Diagonalisering

En **diagonal matris** är en matris som bara har nollor utanför diagonalen. Sådana matriser är väldigt enkla att hantera på alla möjliga sätt och vi ska se hur man kan utnyttja detta för bl a symmetriska matriser.

Definition 1 *Låt A vara en $n \times n$-matris. Vi säger att A är* **diagonaliserbar** *om det finns inverterbar matris P och diagonal matris D sådana att*

$$A = PDP^{-1}.$$

En direkt tillämpning av diagonaliseringen är att det blir snabbt att beräkna en godtycklig potens av A. Vi får

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1},$$

vilket med en uppenbar induktion ger

$$A^n = PD^nP^{-1}.$$

Det är trivialt att beräkna D^n (varför?) och sedan behöver vi bara multiplicera med P och P^{-1}.
Genom att multiplicera med P från höger i båda leden så får vi den ekvivalenta likheten

$$AP = PD.$$

Antag nu att A har n stycken linjärt oberoende egenvektorer $\mathbf{v}_1, \ldots, \mathbf{v}_n$ med motsvarande egenvärden $\lambda_1, \ldots, \lambda_n$, d v s $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ för $i = 1, \ldots, n$. Låt P vara matrisen som har $\mathbf{v}_1, \ldots, \mathbf{v}_n$ som kolonner och låt D vara den diagonalmatris som har $\lambda_1, \ldots, \lambda_n$ på diagonalen. Eftersom kolonnerna är linjärt oberoende så kommer P att vara inverterbar. Vi får

$$AP = A\begin{pmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} A\mathbf{v}_1 & \ldots & A\mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \lambda_1\mathbf{v}_1 & \ldots & \lambda_n\mathbf{v}_n \end{pmatrix} = PD.$$

Vi har alltså visat att om A har n stycken linjärt oberoende egenvektorer så är den diagonaliserbar. Omvänt så antag att $A = PDP^{-1}$ är diagonaliserbar. Då ser man direkt från $AP = PD$ att kolonnerna i P är egenvektorer till A med elementen i D:s diagonal som egenvärden. Vi har alltså visat följande:

Sats 1 *En $n \times n$-matris är diagonaliserbar om och endast om den har n stycken linjärt oberoende egenvektorer.*

Genom att utnyttja vad vi vet om symmetriska matriser så får vi följande följdsats:

Sats 2 *Till varje symmetrisk matris A så finns det diagonal matris D och ON-matris P sådana att*

$$A = PDP^t.$$

Exempel 1 *Vi såg senast att*

$$A = \begin{pmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{pmatrix},$$

hade egenvärdena 1 *respektive* 1/2 *med motsvarande (normerade) egenvektorer* $(1/\sqrt{2})(1\ 1)^t$ *och* $(1/\sqrt{2})(1\ -1)^t$. *Vi sätter då enligt receptet ovan*

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ och } D = \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix},$$

vilket ger

$$PDP^t = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1/2 \\ 1 & -1/2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = A$$

som satsen lovade oss.

Vi ska nu relatera diagonaliseringen av en matris till basbyten för linjära avbildningar. Antag att vi har en linjär avbildning som har matrisen A i standardbasen. Vi såg tidigare att om $F = (\mathbf{f}_1\, \mathbf{f}_2\, \ldots\, \mathbf{f}_n)$ är en annan bas så ges matrisen för den linjära avbildningen i basen F av matrisen A_F där

$$A = FA_F F^{-1}.$$

(Vi formulerade och visade bara satsen för $n = 3$, men samma argument fungerar för godtyckligt n.) Om vi jämför detta med diagonaliseringen

$$A = PDP^{-1}$$

så får vi att D helt enkelt är matrisen A_P för den linjära avbildningen i basen P, d v s basen som består av egenvektorerna. Men om vi tänker efter lite så kan vi inse att detta bara är en omformulering av villkoret att vara egenvektorer.
Antag t ex att $\mathbf{v}_1$ är den första egenvektorn i P. Då kommer den att ha koordinaterna $(1\,0\,0\,\ldots\,0)^t$ i basen P. Men $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ och $\lambda_1 \mathbf{v}_1$ har koordinaterna $(\lambda_1\,0\,0\,\ldots\,0)^t$ i basen P, så första kolumnen i A_P kommer att vara $(\lambda_1\,0\,0\,\ldots\,0)^t$ vilket är samma som för D. På samma sätt motiverar man att de andra kolumnerna också stämmer överens.
Summerar vi så får vi alltså att den linjära avbildningen som har A som matris (i standardbasen) har en väldigt enkel matris i basen av egenvektorer: Den är diagonal.

