

8 - OM DØDSLENGSEL OG LANDEGRENSER - LF

- 1 Vi setter gradienten lik null:

$$f'(x) = (-2x_1 - x_2 + 4, -x_1 - 4x_2 + 6) = (0, 0).$$

Dette er likningssystemet

$$\begin{aligned} 2x_1 + x_2 &= 4 \\ x_1 + 4x_2 &= 6, \end{aligned}$$

som har løsningen $x_1 = 10/7$ og $x_2 = 8/7$. Høyden er $f(10/7, 8/7) = 100/7$. At dette faktisk er en topp og ikke en bunn eller en sadel, viser vi i oppgave 8.

- 2 Høyden i a er $f(1, 1/2) = 4 - 1 - 1/2 = 5/2$. Vi har $f'(x) = (-2x_1, -4x_2)$, så $f'(a) = (-2, -2)$, og tangentlinja blir

$$-2(x_1 - 1) - 2(x_2 - \frac{1}{2}) = 0 \quad \Longleftrightarrow \quad x_1 + x_2 = \frac{3}{2}.$$

- 3 Med $-\nabla f(x) = (2x_1, 4x_2)^T$ er likningssystemet

$$x'_1 = 2x_1, \quad x'_2 = 4x_2,$$

og de to likningene har ingenting med hverandre å gjøre. Med startpunktet $(1, 1/2)^T$ blir

$$x(t) = \begin{pmatrix} e^{2t} \\ \frac{1}{2}e^{4t} \end{pmatrix},$$

og siden $e^{4t} = (e^{2t})^2$, er $x_2 = x_1^2/2$. Når $t \rightarrow -\infty$, går $x(t) \rightarrow 0$, så dråpen kommer fra toppen. Starter du i $(a, b)^T$, får du $x_2 = (b/a^2)x_1^2$. Fallinjene er altså parablene $x_2 = Cx_1^2$ og koordinataksene. Alle starter på toppen, og alle krysser nivåkurvene vinkelrett.

- 4 Gradienten er

$$f'(x) = -\frac{3}{10}(x_1, 2x_2) - 12e^{-2|x-a|^2}(x-a)^T - 8e^{-2|x+a|^2}(x+a)^T,$$

og $f'(x) = 0$ kan ikke løses med penn og papir. Figuren forteller oss likevel omtrent hvor vi skal lete, så vi starter en numerisk likningsløser i nærheten av a , $-a$ og origo:

```
import numpy as np
from scipy.optimize import fsolve

a = np.array([0.9, 0.3])

def df(x):
    e1 = np.exp(-2*np.sum((x - a)**2))
    e2 = np.exp(-2*np.sum((x + a)**2))
    return (-0.3*np.array([x[0], 2*x[1]])
            - 12*e1*(x - a) - 8*e2*(x + a))

for x0 in [a, -a, np.zeros(2)]:
    print(fsolve(df, x0))
```

Vi får

	punkt	høyde
høy topp	$(0.877, 0.285)^T$	3.46
lav topp	$(-0.865, -0.278)^T$	2.46
sadel	$(-0.074, -0.023)^T$	1.40

Hessematrisen har to negative egenverdier i toppene og én positiv og én negativ i sadelen, se oppgave 8.

5 Vi ganger ut:

$$f(\beta) = y^T y - 2y^T X\beta + \beta^T X^T X\beta.$$

Det første leddet er konstant, det andre er lineært i β med derivert $-2y^T X$, og skriver vi ut komponentene, ser vi at $(\beta^T A\beta)' = \beta^T (A + A^T)$ (dette ser vi nærmere på i økt 10). Siden $X^T X$ er symmetrisk, får vi

$$f'(\beta) = -2y^T X + 2\beta^T X^T X.$$

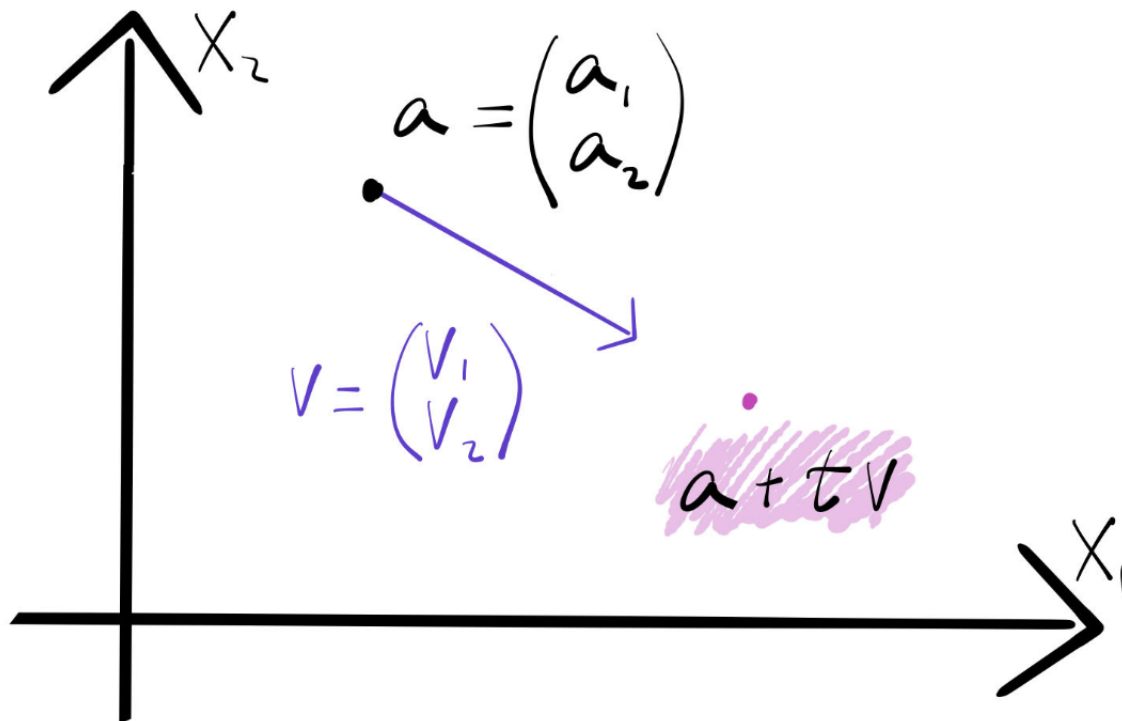
Setter vi denne lik null og transponerer, står normallikningene der:

$$X^T X\beta = X^T y.$$

6 Vi har $f'(x) = (-2x_1 - x_2 + 4, -x_1 - 4x_2 + 6)$, så

$$\partial_{x_1}^2 f = -2, \quad \partial_{x_2}^2 f = -4, \quad \partial_{x_1} \partial_{x_2} f = \partial_{x_2} \partial_{x_1} f = -1.$$

