Definisjon 6.9.1** En begrenset funksjon  $f : R \rightarrow \mathbb{R}$  er integrerbar over  $R$  dersom

$$\iiint_R f(x, y, z) \, dx dy dz = \overline{\iiint_R} f(x, y, z) \, dx dy dz$$

I så fall definerer vi integralet av  $f$  over  $R$  til å være

$$\iiint_R f(x, y, z) \, dx dy dz = \underline{\iiint_R} f(x, y, z) \, dx dy dz = \overline{\iiint_R} f(x, y, z) \, dx dy dz$$

Som i det to-dimensjonale tilfellet har vi nå to grunnleggende spørsmål — hvilke funksjoner er integrerbare, og hvordan regner vi ut integraler? Vi begynner med det første spørsmålet.

**Setning 6.9.2** Anta at  $R = [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2]$  er en rektangulær boks i  $\mathbb{R}^3$  og at  $f : R \rightarrow \mathbb{R}$  er en kontinuerlig funksjon. Da er  $f$  integrerbar over  $R$ .

*Bevis:* Akkurat som beviset for teorem 6.1.5. □

La oss nå se på det neste spørsmålet — hvordan vi regner ut trippel-integraler. Akkurat som for dobbeltintegraler kan vi gjøre dette ved *iterert integrasjon*, det vil si ved å regne ut flere enkeltintegraler etter hverandre.

**Setning 6.9.3** Anta at  $R = [a_1, a_2] \times [b_1, b_2] \times [c_1, c_2]$  er en rektangulær boks i  $\mathbb{R}^3$  og at  $f : R \rightarrow \mathbb{R}$  er kontinuerlig. Da er

$$\begin{aligned} \iiint_R f(x, y, z) \, dx dy dz &= \iint_A \left[ \int_{c_1}^{c_2} f(x, y, z) \, dz \right] dx dy = \\ &= \int_{a_1}^{a_2} \left[ \int_{b_1}^{b_2} \left[ \int_{c_1}^{c_2} f(x, y, z) \, dz \right] dy \right] dx \end{aligned}$$

der  $A = [a_1, a_2] \times [b_1, b_2]$  er projeksjonen av  $R$  ned i  $xy$ -planet.

Setningen ovenfor gjelder også for ikke-kontinuerlige, integrerbare funksjoner dersom vi er litt forsiktigere med hvordan vi formulerer oss (sammenlign med setning 6.1.7). Rekkefølgen av de itererte integralene spiller ingen rolle, så vi kan også regne ut trippelintegraler slik

$$\iiint_R f(x, y, z) \, dx dy dz = \int_{c_1}^{c_2} \left[ \int_{b_1}^{b_2} \left[ \int_{a_1}^{a_2} f(x, y, z) \, dx \right] dy \right] dz$$

eller slik

$$\iiint_R f(x, y, z) \, dx dy dz = \int_{b_1}^{b_2} \left[ \int_{c_1}^{c_2} \left[ \int_{a_1}^{a_2} f(x, y, z) \, dx \right] dz \right] dy$$

(i alt er det seks muligheter og vi skriver ikke opp alle). Vær oppmerksom på at selv om resultatet blir det samme, kan det være mye lettere å regne ut trippelintegralet i én rekkefølge enn en annen.

La oss se på et eksempel:

**Eksempel 1:** Vi skal regne ut  $\iiint_R (x + ye^{2z}) \, dx dy dz$ , der  $R = [0, 1] \times [1, 3] \times [0, 2]$ . Legg merke til hvordan vi integrerer med hensyn på den aktuelle variabelen som om alle andre variable var konstanter.

$$\begin{aligned} \iiint_R (x + ye^{2z}) \, dx dy dz &= \int_0^2 \left[ \int_1^3 \left[ \int_0^1 (x + ye^{2z}) \, dx \right] dy \right] dz = \\ &= \int_0^2 \left[ \int_1^3 \left[ \frac{x^2}{2} + xye^{2z} \right]_{x=0}^{x=1} dy \right] dz = \int_0^2 \left[ \int_1^3 \left( \frac{1}{2} + ye^{2z} \right) dy \right] dz = \\ &= \int_0^2 \left[ \frac{y}{2} + \frac{y^2}{2} e^{2z} \right]_{y=1}^{y=3} dz = \int_0^2 (1 + 4e^{2z}) \, dz = \left[ z + 2e^{2z} \right]_0^2 = 2e^4 \quad \clubsuit \end{aligned}$$

**Trippelintegraler i MATLAB:** Trippelintegraler i MATLAB utføres på samme måte som dobbeltintegraler, bortsett fra at vi må bruke kommandoen `triplequad` istedenfor `doublequad`. Vil vi for eksempel ha MATLAB til å regne ut integralet i eksemplet ovenfor, skriver vi

```
>> triplequad(@(x,y,z)(x+y.*exp(2*z)),0,1,1,3,0,2)
```

Vær oppmerksom på at MATLAB kan bruke lang tid på å regne ut trippelintegraler. Dersom én av integrasjonene er enkel, kan det derfor være greit å utføre den for hånd før du kobler inn MATLAB.

### Trippelintegraler over mer generelle områder

Ofte ønsker vi å regne ut trippelintegraler over områder som ikke er rektangulære bokser. Vi bruker da det samme trikset som for dobbeltintegraler; vi putter området inni en tilstrekkelig stor rektangulær boks og “nuller ut” funksjonen utenfor det området vi er interessert i:

**Definisjon 6.9.4** Anta at  $S$  er et begrenset område i  $\mathbb{R}^3$ , og la  $R$  være en rektangulær boks som inneholder  $S$  i sitt indre. Hvis  $f : S \rightarrow \mathbb{R}$  er en begrenset funksjon, sier vi at  $f$  er integrerbar over  $S$  dersom funksjonen

$$f_S(x, y, z) = \begin{cases} f(x, y, z) & \text{hvis } (x, y, z) \in S \\ 0 & \text{ellers} \end{cases}$$

er integrerbar over  $R$ . I så fall setter vi

$$\iiint_S f(x, y, z) \, dx dy dz = \iiint_R f_S(x, y, z) \, dx dy dz$$

Integrerbarhet over begrensede områder er mer komplisert enn integrerbarhet over rektangler siden vi også må ta hensyn til hvordan området ser ut. Den neste setningen er nyttig i mange situasjoner.

**Setning 6.9.5** *Anta at  $A$  er en lukket, begrenset, Jordan-målbar mengde i  $xy$ -planet, og at  $g, h : A \rightarrow \mathbb{R}$  er to kontinuerlige funksjoner slik at  $g(x, y) \leq h(x, y)$  for alle  $(x, y) \in A$ . La  $S$  være området mellom de to funksjonsgrafene, dvs.*

$$S = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in A \text{ og } g(x, y) \leq z \leq h(x, y)\}$$

Da er enhver kontinuerlig funksjon  $f : S \rightarrow \mathbb{R}$  integrerbar over  $S$ , og

$$\iiint_S f(x, y, z) \, dx dy dz = \iint_A \left[ \int_{g(x, y)}^{h(x, y)} f(x, y, z) \, dz \right] \, dx dy$$

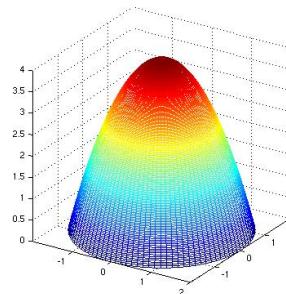
Legg merke til at når vi bruker resultatet ovenfor, regner vi først ut et enkeltintegral  $\int_{g(x, y)}^{h(x, y)} f(x, y, z) \, dz$  og deretter et dobbeltintegral. Når vi regner ut dobbeltintegralet, har vi selvfølgelig lov til å bruke alle de triksene vi kjenner — vi kan regne det ut som et iterert integral, men vi kan også bytte variable hvis det er lønnsomt. Her er et eksempel.

**Eksempel 2:** Beregn integralet

$$\iint_A x \, dx dy dz$$

der  $A$  er området som ligger over  $xy$ -planet, men under flaten  $z = 4 - x^2 - y^2$ .

Figur 1 viser området  $S$ . Projeksjonen  $A$  av  $S$  ned i  $xy$ -planet



Figur 1: Paraboloiden  $z = 4 - x^2 - y^2$

er sirkelen om origo med radius 2. I dette tilfellet er den nedre avgrensningen av området konstanten 0 og den øvre funksjonen  $z = h(x, y) = 4 - x^2 - y^2$ . Vi får dermed (legg merke til at vi i den første integrasjonen får  $xz$  og ikke  $\frac{x^2}{2}$  — vi integrerer med hensyn på  $z$  og ikke  $x$ !)

$$\begin{aligned} I &= \iiint_S x \, dx \, dy \, dz = \iint_A \left[ \int_0^{4-x^2-y^2} x \, dz \right] \, dx \, dy \\ &= \iint_A \left[ xz \right]_{z=0}^{z=4-x^2-y^2} \, dx \, dy = \iint_A (4x - x^3 - y^2 x) \, dx \, dy \end{aligned}$$

Siden  $A$  er en sirkel, lønner det seg å skifte til polarkoordinater. Vi får

$$\begin{aligned} I &= \int_0^{2\pi} \left[ \int_0^2 (4r \cos \theta - r^3 \cos^3 \theta - r^3 \cos \theta \sin^2 \theta) r \, dr \right] \, d\theta \\ &= \int_0^{2\pi} \left[ \int_0^2 (4r^2 \cos \theta - r^4 \cos^3 \theta - r^4 \cos \theta \sin^2 \theta) \, dr \right] \, d\theta \\ &= \int_0^{2\pi} \left( \frac{32}{3} \cos \theta - \frac{32}{5} \cos^3 \theta - \frac{32}{5} \cos \theta \sin^2 \theta \right) \, d\theta \end{aligned}$$

Det første og det siste av disse integralene er greie, vi har

$$\int_0^{2\pi} \frac{32}{3} \cos \theta \, d\theta = \left[ \frac{32}{3} \sin \theta \right]_0^{2\pi} = 0$$

og

$$\int_0^{2\pi} \frac{32}{5} \cos \theta \sin^2 \theta \, d\theta = \left[ \frac{32}{15} \sin^3 \theta \right]_0^{2\pi} = 0$$

Det midterste integralet skriver vi om

$$\begin{aligned} \int_0^{2\pi} \frac{32}{5} \cos^3 \theta \, d\theta &= \int_0^{2\pi} \frac{32}{5} (1 - \sin^2 \theta) \cos \theta \, d\theta \\ &= \left[ \frac{32}{5} \sin \theta - \frac{32}{15} \sin^3 \theta \right]_0^{2\pi} = 0. \end{aligned}$$

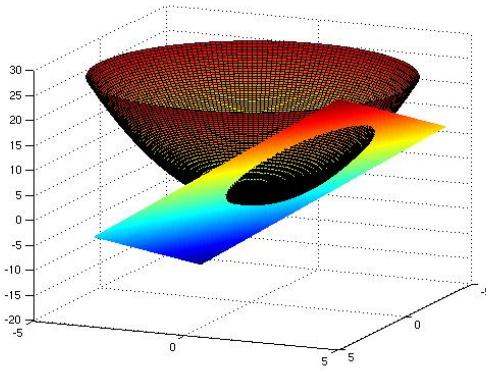
Altså er  $I = 0 - 0 - 0 = 0$ .

Når svaret blir så enkelt som dette, er det grunn til å stoppe opp og tenke seg om. Kunne vi ha kommet frem på en enklere måte? I dette tilfellet

er svaret ja — på grunn av symmetrien til integrasjonsområdet er det klart at bidraget fra positive og negative  $x$ -verdier vil oppveie hverandre. Det kan være lurt å se seg om etter slike symmetrier før man begynner – ofte kan de forenkle regnearbeidet betraktlig.  $\square$

Vi tar med et eksempel til.

**Eksempel 3:** Vi skal integrere funksjonen  $f(x, y, z) = xy$  over området  $R$  som ligger under planet  $z = 2x + 4y$  og over paraboloiden  $z = x^2 + y^2$ . Figur 2 viser hvordan planet skjærer en skalk av paraboloiden



Figur 2: Planet  $z = 2x + 4y$  skjærer gjennom paraboloiden  $z = x^2 + y^2$

Vi må først finne ut hvor de to flatene skjærer hverandre. Setter vi de to  $z$ -verdiene lik hverandre, får vi

$$x^2 + y^2 = 2x + 4y$$

For å finne ut hva slags kurve dette er, flytter vi førstegradsleddene over på den andre siden og fullfører kvadratene:

$$x^2 + y^2 = 2x + 4y \iff x^2 - 2x + y^2 - 4y = 0 \iff (x - 1)^2 + (y - 2)^2 = 5$$

Dette er ligningen til en sirkel med sentrum i  $(1, 2)$  og radius  $\sqrt{5}$ . Innsiden av denne sirkelen er altså projeksjonen  $A$  av området  $R$  ned i  $xy$ -planet. Vi har dermed

$$\iiint_R xy \, dx dy dz = \iint_A \left[ \int_{x^2+y^2}^{2x+4y} xy \, dz \right] dx dy$$

Utfører vi den innerste integrasjonen, får vi

$$I = \iint_A \left[ xyz \right]_{z=x^2+y^2}^{z=2x+4y} dx dy = \iint_A xy(2x + 4y - (x^2 + y^2)) \, dx dy$$

Det er fristende å gange ut parentesene, men la oss tenke litt fremover først. Neste trinn i utregningen er å integrere over sirkelskiven

$$A = \{(x, y) \in \mathbb{R}^2 \mid (x - 1)^2 + (y - 2)^2 \leq \sqrt{5}\}$$

Den enkleste måten å gjøre dette på, er å bytte til polarkoordinater med sentrum i  $(1, 2)$ , altså å sette

$$x - 1 = r \cos \theta \quad \text{og} \quad y - 2 = r \sin \theta$$

Siden  $2x + 4y - x^2 - y^2 = 5 - (x - 1)^2 - (y - 2)^2 = 5 - r^2$  (dette er samme kvadratkompletteringen som ovenfor), ser vi at integralet nå kan skrives (husk Jacobi-determinanten  $r$ ):

$$I = \int_0^{\sqrt{5}} \left[ \int_0^{2\pi} (1 + r \cos \theta)(2 + r \sin \theta)(5 - r^2)r \, d\theta \right] dr$$

Ganger vi ut, gir dette

$$\begin{aligned} I &= \int_0^{\sqrt{5}} (5 - r^2)r \left[ \int_0^{2\pi} (2 + r \sin \theta + 2r \cos \theta + r^2 \sin \theta \cos \theta) \, d\theta \right] dr = \\ &= \int_0^{\sqrt{5}} (5 - r^2)r \left[ 2\theta - r \cos \theta + 2r \sin \theta + \frac{r^2}{2} \sin^2 \theta \right]_{\theta=0}^{\theta=2\pi} dr = \\ &= 4\pi \int_0^{\sqrt{5}} (5 - r^2)r \, dr = 4\pi \int_0^{\sqrt{5}} (5r - r^3) \, dr = \\ &= 4\pi \left[ \frac{5}{2}r^2 - \frac{r^4}{3} \right]_0^{\sqrt{5}} = 50\pi - \frac{20\pi\sqrt{5}}{3} \end{aligned}$$

Som vi ser, blir utregningene ganske lange og kompliserte selv om integranden og integrasjonsområdet i utgangspunktet ikke ser altfor kompliserte ut. Dette er slett ikke uvanlig for trippelintegraler. ♣

Setning 6.9.5 gjelder selvfølgelig også med variablene byttet om. Ønsker vi for eksempel å integrere over et område med formen

$$S = \{(x, y, z) \in \mathbb{R}^3 \mid (y, z) \in A \text{ og } g(y, z) \leq x \leq h(y, z)\},$$

bruker vi formelen

$$I = \iiint_S f(x, y, z) \, dx \, dy \, dz = \iint_A \left[ \int_{g(y, z)}^{h(y, z)} f(x, y, z) \, dx \right] dy \, dz$$

**Mer om trippelintegraler i MATLAB:** Skal vi bruke MATLAB til å integrere over områder som ikke er rektangulære bokser, må vi bruke det samme trikset som vi brukte for dobbeltintegraler; vi må “nulle ut” funksjonen over de delene vi ikke er interessert i. Ønsker vi for eksempel å integrere funksjonen  $f(x, y, z) = x^2y + z^2$  over kulen med sentrum i origo og radius 1, skriver vi

```
>> triplequad(@(x,y,z)((x.^2.*y+z.^2).*(x.^2+y.^2+z.^2<=1)), -1,1,-1,1,-1,1)
```

Vi minner om at trippelintegraler tar tid (selv for MATLAB), og at det kan hende du må vente på svaret noen minutter.

