SINGULÆRVERDI-DEKOMPOSISJON (SEC. 7.4)

Vi har lært om spektralteoremet for symmetriske matriser: Hvis en $n \times n$ matrise A er symmetrisk, dvs. hvis $A^T = A$, så finnes der en ortogonal basis $\{\vec{v}_1, \dots, \vec{v}_n\}$ for $\mathbb{R}^n$ som består av egenvektorer til matrisen A. Desuden, kan vi skrive

$$A = UDU^T,$$

for den ortogonale matrise

$$U = (\vec{v}_1, \dots, \vec{v}_n)$$

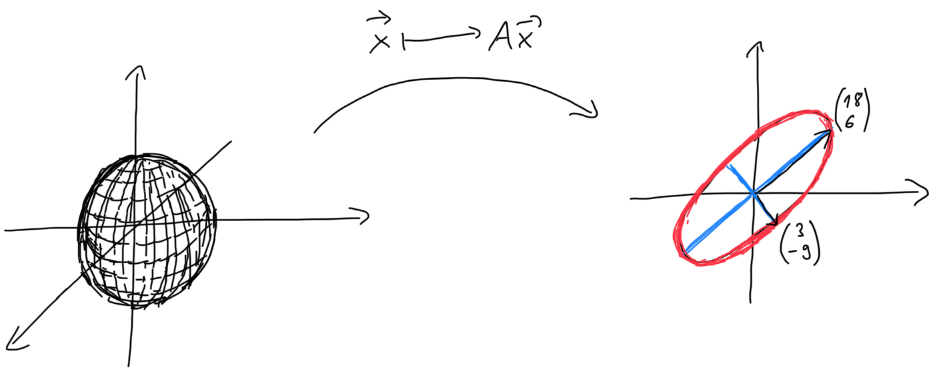
som har vektorerne $\vec{v}_1, \dots, \vec{v}_n$ som søyler, og hvor

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

er denne diagonalmatrise som har egenverdier $\lambda_1, \dots, \lambda_n$ på diagonalen. Spektralteoremet er vigtigt for mange anvendelser, men det har et problem: Ikke alle matriser er symmetriske! Vi skal nu lære om singulærverdi-dekomposisjonen som er en slaks spektralteorem men som gjælder for generelle $m \times n$ matriser.

Matriser og ellipser

Vi har set eksempler av $m \times n$ matriser A, hvor enhetssfæren $\{\vec{x} \in \mathbb{R}^n : \|\vec{x}\| = 1\}$ bliver transformert til ellipsoider under transformasjonen $\vec{x} \mapsto A\vec{x}$. Men det var ikke tilfald! Vi kommer til at vise at for hvert $m \times n$ matrise A enhetssfæren bliver transformert til en ellipsoid under transformasjonen $\vec{x} \mapsto A\vec{x}$. For at forstår hvordan det virker er det hjælpsom at se på et konkret eksempel:



Figur 1. Enhetssfæren transformerer til en ellipse.

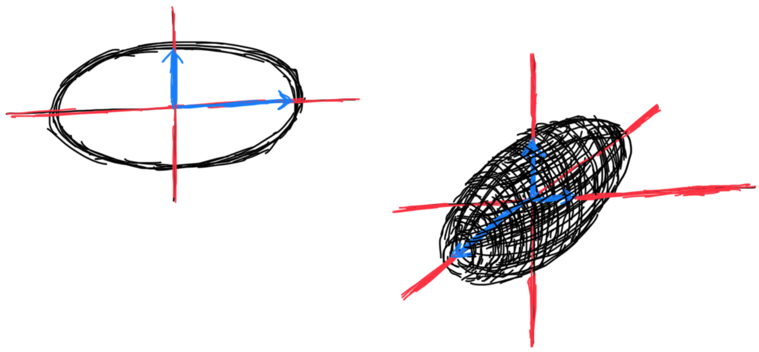
Eksempel. For matrisen

$$A = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix},$$

bliver enhetssfæren

$$\{\vec{x} \in \mathbb{R}^3 \ : \ \|\vec{x}\| = 1\}$$

transformert til en ellipse i $\mathbb{R}^2$ (Se Figur 1). Husk at hvert ellipsoid i $\mathbb{R}^n$ (eller ellipse i $\mathbb{R}^2$) har n akser som treffes i rette vinkler. De segmenter som går fra sentrum av en ellipsoid langs akserne til randen kælles også for *halvakser* av ellipsoiden. Se Figur 2.



FIGUR 2. Ellipsoider med deres akser (i rød). I blå har vi de så kællte halvakser av ellipsoiden.