7 La oss sende en rett linje gjennom punktet a slik:



Parametriseringen for den rette linjen er

$$x(t) = a + tv = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + t \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad t \in \mathbb{R}$$

La oss definere

$$g(t) = f(x(t)) = f(a + tv)$$

og så bruke kjerneregelen på g , og få

$$\begin{aligned} g'(t) &= f'(x(t))x'(t) \\ &= f'(a + tv)v \\ &= (\partial_{x_1} f(a + tv), \partial_{x_2} f(a + tv)) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= v_1 \partial_{x_1} f(a + tv) + v_2 \partial_{x_2} f(a + tv). \end{aligned}$$

Nå kan vi gjenta denne suksessen og derivere dette uttrykket med kjerneregelen, og få

$$\begin{aligned} g''(t) &= v_1 \frac{d}{dt} (\partial_{x_1} f(a + tv)) + v_2 \frac{d}{dt} (\partial_{x_2} f(a + tv)) \\ &= v_1 (v_1 \partial_{x_1}^2 f(a + tv) + v_2 \partial_{x_2} \partial_{x_1} f(a + tv)) + v_2 (v_1 \partial_{x_1} \partial_{x_2} f(a + tv) + v_2 \partial_{x_2}^2 f(a + tv)) \end{aligned}$$

Hvis du nå skriver ut

$$\begin{aligned} v^T A v &= \begin{pmatrix} v_1 & v_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ &= v_1^2 a_{11} + v_1 v_2 a_{12} + v_1 v_2 a_{21} + v_2^2 a_{22}, \end{aligned}$$

ser du at vi kan skrive

$$g''(t) = v^T f''(a + tv)v.$$

Vi kan nå taylorutvikle g :

$$\begin{aligned} g(t) &= g(0) + g'(0)t + g''(0) \frac{t^2}{2} + \dots \\ &= f(a) + (f'(a)v)t + (v^T f''(a)v) \frac{t^2}{2} + \dots \end{aligned}$$

Med $a = x$, $v = h$ og $t = 1$ er dette formelen i oppgaven. Merk ellers at

$$g'(0) = f'(a)v$$

som vi kjenner igjen som den retningsderiverte til f i a .

- 8 La a være et av f sine kritiske punkt. Vi tar en motorsag og snitter flaten til $f(x)$ langs den parametriserte linjen $a + tv$, og så taylorutvikler vi motorsagsnittfunksjonen $g(t) = f(a + tv)$ om $t = 0$, slik som i forrige oppgave. Siden $f'(a) = 0$ dersom a er et kritisk punkt, blir taylorutviklingen

$$f(a + tv) = f(a) + (v^T f''(a)v) \frac{t^2}{2} + \dots$$

Dersom t er bitteliten og f oppfører seg bra nok til at $(v^T f''(a)v) \frac{t^2}{2}$ fullstendig dominerer alle leddene som kommer etter, kan vi nå konkludere med følgende tabell. La λ_1 og λ_2 være egenverdiene til $f''(a)$.

- Dersom λ_1 og λ_2 er positive, er a et bunnpunkt.
- Dersom λ_1 og λ_2 er negative, er a et toppunkt.
- Dersom λ_1 og λ_2 har forskjellig fortegn, er a et sadelpunkt.
- Dersom en av dem er null, gir testen ingen konklusjon.

Hvordan kan vi være så freidige? Siden

$$\partial_{x_2} \partial_{x_1} f = \partial_{x_1} \partial_{x_2} f$$

for alle funksjoner du kommer til å plages med,¹ er $f''(a)$ symmetrisk, og følgelig ortogonalt diagonaliserbar. La u_1 og u_2 være ortogonale egenvektorer til $f''(a)$. Disse utgjør en basis for $\mathbb{R}^2$, så vi kan alltid skrive

$$v = c_1 u_1 + c_2 u_2$$

og regne ut den kvadratiske formen

$$\begin{aligned} v^T f''(a) v &= (c_1 u_1 + c_2 u_2)^T f''(a) (c_1 u_1 + c_2 u_2) \\ &= (c_1 u_1 + c_2 u_2)^T (\lambda_1 c_1 u_1 + \lambda_2 c_2 u_2) \\ &= \lambda_1 c_1^2 |u_1|^2 + \lambda_2 c_2^2 |u_2|^2 \end{aligned}$$

Dette uttrykket er helt klart positivt for alle $v \neq 0$ i det første tilfellet, og konklusjonen følger siden vi følgelig alltid legger noe til $f(a)$ når vi beveger oss bittelitt vekk fra a . Samme resonnement dersom egenverdiene er negative. Dersom egenverdiene har forskjellig fortegn, kan vi bruke samme resonnement på $v = u_1$ og $v = u_2$ og konkludere med at det finnes en retning der vi går nedover og en retning der vi går oppover, slik at a er en sadel. Hvis en av egenverdiene er null, finnes det en retning der $v^T f''(a) v = 0$, og i så fall trenger vi flere ledd i taylorrekken for å konkludere.

Dersom alle egenverdier er positive, sier vi at matrisen er positivt definit. Tilsvarende for negative egenverdier.

Hessematrisen til fjellet i oppgave 1 er

$$f''(x) = \begin{pmatrix} -2 & -1 \\ -1 & -4 \end{pmatrix}$$

som har karakteristisk polynom

$$\det \begin{pmatrix} -2 - \lambda & -1 \\ -1 & -4 - \lambda \end{pmatrix} = (2 + \lambda)(4 + \lambda) - 1 = \lambda^2 + 6\lambda + 7$$

og egenverdier

$$\lambda = -3 \pm \sqrt{2}$$

som begge er negative. Følgelig er $(10/7, 8/7)^T$ en fjelltopp og ikke en dalbunn.

¹Det finnes patologiske unntak, slik som for eksempel $\frac{x_1 x_2 (x_1^2 - x_2^2)}{x_1^2 + x_2^2}$, der de blandede andreordens partiellderiverte ikke er like, men disse er konstruert av matematikere for å krangle med folk som er uforsiktige nok til å påstå at de blandede andreordens partiellderiverte alltid er like. Se oppgave 28.

