

Algorithmik kontinuierlicher Systeme

SVD und PCA: Singulärwertzerlegung und Hauptkomponentenanalyse



- SVD = *Singular Value Decomposition*
PCA = *Principal Component Analysis*
- Behandlung singulärer Matrizen
- über- / unterbestimmte Gleichungssysteme
- weitere Anwendungen
 - ▶ Kompression großer Matrizen
 - ▶ Hauptachsen -Transformation
 - ▶ Data Mining : LSI - *Latent Semantic Indexing*

- Vektorräume, lineare Abbildungen, Matrizen
- Eigenwerte
- Spezielle Matrizen
 - orthogonale Matrizen,
 - symmetrische Matrizen, positiv/negativ (semi-)definite M.
- (Vektor-)Normen, Matrixnorm, Konditionszahl
- Diagonalisierbarkeit, Spektralsatz für sym. Matrizen

Vektorräume & lineare Abbildungen

- Kontinuierliche Funktionen bilden einen unendlich dimensionalen Vektorraum;
- Nach Diskretisierung (endlich viele samples) ein endlich dimensionaler Raum: aber Dimension uU sehr groß (sozusagen $\dim(V) \rightarrow \infty$)
- Endlich dimensionale Vektorräume kann man (nach Wahl einer Basis) immer mit \mathbb{R}^m identifizieren
- Standardnotation:
Spaltenvektoren $\mathbf{x} \in \mathbb{R}^m$,
$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = [x_1, \dots, x_m]^T$$
- **Lineare Abbildungen** werden mit Matrizen beschrieben:

Vektorräume & lineare Abbildungen

- Die **linearen Abbildungen** werden mit Matrizen beschrieben

- $$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$
dann wird durch $\mathbf{x} \mapsto A\mathbf{x}$ eine

lineare Abbildung von \mathbb{R}^m nach mit \mathbb{R}^n definiert
(Darstellung bzgl. der kanonischen Basis!)

- $B = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ und $D = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}$ Basen von \mathbb{R}^m bzw. \mathbb{R}^n

$$A_D^B = \begin{bmatrix} | & & | \\ \mathbf{d}_1 & \cdots & \mathbf{d}_n \\ | & & | \end{bmatrix}^{-1} \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \cdot \begin{bmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_m \\ | & & | \end{bmatrix}$$

- Bezeichnungen

- ▶ $n \times m$ – Matrix
 - n Zeilen,
 - m Spalten

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$$

m – Vektor

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

- ▶ Transponierte

$$A^T = \begin{bmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & \ddots & \vdots \\ a_{1m} & \cdots & a_{nm} \end{bmatrix}$$

$$\mathbf{x}^T = [x_1, \dots, x_m]$$

- ▶ Produkte: A $n \times m$ – Matrix, B $m \times k$ – Matrix:
das Produkt $C = AB$ ist eine $n \times k$ – Matrix mit

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj}$$

- ▶ Spezialfälle $A\mathbf{x}$, $\mathbf{y}^T A$, $\mathbf{x}^T \mathbf{y}$, $\mathbf{x}\mathbf{y}^T$

- ▶ Beachte $\mathbf{x}^T \mathbf{y}$ ist ein Skalar (das **Skalarprodukt**) $\mathbf{x}^T \mathbf{y} = \mathbf{x} \circ \mathbf{y} = \sum_{i=1}^n x_i y_i$

▶

$\mathbf{x}\mathbf{y}^T$ hingegen ist die $n \times m$ – Matrix (falls $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$)

$$(\mathbf{x}\mathbf{y}^T)_{ij} = x_i y_j$$

Eigenvektoren und -werte (einer quadratischen Matrix A)

$$A\mathbf{v} = \lambda \cdot \mathbf{v}$$

Eigenvektor

$$\mathbf{v} \in \mathbb{C}^m, \mathbf{v} \neq 0$$

Eigenwert

$$\lambda \in \mathbb{C}$$

Beispiel

$$\begin{pmatrix} 6 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$A\mathbf{v} = \lambda \cdot \mathbf{v} \Leftrightarrow (A - \lambda \cdot Id)\mathbf{v} = 0 \Leftrightarrow \det(A - \lambda \cdot Id) = 0$$

- EW sind Nullstellen des **char. Polynoms!** Wieviele? $p_A(\lambda) = \det(A - \lambda \cdot Id)$
- Polynom m -ten Grades hat **höchstens** m **verschiedene Lösungen**
 - ▶ können auch komplex sein,
 - ▶ mit Vielfachheit gezählt gibt es genau m Lösungen !

