7.3 Optimering med bibetingelser

- I MAT1110 behandles Lagranges metode for å finne maksimum og/eller minimum av funksjoner av flere variable under gitte bibetingelser.
- Vi skal se på en alternativ måte å gjøre dette på for kvadratiske former under nokså spesielle bibetingelser.
- Hovedpoenget er at dette gir en interessant tolkning av hva egenverdiene til en symmetrisk matrise betyr.
- Til slutt ser vi på *singulærverdiene* til en matrise.

Teorem 6. *La A være en symmetrisk $n \times n$ matrise.*

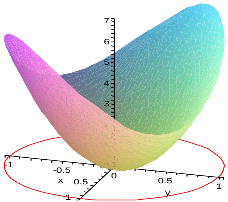
Betrakt den kvadratiske formen $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ på enhetssfæren

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}.$$

Da har Q et (globalt) maksimum M og et (globalt) minimum m på S.

- *M er A's største egenverdi, og $Q(\mathbf{x}) = M$ når $\mathbf{x}$ er en enhetsegenvektor tilhørende egenverdien M.*
- *m er A's minste egenverdi, og $Q(\mathbf{x}) = m$ når $\mathbf{x}$ er en enhetsegenvektor tilhørende egenverdien m.*

En illustrasjon med $Q(\mathbf{x}) = 5\,x_1^2 - 4\,x_1x_2 + 5\,x_2^2$. Her er $M = 7$ og $m = 3$.



Hva med de andre egenverdiene til en symmetrisk matrise A ?

Teorem 7. *La A være en symmetrisk $n \times n$ matrise (der $n \geq 2$) og betrakt*

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}\,.$$

Skriv egenverdiene til A i avtagende rekkefølge:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

så $M = \lambda_1$ og $m = \lambda_n$ i notasjonen fra Teorem 6.

La $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ være en o.n.b. for $\mathbb{R}^n$ som består av tilhørende egenvektorer (så $A\,\mathbf{u}_j = \lambda_j\,\mathbf{u}_j$ for $j = 1, \ldots, n$).

Sett $S_2 = \{\mathbf{x} \in S \mid \mathbf{x}$ er ortogonal på $\mathbf{u}_1\}$, m.a.o.

$$S_2 = \{\mathbf{x} \in S \mid \mathbf{x}^T \mathbf{u}_1 = 0\}\,.$$

Da har Q et (globalt) maksimum på S_2, nemlig λ_2, som oppnåes når $\mathbf{x} = \pm\,\mathbf{u}_2$.

Behold de samme antagelsene som i Teorem 7, og anta at $n \geq 3$. Vi kan da betrakte

$$S_3 = \left\{ \mathbf{x} \in S \mid \mathbf{x}^T \mathbf{u}_1 = \mathbf{x}^T \mathbf{u}_2 = 0 \right\},$$

eller mer generelt, for $k \in \{3, \ldots, n\}$,

$$S_k = \left\{ \mathbf{x} \in S \mid \mathbf{x}^T \mathbf{u}_1 = \cdots = \mathbf{x}^T \mathbf{u}_{k-1} = 0 \right\}.$$

Da har vi at Q et (globalt) maksimum på S_k. Dette maksimumet er egenverdien λ_k og det oppnåes når $\mathbf{x} = \pm \mathbf{u}_k$.

Dette er Teorem 8 i boka.

Preludium til singulærverdi dekomposisjonen av en matrise

La A være en $m \times n$ matrise og betrakt $S = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1\}$.

Hva er maksimum av funksjonen $f(\mathbf{x}) := \|A\,\mathbf{x}\|$ på S ?

Dette maksimumet er en viktig størrelse knyttet til A, som kalles *operatornormen til A*.

Funksjonen $f(\mathbf{x}) = \|A\,\mathbf{x}\|$ er *ikke* en kvadratisk form, så vi kan ikke bruke Teorem 6 på $f(\mathbf{x})$. Men vi har at

$$\|A\,\mathbf{x}\|^2 = (A\mathbf{x})^T A\,\mathbf{x} = \mathbf{x}^T A^T A\,\mathbf{x}$$

for alle $\mathbf{x} \in \mathbb{R}^n$.

Så $K(\mathbf{x}) := f(\mathbf{x})^2 = \|A\,\mathbf{x}\|^2$ er en kvadratisk form på $\mathbb{R}^n$ assosiert med den *symmetriske* $n \times n$ matrisen $A^T A$.