### Oppgaver til seksjon 6.9

**1.** Beregn trippelintegralet:

- a)  $\iiint_A xyz \, dxdydz$  når  $A = [0, 1] \times [0, 1] \times [0, 1]$
- b)  $\iiint_A (x + ye^z) \, dxdydz$  når  $A = [-1, 1] \times [0, 1] \times [1, 2]$
- c)  $\iiint_A zy \cos(xy) \, dxdydz$  når  $A = [1, 2] \times [\pi, 2\pi] \times [0, 1]$
- d)  $\iiint_A (x + y + z) \, dxdydz$  når  $A = [0, 1] \times [0, 2] \times [0, 3]$
- e)  $\iiint_A (\sqrt{y} - 3z) \, dxdydz$  når  $A = [2, 3] \times [0, 1] \times [-1, 1]$

**2.** Beregn trippelintegralet:

- a)  $\iiint_A (xy + z) \, dxdydz$  når  $A = \{(x, y, z) : 0 \leq x \leq 1, 0 \leq y \leq 2, 0 \leq z \leq x^2y\}$
- b)  $\iiint_A z \, dxdydz$  når  $A = \{(x, y, z) : 0 \leq x \leq 2, 0 \leq y \leq \sqrt{x}, -y^2 \leq z \leq xy\}$
- c)  $\iiint_A (x + y)z \, dxdydz$  når  $A = \{(x, y, z) : 0 \leq y \leq 4, 0 \leq x \leq \sqrt{y}, 0 \leq z \leq 4\}$
- d)  $\iiint_A (3y^2 - 3z) \, dxdydz$  når  $A$  er området avgrenset av koordinatplanene og planet  $3x + 2y - z = 6$
- e)  $\iiint_A xy \, dxdydz$  når  $A$  er pyramiden med hjørner i  $(0, 0, 0)$ ,  $(1, 0, 0)$ ,  $(0, 1, 0)$  og  $(0, 0, 1)$ .

**3.** Bruk MATLAB til å regne ut integralene i oppgave 1 (det kan ta tid!).

**4.** Bruk MATLAB til å regne ut integralene i oppgave 2 (det kan ta tid!).

**5.** Anta at  $R$  er en rektangulær boks, og at  $f, g : R \rightarrow \mathbb{R}$  er kontinuerlige funksjoner. Vis at

- (i)  $\iiint_R kf(x) \, dxdydz = k \iiint_R f(x) \, dxdydz$  for alle konstanter  $k$
- (ii)  $\iiint_R (f(x) + g(x)) \, dxdydz = \iiint_R f(x) \, dxdydz + \iiint_R g(x) \, dxdydz$
- (iii)  $\iiint_R (f(x) - g(x)) \, dxdydz = \iiint_R f(x) \, dxdydz - \iiint_R g(x) \, dxdydz$

## 6.10 Skifte av variable i trippelintegraler

Teorien for skifte av variable i trippelintegraler er nesten identisk med den tilsvarende teorien for dobbeltintegraler, og vi gir derfor bare en rask gjennomgang av de viktigste resultatene før vi går løs på eksemplene. En begrenset delmengde  $A$  av  $\mathbb{R}^3$  kalles *Jordan-målbar* dersom integralet

$$\iiint_A 1 \, dx dy dz$$

eksisterer — altså hvis funksjonen  $\mathbf{1}_A$  definert ved

$$\mathbf{1}_A(x, y, z) = \begin{cases} 1 & \text{hvis } (x, y, z) \in A \\ 0 & \text{ellers} \end{cases}$$

er integrerbar. Vi kan nå formulere teoremet for skifte av variabel på akkurat samme måte som for dobbeltintegraler.

**Teorem 6.10.1 (Skifte av variabel i trippelintegral)** *La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^3$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^3$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at det  $\mathbf{T}' \neq 0$  på hele  $U$ . Hvis  $D \subset U$  er en lukket, Jordan-målbar mengde, og  $f : \mathbf{T}(D) \rightarrow \mathbb{R}$  er en kontinuerlig funksjon, så er*

$$\iiint_{\mathbf{T}(D)} f(x, y, z) \, dx dy dz = \iint_D f(\mathbf{T}(u, v, w)) |\det \mathbf{T}'(u, v, w)| \, du dv dw$$

La oss forsøke å forklare hva dette betyr. Vi tenker oss at vi starter med et trippelintegral  $\iiint_A f(x, y, z) \, dx dy dz$  der området  $A$  er komplisert å integrere over. Det finnes imidlertid et mye penere område  $D$  og en funksjon  $\mathbf{T} : D \rightarrow A$  som avbilder hvert punkt  $(u, v, w) \in D$  på et korresponderende punkt  $\mathbf{T}(u, v, w) = (x(u, v, w), y(u, v, w), z(u, v, w))$  i  $A$ . Dersom avbildningen  $\mathbf{T}$  “fyller ut” hele  $A$ , har vi likheten

$$\begin{aligned} \iiint_A f(x, y, z) \, dx dy dz &= \\ &= \iiint_D f(x(u, v, w), y(u, v, w), z(u, v, w)) \left| \frac{\partial(x, y, z)}{\partial(u, v, w)} \right| \, du dv dw \end{aligned}$$

der  $\left| \frac{\partial(x, y, z)}{\partial(u, v, w)} \right|$  er tallverdien til Jacobi-determinanten

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} & \frac{\partial y}{\partial w} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} & \frac{\partial z}{\partial w} \end{vmatrix}$$

Vi kan altså erstatte integralet over det kompliserte området  $A$  med et integral over det enklere området  $D$ , men prisen vi må betale er å få inn en Jacobi-faktor  $\left| \frac{\partial(x,y,z)}{\partial(u,v,w)} \right|$  i integralet.

**Eksempel 1:** Vi skal integrere funksjonen  $f(x, y, z) = 4x - y + 2z$  over parallellogrammet  $A$  utspent av vektorene  $\mathbf{a} = (1, 0, -1)$ ,  $\mathbf{b} = (0, 1, 2)$  og  $\mathbf{c} = (2, 3, 1)$ .

Siden parallellogrammet ligger skjevt i forhold til aksene, blir det kompliserte integrasjonsgrenser hvis vi forsøker å integrere direkte. Vi skal derfor prøve å “rette opp” parallellogrammet før vi integrerer. Ideen er å bruke en funksjon  $\mathbf{T}$  som avbilder boksen  $D$  utspent av enhetsvektorene  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 = (0, 1, 0)$  og  $\mathbf{e}_3 = (0, 0, 1)$  på den opprinnelige boksen  $A$ . Dette er ikke så vanskelig; vi bruker bare lineæravbildningen som avbilder  $\mathbf{e}_1$  på  $\mathbf{a}$ ,  $\mathbf{e}_2$  på  $\mathbf{b}$  og  $\mathbf{e}_3$  på  $\mathbf{c}$  (husk setning 2.8.4), dvs.

$$\mathbf{T}(u, v, w) = (u + 2w, v + 3w, -u + 2v + w)$$

Vi har altså  $x = u + 2w$ ,  $y = v + 3w$ ,  $z = -u + 2v + w$ , og Jacobi-determinanten blir

$$\frac{\partial(x, y, z)}{\partial(u, v, w)} = \begin{vmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ -1 & 2 & 1 \end{vmatrix} = -3$$

(vanligvis blir Jacobi-determinanten en funksjon av  $u, v, w$ , men siden avbildningen  $\mathbf{T}$  er lineær, blir den et tall i dette tilfellet). Ifølge teoremet ovenfor er nå

$$\begin{aligned} I &= \iiint_A (4x - y + 2z) dx dy dz = \\ &= \iiint_D (4(u + 2w) - (v + 3w) + 2(-u + 2v + w)) \cdot |-3| du dv dw = \\ &= \iiint_D (2u + 3v + 7w) \cdot 3 du dv dw = \iiint_D (6u + 9v + 21w) du dv dw \end{aligned}$$

Det gjenstår å integrere over  $D$ , men det er enkelt:

$$\begin{aligned} I &= \int_0^1 \left[ \int_0^1 \left[ \int_0^1 (6u + 9v + 21w) du \right] dv \right] dw = \\ &= \int_0^1 \left[ \int_0^1 \left[ (3u^2 + 9uv + 21uw) \right]_{u=0}^{u=1} dv \right] dw = \\ &= \int_0^1 \left[ \int_0^1 (2 + 9v + 21w) dv \right] dw = \int_0^1 \left[ 3v + \frac{9}{2}v^2 + 21vw \right]_{v=0}^{v=1} dw = \\ &= \int_0^1 \left( \frac{15}{2} + 21w \right) dw = \left[ \frac{15}{2}w + \frac{21}{2}w^2 \right]_{w=0}^{w=1} = 18 \quad \clubsuit \end{aligned}$$

De vanligste koordinatsystemene å bytte til er sylinderkoordinater og kulekoordinater. Vi skal nå ta en grundig kikk på disse.

### Integrasjon i sylinderkoordinater

Fra seksjon 3.7 kjenner vi sammenhengen mellom sylinderkoordinater og vanlige (kartesiske) koordinater:

$$x = r \cos \theta, \quad y = r \sin \theta \quad \text{og} \quad z = z$$

Dette gir Jacobi-determinanten

$$\frac{\partial(x, y, z)}{\partial(r, \theta, z)} = \begin{vmatrix} \cos \theta & -r \sin \theta & 0 \\ \sin \theta & r \cos \theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r$$

Jacobi-determinanten for sylinderkoordinater er altså den samme som for polarkoordinater i planet (dette er ingen tilfeldighet — sylinderkoordinater er bare polarkoordinater med en ekstra koordinat som ikke endrer seg). Formelen for skifte av variabel blir nå seende slik ut

$$\iiint_A f(x, y, z) dx dy dz = \iiint_D f(r \cos \theta, r \sin \theta, z) r dr d\theta dz$$

der  $D$  er området beskrevet i sylinderkoordinater.

Ikke unaturlig egner sylinderkoordinater seg godt når vi skal integrere over et område som ligner en sylinder. Her er et eksempel.

**Eksempel 2:** Vi skal integrere funksjonen  $f(x, y, z) = x^2 z$  over det området  $A$  som ligger over  $xy$ -planet og inni både sylinderen  $x^2 + y^2 = 1$  og kulen  $x^2 + y^2 + z^2 = 4$ . Dette området ser ut som en stående sylinder avrundet av en kuleflate på toppen. Skifter vi variable til sylinderkoordinater, får vi

$$I = \iiint_A x^2 z dx dy dz = \iiint_D (r \cos \theta)^2 z r dr d\theta dz = \iiint_D r^3 z \cos^2 \theta dr d\theta dz$$

der  $D$  er området beskrevet i sylinderkoordinater. Det er ikke så vanskelig å finne denne beskrivelsen: Siden vi hele tiden skal befinner oss inni en sylinder med radius 1, må vi ha  $0 \leq r \leq 1$  og  $0 \leq \theta \leq 2\pi$ , og siden vi må holde oss over  $xy$ -planet og innenforkulen med radius 4, må vi ha  $0 \leq z \leq \sqrt{4 - r^2}$  (dette uttrykket får vi ved å løse ligningen  $x^2 + y^2 + z^2 = 4$  for  $z$  og bruke at  $x^2 + y^2 = r^2$ .) Dermed har vi

$$I = \int_0^{2\pi} \left[ \int_0^1 \left[ \int_0^{\sqrt{4-r^2}} r^3 z \cos^2 \theta dz \right] dr \right] d\theta$$

Det gjenstår å utføre integrasjonene:

$$I = \int_0^{2\pi} \left[ \int_0^1 \left[ r^3 \frac{z^2}{2} \cos^2 \theta \right]_0^{\sqrt{4-r^2}} dr \right] d\theta =$$

$$\begin{aligned}
&= \frac{1}{2} \int_0^{2\pi} \left[ \int_0^1 r^3 (4 - r^2) \cos^2 \theta \, dr \right] d\theta = \frac{1}{2} \int_0^{2\pi} \cos^2 \theta \left[ \int_0^1 (4r^3 - r^5) dr \right] d\theta = \\
&= \frac{1}{2} \int_0^{2\pi} \cos^2 \theta \left[ \left( r^4 - \frac{r^6}{6} \right) \Big|_0^1 \right] d\theta = \frac{5}{12} \int_0^{2\pi} \cos^2 \theta \, d\theta
\end{aligned}$$

Det gjenstående integralet kan løses på mange måter. Vi velger å bruke formelen  $\cos 2\theta = 2\cos^2 \theta - 1$ , som gir  $\cos^2 \theta = \frac{1+\cos 2\theta}{2}$ . Dermed er

$$I = \frac{5}{24} \int_0^{2\pi} (1 + \cos 2\theta) \, d\theta = \frac{5}{24} \left[ \theta + \frac{1}{2} \sin 2\theta \right]_0^{2\pi} = \frac{5\pi}{12} \quad \clubsuit$$

I noen integraler lønner det seg å bruke sylinderkoordinater med sentrum i et annet punkt enn origo. Ønsker vi å ha sentrum i punktet  $(a, b, c)$ , bruker vi substitusjonen

$$x = a + r \cos \theta, \quad y = b + r \sin \theta \quad \text{og} \quad z = c + z$$

Siden konstantene blir borte når vi deriverer, er Jacobi-determinanten fortsatt  $r$ .

### Integrasjon i kulekoordinater

Fra seksjon 3.7 kjenner vi sammenhengen mellom kulekoordinater og vanlige koordinater:

$$x = \rho \sin \phi \cos \theta, \quad y = \rho \sin \phi \sin \theta \quad \text{og} \quad z = \rho \cos \phi$$

Jacobi-determinanten blir dermed

$$\frac{\partial(x, y, z)}{\partial(\rho, \phi, \theta)} = \begin{vmatrix} \sin \phi \cos \theta & \rho \cos \phi \cos \theta & -\rho \sin \phi \sin \theta \\ \sin \phi \sin \theta & \rho \cos \phi \sin \theta & \rho \sin \phi \cos \theta \\ \cos \phi & -\rho \sin \phi & 0 \end{vmatrix} = \rho^2 \sin \phi$$

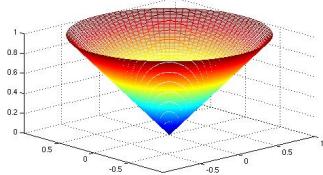
der vi i det siste skrittet har brukt formelen  $\cos^2 v + \sin^2 v = 1$  både for  $v = \theta$  og for  $v = \phi$ . Legg merke til at siden  $0 \leq \phi \leq \pi$ , så er Jacobi-determinanten  $\rho^2 \sin \phi$  aldri negativ. Formelen for skifte av variabel blir dermed

$$\begin{aligned}
&\iiint_A f(x, y, z) \, dx dy dz = \\
&= \iiint_D f(\rho \sin \phi \cos \theta, \rho \sin \phi \sin \theta, \rho \cos \phi) \rho^2 \sin \phi \, d\rho d\phi d\theta
\end{aligned}$$

der  $D$  er området  $A$  beskrevet i kulekoordinater.

Kulekoordinater eigner seg godt når vi integrerer over et område som er (delvis) begrenset av en kuleflate. Her er et eksempel:

**Eksempel 3:** Vi skal integrere funksjonen  $f(x, y, z) = z$  over området  $R$  som ligger inni både kjeglen  $z = \sqrt{x^2 + y^2}$  ogkulen  $x^2 + y^2 + z^2 = 1$ .



Figur 1: Kjeglen  $z = \sqrt{x^2 + y^2}$

Figur 1 viser kjeglen — kuleflaten ville ligge over denne som et krumt lokk.