Eigenvektoren und –werte: Beispiel

$$S = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

- dann $S - \lambda \cdot Id = \begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix} \Rightarrow p_S(\lambda) = (2-\lambda)^2 - 1 = 0$
- Die Eigenwerte: $\lambda_1 = 3$ und $\lambda_2 = 1$
- Die Eigenvektoren: $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ und $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ bzw. $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

Spezielle Matrizen

- Orthogonale Matrizen: (äquivalente Charakterisierungen)

- ▶ $U^T U = U U^T = Id$, $U^{-1} = U^T$
- ▶ die Transformation $x \mapsto Ux$ ist längen- und winkel-treu (bzgl. der Euklidischen Norm $\|\cdot\|_2$)
- ▶ die Spalten/Zeilen von U bilden eine ONB

- Symmetrische Matrizen: $A = A^T$

Eine symmetrische Matrix A heißt

- ▶ **positiv definit**, wenn $x^T A x > 0 \quad \forall x \neq 0$
- ▶ **positiv semi-definit**, wenn $x^T A x \geq 0 \quad \forall x$
- ▶ **negativ semi-definit**, wenn $x^T A x \leq 0 \quad \forall x$
- ▶ **negativ definit**, wenn $x^T A x < 0 \quad \forall x \neq 0$
- ▶ **indefinit** sonst

Spezielle Matrizen

- Warum interessieren uns (positiv) definite Matrizen?
 - sie treten immer auf, wenn Minimierungsprobleme zu lösen sind
 - viele Aufgaben lassen sich als Minimierungsprobleme beschreiben
- Charakterisierung durch Eigenwerte
 - eine symmetrische Matrix ist genau dann positiv definit wenn alle $\text{EW} > 0$ sind, und pos. semi-definit wenn alle $\text{EW} \geq 0$ sind.
- Beispiele

- $$\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \lambda_1 = 3, \lambda_2 = 1$$

- Ist \mathbf{B} eine beliebige $n \times m$ -Matrix, dann sind $\mathbf{B}\mathbf{B}^T$ und $\mathbf{B}^T\mathbf{B}$ positiv semi-definit. Die kleinere der beiden (beide, wenn $n=m$) ist positiv definit wenn die Zeilen bzw. die Spalten von \mathbf{B} linear unabhängig sind (gleichwertig: die Matrix \mathbf{B} hat maximalen Rang $= \min\{n, m\}$)

$$\begin{bmatrix}
 4 & -1 & & & & \\
 -1 & 4 & & & & \\
 & & \ddots & & & \\
 & & & 4 & -1 & \\
 & & & -1 & 4 & \\
 & & & & & \ddots & \\
 & & & & & & 4 & -1 \\
 & & & & & & -1 & 4
 \end{bmatrix}
 \begin{bmatrix}
 u_{11} \\
 \vdots \\
 u_{1n} \\
 u_{21} \\
 \vdots \\
 u_{2n} \\
 \vdots \\
 u_{nn}
 \end{bmatrix}
 =
 \begin{bmatrix}
 * \\
 \vdots \\
 * \\
 * \\
 \vdots \\
 * \\
 \vdots \\
 *
 \end{bmatrix}$$

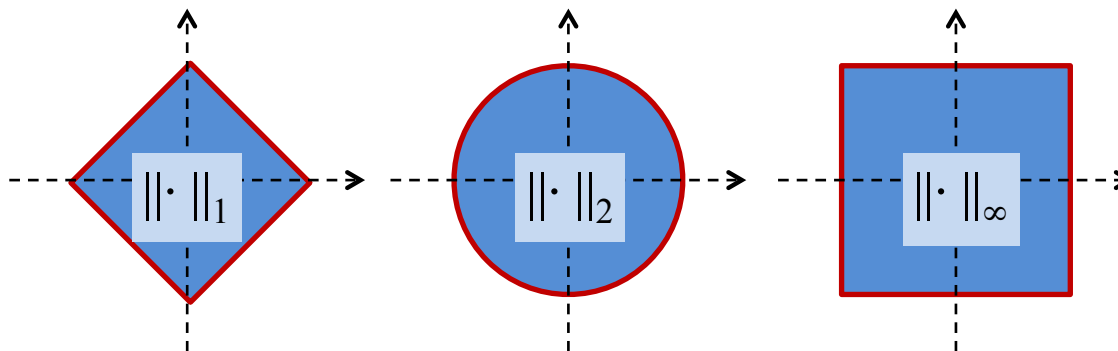
$4u_{ij} - (u_{i+1,j} + u_{i,j+1} + u_{i-1,j} + u_{i,j-1}) = 0$