Teorem 6, anvendt på K, gir at

- *Maksimum av $\|A\mathbf{x}\|^2$ på S eksisterer og er lik A^TA's største egenverdi.*
- *Dette maksimumet oppnåes på enhetsegenvektorene til A^TA tilhørende denne egenverdien.*

La λ_1 betegne den største egenverdien til A^TA og la $\mathbf{v}_1$ være en tilhørende enhetsegenvektor.

Sett $\boxed{\sigma_1 = \sqrt{\lambda_1}}$.

Siden $\|A\mathbf{x}\| = \sqrt{\|A\mathbf{x}\|^2} = \sqrt{K(\mathbf{x})}$, får vi at

- *Maksimum av $\|A\mathbf{x}\|$ på S eksisterer og er lik σ_1.*
- *Dette maksimumet oppnåes på $\pm\mathbf{v}_1$.*

Dette betyr at $\boxed{\text{operatornormen til } A \text{ er lik } \sigma_1}$.

7.4 Singulærverdi dekomposisjonen

- Singulærverdi dekomposisjon til en rektangulær matrise A er en av de viktigste faktoriseringene av A (dvs. A skrives som et produkt av matriser).
- Den inneholder nyttig informasjon om A og de fundamentale underrommene knyttet til A, f.eks. nullrommet.
- Den brukes i mange praktiske anvendelser, men er også av stor betydning i teoretiske studier og ved numeriske betraktninger.

La A være en (rektangulær) $m \times n$ matrise. Da er $A^T A$ en reell *symmetrisk* $n \times n$ matrise, så den kan ortogonal diagonaliseres:

La $\mathbf{v}_1, \ldots, \mathbf{v}_n$ være en ortonormal basis for $\mathbb{R}^n$ som består av egenvektorer for $A^T A$ og la $\lambda_1, \ldots, \lambda_n$ være de tilhørende egenverdiene til $A^T A$.

- For hver $1 \leq i \leq n$ har vi at $\lambda_i = \|A\,\mathbf{v}_i\|^2$.
- Dette medfører at *alle egenverdiene til* $A^T A$ *er ikke-negative*. Så matrisen $A^T A$ er positiv semidefinit.

Ved å bytte om på rekkefølgen om nødvendig, kan vi anta at

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

Singulærverdiene til A, $\sigma_1, \ldots, \sigma_n$, defineres ved

$$\sigma_i = \sqrt{\lambda_i}\,, \quad i = 1, \ldots, n\,.$$

Vi har da at $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ og

$$\sigma_i = \sqrt{\lambda_i} = \sqrt{\|A\,\mathbf{v}_i\|^2} = \|A\,\mathbf{v}_i\|\,,\ i = 1, \ldots, n$$

Teorem 9. *La A være en $m \times n$ matrise ($A \neq O$).*

La $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ være en o.n.b. for $\mathbb{R}^n$ av egenvektorer for $A^T A$ med tilhørende egenverdier $\lambda_1, \ldots, \lambda_n$ ordnet i avtagende rekkefølge.

La $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ angi alle singulærverdiene til A som ikke er null.

Da er $\{A\,\mathbf{v}_1, \ldots, A\,\mathbf{v}_r\}$ en ortogonal basis for $\mathrm{Col}\,A$ og rang $A = r$.

Videre, hvis $r < n$, så er $A\,\mathbf{v}_{r+1} = \ldots = A\,\mathbf{v}_n = \mathbf{0}$.

Teorem 10 (Singulærverdi dekomposisjonen)

La A være en $m \times n$ reell matrise, $A \neq O$.

La $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ angi de singulærverdiene til A som ikke er null (så $r = \texttt{rank}\, A$). Da kan A faktoriseres på formen

$$A = U \Sigma V^T$$

der U er en $m \times m$ ortogonal matrise, V er en $n \times n$ ortogonal matrise, mens $\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$ er $m \times n$ "diagonal" matrisen der $D = diag\,(\sigma_1, \sigma_2, \ldots, \sigma_r)$.

- Faktoriseringen ovenfor kalles en singulærverdi dekomposisjon (eller SVD) til A.
- Merk at matrisene U og V er ikke entydig bestemte av A, så det finnes flere SVD'er for A.
- Kolonnene til U kalles ofte *venstre singulære vektorer for A*, mens kolonnene til V kalles *høyre singulære vektorer for A*.