La oss prøve å finne ut hvilke verdier av  $\rho$ ,  $\theta$ , og  $\phi$  som svarer til punkter i integrasjonsområdet. Siden kjegleflaten danner en vinkel på  $45^\circ$  med  $z$ -aksen, må  $\phi$  gå fra  $0^\circ$  til  $45^\circ$ , altså fra 0 til  $\frac{\pi}{4}$  radianer. Radianen  $\rho$  skal løpe fra origo til kuleskallet, altså fra 0 til 1, mens  $\theta$  kan ha en hvilken som helst verdi, og går altså fra 0 til  $2\pi$ . Integralet blir nå

$$\begin{aligned} I &= \iiint_R z \, dx dy dz = \int_0^{\frac{\pi}{4}} \left[ \int_0^1 \left[ \int_0^{2\pi} \rho \cos \phi \rho^2 \sin \phi \, d\theta \right] d\rho \right] d\phi = \\ &= \int_0^{\frac{\pi}{4}} \left[ \int_0^1 \left[ \int_0^{2\pi} \rho^3 \sin \phi \cos \phi \, d\theta \right] d\rho \right] d\phi = \\ &= 2\pi \int_0^{\frac{\pi}{4}} \left[ \int_0^1 \rho^3 \sin \phi \cos \phi \, d\rho \right] d\phi = 2\pi \int_0^{\frac{\pi}{4}} \left[ \frac{\rho^4}{4} \sin \phi \cos \phi \right]_{\rho=0}^{\rho=1} d\phi = \\ &= \frac{\pi}{2} \int_0^{\frac{\pi}{4}} \sin \phi \cos \phi \, d\phi = \frac{\pi}{2} \left[ \frac{\sin^2 \phi}{2} \right]_0^{\frac{\pi}{4}} = \frac{\pi}{2} \cdot \frac{(\sqrt{2}/2)^2}{2} = \frac{\pi}{8} \end{aligned}$$

♣

La oss se på et eksempel til:

**Eksempel 4:** Anta at  $A$  er et område i rommet. Tenk deg at  $P$  er et punkt utenfor  $A$ , og at vi ønsker å regne ut den gjennomsnittlige avstanden fra et punkt i  $A$  til  $P$ . Dersom  $P$  har koordinatene  $(a, b, c)$  er denne gjennomsnittsavstanden gitt ved

$$\frac{1}{\text{Vol}(A)} \iiint_A \sqrt{(x-a)^2 + (y-b)^2 + (z-c)^2} \, dx dy dz$$

der  $\text{Vol}(A)$  er volumet til  $A$  (tenk gjennom hvorfor vi må dele med volumet!)

La oss bruke denne formelen på en kule  $A$  med radius 1 og et punkt  $P$  utenfor kulen. Vi ønsker altså å finne gjennomsnittsavstanden fra et punkt innikulen til  $P$ .

Vi begynner med å legge kulen i origo og  $P$  i punktet  $(0, 0, a)$  der  $a > 1$ . Siden kulen har volum  $\frac{4\pi}{3}$ , er gjennomsnittsavstanden da gitt ved  $\frac{I}{\frac{4\pi}{3}}$  der  $I$  er integralet

$$I = \iiint_A \sqrt{x^2 + y^2 + (z - a)^2} dx dy dz$$

og  $A$  er kulen. Bytter vi til kulekoordinater, får vi

$$\int_0^1 \left[ \int_0^\pi \left[ \int_0^{2\pi} \sqrt{(\rho \sin \phi \cos \theta)^2 + (\rho \sin \phi \sin \theta)^2 + (\rho \cos \phi - a)^2} d\theta \right] \rho^2 \sin \phi d\phi \right] d\rho$$

Ganger vi ut parentesene og bruker vår gamle venn  $\sin^2 v + \cos^2 v = 1$  et par ganger, sitter vi igjen med

$$I = \int_0^1 \left[ \int_0^\pi \left[ \int_0^{2\pi} \sqrt{\rho^2 - 2a\rho \cos \phi + a^2} \rho^2 \sin \phi d\theta \right] d\phi \right] d\rho$$

Siden integranden ikke inneholder  $\theta$ , er den innerste integrasjonen enkel, og vi får

$$I = 2\pi \int_0^1 \left[ \int_0^\pi \sqrt{\rho^2 - 2a\rho \cos \phi + a^2} \rho^2 \sin \phi d\phi \right] d\rho$$

Det neste integralet er verre, men setter vi  $u = \rho^2 - 2a\rho \cos \phi + a^2$ , får vi  $du = 2a\rho \sin \phi d\phi$ . Dermed er

$$\begin{aligned} \int \sqrt{\rho^2 - 2a\rho \cos \phi + a^2} \rho^2 \sin \phi d\phi &= \frac{\rho}{2a} \int \sqrt{u} du = \frac{\rho}{3a} u^{\frac{3}{2}} + C = \\ &= \frac{\rho}{3a} (\rho^2 - 2a\rho \cos \phi + a^2)^{\frac{3}{2}} + C \end{aligned}$$

Setter vi inn grensene, ser vi at

$$\begin{aligned} I &= 2\pi \int_0^1 \left[ \frac{\rho}{3a} (\rho^2 - 2a\rho \cos \theta + a^2)^{\frac{3}{2}} \right]_{\phi=0}^\pi d\rho = \\ &= \frac{2\pi}{3a} \int_0^1 \rho \left[ (\rho^2 + 2a\rho + a^2)^{\frac{3}{2}} - (\rho^2 - 2a\rho + a^2)^{\frac{3}{2}} \right] d\rho = \\ &= \frac{2\pi}{3a} \int_0^1 \rho \left[ ((\rho + a)^2)^{\frac{3}{2}} - ((\rho - a)^2)^{\frac{3}{2}} \right] d\rho = \\ &= \frac{2\pi}{3a} \int_0^1 \rho [(\rho + a)^3 - (a - \rho)^3] d\rho \end{aligned}$$

der vi i siste skritt har brukt at  $((\rho - a)^2)^{\frac{3}{2}} = (a - \rho)^3$  siden  $a \geq \rho$ . Ganger vi ut parentesene inni integralet og forkorter, får vi

$$I = \frac{2\pi}{3a} \int_0^1 (6a^2\rho^2 + 2\rho^4) d\rho = \frac{2\pi}{3a} \left[ 2a^2\rho^3 + \frac{2\rho^5}{5} \right]_0^1 = \frac{4\pi(5a^2 + 1)}{15a}$$

Gjennomsnittsavstanden er dermed  $\frac{I}{\frac{4\pi}{3}} = a + \frac{1}{5a}$ . Overbevis deg selv om at det er rimelig at gjennomsnittsavstanden er litt større enn avstanden  $a$  fra  $P$  til sentrum i kulen.



I noen integraler lønner det seg å bruke kulekoordinater med sentrum i et annet punkt enn origo. Ønsker vi å ha sentrum i punktet  $(a, b, c)$ , bruker vi substitusjonen

$$x = a + \rho \sin \phi \cos \theta, \quad y = b + \rho \sin \phi \sin \theta \quad \text{og} \quad z = c + \rho \cos \phi$$

Siden konstantene blir borte når vi deriverer, er Jacobi-determinanten fortsatt  $\rho^2 \sin \phi$ . Vi skal se flere eksempler på bruk av kulekoordinater i neste seksjon.

### Integrasjon i $\mathbb{R}^n$

Det går selvfølgelig an å integrere funksjoner  $f(x_1, x_2, \dots, x_n)$  med flere enn tre variable. Disse integralene betegnes med

$$\int \cdots \int_S f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Den grunnleggende teorien er den samme, men det er vanskeligere å visualisere områdene vi skal integrere over.

Også  $n$ -dimensjonale integraler kan regnes ut ved iterasjon. Dersom området  $S$  er på formen

$$S = \{(x_1, \dots, x_{n-1}) \in A \mid g(x_1, \dots, x_{n-1}) \leq x_n \leq h(x_1, \dots, x_{n-1})\}$$

så er (under rimelige betingelser)

$$\begin{aligned} & \int \cdots \int_S f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \\ &= \int \cdots \int_A \left[ \int_{g(x_1, \dots, x_{n-1})}^{h(x_1, \dots, x_{n-1})} f(x_1, x_2, \dots, x_n) dx_n \right] dx_1 dx_2 \dots dx_{n-1} \end{aligned}$$

Kjerneregelen er også den samme som før:

**Teorem 6.10.2 (Skifte av variable i  $\mathbb{R}^n$ )** La  $U$  være en åpen, begrenset mengde i  $\mathbb{R}^n$  og anta at  $\mathbf{T} : U \rightarrow \mathbb{R}^n$  er en injektiv funksjon med kontinuerlige partiellderiverte slik at det  $\mathbf{T}' \neq 0$  på hele  $U$ . Hvis  $D \subset U$  er en lukket, Jordan-målbar mengde, og  $f : \mathbf{T}(D) \rightarrow \mathbb{R}$  er en kontinuerlig funksjon, så er

$$\begin{aligned} & \int \cdots \int_{\mathbf{T}(D)} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \\ &= \int \cdots \int_D f(\mathbf{T}(u_1, u_2, \dots, u_n)) |\det \mathbf{T}'(u_1, u_2, \dots, u_n)| du_1 du_2 \dots du_n \end{aligned}$$

Vi skal ikke komme nærmere inn på integrasjon i mer enn tre variable her — skulle du støte på slike integraler i andre fag (og det er slett ikke utenkelig!), greier du deg som regel godt med formlene ovenfor og den intuisjonen du har fra to og tre dimensjoner.

### Oppgaver til seksjon 6.10

**1.** Bruk sylinderkoordinater til å beregne:

- $\iiint_A x \, dx dy dz$  når  $A = \{(x, y, z) : x, y \geq 0 \text{ og } x^2 + y^2 \leq 9, \text{ og } 0 \leq z \leq 2\}$
- $\iiint_A xy \, dx dy dz$  når  $A = \{(x, y, z) : x^2 + y^2 \leq 1 \text{ og } 0 \leq z \leq 4 - x - y\}$
- $\iiint_A z \sqrt{x^2 + y^2} \, dx dy dz$  når  $A = \{(x, y, z) : x^2 + (y - 1)^2 \geq 1 \text{ og } 0 \leq z \leq 2\}$ ,

**2.** Bruk kulekoordinater til å beregne trippelintegralet:

- $\iiint_A (x^2 + y^2) \, dx dy dz$  når  $A$  er kulen om origo med radius 1
- $\iiint_A x \, dx dy dz$  når  $A = \{(x, y, z) : x, y \geq 0, z \geq \frac{1}{2} \text{ og } x^2 + y^2 + z^2 \leq 1\}$
- $\iiint_A 1 \, dx dy dz$  når  $A = \{(x, y, z) : z^2 \leq x^2 + y^2 \leq 1, 3y^2 \leq x^2, x \geq 0\}$

**3.** Regn ut trippelintegralet.

- $\iiint_A z \, dx dy dz$  der  $A$  er området over paraboloiden  $z = x^2 + y^2$  og under kuleflaten  $x^2 + y^2 + z^2 = 2$ .
- $\iiint_A x \, dx dy dz$  der  $A = \{(x, y, z) | x^2 + y^2 \leq z \leq 4\}$ .
- $\iiint_A e^{-\sqrt{x^2+y^2+z^2}} \, dx dy dz$  der  $A$  er kulen med sentrum i origo og radius 1.
- $\iiint_A \sqrt{x^2 + y^2} \, dx dy dz$  der  $A$  er området inni sylinderen  $x^2 + y^2 = 1$ , mellom  $xy$ -planet og flaten  $z = (x^2 + y^2)^{\frac{3}{2}}$ .
- $\iiint_A (x^2 + y^2) \, dx dy dz$  der  $A$  er området begrenset av sylinderen  $x^2 - 2x + y^2 = 1$  og planene  $z = 0$  og  $z = 2$ .
- $\iiint_A (x - 3y + 4z) \, dx dy dz$  der  $A$  er parallelepipedet utspent av vektorene  $(1, -1, 0)$ ,  $(2, 0, 3)$  og  $(0, 0, 4)$ .

- La  $A$  være parallelepipedet utspent av tre vektorer  $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$  og  $\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$  som ikke ligger i samme plan, og la  $M$  være matrisen  $M = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$

- a) Vis at avbildningen  $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \mathbf{T} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = M \begin{pmatrix} u \\ v \\ w \end{pmatrix}$  avbilder enhetskuben  $K$  utspent av  $\mathbf{e}_1, \mathbf{e}_2$  og  $\mathbf{e}_3$  på  $A$ .

- b) Vis at for alle kontinuerlige funksjoner  $f$  er

$$\iiint_A f(x, y, z) dx dy dz = \det M \int_0^1 \int_0^1 \int_0^1 f(a_1 u + b_1 v + c_1 w, a_2 u + b_2 v + c_2 w, a_3 u + b_3 v + c_3 w) du dv dw$$

- c) Regn ut  $\iint_A (2x - 3z) dx dy$  der  $A$  er parallelogrammet utspent av  $\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix}$  og  $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ .

**5.** (Eksamens i MAT1110 14/6, 2004) La  $D$  være området som både ligger på innsiden av kjeglen  $z = \sqrt{x^2 + y^2}$  og på innsiden av kuleflaten  $x^2 + y^2 + z^2 = 1$ . Beregn  $\iiint_D z dx dy dz$ .

**6.** (Kontinuasjonseksemene i MAT1110 2006) Regn ut

$$\iiint_R \sqrt{x^2 + y^2} dx dy dz$$

der  $R = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq 4\}$ .

**7.** La  $A$  være kulen med sentrum i origo og radius  $R$ , og anta at  $a > R$ . Vis at

$$\iiint_A \frac{1}{\sqrt{x^2 + y^2 + (z - a)^2}} dx dy dz = \frac{4\pi R^3}{3a}$$

Dette resultatet er viktig i fysikk der det kan brukes til å vise at gravitasjonskraften fra en homogen kule er den samme som om all massen var samlet i sentrum.

**8.** En begrenset mengde  $A \subset \mathbb{R}^3$  har *innhold null* dersom det for enhver  $\epsilon > 0$  finnes en endelig samling  $R_1, R_2, \dots, R_n$  av rektangulære bokser med sider parallelle med koordinataksene slik at  $A \subset R_1 \cup R_2 \cup \dots \cup R_n$ , og summen av volumene til  $R_1, R_2, \dots, R_n$  er mindre enn  $\epsilon$ . Vis at hvis  $K$  er en lukket, begrenset delmengde av  $\mathbb{R}^3$  og  $f : K \rightarrow \mathbb{R}$  er kontinuerlig, så har grafen til  $f$  innhold null.

## 6.11 Anvendelser av trippelintegraler

Vi skal nå se på noen praktiske anvendelser av trippelintegraler. Mange av disse anvendelsene ligner på de anvendelsene av enkelt- og dobbeltintegraler som du kjenner fra tidligere.

La oss tenke oss at vi har et tredimensjonalt legeme som utgjør et område  $S \subset \mathbb{R}^3$ . Anta videre at dette legemet har en varierende tetthet beskrevet ved en funksjon  $f(x, y, z)$  (dette betyr at en liten del av legemet rundt punktet  $(x, y, z)$  med volum  $V$  har masse tilnærmet lik  $f(x, y, z)V$ ). Den totale massen til dette legemet er da gitt ved trippelintegralet

$$M = \iiint_S f(x, y, z) \, dx \, dy \, dz$$

Er tettheten konstant lik 1, er massen og volumet like, og vi får

$$\text{Volum}(S) = \iiint_S 1 \, dx \, dy \, dz$$

**Eksempel 1:** Vi skal finne volumet til området  $S$  avgrenset at paraboloiden  $z = x^2 + y^2$  og planet  $z = 2x - 4y + 4$ . Lager du en figur, ser du at planet ligger som et lokk over paraboloiden, og at volumet derfor er

$$\begin{aligned} V &= \iiint_S 1 \, dx \, dy \, dz = \iint_A \left[ \int_{x^2+y^2}^{2x-4y+4} 1 \, dz \right] dx \, dy = \\ &= \iint_A (2x - 4y + 4 - x^2 - y^2) \, dx \, dy \end{aligned}$$

der  $A$  er projeksjonen av  $S$  ned i  $xy$ -planet. For å se hvilket område dette er, finner vi skjæringskurven mellom de to flatene:

$$\begin{aligned} x^2 + y^2 &= 2x - 4y + 4 \iff x^2 - 2x + 1 + y^2 + 4y + 4 = 4 + 1 + 4 \iff \\ &\iff (x - 1)^2 + (y + 2)^2 = 3^2 \end{aligned}$$

$A$  er altså sirkelskiven i  $xy$ -planet med sentrum i  $(1, -2)$  og radius 3. For å komme videre, bytter vi til polarkoordinater med sentrum i  $(1, -2)$ , dvs. vi setter

$$x = 1 + r \cos \theta \quad \text{og} \quad y = -2 + r \sin \theta$$

Jacobi-determinanten er  $r$ , og vi får

$$\begin{aligned} V &= \int_0^3 \left[ \int_0^{2\pi} \left( 2(1 + r \cos \theta) - 4(-2 + r \sin \theta) + \right. \right. \\ &\quad \left. \left. + 4 - (1 + r \cos \theta)^2 - (-2 + r \sin \theta)^2 \right) r \, d\theta \right] dr = \\ &= \int_0^3 \left[ \int_0^{2\pi} (9r - r^3) \, d\theta \right] dr = 2\pi \int_0^3 (9r - r^3) \, dr = \\ &= 2\pi \left[ \frac{9r^2}{2} - \frac{r^4}{4} \right]_0^3 = \frac{81\pi}{2} \end{aligned}$$

(du får litt enklere regninger om du fullfører kvadratet i integranden før du setter inn de nye koordinatene). ♣

**Bemerkning:** I eksemplet ovenfor kunne vi like godt ha regnet ut volumet ved hjelp av et dobbeltintegral. Det finnes imidlertid eksempler der det er enklere å bruke trippelintegraler; det gjelder spesielt i tilfeller der vi må gjøre et ekte tre-dimensjonalt variabelskifte.