**Poissonsgleichung
(diskret)
pos. def. Matrix**

Rechte Seite ergibt sich aus den Randbedingungen

Normen auf Vektorräumen

- Beispiele:
 - Summennorm: $\| \mathbf{x} \|_1 := \sum_i |x_i|$
 - Euklidische Norm: $\| \mathbf{x} \|_2 := (\sum_i |x_i|^2)^{(1/2)}$
 - Maximumnorm: $\| \mathbf{x} \|_\infty := \max_i |x_i|$
- Eigenschaften von Normen: formale Definition
 - definit**, d.h. $\mathbf{x} \neq 0 \Rightarrow \| \mathbf{x} \| > 0$
 - homogen**, d.h. $\| \lambda \mathbf{x} \| = | \lambda | \| \mathbf{x} \| \quad \forall \lambda \text{ komplex bzw. reell}$
 - sub-additiv**, d.h. $\| \mathbf{x} + \mathbf{y} \| \leq \| \mathbf{x} \| + \| \mathbf{y} \|$ (*Dreiecksungleichung*)
- Einheitskugeln/~sphären: $\{ \mathbf{x} : \| \mathbf{x} \| \leq 1 \}$ bzw. $\{ \mathbf{x} : \| \mathbf{x} \| = 1 \}$



Äquivalenz von Normen

- Zwei Normen $\|\cdot\|_p$ und $\|\cdot\|_q$ in einem Vektorraum V heißen äquivalent, wenn es Konstanten c_1 und c_2 gibt, so dass

$$c_1\|\mathbf{x}\|_p \leq \|\mathbf{x}\|_q \leq c_2\|\mathbf{x}\|_p \quad \text{für alle } \mathbf{x} \in V$$
- Im endlich dimensional sind alle Normen äquivalent (d.h. es lassen sich immer solche Konstanten finden)
- Leider gilt dies nicht für Normen in unendlich dimensional Vektorräumen (Funktionenräumen) die wir ja für kontinuierliche Probleme approximieren müssen.
- Beispiel 1- und ∞ -Norm (endlich- und unendlich-dim.)

Matrix-Norm bzw. Operator-Norm

- Die zu einer Vektornorm $\|\cdot\|$ **assoziierte Matrix-Norm** (= **Operator-Norm**) einer Matrix A wird definiert durch

$$\|A\| = \max \left\{ \frac{\|Ax\|}{\|x\|} : x \neq 0 \right\} = \max \{ \|Ax\| : \|x\| = 1 \}$$

- jede Vektornorm definiert eine Matrixnorm
- eine assoz. Matrix-Norm ist eine Norm (und somit per Definition definit, homogen und sub-additiv) aber **mehr als das**:
 - ★ $\|Id\| = 1$
 - ★ eine Matrixnorm ist **sub-multiplikativ** : $\|AB\| \leq \|A\| \cdot \|B\|$
 - ★ mit der Vektornorm **kompatibel** $\|Ax\| \leq \|A\| \cdot \|x\|$ für alle x
 - ★ $\|A\| \geq |\lambda|$ für jeden Eigenwert λ von A

- Matrix-Norm bzw. Operatornorm: Beispiele

- ▶ die zur Summennorm $\| \mathbf{x} \|_1 := \sum_i |x_i|$ assoziierte Norm:

$$\| \| \mathbf{A} \| \|_1 = \max_j \left\{ \sum_i |a_{ij}| \right\} \quad (\text{größte } \mathbf{Spalten-Betrags-Summe})$$

- ▶ die zur Maximumnorm $\| \mathbf{x} \|_\infty := \max_i |x_i|$ assoziierte Norm:

$$\| \| \mathbf{A} \| \|_\infty = \max_i \left\{ \sum_j |a_{ij}| \right\} \quad (\text{größte } \mathbf{Zeilen-Betrags-Summe})$$