Beregning av SVD for hånd.

1. [Finn V] Beregn A^TA og ortogonal diagonaliser denne.
   - La $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ være en o.n.b. for $\mathbb{R}^n$ av egenvektorer for A^TA tilhørende egenverdiene $\lambda_1, \dots, \lambda_n$, ordnet i avtagende rekkefølge.
   - Sett $V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$. Da er V ortogonal.
2. [Finn Σ] Sett $\sigma_j = \sqrt{\lambda_j}$ for $j = 1, \dots, n$.
   - Da er σ_j-ene singulærverdiene til A.
   - Sett $D = \mathrm{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$, der $\sigma_1, \sigma_2, \dots, \sigma_r$ angir alle singulærverdiene som ikke er null (så $\mathrm{rank}\, A = r$).
   - Dann $m \times n$ "diagonalmatrisen" Σ basert på D.
3. [Finn U] Beregn $\mathbf{u}_j = \frac{1}{\sigma_j} A\mathbf{v}_j$ for $j = 1, \dots, r$.
   - Mengden $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ er da ortonormal.
   - Dersom $r < m$: utvid $\{\mathbf{u}_1, \dots, \mathbf{u}_r\}$ til en o.n.b. $\{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ for $\mathbb{R}^m$.
   - Dann den ortogonale matrisen $U = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_m \end{bmatrix}$.
4. Da er $\boxed{A = U\Sigma V^T}$ en SVD for A.

SVD'en til A inneholder mye informasjon. F.eks. har vi at :

a. $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ er en o.n. basis for $\mathrm{Col}\, A$.

b. $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$ er en o.n. basis for $\mathrm{Nul}\, A^T$ (når $r < m$).

c. $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ er en o.n. basis for $\mathrm{Row}\ A$.

d. $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ er en o.n. basis for $\mathrm{Nul}\, A$ (når $r < n$).

Merk: Punkt b. kan utnyttes til å konstruere $\{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m\}$.

Invertibel Matrise Teoremet - Siste tillegg !

La A være en $n \times n$ matrise. Følgende er ekvivalent :

- *A er invertibel.*
- $(\mathrm{Col}\, A)^\perp = \{\mathbf{0}\}$.
- $(\mathrm{Nul}\, A)^\perp = \mathbb{R}^n$.
- $\mathrm{Row}\ A = \mathbb{R}^n$.
- *A har n singulærverdier som ikke er null.*

Kort om kondisjonstallet til en invertibel matrise

La A være en invertibel matrise og betrakt et system $A\mathbf{x} = \mathbf{b}$.

Vi vet at systemet har da en entydig løsning. I noen tilfeller vil avrundingsfeil i beregningene føre til at løsningen man beregner er ganske forskjellig fra den entydige løsningen man ønsker å finne.

Når er A av en slik type ? (Man sier at A er ille-kondisjonert).

La $A = U\Sigma V^T$ være en SVD til A. Siden U og V er ortogonale matriser vil årsaken til problemet ligge i Σ.

Man kan da innføre kondisjonstallet til A ved $\boxed{\text{cond } A = \frac{\sigma_1}{\sigma_n}}$.

Her er σ_1 er den største singulære verdien mens σ_n er den minste.

Det kan vises at et høyt kondisjonstall for A vil som regel føre til numerisk ustabilitet.

Redusert SVD og pseudoinvers

La $A = U\Sigma V^T$ være en SVD til en matrise A med rang r.
Utregning gir

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T \ .$$

Dette kan omskrives som

$$A = U_r \, D \, V_r^T$$

der $U_r = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{bmatrix}$, $D = \text{diag}\left(\sigma_1, \ldots, \sigma_r\right)$, $V_r = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_r]$.

En slik faktorisering av A kalles en *redusert* SVD til A.

Siden D er invertibel kan vi danne matrisen

$$A^+ = V_r \, D^{-1} \, U_r^T$$

Man kaller A^+ for pseudoinversen (eller Moore-Penrose inversen) til A.

Hva kan den pseudoinverse matrisen A^+ brukes til?

Betrakt et (inkonsistent) system $A\mathbf{x} = \mathbf{b}$. Da har vi at

$\hat{\mathbf{x}} := A^+\mathbf{b}$ er en minstekvadraters løsning av $A\mathbf{x} = \mathbf{b}$.