La oss også regne ut massen til et romlegeme.

**Eksempel 2:** Vi skal regne ut massen til den begrensede kjeglen  $z = \sqrt{x^2 + y^2}$ ,  $0 \leq z \leq 1$ , når tettheten er proporsjonal med avstanden til origo.

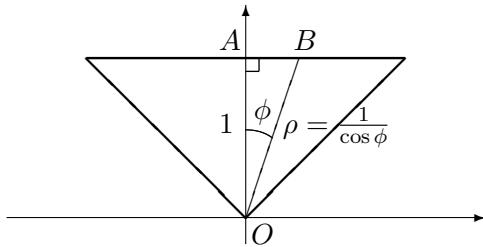
At tettheten er proporsjonal med avstanden til origo, betyr at den har formen  $f(x, y, z) = k\sqrt{x^2 + y^2 + z^2}$ . Vi har dermed

$$M = \iiint_A k\sqrt{x^2 + y^2 + z^2} dx dy dz$$

Det er mulig å løse dette integralet både med sylinderkoordinater og kulekoordinater. Vi velger kulekoordinater som viser seg å gi de enkleste regnogene. Vi ser at  $f(x, y, z) = k\rho$ , og vi vet at Jacobi-determinanten er  $\rho^2 \sin \phi$ . Dermed har vi

$$M = k \iiint_S \rho^3 \sin \phi d\rho d\phi d\theta$$

Grensene for  $\phi$  og  $\theta$  er greie å finne. Siden sidekanten i kjeglen danner en vinkel  $\frac{\pi}{4}$  med  $z$ -aksen, må  $\phi$  gå fra 0 til  $\frac{\pi}{4}$ . Alle  $\theta$ -verdier er mulige, så  $\theta$  går fra 0 til  $2\pi$ . For å finne grensene til  $\rho$  ser vi på figur 1 som viser et snitt gjennom kjeglen (de tykke strekene er kjegleveggene). Enkel trigonometri på



Figur 1: Tverrsnitt gjennom kjeglen

trekanten  $OAB$  viser at  $|OB| = \frac{1}{\cos \phi}$ . Dermed går  $\rho$  fra 0 til  $\frac{1}{\cos \phi}$  og vi har

$$M = k \int_0^{\frac{\pi}{4}} \left[ \int_0^{\frac{1}{\cos \phi}} \left[ \int_0^{2\pi} \rho^3 \sin \phi d\theta \right] d\rho \right] d\phi$$

Integrasjonene blir nå ganske greie:

$$\begin{aligned} M &= 2k\pi \int_0^{\frac{\pi}{4}} \left[ \int_0^{\frac{1}{\cos\phi}} \rho^3 \sin\phi \, d\rho \right] d\phi = 2k\pi \int_0^{\frac{\pi}{4}} \left[ \frac{\rho^4}{4} \sin\phi \right]_{\rho=0}^{\rho=\frac{1}{\cos\phi}} d\phi = \\ &= \frac{k\pi}{2} \int_0^{\frac{\pi}{4}} \frac{\sin\phi}{\cos^4\phi} d\phi = \frac{k\pi}{2} \left[ -\frac{\cos^{-3}\phi}{-3} \right]_0^{\frac{\pi}{4}} = \frac{k\pi}{6} \left[ \frac{1}{\cos^3\phi} \right]_0^{\frac{\pi}{4}} \end{aligned}$$

der vi i siste integrasjon har brukt substitusjonen  $u = \cos\phi$ . Setter vi inn grensene, får vi

$$M = \frac{k\pi}{6} \left( \frac{1}{\left(\frac{\sqrt{2}}{2}\right)^3} - 1 \right) = \frac{k\pi}{6} (2\sqrt{2} - 1) \quad \clubsuit$$

Vi kan også bruke trippelintegraler til å regne ut massemiddelpunktet til et legeme. Dersom tettheten er  $f(x, y, z)$ , er koordinatene  $(\bar{x}, \bar{y}, \bar{z})$  til massemiddelpunktet gitt ved

$$\begin{aligned} \bar{x} &= \frac{\iiint_S x f(x, y, z) \, dx dy dz}{M} \\ \bar{y} &= \frac{\iiint_S y f(x, y, z) \, dx dy dz}{M} \\ \bar{z} &= \frac{\iiint_S z f(x, y, z) \, dx dy dz}{M} \end{aligned}$$

der  $M = \iiint_S f(x, y, z) \, dx dy dz$  er massen. Det tar mye tid å regne ut et massemiddelpunkt siden det er 4 integraler man skal beregnes. Ofte er det imidlertid mulig å slippe unna med færre beregninger dersom man er litt smart. Ønsker man f.eks. å finne massemiddelpunktet til legemet i eksemplet ovenfor, ser man at  $\bar{x}$  og  $\bar{y}$  er 0 av symmetrirunner. For å finne massemiddelpunktet er det derfor nok å regne ut  $\bar{z}$ . Dette er ikke så vanskelig nå som vi vet hva grensene skal være (husk at  $z = \rho \cos\phi$ ):

$$\begin{aligned} I &= \iiint_S z f(x, y, z) \, dx dy dz = \\ &= k \int_0^{\frac{\pi}{4}} \left[ \int_0^{\frac{1}{\cos\phi}} \left[ \int_0^{2\pi} \rho^4 \cos\phi \sin\phi \, d\theta \right] d\rho \right] d\phi \end{aligned}$$

Vi regner som ovenfor

$$\begin{aligned} I &= 2k\pi \int_0^{\frac{\pi}{4}} \left[ \int_0^{\frac{1}{\cos\phi}} \rho^4 \cos\phi \sin\phi \, d\rho \right] d\phi = \\ &= 2k\pi \int_0^{\frac{\pi}{4}} \left[ \frac{\rho^5}{5} \cos\phi \sin\phi \right]_{\rho=0}^{\rho=\frac{1}{\cos\phi}} d\phi = \frac{2k\pi}{5} \int_0^{\frac{\pi}{4}} \frac{\sin\phi}{\cos^4\phi} d\phi = \end{aligned}$$

$$= \frac{2k\pi}{5} \left[ -\frac{\cos^{-3}\phi}{-3} \right]_0^{\frac{\pi}{4}} = \frac{2k\pi}{15} \left[ \frac{1}{\cos^3\phi} \right]_0^{\frac{\pi}{4}} = \frac{2k\pi}{15}(2\sqrt{2}-1)$$

Dermed har vi

$$\bar{z} = \frac{\frac{2k\pi}{15}(2\sqrt{2}-1)}{\frac{k\pi}{6}(2\sqrt{2}-1)} = \frac{4}{5}$$

Massemiddelpunktet er altså  $(0, 0, \frac{4}{5})$ .

### Oppgaver til seksjon 6.11

1. Bruk et trippelintegral til å regne ut volumet til en kule med radius  $R$ .
2. Finn volumet til det området som ligger under grafen  $z = e^{x+y}$  og over trekanten med hjørner i  $(0, 0, 0)$ ,  $(1, 0, 0)$  og  $(\frac{1}{2}, \frac{1}{2}, 0)$ .
3. Finn volumet av den delen avkulen  $x^2 + y^2 + z^2 = R^2$  som ligger over kjegleflaten  $z = \sqrt{\frac{x^2+y^2}{3}}$ .
4. Finn volumet til området som ligger inni både sylinderen  $x^2 + y^2 = 1$  og kulen  $x^2 + y^2 + z^2 = 4$ .
5. Finn volumet til ellipsoiden  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$  ved å innføre nye variable  $u = \frac{x}{a}$ ,  $v = \frac{y}{b}$ ,  $w = \frac{z}{c}$ .
6. Finn massen til sylinderen  $x^2 + y^2 \leq 1$ ,  $0 \leq z \leq 1$ , når tettheten er  $f(x, y, z) = \frac{1}{(x^2+y^2+z^2)}$ .
7. Finn massemiddelpunktet til området

$$A = \{(x, y, z) \mid x^2 + y^2 \leq z \leq 4\}$$

når tettheten er 1.

8. Vis at massemiddelpunktet til en homogen pyramide med hjørner i  $(0, 0, 0)$ ,  $(a, 0, 0)$ ,  $(0, b, 0)$  og  $(0, 0, c)$  er  $\frac{1}{4}(a, b, c)$ .

9. (Eksamens i MAT1110 13/6, 2005) La  $D$  være det begrensede området i  $\mathbb{R}^3$  som er gitt ved ulikheterne  $x^2 + y^2 \leq z \leq \sqrt{2 - x^2 - y^2}$ . Finn volumet av  $D$ .

10. (Eksamens i MAT1110 16/8, 2007) Finn volumet til området over  $xy$ -planet som ligger under kuleflaten  $x^2 + y^2 + z^2 = 1$  og over kjegleflaten  $z^2 = 3x^2 + 3y^2$

11. (Prøveeksamen i MAT1110 V-06)  $R$  er området i rommet avgrenset av flatene  $z = 6 - x^2 - y^2$  og  $z = x^2 - 4x + y^2$ .

- a) Vis at integralet  $I = \iiint_R y \, dx dy dz$  er lik

$$\iint_S (6y - 2x^2y - 2y^3 + 4xy) \, dx dy$$

der  $S = \{(x, y) \in \mathbb{R}^2 \mid (x-1)^2 + y^2 \leq 4\}$ .

- b) Regn ut integralet i a).
- c)  $C$  er skjæringskurven mellom flatene  $z = 6 - x^2 - y^2$  og  $z = x^2 - 4x + y^2$ , og den er orientert mot klokken sett ovenfra. Vis at  $C$  har parametriseringen

$$\mathbf{r}(t) = (1 + 2 \cos t) \mathbf{i} + 2 \sin t \mathbf{j} + (1 - 4 \cos t) \mathbf{k}$$

og regn ut kurveintegralet  $\int_C \mathbf{F} \cdot d\mathbf{r}$  der  $\mathbf{F}(x, y, z) = z \mathbf{i} + y \mathbf{j} + x \mathbf{k}$ .

# Fasit

Finner du feil i fasiten, så send en e-post til [lindstro@math.uio.no](mailto:lindstro@math.uio.no).

## Kapittel 1

### Seksjon 1.1

- 1:**  $\mathbf{a} + \mathbf{b} = (-2, 3, 9, -5, -2)$ ,  $\mathbf{a} - \mathbf{b} = (4, -7, -1, -5, 4)$ ,  $s\mathbf{a} = (3, -6, 12, -15, 3)$ ,  
 $\mathbf{a} \cdot \mathbf{b} = 4$
- 2:**  $\mathbf{a} + \mathbf{b} = (7, 2, 5, -8, -5, 3)$ ,  $\mathbf{a} - \mathbf{b} = (7, -2, 3, 4, -5, 5)$ ,  $s\mathbf{a} = (-28, 0, -16, 8, 20, -16)$ ,  
 $\mathbf{a} \cdot \mathbf{b} = 12$
- 5:**  $\mathbf{m} \cdot \mathbf{p}$  der  $\mathbf{m} = (m_1, m_2, \dots, m_n)$  og  $\mathbf{p} = (p_1, p_2, \dots, p_n)$ .

### Seksjon 1.2

**1:** Skalarprodukt:  $-5$ , vinkel:  $\theta \approx 109.65^\circ$

**2:**  $10\sqrt{2}$

**3:** ca.  $67.8^\circ$

**4:** ca.  $31.8^\circ$ .

**5:** ca.  $70^\circ$ ,  $\mathbf{p} = \frac{1}{2}\mathbf{b}$

**6:**  $\frac{12}{7}\sqrt{14}$

**7:**  $(4, 3) = (2, 4) + (2, -1)$

**8:**  $(2, 2, 1) = (\frac{1}{2}, 0, -\frac{1}{2}) + (\frac{3}{2}, 2, \frac{3}{2})$

**9:**  $30^\circ$  og  $45^\circ$ ,  $\cos 15^\circ = \frac{\sqrt{6}+\sqrt{2}}{4}$

**10:** Utallige muligheter, f.eks.  $(1, 0, 3)$  og  $(0, 1, 2)$ . Sjekk dine svar ved å ta skalarproduktet med  $(3, 2, -1)$ .

**12:** b)  $-\frac{43}{2}$ , c)  $90^\circ$

**13:** Umulig ifølge trekantulikheten.

**14:** Umulig ifølge Schwarz' ulikhet.

**18:** For eksempel  $\mathbf{r}(t) = (-1 + 2t, -1 + 3t, 2 + t)$  (det er mange muligheter)

**19:** For eksempel  $\mathbf{r}(t) = (-3 + t, -2 - 2t, 5 - t, 8 + 3t)$  (det er mange muligheter).

Nei

**20:** For eksempel  $\mathbf{r}(t) = (2 + t, -1 + 9t, 3 - 5t)$  (det er mange muligheter).

**21:** For eksempel  $\mathbf{r}(t) = (7 - 5t, -3 + 4t, 2 - 3t, 4 - 5t, -2 + 7t)$  (det er mange muligheter).

**22:** For eksempel  $\mathbf{r}(t) = (5 + 2t, -2 + t)$  (det er mange muligheter)

**23:** For eksempel  $\mathbf{r}(t) = (3t, 2 - 2t)$  (det er mange muligheter)

**24:**  $y = -\frac{1}{2}x + \frac{1}{2}$

**25:** a) Skjæringspunkt (15, 24).

b) Skipene kolliderer ikke.

**26:**

a) Ja, kursene til flyene krysser hverandre i punktet (4000, 4000, 4000).

b) Nei, flyene kolliderer ikke. De kommer til møtestedet etter henholdsvis 40 sekunder og  $\frac{100}{3}$  sekunder.

**27:** c) ca.  $49^\circ$  og ca. 12.7 m.

### Seksjon 1.3

**1:**  $\begin{pmatrix} 4+5i \\ -3+4i \end{pmatrix}$ .

**2:**  $|\mathbf{a}| = \sqrt{15}$ ,  $|\mathbf{b}| = \sqrt{19}$

**3:**  $-1-i$

### Seksjon 1.4

**1:** a)  $(19, 3, 5)$  b)  $(-1, -6, -14)$

**2:**  $6\sqrt{5}$

**3:**  $\sqrt{3}$

**4:**  $(9, -5, 6)$

**5:**  $\mathbf{i} \times \mathbf{j} = \mathbf{k}$ ,  $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ ,  $\mathbf{k} \times \mathbf{i} = \mathbf{j}$

**6:** 0

**7:**  $\frac{7}{2}$

**8:**  $2x + y - z = 4$

**9:**  $3x - y + 2z = 3$

**10:** Hint: Bruk setning 1.4.4.

### Seksjon 1.5

**1:**  $2A = \begin{pmatrix} 6 & -4 & 8 \\ -2 & 0 & 12 \end{pmatrix}$ ,  $-3B = \begin{pmatrix} -6 & -6 & 3 \\ -12 & 0 & 18 \end{pmatrix}$

$A + B = \begin{pmatrix} 5 & 0 & 3 \\ 3 & 0 & 0 \end{pmatrix}$ ,  $A - B = \begin{pmatrix} 1 & -4 & 5 \\ -5 & 0 & 12 \end{pmatrix}$

**2:**  $4A - 3B = \begin{pmatrix} -5 & -18 \\ 14 & -1 \\ 3 & -4 \end{pmatrix}$ .

**3:**  $A^T = \begin{pmatrix} 1 & -1 \\ -2 & 8 \\ 6 & 6 \end{pmatrix}$ ,  $B^T = \begin{pmatrix} 2 & 4 & 1 \\ 2 & 0 & 7 \\ -1 & -6 & -3 \end{pmatrix}$

**4:**  $A^T = \begin{pmatrix} 1 & -5 & 2 \\ 0 & 7 & -1 \end{pmatrix}$ ,  $B^T = \begin{pmatrix} 4 & -1 & 2 \\ -1 & 3 & -6 \end{pmatrix}$

$(4A - B)^T = \begin{pmatrix} 0 & -19 & 6 \\ 1 & 25 & 2 \end{pmatrix}$

**5:** a)  $\begin{pmatrix} 1 \\ -7 \end{pmatrix}$

b)  $\begin{pmatrix} 6 \\ 7 \\ 22 \end{pmatrix}$   
c)  $\begin{pmatrix} 8 \\ -18 \\ 10 \end{pmatrix}$

**6:** I: 9.7 tonn, II: 11.8 tonn, III: 7.5 tonn

**7:** X: 88 vogner, Y: 56 vogner, Z: 56 vogner

**10:**  $\begin{pmatrix} 0.1 & 0 & 0.2 \\ 0.2 & 0.3 & 0.2 \\ 0.3 & 0.3 & 0.1 \\ 0.3 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{pmatrix}$

A: 2.6 enheter, B: 7.2 enheter, C: 7.4 enheter, D: 7.4 enheter, E: 5.4 enheter

**11: a)**  $\begin{pmatrix} 0.94 & 0 & 0.01 \\ 0.05 & 0.2 & 0 \\ 0.01 & 0.8 & 0.99 \end{pmatrix}$

b)  $\mathbf{v}_1 = \begin{pmatrix} 0.846 \\ 0.065 \\ 0.089 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0.79613 \\ 0.0553 \\ 0.14857 \end{pmatrix}$  (vi tar med flere siffer enn det som er rimelig for å gjøre svaret lett å kontrollere).