Singulärvärdesuppdelning

Nu ska vi titta på vad man kan göra om man har en matris som inte är diagonaliserbar. Om man släpper på villkoret $P = Q$ i faktoriseringen $A = PDQ^t$ av en symmetrisk matris så visar det sig att man alltid kan finna en sådan faktorisering av en godtycklig (inte nödvändigtvis kvadratisk) matris. Beviset av följande sats ligger utanför den här kursen.

Sats 3 *Låt A vara en $m \times n$-matris. Då finns det matriser U, V och Σ sådana att:*

1. $A = U\Sigma V^t$.
2. U *är en ON-matris av storlek* $m \times m$.
3. V *är en ON-matris av storlek* $n \times n$.
4. Σ *är en 'diagonal' matris av formen*

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ldots & \sigma_r & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

där $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

Faktoriseringen $A = U\Sigma V^t$ kallas för en **singulärvärdesuppdelning (SVD)** *av A och $(\sigma_1, \ldots, \sigma_r)$ kallas för A:s* **singulärvärden**.

Anmärkning 1 *Matrisen Σ är unikt bestämd (och därmed är singulärvärdena unikt bestämda), men däremot är matriserna U och V inte unikt bestämda.*

Det finns ingen enkel algoritm att beräkna SVD av en matris för hand, däremot finns det bra numeriska metoder som lämpar sig för datorberäkningar. Vi ska dock se att faktoriseringen hänger ihop med de symmetriska matriserna AA^t och A^tA. Antag att $A = U\Sigma V^t$ är en SVD av A. Vi får då

$$AA^t = (U\Sigma V^t)(U\Sigma V^t)^t = U\Sigma V^t V \Sigma^t U^t = U(\Sigma\Sigma^t)U^t.$$

Observera att $\Sigma\Sigma^t$ är diagonal med σ_i^2 på diagonalen så vi ser att vi har en diagonalisering av AA^t. Detta ger att kolonnerna i U är egenvektorer till AA^t med motsvarande egenvärden σ_i^2. På motsvarande sätt får vi att

$$A^tA = V(\Sigma^t\Sigma)V^t.$$

Detta ger att kolonnerna i V (det är därför man envisas med att ha V^t istället för V) är egenvektorer till A^tA.
En viktig tillämpning på SVD är att man kan utnyttja den till att ta ut det viktiga ur en matris som innehåller någon typ av information. Exempel finns inom signalbehandling, om A är en term/dokument-matris för informationssökning (så kallad LSI som ni hörde om på temaföreläsningen) och om A innehåller pixlarna till en bild. Vi ska nu antyda hur det här fungerar.
Om vi låter kolonn i i U vara $\mathbf{u}_i$ och kolonn i i V vara $\mathbf{v}_i$ så får vi

$$A = U\Sigma V^t = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ldots & \sigma_r & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^t \\ \mathbf{v}_2^t \\ \vdots \\ \mathbf{v}_n^t \end{pmatrix} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^t.$$

Observera först att $\mathbf{u}_i\mathbf{v}_i^t$ är $m \times n$-matriser, så att summan i högerledet är en uppdelning av A som summan av r stycken matriser. Eftersom kolonnerna i U och V är normerade så kommer alla element i $\mathbf{u}_i\mathbf{v}_i^t$ att vara små (≤ 1). Därmed kommer det att vara singulärvärdena som avgör hur stora matriserna i summan är. Om vi nu tar en del av summan

$$\sum_{i=1}^{s} \sigma_i \mathbf{u}_i \mathbf{v}_i^t$$

med $s < r$ så kommer denna att vara nästan lika med A om σ_i för $i > s$ är mycket mindre än det största singulärvärdet σ_1.
För signalbehandling och för LSI fungerar det så att man genom att bara ta med de största singulärvärdena filtrerar bort skräp i informationen. När det gäller bilder så kan man på det här sättet komprimera lagringen av dem genom att försumma termerna då $i > s$. (Det finns idag bättre metoder för komprimering, men detta exempel är en bra illustration hur SVD fungerar.)