Poenget er at da er $A\hat{\mathbf{x}} = U_r U_r^T\mathbf{b} = \mathrm{Proj}_{\mathrm{Col}\,A}(\mathbf{b})$

og vi har sett tidligere at dette betyr at $\hat{\mathbf{x}}$ er en minstekvadraters løsning av systemet.

Merk: Det kan vises at $\boxed{\hat{\mathbf{x}} = A^+\mathbf{b}}$ har minst avstand til origo blant alle minstekvadraters løsninger til systemet.

Det er også verdt å poengtere at her antaes det *ikke* at A har lineært uavhengige kolonner; likevel angies det en formel for en minstekvadraters løsning, og det er ofte nyttig.

Lyst på mer lineær algebra ?!

Det fins flere tilbud!

- MAT3100 Lineær optimering (vår)
- MAT-IN4130 - Numerisk lineær algebra (høst)
- MAT3400/4400 - Lineær analyse med anvendelser (vår) – NB: bør ha tatt først MAT2400 - Reell analyse (vår)

Teorem 8

La A være en $n \times n$ matrise ($n \geq 2$)
og betrakt $Q(x) = x^T A x$. La egenverdiene
til A være $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, så:
⋮

Oppgave 7.3.6 a) Finn max val av
$Q(x) = 3x_1^2 + 9x_2^2 + 8x_1x_2$

$$A = \begin{bmatrix} 3 & 4 \\ 4 & 9 \end{bmatrix} \Rightarrow Q(x) = x^T A x$$

a) Ved teorem 6 er $\max\{Q(x) : x^T x = 1\}$
$= \lambda_1 = 11$

b) Kan velge u som enhets egenvektor
tilhørende $\lambda_1 = 11$. Regnesningene gir
$(1, 2)$ en egenvektor for λ_{11}, så $u = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$
oppnå maks verdi.

c) Ved Teorem 7 er maksimum av Q når $x^Tx = 1$ og $x^Tu = 0$ lik ......

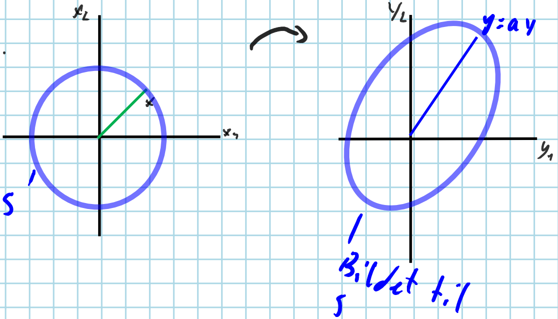
Preludium til singulær verdi dekomposisjonen til en matrise.

La A være en $m \times n$ matrise og betrakt $S = \{x \in \mathbb{R}^n : \|x\| = 1\}$ $(x^Tx = 1)$

Hva er maksimum av funksjonen $f(x) = \|Ax\|$ på S?

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

$A : 2 \times 2$



Operatornormet, hvor mye kan vektoren øke $(x \to Ax)$ lengden.

Funksjonen $f(x) = \|Ax\|$ er <u>ikke</u> en kvadratisk form, så vi kan ikke bruke teorem 6 på $f(x)$.

Men vi har at

$$\|Ax\|^2 = (Ax)^T(Ax) = x^T A^T A x \quad \text{for alle } x \in \mathbb{R}^n$$

Så $K(x) := f(x)^2 = \|Ax\|^2$ så det er en kvadratisk form på $\mathbb{R}^n$ assosiert med den symetriske $n \times n$ matrisen A^TA (Husk: $(A^TA)^T = A^T(A^T)^T = A^TA$)

Teorem 6, så anvendt på K, gir at

- Max av $\|Ax\|^2$ det eksisterer og er lik A^TA's største egenverdi
- Dette max oppnås for enhets egenvektorene til A^TA tilhørende denne egenverdien

La λ_1 være den største egenverdien til A^TA og la v_1 være en tilhørende egenvektoren.

Sett $\sigma_1 = \sqrt{\lambda_1}$

Siden $\|Ax\| = \sqrt{\|Ax\|^2} = \sqrt{K(x)}$, får vi at

- max av $\|Ax\|$ på S eksisterer og er lik σ_1
- Dette maksimumet oppnås på $\pm v_1$

Dette betyr at operatornormen til A er lik σ_1