**12: a)**  $\begin{pmatrix} 0 & 20 & 50 & 10 \\ 0.05 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.3 & 0.1 \end{pmatrix}$

b) Unge: 300, unge voksne 250, voksne 0, eldre 3.

**13: a)**  $\begin{pmatrix} 0 & 0.1 & 0.5 \\ 0.1 & 0 & 0.5 \\ 0.9 & 0.9 & 0 \end{pmatrix}$

b) Per: 45%, Pål: 46%, Espen 9%.

## Seksjon 1.6

**1. a)**  $AB = \begin{pmatrix} 0 & -5 \\ 7 & -1 \end{pmatrix}, BA = \begin{pmatrix} -1 & -5 \\ 7 & 0 \end{pmatrix}$

b)  $AB = \begin{pmatrix} 1 & 4 & 1 \\ 3 & -5 & 0 \\ 1 & -7 & 1 \end{pmatrix}, BA = \begin{pmatrix} -5 & 2 & 3 \\ 3 & 1 & -2 \\ 3 & 1 & 1 \end{pmatrix}$

**2.**  $AB = \begin{pmatrix} 5 & 7 \\ 2 & 3 \end{pmatrix}.$

**3.**  $AB = \begin{pmatrix} 8 & -1 & -2 \\ 6 & 3 & 0 \\ -8 & 1 & 2 \end{pmatrix}.$

**4. a)**  $AB$  blir en  $8 \times 9$ -matrise.

- b)  $B$  er en  $3 \times 5$ -matrise  
c)  $B$  har 7 søyler

5. a)  $AB = \begin{pmatrix} 83 & 62 \\ 72 & 48 \\ 50 & 80 \end{pmatrix}$

b)  $AC = \begin{pmatrix} 3 & 81 & 41 \\ 0 & 72 & 32 \\ 21 & 36 & 51 \end{pmatrix}$

c) Uttrykket er ikke definert siden matrisene  $B$  og  $C$  ikke har samme dimensjon (dermed er ikke summen  $B + C$  definert).

d)  $(BC)^T = \begin{pmatrix} 6 & 27 \\ 72 & 54 \\ 42 & 69 \end{pmatrix}$

e)  $B^T C^T t$  er ikke definert siden  $B^T$  er en  $2 \times 2$ -matrise og  $C^T$  er en  $3 \times 2$ -matrise, så dimensjonene stemmer ikke overens.

f)  $(A + C^T)B = \begin{pmatrix} 89 & 86 \\ 153 & 102 \\ 96 & 144 \end{pmatrix}$

g)  $B(A^T - 2C) = \begin{pmatrix} 62 & -80 & -38 \\ 9 & -60 & -51 \end{pmatrix}$

7. Vi finner første rad i  $DE$  ved å multiplisere første rad i  $D$  med matrisen  $E$ :

$$(3, 1, 4) \begin{pmatrix} 2 & 7 & 1 \\ 8 & 2 & 8 \\ 1 & 8 & 2 \end{pmatrix} = (18, 55, 19)$$

Vi finner andre søyle i  $DE$  ved å multiplisere matrisen  $D$  med andre søyle i  $E$ :

$$\begin{pmatrix} 3 & 1 & 4 \\ 1 & 5 & 9 \\ 2 & 6 & 5 \end{pmatrix} \begin{pmatrix} 7 \\ 2 \\ 8 \end{pmatrix} = \begin{pmatrix} 55 \\ 89 \\ 66 \end{pmatrix}$$

Regner du ut hele matriseproduktet på vanlig måte, får du:

$$DE = \begin{pmatrix} 18 & 55 & 19 \\ 51 & 89 & 59 \\ 57 & 66 & 60 \end{pmatrix}$$

8. a) Første og andre søyle i produktmatrisen  $AB$  er også like.  
b) Andre søyle i produktmatrisen  $AB$  består også bare av nuller.

#### 9. 4

10.  $A^2 = \begin{pmatrix} 3 & -4 \\ -2 & 3 \end{pmatrix}, A^3 = \begin{pmatrix} 7 & -10 \\ -5 & 7 \end{pmatrix}.$

11. a)  $\mathbf{y} = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$

b)  $C = \begin{pmatrix} 9 & -3 \\ 8 & -3 \end{pmatrix}$

c)  $A\mathbf{y} = C\mathbf{x} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$

**12.** a)  $M = \begin{pmatrix} 0.7 & 0.05 & 0.1 \\ 0.15 & 0.75 & 0.15 \\ 0.1 & 0.1 & 0.7 \end{pmatrix}$

b)  $N = \begin{pmatrix} 1 & 0.2 & 0 \\ 0 & 0.8 & 0.05 \\ 0 & 0 & 0.95 \end{pmatrix}$

c)  $K = \begin{pmatrix} 0.73 & 0.2 & 0.13 \\ 0.125 & 0.605 & 0.155 \\ 0.095 & 0.095 & 0.665 \end{pmatrix}$

Fordeling: A: 258, B: 268.5, C: 313.5

d) Fordeling: A: 282.8, B: 243.3, C: 258.5 (vi tar med flere siffer enn det som er rimelig for å gjøre svaret lett å kontrollere).

## Seksjon 1.7

**3.** Matrisene  $A$ ,  $B$  og  $D$  er inverterbare med inverse matriser

$$A^{-1} = \begin{pmatrix} -2 & 7 \\ -1 & 3 \end{pmatrix}, B^{-1} = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.25 \end{pmatrix}, D^{-1} = \begin{pmatrix} -1 & 1 \\ 0.75 & -0.5 \end{pmatrix}$$

Matrisen  $C$  er singulær (siden den har en søyle som bare består av nuller).

**4.**  $A = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}$

**5.**  $(AB)^{-1} = B^{-1}A^{-1} = \begin{pmatrix} 14 & 59 \\ 4 & 17 \end{pmatrix}$

**10.** b)  $A^{-1} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$

c) Multipliserer vi med  $A^{-1}$  på begge sider av matriseligningen, får vi

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 19 \\ 7 \end{pmatrix}$$

## Seksjon 1.8

**1:** a) 14 b) 38 c) 0

**2:** 11

**3:**  $\frac{23}{2}$

**4:** 27

**5:** a) negativt, b) negativt

**6:** Hint: Tolk determinanten som et areal.

**8:** Hint: Utrykk arealet som en determinant

**9:** b) Ligningssystemet har enten ingen eller uendelig mange løsninger avhengig av konstantene  $c_1$  og  $c_2$ . Linjene  $a_1x + b_1y = c_1$  og  $a_2x + b_2y = c_2$  er nemlig enten parallelle (ingen løsninger) eller sammenfallende (uendelig mange løsninger).

**10:** a) 84 b) 20 c) 0

**11:** 7

**12:**  $\frac{5}{3}$

**13:** Positivt

**15:** Hint: Tolk determinanten som et volum.

**16:** -127

## Kapittel 2

### Seksjon 2.1

- 1:** a)  $D_f = \mathbf{R}^2 \setminus \{(0, 0)\}$   
 b)  $D_f = \{(x, y) \in \mathbf{R}^2 | x \neq y \text{ og } x \neq -y\}$   
 c)  $D_f = \{(x, y) \in \mathbf{R}^2 | x + y > 0\}$   
 d)  $D_f = \{(x, y) \in \mathbf{R}^2 | \text{det ikke finnes et heltall } k \text{ slik at } x - y = \frac{\pi}{2} + k\pi\}$   
 e)  $D_f = \mathbf{R}^3 \setminus \{\text{kuleflaten med sentrum i origo og radius } 5\}$

### Seksjon 2.3

- 1:** a) 20  
 b) 1  
 c) e  
 d) 1

### Seksjon 2.4

- 1:** a)  $\frac{\partial f}{\partial x} = 3x^2y + 3y^4$  og  $\frac{\partial f}{\partial y} = x^3 + 12xy^3$   
 b)  $\frac{\partial f}{\partial x} = \frac{2x+3x^2}{y}$  og  $\frac{\partial f}{\partial y} = -\frac{x^2+x^3}{y^2}$   
 c)  $\frac{\partial f}{\partial x} = -\sin(x + y^2)$  og  $\frac{\partial f}{\partial y} = -2y \sin(x + y^2)$   
 d)  $\frac{\partial f}{\partial x} = 2x \ln(xy^2) + x$  og  $\frac{\partial f}{\partial y} = \frac{2x^2}{y}$   
 e)  $\frac{\partial f}{\partial x} = e^{-z}$ ,  $\frac{\partial f}{\partial y} = e^{-z}$  og  $\frac{\partial f}{\partial z} = -(x + y)e^{-z}$   
 f)  $\frac{\partial f}{\partial x} = \frac{z^2}{(1+y^2) \cos^2(x)}$ ,  $\frac{\partial f}{\partial y} = -\frac{2yz^2 \tan x}{(1+y^2)^2}$  og  $\frac{\partial f}{\partial z} = \frac{2z \tan x}{1+y^2}$   
 g)  $\frac{\partial f}{\partial x} = \frac{z}{1+(x+y)^2}$ ,  $\frac{\partial f}{\partial y} = \frac{z}{1+(x+y)^2}$  og  $\frac{\partial f}{\partial z} = \arctan(x + y)$   
 h)  $\frac{\partial f}{\partial x} = -(z^2 + u)e^{-x+3y}$ ,  $\frac{\partial f}{\partial y} = 3(z^2 + u)e^{-x+3y}$ ,  $\frac{\partial f}{\partial z} = 2ze^{-x+3y}$  og  $\frac{\partial f}{\partial u} = e^{-x+3y}$

- 2:** a)  $\nabla f = (2xy, x^2)$   
 b)  $\nabla f = (\cos(xy^2z) - xy^2z \sin(xy^2z), -2x^2yz \sin(xy^2z), -x^2y^2 \sin(xy^2z))$   
 c)  $\nabla f = (w \cos ve^{u \cos v}, -wue^{u \cos v} \sin v, e^{u \cos v})$   
 d)  $\nabla f = (\frac{z_2 z_3}{1+z_1^2 z_2^2}, \frac{z_1 z_3}{1+z_1^2 z_2^2}, \arctan(z_1 z_2) + e^{z_3})$

- 3:** a)  $f'((1, 2); (3, -1)) = 11$   
 b)  $f'((1, 0); (-1, 1)) = -1$   
 c)  $f'((1, 0, 1); (1, 1, -1)) = -1$   
 d)  $f'((\frac{\pi}{2}, 1, 0); (2, 0, -1)) = -1$

- 4:** a) I retningen gitt av vektoren  $(24, 173)$   
 b) I retningen gitt av vektoren  $(1, 1, 0)$   
 c) I retningen gitt av vektoren  $(3, 0, -3, 1)$

**5:** Hint: Bruk at  $\Delta V \approx V'((r, h); (\Delta r, \Delta h))$  Anslått usikkerhet:  $\Delta V \approx 3.8m^3$

**6:** Tommelfingerregelen gjelder når høyden i centimeter er det dobbelte av vekten i kilo (f.eks høyde 180 cm og vekt 90 kilo). Er man lettare enn dette, er regelen "for snill" i den forstand at BMI går opp hvis man legger på en kilo og en centimeter.

**Seksjon 2.5**

- 1:** a)  $\frac{\partial^2 f}{\partial x^2} = 6y, \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 6x + 4y, \frac{\partial^2 f}{\partial y^2} = 4x$   
 b)  $\frac{\partial^2 f}{\partial x^2} = 0, \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = \cos y, \frac{\partial^2 f}{\partial y^2} = -x \sin y$   
 c)  $\frac{\partial^2 f}{\partial x^2} = (x^2 + 4x + 2)e^{x-y}, \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = -(x^2 + 2x)e^{x-y}, \frac{\partial^2 f}{\partial y^2} = x^2 e^{x-y}$   
 d)  $\frac{\partial^2 f}{\partial x^2} = 2z, \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 0, \frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x} = 2x, \frac{\partial^2 f}{\partial y^2} = -2z^2,$   
 $\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y} = -4yz, \frac{\partial^2 f}{\partial z^2} = -2y^2$

- 2:** a)  $(6xy + 6x^2yz + x^3yz^2)e^{xz}$ .  
 b)  $-20x^3y^4 \cos(xyz) + 10x^4y^5z \sin(xyz) + x^5y^6z^2 \cos(xyz)$

**Seksjon 2.6**

**1:** a)  $\mathbf{F}'(x, y) = \begin{pmatrix} 2xy & x^2 \\ 1 & 2y \end{pmatrix}$

b)  $\mathbf{F}'(x, y) = \begin{pmatrix} 2xye^{x^2y+z} & x^2e^{x^2y+z} & e^{x^2y+z} \\ yz^2 & xz^2 & 2xyz \end{pmatrix}$

c)  $\mathbf{F}'(x, y) = \begin{pmatrix} \arctan(xy) + \frac{xy}{1+x^2y^2} & \frac{x^2}{1+x^2y^2} \\ \ln y & \frac{x}{y} \\ y \cos y^2 & x \cos y^2 - 2xy^2 \sin y^2 \end{pmatrix}$

d)  $\mathbf{F}'(x, y, z, u) = \begin{pmatrix} y \sin(xu^2) + xyu^2 \cos(xu^2) & x \sin(xu^2) & 0 & 2x^2yu \cos(xu^2) \\ 0 & 0 & 2zu & z^2 \end{pmatrix}$

**Seksjon 2.7**

**1:**  $\frac{\partial k}{\partial x} = 8xy^2 + 1, \frac{\partial k}{\partial y} = 8x^2y + 2y$

**2:**  $\frac{\partial k}{\partial x} = 2ye^{-2y(x+z)} - (4xy^2 + 2yz)e^{-2y(x+z)},$   
 $\frac{\partial k}{\partial y} = 2xe^{-2y(x+z)} - 2(2xy + z)(x + z)e^{-2y(x+z)},$   
 $\frac{\partial k}{\partial z} = e^{-2y(x+z)} - 2(2xy^2 + yz)e^{-2y(x+z)}$

**3:**  $\frac{\partial h}{\partial x_2} = 54x_1^5x_2^2x_3^2 \sin x_3, \frac{\partial h}{\partial x_3} = 36x_1^5x_2^3x_3 \sin x_3 + 18x_1^5x_2^3x_3^2 \cos x_3$