Hvis vi nu kommer tilbake til ellipsoider som er transformasjoner av enhetssfæren under matrisetransformasjoner så kan vi forstår følgende punkter:

- I hvilke retninger peger akser av ellipsoiden for en matrise A?
- Hvor langt er halvakser av ellipsoiden for en matrise A?

Svarene på disse spørgesmål hænger tæt sammen med singulærverdi-dekomposisjon. La oss starte med en delvis svar på den annen spørgesmål.

Længden av den længste halvakse. La A være en $m \times n$ matrise. Hvis enhetssfæren i $\mathbb{R}^n$ bliver transformert til en ellipsoid, hvordan kan vi så regne ut hvor langt den længste halvakse er? Længden av den længste halvakse svarer til

$$\max_{\vec{x}:\|\vec{x}\|=1} \|A\vec{x}\|,$$

dvs. det er den største længde av alle vektorer $A\vec{x}$ for vektorer $\vec{x}$ på enhetssfæren. Normen $\|A\vec{x}\|$ og funksjonen $\|A\vec{x}\|^2$ er maksimert for den sammen vektor $\vec{x}$ og derfor kan vi like så godt maksimere $\|A\vec{x}\|^2$. Vi regner ut at

$$\|A\vec{x}\|_2^2 = \langle A\vec{x}, A\vec{x}\rangle = \langle \vec{x}, A^TA\vec{x}\rangle. \tag{1}$$

Det er måske ikke helt åbenlyst hvordan man maksimerer denne funksjon over alle $\vec{x}$ i enhetssfæren men vi kan i første omgang bruke spektralteoremet for at gør funksjonen lidt nemmer. Husk at der finnes en ortogonale $n \times n$ matrise U og en $n \times n$ diagonalmatrise D slik at

$$A^TA = UDU^T.$$

Vi har også set at D har formen

$$D = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

hvor $\lambda_1, \dots, \lambda_n$ er egenverdierne av matrise $A^T A$, og hvor vi kan anta at

$$\lambda_1 \geq \cdots \geq \lambda_n.$$

La oss nu simplifisere funksjonen fra (1). Vi har

$$\|A\overrightarrow{x}\|^2 = \langle \overrightarrow{x}, A^T A\overrightarrow{x} \rangle = \langle \overrightarrow{x}, UDU^T \overrightarrow{x} \rangle = \langle U^T \overrightarrow{x}, DU^T \overrightarrow{x} \rangle.$$

Hvis vi maksimerer denne funksjon kan vi bruke at transformasjonen $\overrightarrow{x} \mapsto U^T \overrightarrow{x}$ transformerer enhetssfæren til sigselv. Derfor har vi

$$\max_{\overrightarrow{x} : \|\overrightarrow{x}\|=1} \|A\overrightarrow{x}\|^2 = \max_{\overrightarrow{x} : \|\overrightarrow{x}\|=1} \langle U^T \overrightarrow{x}, DU^T \overrightarrow{x} \rangle = \max_{\overrightarrow{y} : \|\overrightarrow{y}\|=1} \langle \overrightarrow{y}, D\overrightarrow{y} \rangle,$$

hvor vi har skiftet variabler til $\overrightarrow{y} = U^T \overrightarrow{x}$. Til sidst ser vi at

$$\max_{\overrightarrow{x} : \|\overrightarrow{x}\|=1} \|A\overrightarrow{x}\|^2 = \max_{\overrightarrow{y} : \|\overrightarrow{y}\|=1} \langle \overrightarrow{y}, D\overrightarrow{y} \rangle = \max_{\overrightarrow{y} : \|\overrightarrow{y}\|=1} \left(\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \right).$$

Bemærk at vi maksimerer over alle indganger $y_1, \dots, y_n$ slik at

$$y_1^2 + \cdots + y_n^2 = 1.$$

Derfor har vi

$$\max_{\overrightarrow{x} : \|\overrightarrow{x}\|=1} \|A\overrightarrow{x}\|^2 = \max_{\overrightarrow{y} : \|\overrightarrow{y}\|=1} \left(\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \right) = \lambda_1,$$

og vi konkluderer at længden av den længste halvakse er

$$\max_{\overrightarrow{x} : \|\overrightarrow{x}\|=1} \|A\overrightarrow{x}\| = \sqrt{\lambda_1}.$$