- 9 Fra oppgave 5 har vi $f'(\beta) = -2y^T X + 2\beta^T X^T X$, så hessematrisen er

$$f''(\beta) = 2X^T X.$$

For $v \neq 0$ er $v^T X^T X v = |Xv|^2$, og dette er positivt dersom X har full rang, for da er $Xv \neq 0$. Hessematrisen er altså positivt definit, og det kritiske punktet er et minimum.

- 10 Du må lete etter maksima og minima til de tre envariable funksjonene

$$f(x_1, 0) = -x_1^2 + 4x_1 + 8 \quad x_1 \in [0, 4]$$

$$f(0, x_2) = -2x_2^2 + 6x_2 + 8 \quad x_2 \in [0, 3]$$

$$f(4 \cos \theta, 3 \sin \theta) = -16 \cos^2 \theta - 12 \cos \theta \sin \theta - 18 \sin^2 \theta + 16 \cos \theta + 18 \sin \theta + 8 \quad \theta \in [0, \pi/2]$$

De to første er parabler med toppunkt i $x_1 = 2$ og $x_2 = 3/2$, der høyden er henholdsvis 12 og $25/2$. Den siste er uhåndterlig med penn og papir, men setter vi den deriverte lik null numerisk, finner vi et maksimum på 9.16 i $\theta \approx 28.2^\circ$, altså i punktet $(3.53, 1.42)^T$. Turens høyeste punkt er derfor $(0, 3/2)^T$ på x_2 -aksen, med høyde $25/2$.

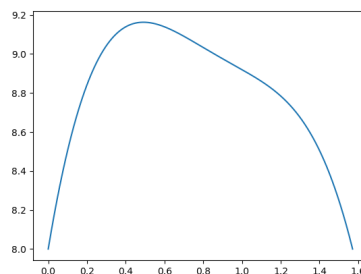
- 11 Siden f er et andreordens polynom, kan vi skrive

$$f(a + tv) = f(a) + (v^T f''(a) v) \frac{t^2}{2},$$

der $a = (10/7, 8/7)^T$ er det kritiske punktet, og siden $f''(a)$ er negativt definit, vil funksjonsverdien i et tilfeldig punkt i $\mathbb{R}^2$ være funksjonsverdien i det kritiske punktet pluss noe negativt. Derfor er det kritiske punktet et globalt maksimum. Det ligger inne i Ω , siden $\frac{(10/7)^2}{16} + \frac{(8/7)^2}{9} < 1$, så det globale maksimumet er $f(10/7, 8/7) = 100/7$. Minimumspunktene må sitte på $\partial\Omega$, så la oss sjekke. Funksjonsverdiene på de rette delene har åpenbart minimum i endepunktene, og funksjonsverdiene på den krumme delen av $\partial\Omega$ kan skrives

$$f(4 \cos \theta, 3 \sin \theta) = -12 \cos \theta \sin \theta - 2 \sin^2 \theta + 16 \cos \theta + 18 \sin \theta - 8 \quad \theta \in [0, \pi/2]$$

så her må vi nesten plote litt:

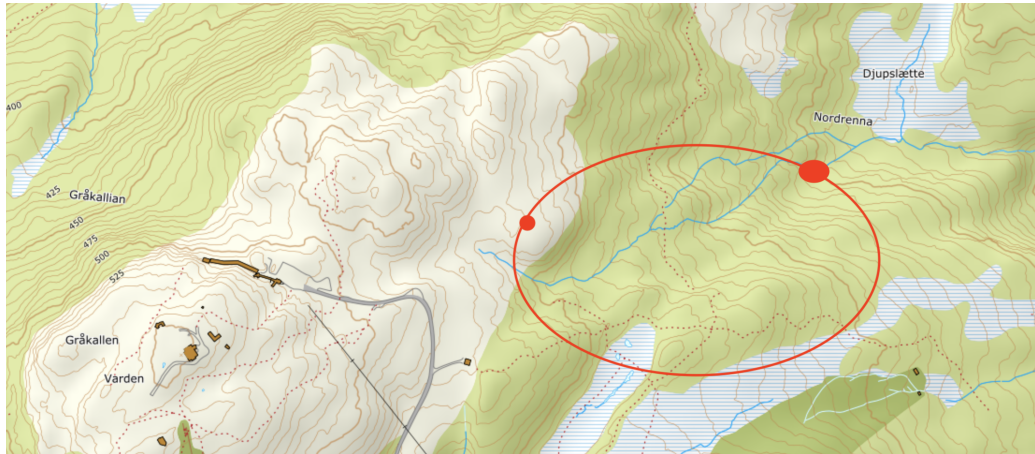


Det er klart fra figuren at de lokale minima finnes i endepunktene også på den krumme delen, så da må vi sjekke

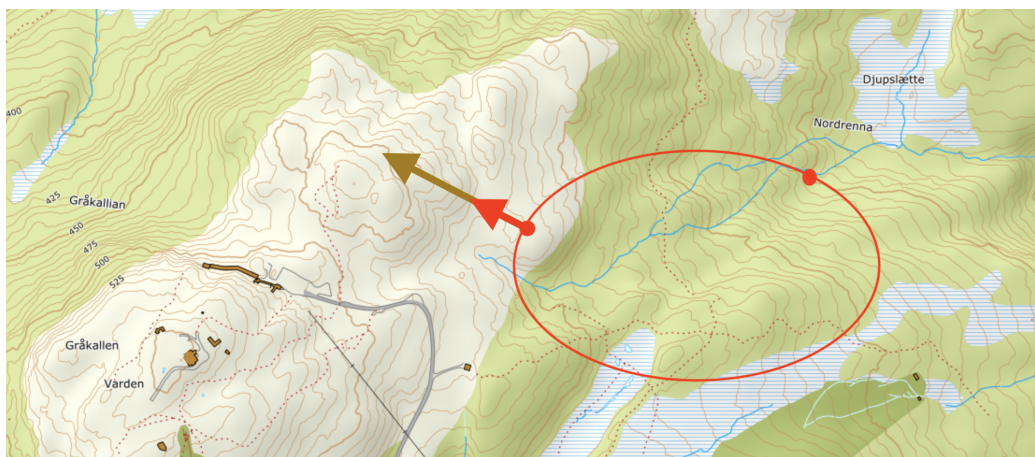
$$f(0, 0) = f(4, 0) = f(0, 3) = 8,$$

og globalt minimum er 8.