**4:**  $\frac{\partial h}{\partial v} = e^{3uv^3}u \cos(uv) \left( 9v^2 \cos(uv) - 2 \sin(uv) \right)$

**5:**  $\mathbf{H}'(1, -2) = \begin{pmatrix} 13 & -7 \\ 10 & 0 \end{pmatrix}$

**6:**  $\mathbf{H}'(-1, -2, 1) = \begin{pmatrix} 5 & -13 & 3 \\ 2 & 6 & -2 \end{pmatrix}$

**Seksjon 2.8**

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

**2:**  $\begin{pmatrix} -1 & 0 \\ 2 & -2 \\ -3 & 4 \\ 4 & 7 \end{pmatrix}$

**3:**  $-\begin{pmatrix} 6 \\ 3 \end{pmatrix}$

**4:**  $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

**5:**  $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$

**6:**  $\begin{pmatrix} 2 \cos \theta & -2 \sin \theta \\ 2 \sin \theta & 2 \cos \theta \end{pmatrix}$

**7:**  $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

**8:**  $\begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}$

**10:**  $\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$

**11:**  $x = -\frac{3}{7}, y = \frac{1}{7}, z = \frac{1}{7}, u = \frac{2}{7}$

b)  $\mathbf{T}(\mathbf{e}_1) = \begin{pmatrix} -\frac{2}{7} \\ -\frac{4}{7} \\ \frac{3}{7} \end{pmatrix}, \mathbf{T}(\mathbf{e}_2) = \begin{pmatrix} \frac{3}{7} \\ -\frac{1}{7} \\ \frac{2}{7} \end{pmatrix}$

c)  $\begin{pmatrix} -\frac{2}{7} \\ -\frac{4}{7} \\ -\frac{1}{7} \end{pmatrix}$

**12:** a)  $x = \frac{1}{4}, y = \frac{1}{4}$ .  $\mathbf{T}^4(\mathbf{e}_1) = \begin{pmatrix} 353 \\ -136 \end{pmatrix}$

b)  $u = -\frac{1}{2}, v = \frac{1}{2}$ .  $\mathbf{T}^4(\mathbf{e}_2) = \begin{pmatrix} -544 \\ 353 \end{pmatrix}$

c)  $\begin{pmatrix} 353 & -544 \\ -136 & 353 \end{pmatrix}$

**14:** c)  $x = 1, y = 2$ .  $A^{10}\mathbf{a} = \begin{pmatrix} 3^{10} + 2 \\ 3^{10} - 2 \end{pmatrix} = \begin{pmatrix} 59051 \\ 59047 \end{pmatrix}$

## Seksjon 2.9

**1:**  $A = \begin{pmatrix} 2 & -3 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \mathbf{c} = -\begin{pmatrix} 7 \\ 2 \end{pmatrix}$

**2:**  $\mathbf{r}(t) = \begin{pmatrix} 5t + 11 \\ -4t - 10 \end{pmatrix}$

**3:**  $T_{\mathbf{a}}\mathbf{F} = \begin{pmatrix} -4x + 4y - 8 \\ 2x - 2y + 2 \end{pmatrix}$

**4:**  $T_{\mathbf{a}}\mathbf{F} = \begin{pmatrix} 4y \\ x + 2y \\ 24x + y - 32 \end{pmatrix}$

**5:** Matrise  $\begin{pmatrix} 1 & -2 \\ 4 & 1 \end{pmatrix}$ , konstantledd  $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$

**6:** Matrise  $\begin{pmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$ , konstantledd  $\begin{pmatrix} 3 \\ -1 \end{pmatrix}$

**7:** a) Matrise  $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ , konstantledd  $\begin{pmatrix} 6 \\ 0 \end{pmatrix}$

b) Matrise  $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ , konstantledd  $\begin{pmatrix} 0 \\ -4 \end{pmatrix}$

**8:** a) Matrise  $\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$ , konstantledd  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

## Kapittel 3

### Seksjon 3.1

**1:**

$$\mathbf{v}(t) = (3t^2, 2t)$$

$$v(t) = t\sqrt{9t^2 + 4}$$

$$\mathbf{a}(t) = (6t, 2)$$

$$a(t) = \frac{18t^2 + 4}{\sqrt{9t^2 + 4}}$$

**2:**

$$\mathbf{v}(t) = (-\sin t, \sin t + t \cos t)$$

$$v(t) = \sqrt{2 \sin^2 t + t \sin 2t + t^2 \cos^2 t}$$

$$\mathbf{a}(t) = (-\cos t, 2 \cos t - t \sin t)$$

$$a(t) = \frac{(3-t^2) \sin 2t + 2t(\cos 2t + \cos^2 t)}{2\sqrt{2 \sin^2 t + t \sin 2t + t^2 \cos^2 t}}$$

**3:**

$$\mathbf{v}(t) = (1, -e^{-t}, \cos t)$$

$$\mathbf{a}(t) = (0, e^{-t}, -\sin t)$$

**4:**

$$\mathbf{v}(t) = (\frac{1}{t}, 2t, -\sin t)$$

$$\mathbf{a}(t) = (-\frac{1}{t^2}, 2, -\cos t)$$

**7:**

b)  $\mathbf{v}(t) = (-a \sin t, b \cos t)$

$$v(t) = \sqrt{a^2 \sin^2 t + b^2 \cos^2 t} = \sqrt{(a^2 - b^2) \sin^2 t + b^2}$$

$$\mathbf{a}(t) = (-a \cos t, -b \sin t) = -\mathbf{r}(t)$$

$$a(t) = \frac{(a^2 - b^2) \sin 2t}{2\sqrt{a^2 \sin^2 t + b^2 \cos^2 t}}$$

c)  $s \approx 25.53$

**8:**  $\frac{904^{3/2} - 8}{27} \approx 1006.4$

**9:**

a)  $\mathbf{v}(t) = (1, -\tan t)$

$$v(t) = \frac{1}{\cos t}$$

b)  $\frac{1}{2} \ln \frac{2+\sqrt{2}}{2-\sqrt{2}} = \ln(\sqrt{2} + 1)$

**10:** a)  $\mathbf{v}(t) = (-2 \sin t, \sqrt{2} \cos t, \sqrt{2} \cos t)$

$v(t) = 2$

$\mathbf{a}(t) = -\mathbf{r}(t)$

b)  $4\pi$

c)  $x^2 + y^2 + z^2 = 4$

e) En sirkel med radius 2 om origo i planet y-z=0.

**11:**

a)  $\mathbf{v}(t) = (\cos t - t \sin t, \sin t + t \cos t, 1)$

$v(t) = \sqrt{2 + t^2}$

b)  $22.43$

c)  $\pi\sqrt{2 + 4\pi^2} + \ln(\sqrt{2}\pi + \sqrt{2\pi^2 + 1})$

**12:**

a)  $\mathbf{v}(t) = (r(1 - \cos t), r \sin t)$

$v(t) = r\sqrt{2}\sqrt{1 - \cos t}$

$\mathbf{a}(t) = (r \sin t, r \cos t)$

$a(t) = \frac{r\sqrt{2} \sin t}{2\sqrt{1 - \cos t}}$

d)  $8r$

**15:**

a)  $\mathbf{v}(t) = (v_0 \cos \alpha, v_0 \sin \alpha - gt)$

$\mathbf{a}(t) = (0, -g)$

b)  $\frac{v_0^2}{2g} \sin^2 \alpha$

c)  $\frac{v_0^2}{g}$

**16:**

$t = \frac{m}{k} \ln(1 + \frac{ku_2}{mg})$

$y_{max} = \frac{mu_2}{k} - \frac{m^2 g}{k^2} \ln(1 + \frac{ku_2}{mg})$

**17:**

b)  $\mathbf{r}_2(t) = (t + \frac{1}{\sqrt{1+\cos^2 t}}, \sin t + \frac{\cos t}{\sqrt{1+\cos^2 t}})$

d) Fra venstre mot høyre

**18:** Med origo midt i sirkelen og med startpunkt  $(b, 0)$ , får vi:  $x(t) = (a + b) \cos t - a \cos(\frac{a+b}{a}t)$

$y(t) = (a + b) \sin t - a \sin(\frac{a+b}{a}t)$

**19:**

a)  $\mathbf{r}'(t) = (1, 2t)$ . Finn uttrykket for  $f'(t)$  ved å derivere  $f(t) = |\mathbf{r}(t) - \mathbf{b}|^2 = (\mathbf{r}(t) - \mathbf{b}) \cdot (\mathbf{r}(t) - \mathbf{b})$  etter produktregelen.

b) Geometrisk beskrivelse: Vektoren fra  $\mathbf{b}$  til  $\mathbf{r}(t_0)$  står normalt på tangentvektoren  $\mathbf{r}'(t_0)$ .

c) Dersom  $y > \frac{1}{2}$ , så er  $(\pm\sqrt{y - \frac{1}{2}}, y - \frac{1}{2})$  de nærmeste punktene. Dersom  $y \leq \frac{1}{2}$ ,

så er  $\mathbf{0} = (0, 0)$  det nærmeste punktet.

### Seksjon 3.2

1.  $189t^6$
2.  $e^{\sin t \cos^2 t} (2 \sin t \cos t + \sin^2 t \cos^3 t - 2 \sin^4 t \cos t)$
3.  $2 \cos(t^2) e^{2t} - \sin(t \cos(t^2)) - t \cos(t^2) \cos(t \cos(t^2)) - 2t e^{2t} \sin(t^2) + 2t^3 \sin(t^2) \cos(t \cos(t^2))$
4.  $\frac{6t^6(3t+1)^2}{t^6+1} + 6t(3t+1) \ln(t^6+1) + (3t+1)^2 \ln(t^6+1)$
5.  $\begin{pmatrix} 2 \sin t \cos^2 t - \sin^3 t \\ \cos^2 t - \sin^2 t + \cos t \end{pmatrix}$
6.  $\begin{pmatrix} -4(1-t^2)t^2 + (1-t^2)^2 + \cos t \\ -\sin((1-t^2)\sin t)(-2t \sin t + (1-t^2)\cos t) \\ 2t(1-t^2) - 2t^3 \end{pmatrix}$
7. Avtagende

### Seksjon 3.3

1. 0
2. 160
3. 0
4.  $\frac{864}{35}$
5.  $\frac{1}{3} (2 + 4\pi^2)^{\frac{3}{2}} - \frac{2\sqrt{2}}{3}$
6.  $2\sqrt{2}(e^{-1} - 1)$
10.  $33.7K$
11.  $\frac{89}{45}$  liter
12. c)  $\pi$ , d) 0

### Seksjon 3.4

1. -48
2. 0
3.  $\frac{1544}{21}$
4.  $\cos 2 - \cos 2 + \frac{1}{2}(\ln 2)^2 - \frac{e}{2}(\sin 1 + \cos 1) + \frac{e^2}{2}(\cos 2 + \sin 2)$
5.  $\frac{\pi}{2} + \frac{\ln 2}{2} - 1$
6. 0
7.  $\frac{192}{5}$
8.  $\frac{\pi^2}{2}$
14.  $5K(\sqrt{17} - 1)$
15. c)  $\frac{1+2t+2t^2}{\sqrt{2+2t+t^2}}$ , d)  $K\sqrt{2}$ , e)  $\sqrt{2}$

### Seksjon 3.5

1. Konservert med potensialfunksjon  $\phi(x, y) = x^2y + x^2$
2. Ikke konservert
3. Konservert med potensialfunksjon  $\phi(x, y) = x \sin xy + 3x - 2y$
4. Konservert med potensialfunksjon  $\phi(x, y, z) = xy^2z + xz - 2y$
5. Konservert med potensialfunksjon  $\phi(x, y, z) = xy \sin z + 2xz + 3y$
6. Ikke konservert
7. 0
8.  $-\pi$

**9.** 0**10.** 3**11.** 0**12.** b)  $\frac{\pi}{2}$ 

- e)  $\arctan \frac{y}{x} = \begin{cases} -\arctan \frac{x}{y} + \frac{\pi}{2} & \text{når } (x, y) \text{ ligger i 1. og 3. kvadrant} \\ -\arctan \frac{x}{y} - \frac{\pi}{2} & \text{når } (x, y) \text{ ligger i 2. og 4. kvadrant} \end{cases}$
- e) For eksempel

$$\phi(x, y) = \begin{cases} \arctan(\frac{y}{x}) & \text{når } x > 0 \\ \arctan(\frac{y}{x}) + \pi & \text{når } x < 0 \\ \frac{\pi}{2} & \text{når } x = 0 \text{ og } y > 0 \end{cases}$$

Denne funksjonen har et sprang langs den negative  $y$ -aksen.

### Seksjon 3.6

**1.**  $\frac{(x+4)^2}{9} + \frac{(y-1)^2}{4} = 1$ . Ellipse med sentrum i  $(-4, 1)$  og halvakser  $a = 3$ ,  $b = 2$ .

Brennpunkter i  $(-4 - \sqrt{5}, 1)$  og  $(-4 + \sqrt{5}, 1)$ .

**2.**  $(y-1)^2 = 4(x+2)$ , parabel med toppunkt i  $(-2, 1)$ , brennvidde 1 og brennpunkt i  $(-1, 1)$

**3.**  $\frac{(x-1)^2}{4} - \frac{(y-2)^2}{4} = 1$ . Hyperbel med sentrum i  $(1, 2)$ ,  $a = 2$ ,  $b = 2$ ,  $c = 2\sqrt{2}$ . Brennpunkt  $(1 - 2\sqrt{2}, 2)$  og  $(1 + 2\sqrt{2}, 2)$ . Asymptoter  $(y-2) = \pm(x-1)$

**4.**  $\frac{(y+1)^2}{9} - \frac{(x-3)^2}{16} = 1$ . Hyperbel med sentrum i  $(3, -1)$ ,  $a = 4$ ,  $b = 3$ ,  $c = 5$ . Brennpunkt  $(3, -6)$  og  $(3, 4)$ . Asymptoter  $(y+1) = \pm\frac{3}{4}(x-3)$

**5.**  $\frac{(x+1)^2}{5} + \frac{(y-2)^2}{3} = 1$ . Ellipse med sentrum i  $(-1, 2)$ ,  $a = \sqrt{5}$ ,  $b = \sqrt{3}$ ,  $c = \sqrt{2}$ . Brennpunkt  $(-1 - \sqrt{2}, 2)$  og  $(-1 + \sqrt{2}, 2)$ .

**6.**  $(x+2)^2 = -2(y-4)$ . Parabel med toppunkt  $(-2, 4)$ , brennvidde  $\frac{1}{2}$ , brennpunkt  $(-\frac{5}{2}, 4)$

**7.** Ligningen kan omformes til  $\frac{(x-1)^2}{3} + \frac{(y+2)^2}{9} = -1$  som ikke har noen løsninger,  
**10.** Parallelt med aksen og så brutt gjennom  $F$ .

**11.** Like lang tid

**13.** a)  $3\sqrt{21}$

b)  $(0, -18)$

e)  $(\frac{24}{5}, -9)$

**14.** a) Ellipse med ligning  $\frac{x^2}{16} + \frac{(y-3)^2}{25} = 1$

c)  $S(\frac{4\sqrt{5}}{3}, \frac{19}{3})$ ,  $B(\frac{40\sqrt{5}}{21}, \frac{190}{21})$

**15.** a) Hyperbel med brennpunkter i  $A$  og  $B$ , brennvidde  $c$  og halvakse  $a$ .

### Seksjon 3.7

**5.** a)  $4x - y + z = 4$

b)  $-x + y + z = 0$

c)  $12x - 12y + z = 32$

### Seksjon 3.9

Husk at flater kan parametriseres på mange måter, og at det derfor kan være andre riktige svar enn de som er oppgitt her!

**1.**  $\mathbf{r}(x, y) = x\mathbf{i} + y\mathbf{j} + (x^2 + y^2)\mathbf{k}$  for  $x, y \in \mathbb{R}$  og

$\mathbf{r}(r, \theta) = r \cos \theta \mathbf{i} + r \sin \theta \mathbf{j} + r^2 \mathbf{k}$  for  $r \geq 0$ ,  $0 \leq \theta < 2\pi$ .

**2.**  $\mathbf{r}(\phi, \theta) = 2 \sin \phi \cos \theta \mathbf{i} + 2 \sin \phi \sin \theta \mathbf{j} + 2 \cos \phi \mathbf{k}$ , der  $0 \leq \phi \leq \frac{\pi}{2}$ ,  $0 \leq \theta \leq \frac{\pi}{2}$

3.  $\mathbf{r}(\theta, z) = \cos \theta \mathbf{i} + \sin \theta \mathbf{j} + z \mathbf{k}$ , der  $0 \leq \theta < 2\pi, 0 \leq z \leq 1$
4.  $\mathbf{r}(v, x) = x \mathbf{i} + 3 \cos v \mathbf{j} + 3 \sin v \mathbf{k}$ , der  $0 \leq v < 2\pi, x \in \mathbb{R}$
5.  $\mathbf{r}(r, v) = r \mathbf{i} + r \cos v \mathbf{j} + r \sin v \mathbf{k}$ , der  $0 \leq v < 2\pi, r \geq 0$
6.  $\mathbf{r}(\phi, \theta) = a \sin \phi \cos \theta \mathbf{i} + b \sin \phi \cos \theta \mathbf{j} + c \cos \phi \mathbf{k}$ , der  $0 \leq \phi \leq \pi, 0 \leq \theta < 2\pi$
7. En cylinder med radius 2 og med akse langs  $x$ -aksen. Høyden er 2 (strekker seg fra  $x = 0$  til  $x = 2$ ).
8.  $\mathbf{r}(\phi, \theta) = 2 \sin \phi \cos \theta \mathbf{i} + 2 \sin \phi \cos \theta \mathbf{j} + 2 \cos \phi \mathbf{k}$ , der  $0 \leq \phi \leq \frac{\pi}{6}, 0 \leq \theta < 2\pi$

## Kapittel 4

### Seksjon 4.1

1.  $x = 25, y = -5, z = 12$
2.  $x = -\frac{7}{2}, y = -\frac{11}{2}, z = \frac{1}{2}$
3.  $z$  velges fritt,  $y = 2z - \frac{5}{4}, x = z - \frac{7}{2}$
4. Ingen løsninger
5.  $x = \frac{1}{5}(a + 2b + 2c), y = \frac{1}{5}(2a + b + 2c), z = \frac{1}{5}(2a + 2b + c)$
6.  $x = \frac{9}{13}, y = \frac{57}{65}, z = \frac{27}{65}$

### Seksjon 4.2

1. Alle unntatt  $C$  og  $F$  er på trappeform.
2. a)  $\begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix}$ , b)  $\begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$ , c)  $\begin{pmatrix} 1 & 1 & -2 & 3 \\ 0 & 1 & -7 & 6 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ ,  
d)  $\begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & 0 \end{pmatrix}$
3.  $x = \frac{1}{6}, y = \frac{1}{6}, z = \frac{1}{2}$
4.  $z$  kan velges fritt,  $x = -z, y = -\frac{1}{2} - \frac{z}{2}$ .
5. Ingen løsninger
6.  $u$  kan velges fritt, men da er  $x = 28 + u, y = -5 - u, z = 9 + u$ .
7. Ja
8. Nei
9.  $u$  kan velges fritt, men da er  $x = 3 - 2u, y = 1, z = 2 - u$ .
10. 60 biler i  $A$ , 40 biler i  $B$  og 20 biler i  $C$ .