Hvilken vektor $\overrightarrow{x}$ med $\|\overrightarrow{x}\| = 1$ maksimerer funksjonen $\|A\overrightarrow{x}\|^2$? Det er nu nemt at regne ut. Vi ser at $\overrightarrow{y} = \overrightarrow{e_1}$ maksimerer funksjonen efter koordinateskift $\overrightarrow{y} = U^T \overrightarrow{x}$. Derfor finner vi at $\overrightarrow{x} = U\overrightarrow{y} = U\overrightarrow{e_1}$ maksimerer $\|A\overrightarrow{x}\|^2$ over enhetssfæren. Men $\overrightarrow{x} = U\overrightarrow{e_1}$ er jo den første søyle av U som er egenvektoren av den symmetriske matrise $A^T A$ til egenverdien λ_1. Med denne argument har vi vist den efterfølgende teorem:

Teorem. *La A være en $m \times n$ matrise og la*

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

være egenverdier til matrisen $A^T A$ med ortonormale egenvektorer

$$\{\overrightarrow{v}_1, \overrightarrow{v}_2, \dots, \overrightarrow{v}_n\}.$$

Der er

$$\max_{\overrightarrow{x} : \|\overrightarrow{x}\|=1} \|A\overrightarrow{x}\| = \|A\overrightarrow{v}_1\| = \sqrt{\lambda_1}.$$

Teoremet sier at den længste halvakse i ellipsoiden til matrisen A har længden $\sqrt{\lambda_1}$ og peger i retningen $A\overrightarrow{v}_1$ hvor λ_1 er den største egenverdi av matrisen $A^T A$ og $\overrightarrow{v}_1$ er den tilsvarende egenvektor.

Længde og retning av de andre halvakser. Med en lignende argument som den fra sidste avsnit kan man vise at længder av de andre halvakser er tallene $\sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$ og at de peger i retninger $A\overrightarrow{v}_2, \dots, A\overrightarrow{v}_n$ hvor $\lambda_2, \dots, \lambda_n$ er egenverdier av matrisen $A^T A$ slik at $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ og $\overrightarrow{v}_2, \dots, \overrightarrow{v}_n$ er de tilsvarende egenvektorer. Bemærk at hvis det er sant, så er vektorerne $A\overrightarrow{v}_1, A\overrightarrow{v}_2, \dots, A\overrightarrow{v}_n$ ortogonal, og egenverdier $\lambda_1, \dots, \lambda_n \geq 0$. La oss se i næste avsnit at det er rigtigt.

Singulærverdier og singulærverdi-dekomposisjon

For hvert $m \times n$ matrise A finnes der en ortogonal basis $\{\overrightarrow{v}_1, \dots, \overrightarrow{v}_n\}$ som består av egenvektorer til den symmetriske $n \times n$ matrise $A^T A$. La $\lambda_1, \dots, \lambda_n$ være de tilsvarende egenverdier for $A^T A$. Vi kan regne ut at

$$\|A\overrightarrow{v}_i\|^2 = \langle A\overrightarrow{v}_i, A\overrightarrow{v}_i\rangle = \langle \overrightarrow{v}_i, A^T A\overrightarrow{v}_i\rangle = \langle \overrightarrow{v}_i, \lambda_i\overrightarrow{v}_i\rangle = \lambda_i, \tag{2}$$

og vi konkluderer at $\lambda_i \geq 0$ for alle $i \in \{1, \dots, n\}$. Vi kan altid anta at

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

og vi definerer *singulærverdier* for matrisen A som

$$\sigma_i = \sqrt{\lambda_i}.$$

Bemærk fra (2) at $\sigma_i = \|A\overrightarrow{v}_i\|$.

Teorem 9. *La A være en $m \times n$ matrise og la*

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

være egenverdier til matrisen $A^T A$ med ortonormale egenvektorer

$$\{\overrightarrow{v}_1, \overrightarrow{v}_2, \dots, \overrightarrow{v}_n\}.$$

Anta at $\lambda_1, \dots, \lambda_r > 0$ og at $\lambda_{r+1} = \cdots = \lambda_n = 0$. Da er

- *Mængden $\{A\overrightarrow{v}_1, \dots, A\overrightarrow{v}_r\}$ er en ortogonal basis for Col(A).*
- *Vi har $r =$ rank(A).*

Bevis. La oss se først at vektorerne i mængden $\{A\overrightarrow{v}_1, \dots, A\overrightarrow{v}_r\}$ er ortogonale. Hvis $i \neq j$ så har vi $\langle \overrightarrow{v}_i, \overrightarrow{v}_j\rangle = 0$ (de er jo ortonormale egenvektorer for matrisen $A^T A$). Derfor finner vi at

$$\langle A\overrightarrow{v}_i, A\overrightarrow{v}_j\rangle = \langle \overrightarrow{v}_i, A^T A\overrightarrow{v}_j\rangle = \langle \overrightarrow{v}_i, \lambda_j\overrightarrow{v}_j\rangle = \lambda_j\langle \overrightarrow{v}_i, \overrightarrow{v}_j\rangle = 0.$$

Det viser at $\{A\overrightarrow{v}_1, \dots, A\overrightarrow{v}_r\}$ er en mængde av ortogonale vektorer, og selvfølgelig ligger disse vektorer i Col(A). For at vise at de former en ortogonal basis for Col(A) er det derfor nok at sjekke at de utspenner hele Col(A).