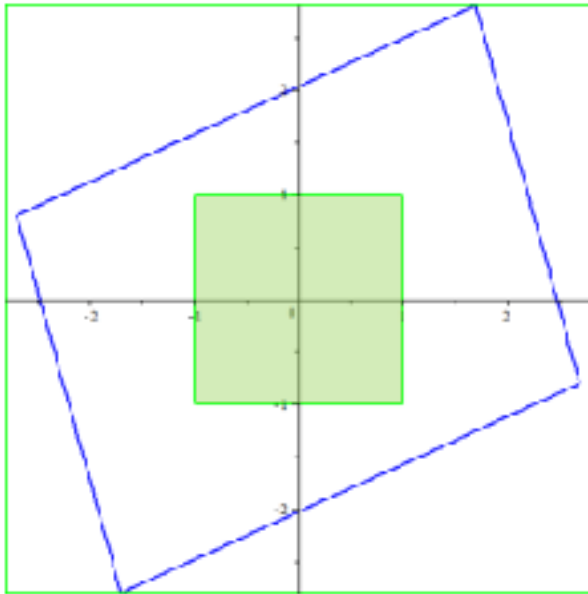
- ▶ die zur Euklidischen Norm $\| \mathbf{x} \|_2 := (\sum_i |x_i|^2)^{(1/2)}$ assoz. Norm:

$$\| \| \mathbf{A} \| \|_2 = \sqrt{\lambda_1} \quad \text{wobei} \quad \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_n \geq 0$$

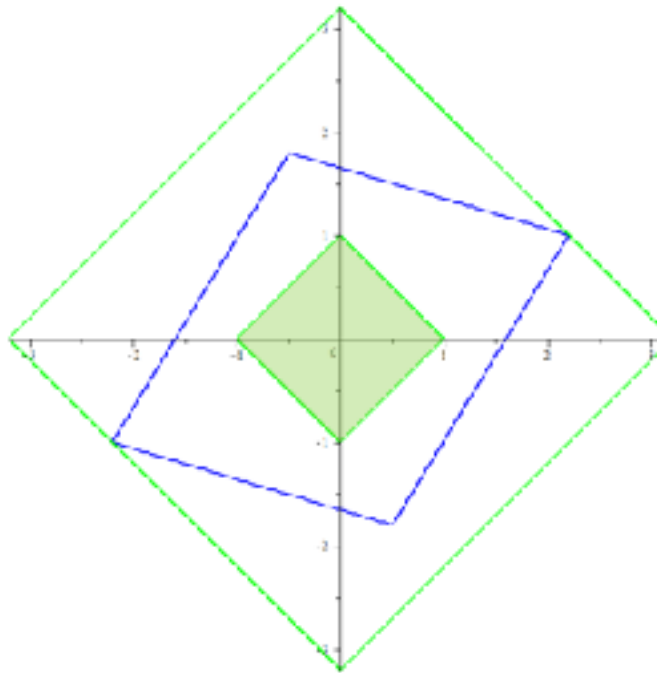
die absteigend angeordneten Eigenwerte der symmetrischen, positiv semi-definiten Matrix $\mathbf{A}^T \mathbf{A}$ sind (siehe unten).

- Geometrische Interpretation: Matrixnorm/Operatornorm

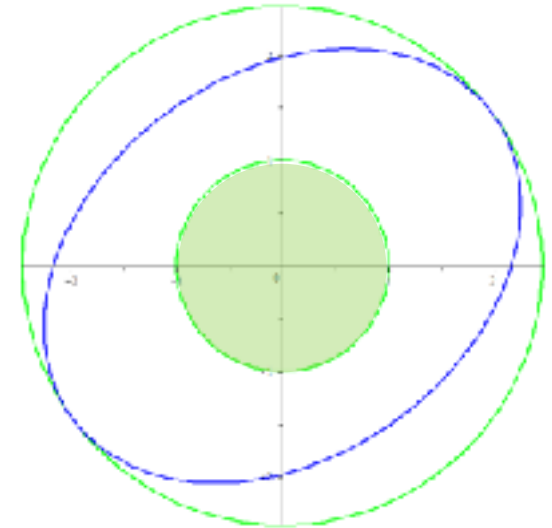
$$A_1 = \begin{bmatrix} -0.5 & 2.2 \\ 1.8 & 1.0 \end{bmatrix}$$