### Seksjon 4.3

**1.** Alle unntatt  $C$  og  $F$  er på redusert trappeform

**2.** a)  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ , b)  $\begin{pmatrix} 1 & 0 & -3 & 1 \\ 0 & 1 & -3 & 2 \end{pmatrix}$ , c)  $\begin{pmatrix} 1 & 0 & 0 & -\frac{8}{3} \\ 0 & 1 & 0 & \frac{5}{3} \\ 0 & 0 & 1 & \frac{5}{3} \end{pmatrix}$  d)  $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

**4.** Ja

**5.** Nei

**6.**  $u$  kan velges fritt, men da må man sette  $x = \frac{1}{2} - \frac{7}{2}u$ ,  $y = \frac{15}{4} - \frac{15}{4}u$ ,  $z = -\frac{5}{4} + \frac{u}{4}$ .

**7.**  $v$  og  $y$  kan velges fritt, de andre variablene er da gitt ved:  $x = 7 - y - 2v$ ,  $z = 26 - 7v$ ,  $u = 10 - 2v$ .

### Seksjon 4.4

**1.** a)  $\begin{pmatrix} 3 \\ -4 \end{pmatrix}$ , b)  $\begin{pmatrix} 2 \\ -\frac{4}{3} \end{pmatrix}$ , c)  $\mathbf{b} = \begin{pmatrix} 4 \\ -\frac{8}{3} \\ \frac{7}{3} \end{pmatrix} + u \begin{pmatrix} 0 \\ -\frac{5}{3} \\ \frac{1}{3} \end{pmatrix}$  der  $u \in \mathbb{R}$ , d) Ingen løsning

**2.**  $\begin{pmatrix} \frac{4}{7} \\ -\frac{5}{7} \end{pmatrix}, \begin{pmatrix} -\frac{6}{7} \\ \frac{11}{7} \end{pmatrix}$

**3.**  $\frac{1}{3} \begin{pmatrix} -11 \\ 2 \\ -7 \end{pmatrix}, \frac{1}{3} \begin{pmatrix} -11 \\ 5 \\ -7 \end{pmatrix}$

**4.** a)  $\begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

b) Løsning for  $h = 7$ : Vi kan velge  $x_3$  fritt, og da blir:  $x_1 = x_3$ ,  $x_2 = -2x_3$ ,  $x_4 = 1$ .

**5.** a)  $\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & a^2 - a - 2 & 1 \\ 0 & 0 & -a^2 + a & a \end{pmatrix}$

b) Entydig løsning for  $a \neq 0$  og  $a \neq 1$ , uendelig mange løsninger for  $a = 0$ , ingen løsninger for  $a = 1$ .

**6.** a)  $\begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & a^2 + 3a - 4 & 3 - 3a \end{pmatrix}$

b) Entydig løsning når  $a \neq 1$  og  $a \neq -4$ , uendelig mange løsninger for  $a = 1$ , ingen løsninger for  $a = -4$ .

**Seksjon 4.5**

**1.** a)  $\frac{1}{7} \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$ , b)  $\frac{1}{3} \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$  c) Ikke inverterbar

**2.** a)  $\begin{pmatrix} 0 & -0.6 & 0.2 \\ -1 & 0.2 & 0.6 \\ 1 & 0.2 & -0.4 \end{pmatrix}$ , b) Ikke inverterbar, c)  $\begin{pmatrix} 0.75 & -1.5 & -0.25 \\ 1.25 & -1.5 & -0.75 \\ -0.5 & 1 & 0.5 \end{pmatrix}$ , d)

$$\begin{pmatrix} -\frac{5}{9} & \frac{7}{9} & -\frac{1}{9} \\ -\frac{1}{9} & \frac{5}{9} & -\frac{2}{9} \\ \frac{2}{9} & -\frac{1}{9} & \frac{4}{9} \end{pmatrix}$$

**6.** a)  $B^{-1} = \frac{1}{3} \begin{pmatrix} 3 & -2 & 2 \\ 0 & 1 & -1 \\ 0 & 2 & 1 \end{pmatrix}$  b)  $x = 5, y = 0, z = 3$

c) Én løsning for  $a \neq -3$  (uansett  $b$ ). Når  $a = -3$ , er det uendelig mange løsninger når  $b = \pm 2$ , og ingen løsninger når  $b \neq \pm 2$ ,

**9. b)**  $A^{-1} = \frac{1}{4} \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

c)  $x = \frac{7}{4}, y = 2, z = \frac{9}{4}$   
d)  $a = -2, b = 2, c = 6$

**Seksjon 4.6**

**1.**  $\mathbf{b} = -\mathbf{a}_1 + \frac{1}{2}\mathbf{a}_2$

**2.**  $\mathbf{b} = (-1)\mathbf{a}_1 + 2\mathbf{a}_2 + (-1)\mathbf{a}_3$

**3.** a) Ja, b) Nei, c) Ja

**4.** Koeffisientene er  $x_1 = \frac{100}{9}, x_2 = \frac{17}{9}, x_3 = \frac{11}{27}, x_4 = -\frac{134}{27}$

**5.** Koeffisientene er:  $x_1 = 1.2047, x_2 = -0.2781, x_3 = 0.0795, x_4 = 0.3222$ .

**6.** Ja

**7.** a) Ja, b) Nei, c) Nei, d) Ja., e) Ja.

**8.** a) F.eks. vektor 1 og 3, b) F.eks de tre første, c) F.eks. vektor 1, 2 og 4

**9.** a) Nei (for mange elementer), b) Nei (for få elementer), c) Ja, d) Ja

**10. a)** Legg f.eks. til  $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$

b) Legg f.eks. til  $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$

**11.** b)  $\mathbf{e}_1 = \frac{1}{2}\mathbf{v}_1 + \frac{1}{2}\mathbf{v}_2$ ,  $\mathbf{e}_2 = \frac{1}{2}\mathbf{v}_1 - \frac{1}{2}\mathbf{v}_2$ , d)  $\frac{1}{2} \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$

### Seksjon 4.7

**1.** Det er mange riktige svar på disse oppgavene.

a) For eksempel  $\mathbf{a}_1, \mathbf{a}_2$

b) For eksempel  $\mathbf{i}, \mathbf{j}$

c) For eksempel  $(0, 1, 2), (2, 1, 0)$

d) For eksempel  $\begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$

e) For eksempel  $(3, -1), (2, 0)$

**2.** a) 2, b) 2, c) 2

**5.**  $\mathbf{v} = \sqrt{2}\mathbf{v}_1 + 2\sqrt{2}\mathbf{v}_2$

### Seksjon 4.8

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

**3.**  $\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

### Seksjon 4.9

**1.** a) 12, b) -5, c) -7

**2.** a) 9, b) -9, c) -9

**3.a)** 1, b) 30, c) -4

**11.** d)  $x = \frac{22}{5}, y = \frac{8}{5}$

### Seksjon 4.10

**1. a)**  $\lambda_1 = 1, \lambda_2 = 3, \mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

b)  $\lambda_1 = 3, \lambda_2 = -1, \mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

c)  $\lambda_1 = 5, \lambda_2 = 1, \mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

d)  $\lambda_1 = 2, \mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

e)  $\lambda_1 = 1 + \sqrt{3}, \lambda_2 = 1 - \sqrt{3}, \mathbf{v}_1 = \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}$

f)  $\lambda_1 = 4 + i, \lambda_2 = 4 - i, \mathbf{v}_1 = \begin{pmatrix} 2 \\ -1+i \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 2 \\ -1-i \end{pmatrix}$

**2.** a)  $\lambda_1 = 4, \lambda_2 = 0, \lambda_3 = 1, \mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -2 \\ 5 \\ 1 \end{pmatrix}$

b)  $\lambda_1 = -2, \lambda_2 = 3, \lambda_3 = 2, \mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 4 \\ 1 \\ -5 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -1 \\ 1 \\ 4 \end{pmatrix}$

c)  $\lambda_1 = 0, \lambda_2 = \sqrt{5}, \lambda_3 = -\sqrt{5}, \mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} \frac{\sqrt{5}-1}{2} \\ \sqrt{5}-1 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} \frac{\sqrt{5}+1}{2} \\ \sqrt{5}+1 \\ -1 \end{pmatrix}$

**4.** a) Egenverdier  $\lambda_1 = 3, \lambda_2 = -2$ , egenvektorer  $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ ,

lineærkombinasjon  $\mathbf{x} = \frac{3}{5}\mathbf{v}_1 - \frac{11}{5}\mathbf{v}_2$

b) Egenverdier  $\lambda_1 = 5, \lambda_2 = 3$ , egenvektorer  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ , lineærkombinasjon  $\mathbf{x} = -4\mathbf{v}_1 + 2\mathbf{v}_2$

c) Egenverdier  $\lambda_1 = 1, \lambda_2 = -2, \lambda_3 = -1$ , egenvektorer  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 6 \\ 5 \end{pmatrix}$ , lineærkombinasjon  $\mathbf{x} = -\frac{1}{2}\mathbf{v}_1 + \mathbf{v}_2 + \frac{1}{2}\mathbf{v}_3$

**6.**  $M = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}, D = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}$

**7.** Egenvektorene er vanligvis ikke de samme. Vektoren  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  er f.eks. en egenvektor for matrisen  $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ , men ikke for  $A^T$ .

**10.** Egenvektorene til  $B$  er på formen  $\mathbf{v} = P^{-1}\mathbf{u}$  der  $\mathbf{u}$  er en egenvektor for  $A$ .

**12.**  $\lambda_1 = 5, \lambda_2 = -3, \lambda_3 = -1, \mathbf{v}_1 = \begin{pmatrix} 5 \\ 8 \\ 11 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$

## Seksjon 4.11

**1.**  $x_n = -4^n + 6 \cdot (-1)^n, y_n = -4^n - 4 \cdot (-1)^n$

**2.**  $x(t) = \frac{13}{2}e^{5t} - \frac{11}{2}e^{-3t}, y(t) = \frac{13}{4}e^{5t} + \frac{11}{4}e^{-3t}$

**3.** a) Egenverdier  $\lambda_1 = 1, \lambda_2 = 0.9$ , egenvektorer  $\mathbf{v}_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

b)  $\mathbf{r}_0 = 2000\mathbf{v}_1 - 1000\mathbf{v}_2$

$x_n = 4000 - 1000 \cdot 0.9^n, y_n = 2000 - 1000 \cdot 0.9^n$ . Bestandene stabiliserer seg på henholdsvis 4000 og 2000 dyr.

**4.** b) Egenverdiene er  $\lambda_1 = 1$  og  $\lambda_2 = 0.8$ . Egenvektorer:  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ .

c)  $\begin{pmatrix} 2 \\ 4 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (-1) \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ ,  $M^n \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 - (0.8)^n \\ 3 + (0.8)^n \end{pmatrix}$ .

d) 12 ganger

**5.** a) Egenverdier  $\lambda_1 = \frac{19}{18}$ ,  $\lambda_2 = -\frac{1}{18}$ , egenvektorer  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

c)  $x_n = 5 \left( \frac{19}{18} \right)^n + 3 \left( -\frac{1}{18} \right)^n$ ,  $y_n = 5 \left( \frac{19}{18} \right)^n - 3 \left( -\frac{1}{18} \right)^n$  der  $x_n$  og  $y_n$  er målt i millioner.

d) 50%

**6.** a)  $M = \begin{pmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{pmatrix}$

b) 500

c)  $\lambda_1 = 1$ ,  $\lambda_2 = \frac{1}{2}$ ,  $\mathbf{v}_1 = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

d)  $x_n = 600 - 100 \cdot \left( \frac{1}{2} \right)^n$ ,  $y_n = 400 + 100 \cdot \left( \frac{1}{2} \right)^n$

e)  $P_n = 10^6 (6 - 0.5^n - 1.6 \cdot 1.1^n - 0.4 \cdot (0.55)^n)$

**7.** a)  $M = \begin{pmatrix} 0 & 3 & 4 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$

b) Egenverdier  $\lambda_1 = 2$ ,  $\lambda_2 = -1$ , egenverdier  $\mathbf{v}_1 = \begin{pmatrix} 8 \\ 4 \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$

c)  $\begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \begin{pmatrix} 16 \cdot 2^n + 8 \cdot (-1)^n \\ 8 \cdot 2^n - 8 \cdot (-1)^n \\ 2 \cdot 2^n + 4(-1)^n \end{pmatrix}$

**8.** a)  $M = \begin{pmatrix} 1 & 0.1 & 0 \\ 0.1 & 1 & 0.1 \\ 0 & 0.1 & 1 \end{pmatrix}$

b)  $\lambda_1 = 1$ ,  $\lambda_2 = 1 + \frac{\sqrt{2}}{10}$ ,  $\lambda_3 = 1 - \frac{\sqrt{2}}{10}$ ,  $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ ,  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix}$ ,  $\mathbf{v}_3 = \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix}$

c)  $\begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \frac{1}{2}(x_0 - z_0) \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + (1 + \frac{\sqrt{2}}{10})^n (\frac{1}{4}(x_0 + z_0) + \frac{\sqrt{2}}{4}y_0) \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix} + (1 - \frac{\sqrt{2}}{10})^n (\frac{1}{4}(x_0 + z_0) - \frac{\sqrt{2}}{4}y_0) \begin{pmatrix} 1 \\ -\sqrt{2} \\ 1 \end{pmatrix}$

d)  $k_1 = \frac{\sqrt{2}}{2}$ ,  $k_2 = \sqrt{2}$

**9.** c)  $x(t) = 6000e^{0.01t} - 1000e^{-0.01t}$ ,  $y(t) = 2000e^{0.01t} - 1000e^{-0.01t}$

- 10.** a)  $x(t) = 500e^t(\cos t + 2 \sin t)$ ,  $y(t) = 500e^t(2 \cos t - \sin t)$   
 b)  $x(t) = 1250 - 750e^{-4t}$ ,  $y(t) = 625 + 375e^{-4t}$   
 c) Modell a) gir  $y(t) = 0$  når  $\tan t = 2$ , dvs. etter ca. 1.1 år. Modellen i b) gir grenseverdier 1250 og 625.

- 11.** En utvidet versjon av denne oppgaven ligger på  
<http://www.uio.no/studier/emner/matnat/math/MAT1110/v06/Oblig1.pdf>  
 med løsningsforslag på  
<http://www.uio.no/studier/emner/matnat/math/MAT1110/v06/Oblig1Los.pdf>

## Kapittel 5

### Seksjon 5.1

- 1.** a) lukket b) åpen c) hverken lukket eller åpen d) lukket e) åpen f) lukket g)  
 hverken lukket eller åpen h) åpen i) lukket

- 2.** a)  $(2, -\frac{3}{2})$  b)  $(1, -2)$  c)  $(1, 1, e^{-\frac{1}{2}})$

### Seksjon 5.4

- 1.** Fikspunkt  $(-1, 1)$ .  
**2.** Likevektpunkt:  $x \approx 297, y \approx 3970$

### Seksjon 5.6

- 1.** a) Nullpunkter:  $0, \pm 1$ .  
**2.** b)  $(\pm 0.7862, 0, 6180)$  c)  $(\pm \sqrt{\frac{\sqrt{5}-1}{2}}, \frac{\sqrt{5}-1}{2})$   
**3.** a) Førstekomponenten er null på alle linjer  $x + y = k\pi$ ,  $k \in \mathbb{Z}$ , mens annenkomponenten er null på alle linjer  $x - y = \frac{\pi}{2} + m\pi$ ,  $m \in \mathbb{Z}$ . Funksjonen er null i alle skjæringspunkter mellom slike linjer.