La oss se på en vektor $\overrightarrow{y} \in$ Col(A) som kan skrives som $\overrightarrow{y} = A\overrightarrow{x}$ for en vektor $\overrightarrow{x} \in \mathbb{R}^n$. Matrisen $A^T A$ er symmetrisk og vi har set før at mængden $\{\overrightarrow{v}_1, \overrightarrow{v}_2, \dots, \overrightarrow{v}_n\}$ av egenvektorer er en ortonormal basis for $\mathbb{R}^n$. Derfor finnes der tal $c_1, \dots, c_n$ slik at

$$\overrightarrow{x} = c_1\overrightarrow{v}_1 + \cdots + c_n\overrightarrow{v}_n,$$

og vi ser at

$$\overrightarrow{y} = A\overrightarrow{x} = c_1 A\overrightarrow{v}_1 + \cdots + c_n A\overrightarrow{v}_n. \tag{3}$$

Det ser ut som at der er lidt for mange vektorer i denne lineærkombinasjon. Vi har også brukt vektorer $A\overrightarrow{v}_{r+1}, \dots, A\overrightarrow{v}_n$. Kan vi si noget om disse vektorer? Vi kan regne ut deres norm! For hvert $i \geq r+1$ har vi

$$\|A\overrightarrow{v}_i\|_2^2 = \langle A\overrightarrow{v}_i, A\overrightarrow{v}_i\rangle = \langle \overrightarrow{v}_i, A^T A\overrightarrow{v}_i\rangle = \lambda_i\langle \overrightarrow{v}_i, \overrightarrow{v}_i\rangle = \lambda_i\|\overrightarrow{v}_i\|_2^2 = 0,$$

fordi $\lambda_i = 0$ hvis $i \geq r+1$. Vi konkluderer at $A\overrightarrow{v}_{r+1} = \cdots = A\overrightarrow{v}_n = \overrightarrow{0}$, og fra (3) finner vi at

$$\overrightarrow{y} = A\overrightarrow{x} = c_1 A\overrightarrow{v}_1 + \cdots + c_n A\overrightarrow{v}_r.$$

Det viser at $\overrightarrow{y} \in$ Span$\{A\overrightarrow{v}_1, \dots, A\overrightarrow{v}_r\}$ og vi konkluderer at mængden

$$\{A\overrightarrow{v}_1, \dots, A\overrightarrow{v}_r\}$$

utspenner Col(A). Mængden $\{A\overrightarrow{v}_1, \dots, A\overrightarrow{v}_r\}$ er en basis for Col(A) og fordi den inneholder r vektorer har vi Rank$(A) = r$.

$\square$

La oss nu komme til singulærverdi-dekomposisjonen.

Teorem 10 (Singulærverdi-dekomposisjon)**.** *La A være en $m \times n$ matrise med rang r. Da finnes der en $m \times m$ ortogonale matrise U og en $n \times n$ ortogonale matrise V slik at*

$$A = U\Sigma V^T, \tag{4}$$

med matrisen

$$\Sigma = \begin{pmatrix} D & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix},$$

hvor

$$D = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{pmatrix}.$$

Dekomposisjonen i (4) hedder *singulærverdi-dekomposisjon* av matrisen A. Ortogonale matriser U og V i denne dekomposijon er *ikke* unik for matrisen A men singulærverdier (og dermed matrisen Σ) er unike. Søylerne i matrisen U kælles *venstre singulærvektorer* av A og søylerne i matrisen V kælles for *høyre singulærvektorer* av A.