- 12 Lagranges multiplikatormetode er basert på en ide som er veldig enkel å forstå om man har det klart for seg hvordan gradienten til en funksjon skal tolkes geometrisk. Den er også til god hjelp dersom man ønsker en enkel huskeregel nettopp for å huske gradientens geometriske tolkning. Hvis jeg sier at man går en tur på den elliptiske kurven i figuren under, klarer de aller fleste uten å tenke seg om merke av turens høyeste og laveste punkt:



Nå gjelder det å tenke på f (altså den funksjonen du skal maksimere eller minimere) som en funksjon som beskriver fjellet du går på. Nivåkurvene til f er det som er printa som ekvidistanselinjene på et vanlig turkart. Den snedige ideen til Lagrange var å se på turtrajektorien som nivåkurven til en eller annen annen funksjon g , og så innse at i turens topp-, bunn- eller sadelpunkter **må gradienten til f være parallell med gradienten til g , siden gradienter og nivåkurver står normalt på hverandre.**



Denne ideen gir tre likninger

$$\begin{aligned}\nabla f(x) &= \lambda \nabla g(x) \\ g(x) &= c\end{aligned}$$

som kan løses for x og λ . Det er litt forskjellige betingelser ute og går; f og g må være deriverbare, og gradientene kan ikke være null, og du må ikke være i endepunktet for trajektorien og litt forskjellig.

Alternativt kan vi parametrisere kurven og bruke kjerneregelen og få

$$\nabla^T f(x(t))\dot{x}(t) = 0$$

siden $f(x(t))$ har et lokalt maksimum og

$$\nabla^T g(x(t))\dot{x}(t) = 0$$

siden $g(x(t))$ er konstant. Begge gradientene står altså normalt på $\dot{x}(t)$, og i planet betyr det at de er parallelle.

13 Gradienten til L er

$$\begin{aligned} L'(x, \lambda) &= (\partial_{x_1} L, \partial_{x_2} L, \partial_{\lambda} L) \\ &= (\partial_{x_1} f(x) - \lambda \partial_{x_1} g(x), \partial_{x_2} f(x) - \lambda \partial_{x_2} g(x), c - g(x)), \end{aligned}$$

og setter vi denne lik null, får vi nøyaktig lagrangelikningene $\nabla f = \lambda \nabla g$ og $g(x) = c$.

Vi kan nå prøve oss på den krumme delen av $\partial\Omega$ i oppgave 11, som har likning

$$\frac{x_1^2}{16} + \frac{x_2^2}{9} = 1,$$

slik at

$$g(x) = \frac{x_1^2}{16} + \frac{x_2^2}{9}.$$

Gradienten til g er

$$g'(x) = \left(\frac{x_1}{8}, \frac{2x_2}{9} \right)$$

mens gradienten til f regnet vi ut i oppgave 1:

$$f'(x) = (-2x_1 - x_2 + 4, -x_1 - 4x_2 + 6).$$

Alt i alt får vi likningssystemet

$$-2x_1 - x_2 + 4 = \lambda \frac{x_1}{8}$$

$$-x_1 - 4x_2 + 6 = \lambda \frac{2x_2}{9}$$

$$\frac{x_1^2}{16} + \frac{x_2^2}{9} = 1$$

De to første likningene kan skrives om til

$$(\lambda + 16)x_1 + 8x_2 = 32$$

$$9x_1 + 2(\lambda + 18)x_2 = 54$$

og i dette tilfellet synes jeg det er kjappest å bare invertere matrisen og skrive

$$\begin{aligned} x &= \frac{1}{2(\lambda + 16)(\lambda + 18) - 72} \begin{pmatrix} 2(\lambda + 18) & -8 \\ -9 & (\lambda + 16) \end{pmatrix} \begin{pmatrix} 32 \\ 54 \end{pmatrix} \\ &= \frac{1}{(\lambda + 16)(\lambda + 18) - 36} \begin{pmatrix} 2(\lambda + 18) & -8 \\ -9 & (\lambda + 16) \end{pmatrix} \begin{pmatrix} 16 \\ 27 \end{pmatrix}. \end{aligned}$$

Setter vi denne inn i ellipselikningen, får vi en fjerdegradslikning i λ som er uhåndterlig med penn og papir, men triviell å løse med en numerisk likningsløser. Den har to reelle løsninger. Den ene, $\lambda \approx -35.1$, gir punktet $(-2.62, -2.27)^T$, som ligger på ellipsen, men ikke i første kvadrant. Den andre, $\lambda \approx -10.14$, gir punktet $(3.53, 1.42)^T$ med høyde 9.16 - det samme maksimumet på buen som vi fant i oppgave 10. Lagranges metode ser ikke hjørnene i $\partial\Omega$, så dem må vi sjekke for seg selv, og der finner vi igjen minimumet 8.

14 Gradienten til fjellet er

$$h'(x) = (-2x_1 - x_2, -x_1 - 2x_2)$$

og setter vi

$$g(x) = x_1^2 + \frac{x_2^2}{4},$$

blir

$$g'(x) = (2x_1, x_2/2).$$

Lagrangelikningene blir

$$\begin{aligned} -2x_1 - x_2 &= 2\lambda x_1 \\ -x_1 - 2x_2 &= \lambda x_2/2 \end{aligned}$$

eller

$$\begin{aligned} 2(1 + \lambda)x_1 + x_2 &= 0 \\ 2x_1 + (4 + \lambda)x_2 &= 0. \end{aligned}$$

Dersom matrisen er inverterbar, er $x = 0$ den eneste løsningen, og dette punktet ligger ikke på ellipsen, så da får vi nesten kreve at determinanten

$$p(\lambda) = 2(1 + \lambda)(4 + \lambda) - 2 = 2\lambda^2 + 10\lambda + 6$$

blir null, noe som skjer når

$$\lambda = \frac{-5 \pm \sqrt{13}}{2}.$$

Likningen $2(1 + \lambda)x_1 + x_2 = 0$ gir i såfall at

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = s \begin{pmatrix} 1 \\ -2(1 + \lambda) \end{pmatrix} = s \begin{pmatrix} 1 \\ 3 \mp \sqrt{13} \end{pmatrix}.$$

Hvis vi nå setter inn disse punktene i ellipselikningen, får vi

$$s^2 \left(1 + \frac{1}{4}(3 \mp \sqrt{13})^2 \right) = 1,$$

slik at

$$s = \pm \frac{2}{\sqrt{4 + (3 \mp \sqrt{13})^2}}.$$

Alt i alt har vi nå de fire punktene

$$\pm \frac{2}{\sqrt{4 + (3 \mp \sqrt{13})^2}} \begin{pmatrix} 1 \\ 3 \mp \sqrt{13} \end{pmatrix}.$$

Setter vi dem inn i h , får vi

$$h(\pm(0.957, -0.580)) = \frac{\sqrt{13} - 3}{2} \approx 0.30$$

$$h(\pm(0.290, 1.914)) = -\frac{\sqrt{13} + 3}{2} \approx -3.30,$$

så turens høyeste punkt har høyde $(\sqrt{13} - 3)/2$, og det nås i de to punktene $\pm(0.957, -0.580)^T$.