### Seksjon 5.7

- 1.**  $\mathbf{G}'(1, -2) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $\mathbf{H}'(1, -2) = -\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$   
**2.**  $\mathbf{G}'(1, -1) = \frac{1}{3} \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$ ,  $\mathbf{H}'(1, -1) = \frac{1}{3} \begin{pmatrix} -1 & 4 \\ -1 & 1 \end{pmatrix}$   
**3.**  $f'(x_0) = -\frac{3x_0^2}{3y_0^2+1}$   
**4.**  $\frac{\partial g}{\partial x} = \frac{4}{3}, \frac{\partial g}{\partial y} = -\frac{4}{3}$

**5.** a)  $A^{-1} = \frac{1}{3} \begin{pmatrix} 2 & 0 & 1 \\ -5 & 3 & -1 \\ 1 & 0 & -1 \end{pmatrix}$

b)  $\mathbf{G}(0, \frac{1}{2}, 2) = A^{-1}$

**6.**  $k = -\frac{b^2 x_0}{a^2 y_0}$

**7.**  $k = \frac{b^2 x_0}{a^2 y_0}$

**8.**  $k = -\frac{2a}{y_0}$

**10.**  $\frac{\partial z}{\partial x} = \frac{1-3yz}{3(xy-z^2)}$  og  $\frac{\partial z}{\partial y} = \frac{2y-3xz}{3(xy-z^2)}$  forutsatt at  $z^2 \neq xy$ .

**11.**  $\frac{\partial z}{\partial x} = -\frac{4x}{2z+e^{-z}}$  og  $\frac{\partial z}{\partial y} = -\frac{6y}{2z+e^{-z}}$  forutsatt at  $2z + e^{-z} \neq 0$ .

**12.**  $\frac{\partial u}{\partial x} = \frac{13}{8}$  og  $\frac{\partial v}{\partial x} = \frac{5}{4}$ .

**13.** b)  $x'(t) = -\frac{\frac{\partial \phi}{\partial y}(x(t), y(t))}{\frac{\partial \phi}{\partial x}(x(t), y(t))} y'(t) - \frac{\frac{\partial \phi}{\partial z}(x(t), y(t))}{\frac{\partial \phi}{\partial x}(x(t), y(t))} z'(t)$

## Seksjon 5.9

**1.** a)  $(2, -1)$  b)  $(0, 0)$  c)  $(-\frac{3}{2}, 1)$  d)  $(-1, 0)$ , e)  $(\frac{1}{4}, -4)$

**2.** a) Min. i  $(1, -2)$  b) Sadelpunkt i  $(-1, 1)$  c) Min. i  $(0, 0)$  d) Maks. i  $(\frac{1}{2}, -\frac{1}{2})$  e) Sadelpunkt i  $(-2, 0)$

**3.**  $(1, -2)$ , lokalt minimum.

**4.**  $(0, 0)$  er et lokalt minimum,  $(-\frac{4}{3}, -\frac{4}{3})$  er et sadelpunkt.

**5.**  $(1, -\frac{1}{2})$ , sadelpunkt.

**6.**  $(-1, 0)$ , lokalt minimum.

**7.** Lokalt minimum i  $(0, 0, 0)$ , sadelpunkter i  $(2, 2, 2), (2, -2, -2), (-2, 2, -2), (-2, -2, 2)$

**8.** a) Stasjonære punkter:  $(0, 0), (-2, 0), (-1, -1)$

b) De to første punktene er sadelpunkter, det siste et lokalt maksimum.

**9.** a)  $(0, 0), (-2, 0)$

b)  $(0, 0)$  er et lokalt (og faktisk et globalt) minimum,  $(-2, 0)$  er et sadelpunkt.

**10.** a)  $(0, 0)$  er et sadelpunkt,  $(\pm\sqrt{2}, 0)$  er lokale (og globale) maksimumspunkter,  $(0, \pm\sqrt{2})$  er lokale (og globale) minimumspunkter.

b) Minimumsverdi:  $-2e^{-1}$  i  $(0, \pm\sqrt{2})$ . Maksimumsverdi:  $e^{-\frac{1}{2}}$  i  $(\pm 1, 0)$

**11.** Sidene i grunnflaten skal være  $\sqrt[3]{2V}$  og høyden  $\frac{\sqrt[3]{2V}}{2}$ .

**12.** a)  $\frac{\partial L}{\partial x} = 4 - \frac{1000}{x^2 y}$ ,  $\frac{\partial L}{\partial y} = 2 - \frac{1000}{x y^2}$   
 b)  $x = 5, y = z = 10$

**13.** a)  $\frac{\partial A}{\partial x} = 28 - 4x - 3y$ ,  $\frac{\partial A}{\partial y} = 28 - 4y - 3x$ , lokalt maksimum for  $x = y = 4$   
 b) Maksimalt areal  $A = 122$  for  $x = y = 4$

**14.**  $l = \frac{60}{\sqrt[3]{78}}$ ,  $h = \frac{50}{\sqrt[3]{78}}$

**15.** a)  $\frac{\partial P}{\partial x} = \frac{60}{x^2} - 60y$ ,  $\frac{\partial P}{\partial y} = \frac{60}{y^2} - 60x$   
 b) Maksimalverdi  $P = 1140$  for  $x = y = 1$

**16.**  $x = y = z = 36$

**17.**  $x = 650, y = 750$

**18.** a)  $x = y = 12000, P = 36000000, Q = 48000000$   
 b)  $x = 9000, y = 8000, P = 51500000, Q = 50500000$   
 c)  $x = 9000, y = 12000, P = 31500000, Q = 58500000$

## Seksjon 5.10

**1.** a) Maks.verdi 5 i punktet  $(\frac{4}{5}, -\frac{3}{5})$ , min.verdi -5 i punktet  $(-\frac{4}{5}, \frac{3}{5})$ .  
 b) Maks.verdi 3 i punktene  $(1, 3)$  og  $(-1, -3)$ , min.verdi -3 i  $(1, -3)$  og  $(-1, 3)$ .  
 c) Min.verdi 17 i punktet  $(2, -3, 2)$ . Ingen maksimalverdi.  
 d) Maks.verdi  $1 + \sqrt{2}$  i punktet  $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, 1 + \sqrt{2})$ , min.verdi  $1 - \sqrt{2}$  i punktet  $(\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, 1 - \sqrt{2})$   
 e) Maks.verdi  $\frac{19}{\sqrt{70}}$  i punktet  $(\frac{4}{\sqrt{70}}, \frac{9}{\sqrt{70}})$ , min.verdi  $-\frac{19}{\sqrt{70}}$  i punktet  $(-\frac{4}{\sqrt{70}}, -\frac{9}{\sqrt{70}})$ .  
 f) Min. verdi  $-\frac{1}{12}$  i punktet  $(2, -\frac{1}{6}, -\frac{5}{6})$ .

**2.**  $(0, 0, 1), (0, 0, -1)$

**3.**  $(\pm 1, 0, 0), (0, \pm 1, 0)$

**8.** a) Lokalt min. i  $(0, 0)$ , sadelpunkter i  $(\pm 1, 0)$   
 b) Min.verdi 0 i  $(0, 0)$ , maks.verdi  $\ln 2 + 1$  i  $(0, \pm 1)$

**9.** Maks. punkt  $(\frac{2\sqrt{5}}{5}, \frac{4\sqrt{5}}{5})$ , min. punkt  $-(\frac{2\sqrt{5}}{5}, \frac{4\sqrt{5}}{5})$

**10.** a) Ellipse med sentrum i  $(1, -2)$ , halvakser  $a = 2, b = 3$ , brennvidde  $c = \sqrt{5}$ , brennpunkter  $(1, 2 \pm \sqrt{5})$ .  
 b) Maks. punkt  $(\frac{13}{5}, -\frac{1}{5})$ , min. punkt  $(-\frac{3}{5}, -\frac{19}{5})$

**11.**  $V = \frac{abc}{3\sqrt{3}}$

**12.**  $x = \frac{R}{\sqrt{3}}, y = \sqrt{\frac{2}{3}}R$

**13.**  $x = \frac{b}{3}, u = \frac{\pi}{3}$

**15.** a)  $\frac{x_0x}{a^2} + \frac{y_0y}{b^2} + \frac{z_0z}{c^2} = 1$   
 b)  $V = \frac{a^2b^2c^2}{6x_0y_0z_0}$   
 c)  $x = \frac{a}{\sqrt{3}}, y = \frac{b}{\sqrt{3}}, z = \frac{c}{\sqrt{3}}$   
 d)  $\frac{\sqrt{3}}{2}abc$

**16.** a)  $P_{\max} = K \frac{\alpha^\alpha \beta^\beta S^{\alpha+\beta}}{(\alpha+\beta)^{\alpha+\beta}}$   
 b)  $P_{\max} = K \frac{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \cdots \alpha_n^{\alpha_n} S^{\alpha_1+\alpha_2+\cdots+\alpha_n}}{(\alpha_1+\alpha_2+\cdots+\alpha_n)^{\alpha_1+\alpha_2+\cdots+\alpha_n}}$  for  $x_1 = \frac{\alpha_1 S}{\alpha_1+\alpha_2+\cdots+\alpha_n}, x_2 = \frac{\alpha_2 S}{\alpha_1+\alpha_2+\cdots+\alpha_n}, \dots$

**18.**  $x = \frac{S\alpha(1-\beta)-\alpha\beta KT}{\alpha-\beta}, y = \frac{S(1-\beta)(1-\alpha)-\beta(1-\alpha)KT}{(\alpha-\beta)K}, z = \frac{S\beta(1-\alpha)-\alpha\beta KT}{\beta-\alpha}$  og  $u = \frac{S(1-\beta)(1-\alpha)-\alpha(1-\beta)KT}{(\beta-\alpha)K}$

## Kapittel 6

### Seksjon 6.1

1. a) 9, b)  $\frac{\pi}{2} + 2$ , c)  $\frac{2}{3}(e-1)$ , d)  $\frac{2}{\pi}$ , e)  $\frac{1}{8}(e^8 - e^4 - 4)$ , f)  $2(e-1)$ , g)  $(1 - \frac{2\sqrt{3}}{3}) \ln 2 + \frac{\pi}{6}$

### Seksjon 6.2

1. a)  $\frac{16}{5}$  b)  $\frac{459}{4}$  c)  $\frac{17}{12}$  d)  $\frac{1}{6}$  e)  $\frac{e-1}{2}$  f)  $\frac{3}{56}$  g)  $-\frac{3\pi}{2}$  h)  $\frac{\pi^2}{8}$  i)  $-\frac{e^2}{12} + \frac{e}{6} + \frac{5}{12}$   
 3. a)  $\frac{e-1}{2}$  b) 1 c)  $\frac{e-1}{3}$

### Seksjon 6.3

1. a)  $\frac{81}{5}$  b)  $\frac{625}{16}\pi$  c)  $\pi(e^{16} - e)$  d)  $\frac{1}{16}$  e) 0 f)  $\frac{\pi}{6}(2\sqrt{2} - 1)$  g)  $\frac{512}{75}$   
 3. b)  $\frac{32}{9}$   
 4.  $\frac{\pi}{8}$

### Seksjon 6.4

1. a)  $\frac{8}{3}$  b)  $\frac{1}{4}$  c) 12 d)  $\frac{128\sqrt{2}}{3}\pi$  e) 0 f)  $8\pi$   
 2.  $\bar{x} = \frac{3}{4}, \bar{y} = \frac{3}{8}$   
 3.  $\bar{x} = \frac{4}{7}, \bar{y} = \frac{3}{4}$   
 4.  $4\pi R^2$   
 5.  $\frac{\pi}{6}(17\sqrt{17} - 1)$

**6.**  $\sqrt{2}\pi$ .

**7.**  $\pi - 2$

**9.**  $\frac{\pi}{120}(25\sqrt{5} + 1)$

**10.**  $\frac{4}{3}$

**11.**  $270\pi^2$

**12.**  $20\pi$

**13.**  $R^2 \sin \phi, \frac{\pi R^2}{8}$

**14.**  $\sqrt{2}u, \frac{31\sqrt{2}\pi}{5}$

**15.**  $(0, 0, \frac{1}{2})$

**16.** a)  $\frac{16}{9}(3\pi - 4)$ , b)  $8\pi - 16$

**17.** a)  $\frac{7\pi}{2}$  b)  $\frac{\pi}{6}(5\sqrt{5} - 1)$

**18.**  $\frac{81\pi}{2}$

**19.** b)  $8\pi$  c)  $-24\pi$

### Seksjon 6.5

**1.** a) 4 b) 0, c)  $\frac{7}{12}$ , d)  $\frac{135}{4}$

**2.**  $4\pi^2$

**3.**  $\frac{10-3\pi}{9}$

**4.**  $\frac{3\pi ab}{8}$

**5.**  $\frac{1}{420}$

**6.**  $4\pi^3 - 24\pi$

**7.**  $2\pi$

**8.** a)  $\frac{4}{3}$ .

**9.** a) og b)  $\frac{7}{6}$

**10.** a) og b)  $\frac{\sqrt{2}}{6} + \frac{\pi}{32} - \frac{1}{16}$

**11.** 4**12.** a) Sentrum  $(1, -2)$ , halvakser  $a = 2$ ,  $b = 3$  b)  $30\pi$  c)  $30\pi$ **Seksjon 6.7**

**1.** a)  $\frac{1}{6}$  b) 6 c)  $\frac{560}{81}$

**3.** a)  $-\frac{110}{27}$ , b)  $e^2 + 1$ , c)  $-\frac{3}{4} \ln 2 + \frac{3}{4}$

**5.** a)  $\frac{\ln 2}{2}(1 - e^{-5})$  b)  $2 \ln 2$  c)  $\frac{1}{2}$  d)  $\frac{5}{2}$  e) 20

**7.** c)  $\frac{1}{7}(e^7 + e^{-7} - 2)$

**8.** a) Beskrivelse av linjen:  $u = 0$  eller  $v = \frac{\pi}{4} + k\pi$ , der  $k \in \mathbb{Z}$ , Areal =  $\frac{\pi}{4}$   
b)  $\frac{\pi}{24}(5\sqrt{5} - 1)$ **9.** 2**Seksjon 6.8**

**1.**  $\frac{\pi}{8}$

**2.** Divergerer**3.** Konvergerer mot  $\frac{1}{4}$ **4.** Divergerer

**5.**  $\frac{\pi}{8}$

**6.** Konvergens for  $p > 1$  mot  $\frac{\pi}{p-1}$ **Seksjon 6.9**

**1.** a)  $\frac{1}{8}$  b)  $e^2 - e$  c) 1 d) 18 e)  $\frac{4}{3}$

**2.** a)  $\frac{14}{15}$  b)  $\frac{344\sqrt{2}}{945}$  c)  $\frac{672}{5}$  d)  $-\frac{216}{5}$  e)  $\frac{1}{120}$

**Seksjon 6.10**

**1.** a) 18 b) 0 c)  $\frac{64}{9}$

**2.** a)  $\frac{8}{15}\pi$  b)  $-\frac{3\sqrt{3}}{64} + \frac{\pi}{24}$  c)  $\frac{\pi^2}{4}$

**3.** a)  $\frac{7\pi}{12}$  b) 0 c)  $8\pi(1 - \frac{5}{e})$  d)  $\frac{\pi}{3}$  e)  $8\pi$  f) 136

**4.** c) 9

**5.**  $\frac{\pi}{8}$

.

**6.**  $4\pi^2$

**Seksjon 6.11**

**1.**  $\frac{4}{3}\pi R^3$

**2.**  $\frac{1}{2}$

**3.**  $\frac{\pi R^3}{3}$

**4.**  $\frac{4\pi}{3}(8 - 3\sqrt{3})$

**5.**  $\frac{4\pi abc}{3}$

**6.**  $\frac{\pi^2}{2} + \pi \ln 2$

**7.**  $(0, 0, \frac{8}{3})$

**9.**  $\frac{4\pi}{3}\sqrt{2} - \frac{7\pi}{6}$

**11.** b) 0 c) 0