$$||| A |||_{\infty} =$$



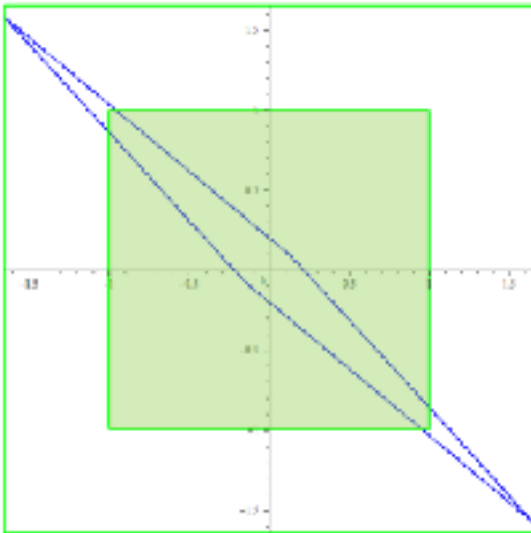
$$||| A |||_1 =$$



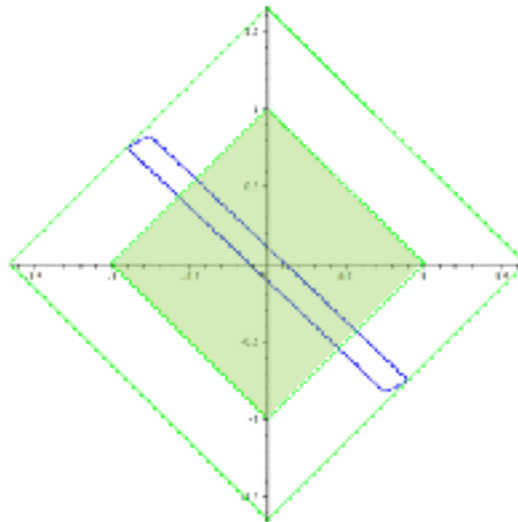
$$||| A |||_2 =$$

- Geometrische Interpretation: Matrixnorm/Operatornorm

$$A_2 = \begin{bmatrix} -0.750 & 0.900 \\ 0.825 & -0.75 \end{bmatrix}$$



$$\|A\|_{\infty} =$$



$$\|A\|_1 =$$



$$\|A\|_2 =$$

Konditionszahl einer Matrix (Wiederholung!)

- Die Kondition einer (nicht singulären) Matrix A bzgl. einer (Vektor-)Norm ist definiert durch

$$\kappa(A) = \frac{\max \{\|Ax\| : \|x\| = 1\}}{\min \{\|Ax\| : \|x\| = 1\}}$$

- Geometrische Interpretation: die Verzerrung der Einheitskugel ($\|x\| = 1$) unter der Abbildung $x \mapsto Ax$
- für nichtsinguläre quadratische Matrizen folgt

$$\kappa(A) = \|A\| \cdot \|A^{-1}\| \quad \text{wobei } \|\cdot\| \text{ die assoz. Matrixnorm}$$