15 Vi setter

$$g(x) = 6(x_1 + x_2)^2 + 2(x_1 - x_2)^2,$$

og beregner

$$f'(x) = (1, 2)$$

og

$$g'(x) = (12(x_1 + x_2) + 4(x_1 - x_2), 12(x_1 + x_2) - 4(x_1 - x_2)),$$

slik at lagrangelikningene blir

$$12(x_1 + x_2) + 4(x_1 - x_2) = \lambda$$

$$12(x_1 + x_2) - 4(x_1 - x_2) = 2\lambda$$

$$6(x_1 + x_2)^2 + 2(x_1 - x_2)^2 = 3$$

Nå er det kanskje ryddigst å sette $u = x_1 + x_2$, $v = x_1 - x_2$, slik at likningssystemet blir

$$12u + 4v = \lambda$$

$$12u - 4v = 2\lambda$$

$$6u^2 + 2v^2 = 3$$

Gausseliminasjon av de to første likningene gir

$$\begin{pmatrix} u \\ v \end{pmatrix} = \frac{\lambda}{8} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

og setter vi denne i ellipselikningen, får vi

$$\lambda^2 \left(\frac{6}{64} + \frac{2}{64} \right) = 3,$$

som gir

$$\lambda = \pm\sqrt{24} = \pm 2\sqrt{6}.$$

Da er $u = -v = \pm\sqrt{6}/4$, og siden $x_1 = (u + v)/2$ og $x_2 = (u - v)/2$, blir punktene

$$x = \pm \begin{pmatrix} 0 \\ \sqrt{6}/4 \end{pmatrix}.$$

Siden f er lineær, gir det ene punktet maksimum og det andre minimum. Den største verdien er $f(0, \sqrt{6}/4) = \sqrt{6}/2 \approx 1.22$.

- 16 Vi setter lengdene i X og vekten i y . Dersom han har målt de samme kolbene to ganger i samme rekkefølge, men med forskjellig enhet, blir de to kolonnene parallelle (den ene er 2.54 ganger den andre), og følgelig får ikke X full rang. Dersom to av kolonnene i X er parallelle, er også to av kolonnene i $X^T X$ parallelle, og da vil python klage på at $X^T X$ ikke er inverterbar når han løser normallikningene $X^T X \beta = X^T y$.

- 17 Singulærverdiene er kvadratrøttene av egenverdiene til

$$A^T A = \begin{pmatrix} 1 + b^2 & a + b \\ a + b & 1 + a^2 \end{pmatrix}.$$

Vi slipper å finne dem eksplisitt, for determinanten og sporet gir

$$\sigma_1^2 \sigma_2^2 = \det A^T A = (1 - ab)^2 \quad \text{og} \quad \sigma_1^2 + \sigma_2^2 = \text{tr } A^T A = 2 + a^2 + b^2,$$

og deler vi det siste på $\sigma_1 \sigma_2 = |1 - ab|$, får vi

$$\kappa + \frac{1}{\kappa} = \frac{2 + a^2 + b^2}{|1 - ab|}.$$

Med $a = b = 0$ får vi enhetsmatrisen og $\kappa = 1$. Med $a = 1$ og $b = 0$ blir $\kappa + 1/\kappa = 3$, altså $\kappa = (3 + \sqrt{5})/2 \approx 2.6$. Med $a = b = 0.9$ er A symmetrisk med egenverdier $1 \pm a$, så $\kappa = 1.9/0.1 = 19$. Når $ab \rightarrow 1$, går determinanten mot null, og $\kappa \rightarrow \infty$. Kondisjonstallet sier altså hvor nær matrisen er å være singulær. Du kan sjekke tallene med `np.linalg.cond(np.array([[1,a],[b,1]]))`.

- 18 La $g(\beta) = |\beta|^2$. Gradienten til

$$f(\beta) = (y - X\beta)^T (y - X\beta) = y^T y - 2y^T X\beta + \beta^T X^T X\beta$$

er fortsatt

$$\nabla f(\beta) = -2X^T y + 2X^T X\beta,$$

men nå skal den settes parallell med

$$\nabla g(\beta) = 2\beta.$$

Vi får

$$-X^T y + X^T X\beta = \lambda\beta,$$

som gir

$$(X^T X - \lambda I) \beta = X^T y.$$

Siden det ubegrensede minimumet ligger utenfor kula $|\beta|^2 \leq r$, peker ∇f innover i kula på randen, altså motsatt av ∇g , så λ er negativ. Skriver vi $\mu = -\lambda > 0$, får vi

$$\beta = (X^T X + \mu I)^{-1} X^T y,$$

der μ velges slik at $|\beta|^2 = r$. Dette er ridgeregresjon. Legg merke til at $X^T X + \mu I$ er inverterbar for alle $\mu > 0$, også når $X^T X$ er singulær, for egenverdiene til $X^T X$ er ikke-negative, og vi legger μ til hver av dem. Kjemikeren i oppgave 16 hadde altså sluppet feilmeldingen.

- 19 Vi skriver punktene som (x_1, x_2, y) . Gjennomsnittene er $\bar{x}_1 = \bar{x}_2 = 1/2$ og $\bar{y} = 3/4$, så de sentrerte datamatrixene blir

$$X = \begin{pmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \\ 1/2 & 1/2 \\ -1/2 & -1/2 \end{pmatrix} \quad \text{og} \quad y = \begin{pmatrix} 1/4 \\ 1/4 \\ 1/4 \\ -3/4 \end{pmatrix}.$$

Vi regner ut

$$X^T X = I \quad \text{og} \quad X^T y = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix},$$

så den vanlige regresjonen gir $\beta = (1/2, 1/2)^T$ med $|\beta|^2 = 1/2$. Det er større enn $r = 1/10$, så begrensningen biter. Fra forrige oppgave er

$$\beta = (I + \mu I)^{-1} X^T y = \frac{1}{2(1 + \mu)} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

og kravet $|\beta|^2 = \frac{1}{2(1+\mu)^2} = \frac{1}{10}$ gir $1 + \mu = \sqrt{5}$. Da er

$$\beta = \frac{1}{2\sqrt{5}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.224 \\ 0.224 \end{pmatrix},$$

og ridgeregresjonen blir

$$y \approx \frac{3}{4} + 0.224 \left(x_1 - \frac{1}{2}\right) + 0.224 \left(x_2 - \frac{1}{2}\right) = 0.526 + 0.224(x_1 + x_2).$$

Til sammenlikning ga eksamensoppgaven $y = \frac{1}{4} + \frac{1}{2}x_1 + \frac{1}{2}x_2$. Ridge krymper stigningstallene mot null.