Bevis. Like som i Teorem 9, la $\overrightarrow{v}_1, \ldots, \overrightarrow{v}_n$ være egenvektorer til matrisen A^TA for egenverdier $\lambda_1, \ldots, \lambda_n$ hvor vi anta at

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \geq \lambda_{r+1} = \cdots = \lambda_n = 0.$$

Vi definerer singulærverdier $\sigma_i = \sqrt{\lambda_i}$ og husker at

$$\|A\overrightarrow{v}_i\|_2 = \sigma_i,$$

for hvert $i \in \{1, \ldots, n\}$. For hvert $i \in \{1, \ldots, r\}$ kan vi definere $\overrightarrow{u}_i$ den normaliserte vektor $A\overrightarrow{v}_i$, dvs.

$$\overrightarrow{u}_i = \frac{A\overrightarrow{v}_i}{\|A\overrightarrow{v}_i\|} = \frac{A\overrightarrow{v}_i}{\sigma_i}. \tag{5}$$

Efter Teorem 9 har vi en ortonormal basis $\{\overrightarrow{u}_1, \ldots, \overrightarrow{u}_r\}$ for $\mathrm{Col}(A)$ som vi kan forlenge til en ortonormalbasis $\{\overrightarrow{u}_1, \ldots, \overrightarrow{u}_r, \overrightarrow{u}_{r+1}, \ldots, \overrightarrow{u}_m\}$ for hele $\mathbb{R}^m$. Vi kan definere to ortogonale matriser

$$U = \left(\overrightarrow{u}_1, \ldots, \overrightarrow{u}_m\right),$$

og

$$V = \left(\overrightarrow{v}_1, \ldots, \overrightarrow{v}_n\right).$$

Fra (5) ser vi at $A\overrightarrow{v}_i = \sigma_i\overrightarrow{u}_i$ for alle $i \in \{1, \ldots, r\}$ og vi ved også at $A\overrightarrow{v}_i = 0 = \sigma_i\overrightarrow{u}_i$ for alle $i \geq r+1$. Derfor finner vi

$$AV = (A\overrightarrow{v}_1, \ldots, A\overrightarrow{v}_n) = \left(\sigma_1\overrightarrow{u}_1, \ldots, \sigma_r\overrightarrow{u}_r, \overrightarrow{0}, \ldots, \overrightarrow{0}\right).$$

Vi finner også at

$$\begin{aligned} U\Sigma &= \left(\overrightarrow{u}_1, \ldots, \overrightarrow{u}_m\right) \begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \mathbf{0} \\ & & & \sigma_r & \\ & & \mathbf{0} & & \mathbf{0} \end{pmatrix} \\ &= \left(\sigma_1\overrightarrow{u}_1, \ldots, \sigma_r\overrightarrow{u}_r, \overrightarrow{0}, \ldots, \overrightarrow{0}\right) \\ &= AV. \end{aligned}$$

Fordi V er ortogonal kan vi regne ut at
$$U\Sigma V^T = AVV^T = A.$$
□

Prøv selv at regne ut singulærverdi-dekomposisjonen for 2×3 matrisen fra den eksempel i starten.

Anvendelser av singulærverdi-dekomposisjon

Singulærverdi-dekomposisjonen har mange anvendelser. Her, vi kan se to av dem:

Basiser av underrom assosiert med en matrise. La oss anta at A er en $m \times n$ matrise med rang r singulærverdi-dekomposisjon
$$A = U\Sigma V^T,$$
med en ortogonal $n \times n$ matrise U, en ortogonal $m \times m$ matrise V. La $\vec{u}_1, \dots, \vec{u}_m$ være de venstre singulærvektorer, dvs. søylerne i U, og $\vec{v}_1, \dots, \vec{v}_n$ være de høyre singulærvektorer, dvs. søylerne i V. Vi har set i Teorem 9 at
$$\{\vec{u}_1, \dots, \vec{u}_r\}$$
er en ortonormalbasis for $\mathrm{Col}(A)$. Fordi de venstre singulærvektorer er en ortonormal basis til $\mathbb{R}^m$ konkluderer vi at
$$\{\vec{u}_{r+1}, \dots, \vec{u}_m\}$$
er en ortonormal basis for $\mathrm{Col}(A)^\perp = \mathrm{Nul}(A^T)$. Vi har også set at $\|A\vec{v}_i\| = \sigma_i$ som er like 0 hvis og kun hvis $i \geq r+1$. Derfor er $\vec{v}_{r+1}, \dots, \vec{v}_n \in \mathrm{Nul}(A)$. Fordi $\dim(\mathrm{Nul}(A)) = n - r$ (efter rangteoremet) ser vi at
$$\{\vec{v}_{r+1}, \dots, \vec{v}_n\}$$
er en ortonormal basis for $\mathrm{Nul}(A)$. Fordi de høyre singulærvektorer former en ortonormalbasis for $\mathbb{R}^n$ ser vi at
$$\{\vec{v}_1, \dots, \vec{v}_r\}$$
er en ortonormal basis for $\mathrm{Nul}(A)^\perp = \mathrm{Col}(A^T) = \mathrm{Row}(A)$.