- 20 Singulærverdiene til enhetsmatrisen er alle lik 1, så $\kappa(I) = 1/1 = 1$. Siden $\sigma_1 \geq \sigma_r$ for alle matriser, er $\kappa \geq 1$, så bedre enn dette blir det ikke.

- 21 Aksiomene er

- $\|x\| > 0$ dersom $x \neq 0$
- $\|ax\| = |a|\|x\|$
- $\|x + y\| \leq \|x\| + \|y\|$

I TMA4106 sjekket vi at (x, x) er et indreprodukt, og $\|x\| = \sqrt{(x, x)}$ er positiv når $x \neq 0$, så det første er greit. Aksiom to følger av lineariteten til indreproduktet, mens aksiom tre er trekantulikheten, også bevist i TMA4106. Enhetssirklene er kurvene

$$|x_1|^p + |x_2|^p = 1.$$

Dersom $p = 1$ får vi fire likninger på formen

$$\pm x_1 \pm x_2 = 1$$

alt etter i hvilken kvadrant vi er i, så vi får en klassisk diamant. Når $p \rightarrow \infty$, får du ∞ -normen

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

for når p blir stor, er

$$\sqrt[p]{\sum_k |x_k|^p} \approx \sqrt[p]{|x_l|^p} = |x_l|$$

der x_l er det største elementet. For $p = \infty$ får vi likningen

$$\max(|x_1|, |x_2|) = 1$$

som rett og slett blir enhetskvadratet. For $p = 2$ får vi den vanlige enhetssirkelen, og for andre p får vi forskjellige mellomting mellom $p = 1$ og $p = \infty$. Med $p = 4$ får vi et avrundet kvadrat, som fontenen på Sergels torg. Se figur her:

[https://en.wikipedia.org/wiki/Norm_\(mathematics\)#Properties](https://en.wikipedia.org/wiki/Norm_(mathematics)#Properties)

- 22 Vi skriver $A = U\Sigma V^T$. Siden ortogonale matriser ikke endrer lengder, er $|Ax| = |\Sigma V^T x|$, og med $z = V^T x$ er $|z| = |x|$ og

$$|Ax|^2 = |\Sigma z|^2 = \sum_k \sigma_k^2 z_k^2 \leq \sigma_1^2 \sum_k z_k^2 = \sigma_1^2 |x|^2,$$

med likhet når $z = (1, 0, \dots, 0)^T$, altså når x er den første kolonnen i V . Normen $|A| = \max_{x \neq 0} |Ax|/|x|$ er derfor σ_1 . Er A kvadratisk og inverterbar, er $A^{-1} = V\Sigma^{-1}U^T$, så singulærverdiene til A^{-1} er $1/\sigma_k$. Den største av dem er $1/\sigma_r$, så

$$|A||A^{-1}| = \frac{\sigma_1}{\sigma_r} = \kappa(A).$$

- 23 Ridge leter etter minimum på en sirkel (kule), og lasso på en diamant $|\beta|_1 = r$. Nivåkurvene til $(y - X\beta)^T(y - X\beta)$ er ellipser, og minimumet på randen ligger der den minste ellipsen akkurat treffer randen. En sirkel har ingen hjørner, så ellipsen treffer den som regel i et punkt der ingen av komponentene i β er null. Diamanten har hjørner på koordinataksene, og ellipsene treffer den ofte nettopp i et hjørne, slik som i lineær programmering senere i økten. I et hjørne er noen av komponentene i β lik null.

- 24 La t_1 og t_2 være timene du leser i hvert av fagene. Da er

$$\Omega : \quad t_1, t_2 \geq 0, \quad t_1 + t_2 \leq 100, \quad 2.1t_1 + 5.0t_2 \leq 200, \quad 2.9t_1 + 4.1t_2 \leq 300.$$

Strevet biter først. Linja $2.1t_1 + 5.0t_2 = 200$ skjærer aksene i $(95.2, 0)$ og $(0, 40)$, og der er både timetallet og møya godt innenfor grensene ($2.9 \cdot 95.2 = 276 < 300$). Ω er derfor trekanten med hjørner i $(0, 0)$, $(95.2, 0)$ og $(0, 40)$, og de to andre begrensningene spiller ingen rolle.

- 25 Vi antar at det første faget gir ett poeng per time og det andre to. Vi skal maksimere $t_1 + 2t_2$ på Ω . Nivåkurvene er rette linjer, så maksimum ligger i et hjørne:

$$(0, 0) : 0, \quad (95.2, 0) : 95.2, \quad (0, 40) : 80.$$

Du bør bruke alt strevet på det første faget og få 95 poeng der. Grunnen er at det første faget gir $1/2.1 \approx 0.48$ poeng per strev, mens det andre gir $2/5.0 = 0.40$.

- 26 Med $s_1 = t_1$ og $s_2 = 2t_2$ som poengscore blir strevgrensen

$$2.1s_1 + 2.5s_2 \leq 200,$$

så ett poeng koster 2.1 strev i det første faget og 2.5 i det andre. Løsningen fra forrige oppgave gir A og F. Vil du stå i begge fagene, må du fordele strevet.

Ved UiO (E fra 40, D fra 46, C fra 58, B fra 77, A fra 92) er $2.1 \cdot 46 + 2.5 \cdot 40 = 196.6 \leq 200$, mens D i begge koster 211.6. Det beste er da D i det ene faget og E i det andre. D i det første faget er billigst, men E og D går også (199 strev).

Ved NTNU (E fra 41, D fra 53, C fra 65, B fra 77, A fra 89) koster E i begge $2.1 \cdot 41 + 2.5 \cdot 41 = 188.6$, mens D og E koster $2.1 \cdot 53 + 2.5 \cdot 41 = 213.8$. Da blir det E og E.

Teller du karakterpoeng ($A = 5$, ..., $E = 1$, $F = 0$), gir A og F likevel mest, men da må du ta det ene faget på nytt. Legg merke til at målfunksjonen nå er en trappefunksjon, og da gjelder ikke hjørneregelen fra lineær programmering lenger.

27 Matrisen

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

er ikke nullmatrisen, men begge egenverdiene er null, så $\rho(A) = 0$. Det bryter med det første aksiomet. Trekantulikheten går også i stykker: med $B = A^T$ er $\rho(A) = \rho(B) = 0$, mens $A + B$ har egenverdier ± 1 , så $\rho(A + B) = 1 > \rho(A) + \rho(B)$.

28 Langs aksene er $f(x_1, 0) = f(0, x_2) = 0$, så de rene andrederiverte i origo er null. For $x_2 \neq 0$ er

$$\partial_{x_1} f(0, x_2) = \lim_{h \rightarrow 0} \frac{f(h, x_2)}{h} = \lim_{h \rightarrow 0} \frac{x_2(h^2 - x_2^2)}{h^2 + x_2^2} = -x_2,$$

og på samme måte er $\partial_{x_2} f(x_1, 0) = x_1$. Da blir

$$\partial_{x_2} \partial_{x_1} f(0, 0) = -1 \quad \text{og} \quad \partial_{x_1} \partial_{x_2} f(0, 0) = 1,$$

så hessematrisen i origo er

$$f''(0) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

som ikke er symmetrisk. Det går an fordi de andrederiverte ikke er kontinuerlige i origo. Er de kontinuerlige, er hessematrisen alltid symmetrisk (Schwarz' teorem).²

²https://en.wikipedia.org/wiki/Symmetry_of_second_derivatives

GAMLE EKSAMENSOPPGAVER

1 Med

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

er $f(x) = \frac{1}{2}x^T A x - b^T x$, og siden A er symmetrisk, er

$$f'(x) = x^T A - b^T = (2x_1 + x_2 - 3, x_1 + 2x_2 - 3).$$

Setter vi denne lik null, får vi $Ax = b$, altså

$$2x_1 + x_2 = 3$$

$$x_1 + 2x_2 = 3,$$

med løsning $x = (1, 1)^T$. Hessematrisen er $f''(x) = A$, som har egenverdier 1 og 3 (egenvektorer $(1, -1)^T$ og $(1, 1)^T$). Begge er positive, så A er positivt definit og $(1, 1)^T$ er et bunnpunkt. Siden f er et andreordens polynom, er det også det globale minimumet, og bunnen ligger på høyden

$$f(1, 1) = \frac{1}{2} \cdot 6 - 6 = -3.$$

2 Grafen til f er et plan, så $f' = (1, 2) \neq 0$ overalt, og største og minste verdi må ligge på randen ∂D , som er nivåkurven $g(x) = 1$ til

$$g(x) = \frac{(x_1 + x_2)^2}{4} + (x_1 - x_2)^2, \quad g'(x) = \left(\frac{x_1 + x_2}{2} + 2(x_1 - x_2), \frac{x_1 + x_2}{2} - 2(x_1 - x_2) \right).$$

Lagrange, $f' = \lambda g'$, gir de to likningene

$$\frac{x_1 + x_2}{2} + 2(x_1 - x_2) = \frac{1}{\lambda} \quad \text{og} \quad \frac{x_1 + x_2}{2} - 2(x_1 - x_2) = \frac{2}{\lambda},$$

og legger vi dem sammen og trekker dem fra hverandre, får vi

$$x_1 + x_2 = \frac{3}{\lambda} \quad \text{og} \quad x_1 - x_2 = -\frac{1}{4\lambda}.$$

Innsatt i $g(x) = 1$ gir dette $\frac{9}{4\lambda^2} + \frac{1}{16\lambda^2} = \frac{37}{16\lambda^2} = 1$, altså $\lambda = \pm\sqrt{37}/4$, og siden $f = \frac{3}{2}(x_1 + x_2) - \frac{1}{2}(x_1 - x_2) + 4$, blir

$$f = \frac{9}{2\lambda} + \frac{1}{8\lambda} + 4 = \frac{37}{8\lambda} + 4 = 4 \pm \frac{\sqrt{37}}{2}.$$

Største verdi er $4 + \sqrt{37}/2 \approx 7.04$ i punktet $(11, 13)^T / (2\sqrt{37})$, og minste er $4 - \sqrt{37}/2 \approx 0.96$ i det motsatte punktet. Raskere: med $y_1 = x_1 + x_2$ og $y_2 = x_1 - x_2$ er randen ellipsen $y_1^2/4 + y_2^2 = 1$, som kan parametriseres $y_1 = 2 \cos t$, $y_2 = \sin t$, og da er $f = 3 \cos t - \frac{1}{2} \sin t + 4$, en sinusoid med amplitude $\sqrt{9 + 1/4} = \sqrt{37}/2$ rundt 4.

3 Jacobimatrisen er

$$f'(x) = (2x_1 + x_2 - 3, x_1 + 2x_2 - 3),$$

så $f'(2, 3) = (4, 5)$ og $f(2, 3) = 7$, og tangentplanet er

$$y = 7 + 4(x_1 - 2) + 5(x_2 - 3).$$

Det kritiske punktet er løsningen av $f'(x) = 0$, altså $2x_1 + x_2 = 3$ og $x_1 + 2x_2 = 3$, som gir $(1, 1)^T$ med $f(1, 1) = 0$. Hessematrixen

$$f''(x) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

har egenverdier 3 og 1, så det er et bunnpunkt - samme matrise som i gammel eksamensoppgave 1, og det er ingen tilfældighet, for $f(x) = \frac{1}{2}x^T Ax - b^T x + 3$ med samme A og b .

Bunnpunktet ligger inni trekanten ($0 \leq 1 \leq 4/3$), så 0 er minste verdi. Største verdi må ligge på randen, og der parametriserer vi: på kateten $x_2 = 0$ er $f = x_1^2 - 3x_1 + 3$, som går fra 3 ned til $3/4$ og opp til 3 igjen; på kateten $x_1 = 3$ er $f = x_2^2 + 3$, som vokser fra 3 til 19; og på hypotenusen $x_2 = 4x_1/3$ er $f = \frac{37}{9}x_1^2 - 7x_1 + 3$, som er 3 i origo, når ned til $3/148$ ved $x_1 = 63/74$, og ender på 19 i $(3, 4)$. Alle tre er parabler som vender opp, så på randen er det hjørnene som konkurrerer om maksimum:

$$f(0, 0) = 3, \quad f(3, 0) = 3, \quad f(3, 4) = 19.$$

Største verdi er 19 i $(3, 4)^T$ og minste er 0 i $(1, 1)^T$. Legg merke til hvor nær hypotenusen kommer bunnpunktet - $3/148 \approx 0.02$ - nivåkurvene er ellipser rundt $(1, 1)^T$, og linja $x_2 = 4x_1/3$ går like forbi.

- 4 Vi har $f'(x) = (2x_1, -2x_2)$, så $f'(1, 2) = (2, -4)$ og $f(1, 2) = -2$, og tangentplanet er

$$y = -2 + 2(x_1 - 1) - 4(x_2 - 2).$$

Eneste kritiske punkt er origo (som ligger i Ω , siden $4 + 1 \leq 9$), og hessematrixen

$$f''(x) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

har egenverdier 2 og -2 , så det er et sadelpunkt: opp langs x_1 -aksen og ned langs x_2 -aksen. Grafen er den klassiske sadelflaten, og f er harmonisk - $\Delta f = 2 - 2 = 0$ - som du får bruk for i økt 9 og 13.

- 5 Igjen er grafen et plan, $f' = (2, 1)$, uten kritiske punkter, så ekstremalverdiene ligger på sirkelen. Med $g(x) = (x_1 - 1)^2 + (x_2 - 1)^2$ gir Lagrange $f' = \lambda g'$:

$$2 = 2\lambda(x_1 - 1) \quad \text{og} \quad 1 = 2\lambda(x_2 - 1),$$

så $x - (1, 1)^T$ er parallell med $(2, 1)^T$, og siden radien er 1, er

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \pm \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Der er $f = 2 + 1 + 1 \pm \frac{4+1}{\sqrt{5}} = 4 \pm \sqrt{5}$: største verdi $4 + \sqrt{5}$ og minste $4 - \sqrt{5}$. Det er akkurat det man venter: planet stiger brattest i gradientens retning $(2, 1)^T$, så det høyeste punktet på sirkelen ligger én radius fra sentrum i den retningen, og stigningen over den radien er $|f'| = \sqrt{5}$.

- 6 Jacobimatrixen er

$$f'(x) = (-2(x_1 - 2x_2), 4(x_1 - 2x_2)),$$

som er null nøyaktig når $x_1 = 2x_2$ - de kritiske punktene utgjør en hel linje. Hessematrixen

$$f''(x) = \begin{pmatrix} -2 & 4 \\ 4 & -8 \end{pmatrix}$$

har egenverdier 0 og -10 , så andrederiverttesten gir ingen konklusjon. Men her trenger vi ingen test: $f = 9 - (\text{noe})^2 \leq 9$, med likhet nøyaktig på linja, så hele linja er toppunkter - grafen er en fjellrygg med flat kam i høyden 9, og egenverdien 0 hører til retningen langs kammen, $(2, 1)^T$.

Kvadratet er $|x_1 - 1| + |x_2 - 1| \leq 1$, stilt på høykant, og kammen $x_1 = 2x_2$ går gjennom det (for eksempel i $(1, 1/2)^T$ og hjørnet $(2, 1)^T$), så største verdi er 9. Minste verdi ligger på randen. Langs hver av de fire sidene er $x_1 - 2x_2$ lineær i parameteren, så f er en parabel som vender ned, og minimum på hver side ligger i et hjørne:

$$f(0, 1) = 5, \quad f(1, 0) = 8, \quad f(2, 1) = 9, \quad f(1, 2) = 0.$$

Minste verdi er 0 i $(1, 2)^T$, det hjørnet som ligger lengst fra kammen.

- 7 Vi minimerer $|x|^2 = x_1^2 + x_2^2 + x_3^2$ under kravet $g(x) = x_1 + 2x_2 + 3x_3 = 1$. Lagrange gir $2x = \lambda(1, 2, 3)^T$, så x er parallell med planets normalvektor $(1, 2, 3)^T$, som det bør være: korteste vei fra origo til planet er rett ned på det. Med $x = \mu(1, 2, 3)^T$ gir kravet $\mu(1+4+9) = 1$, altså

$$x = \frac{1}{14} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix},$$

og avstanden er $|x| = \sqrt{14}/14 = 1/\sqrt{14}$ - som formelen for avstand fra punkt til plan sier.

- 8 Siden A er symmetrisk, er $\partial_x (x^T A x) = 2x^T A$ (jf. økt 10), så

$$V'(x) = x^T A - b^T = (Ax - b)^T.$$

En positivt definit matrise har bare positive egenverdier og er derfor inverterbar, så $V'(x) = 0$ har nøyaktig én løsning, $x = A^{-1}b$. Hessematrisen er $V''(x) = A$, som er positivt definit, så det kritiske punktet er et bunnpunkt, og siden V er et andreordens polynom, er det det globale minimumet. Gammel eksamensoppgave 1 er dette med tall, og normallikningene fra økt 7 er det samme med $A = X^T X$ og $b = X^T y$.

IKKE EKSAMENSRELEVANT

- 1 Av de 25 bordene i margin er 9 på 154, 10 på 155, 5 på 156 og 1 på 157. Vi vet at det var 29 bord totalt, så den relative frekvensen i hver bredde er antallet delt på 29. Disse skal passe med normalfordelingstettheten der vi har målinger, så vi minimerer

$$J(\mu, \sigma) = \sum_k (c_k - f(x_k))^2,$$

der x_k er breddene 154, ..., 158 og c_k de relative frekvensene. Steepest descent betyr å gå rett ned i fallinja:

$$p_{n+1} = p_n - \alpha \nabla J(p_n),$$

der $p = (\mu, \sigma)^T$.

```
import numpy as np
```

```
x = np.array([154, 155, 156, 157, 158])    # cm
c = np.array([9, 10, 5, 1, 0]) / 29        # 29 bord totalt

def phi(x, mu, s):
    return np.exp(-(x - mu)**2 / (2*s**2)) / (np.sqrt(2*np.pi)*s)

def J(p):
    return np.sum((c - phi(x, *p))**2)

def grad(p, h=1e-6):
    e = np.eye(2)
    return np.array([(J(p + h*e[k]) - J(p - h*e[k])) / (2*h)
                      for k in range(2)])

p = np.array([155.0, 1.0])
for n in range(2000):
    p = p - 2*grad(p)
print(p)
```

Vi får $\mu \approx 154.64$ og $\sigma \approx 1.09$. Gjennomsnittet og standardavviket til de 25 bordene er 154.92 og 0.86 - for høyt og for lavt, slik som ventet. Tar vi med de fire bordene som ble fjernet, er de 154.62 og 1.12, så minste kvadraters metode traff godt.