ist

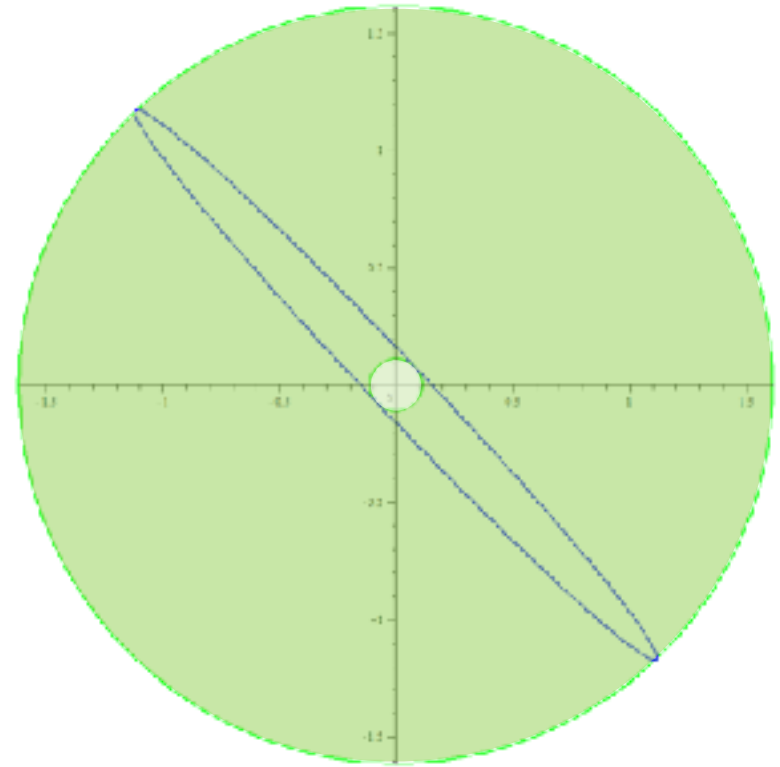
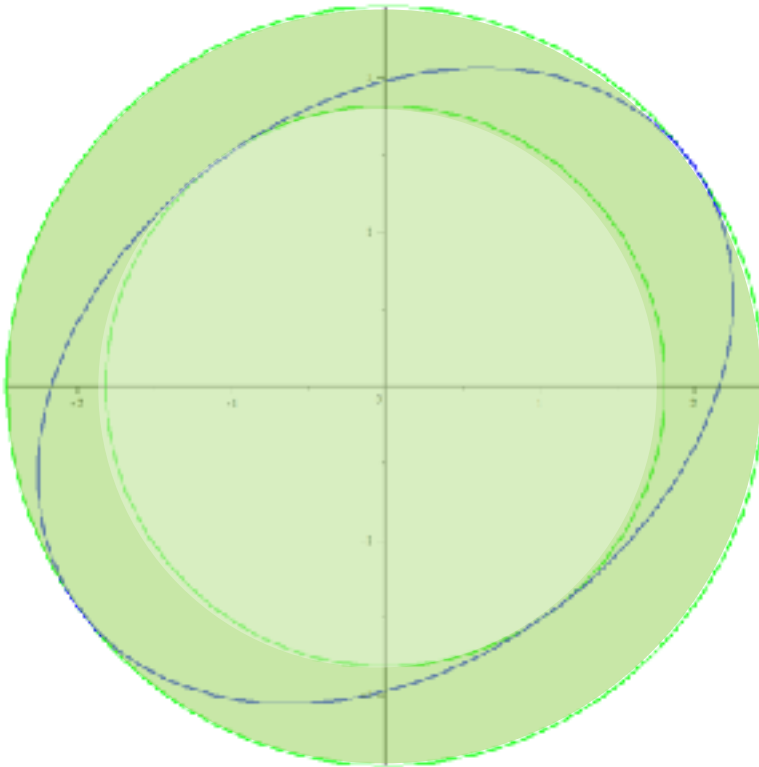
- stets gilt $\kappa(A) \geq \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$ (größter bzw. kleinster EW von A)

- Konditionszahl: Geometrische Interpretation

$$A_1 = \begin{bmatrix} -0.5 & 2.2 \\ 1.8 & 1.0 \end{bmatrix}$$

$$\kappa_2(A) = \frac{\max \{\|Ax\|_2 : \|x\|_2 = 1\}}{\min \{\|Ax\|_2 : \|x\|_2 = 1\}}$$

$$A_2 = \begin{bmatrix} -0.750 & 0.900 \\ 0.825 & -0.75 \end{bmatrix}$$



Kondition linearer Gleichungssysteme und Konditionszahl

- Problemstellung: Gegeben eine invertierbare $n \times n$ -Matrix A

Bestimme zu „rechter Seite“ b die Lösung x des linearen Gleichungssystems $Ax = b$

- **Kondition des Problems** (Erinnerung):
Welchen Effekt haben Fehler in den Eingangsdaten b auf des Ergebnis x ? (bei exakter Rechnung!)
Statt $Ax = b$ löst man $A\tilde{x} = \tilde{b}$,

Wie verhalten sich die **relativen Fehler**? $\frac{\|\tilde{b} - b\|}{\|b\|} \rightarrow \frac{\|\tilde{x} - x\|}{\|x\|}$

- Antwort: $\frac{\|\tilde{x} - x\|}{\|x\|} \leq \kappa(A) \cdot \frac{\|\tilde{b} - b\|}{\|b\|}$

Beispiel: $\begin{bmatrix} 0 & 100 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ \varepsilon \end{bmatrix}$

Kondition linearer Gleichungssysteme und Konditionszahl

- Auch wenn man beim Lösen der Linearen Gleichung $Ax = b$ auch Störungen in der Koeffizienten-Matrix A berücksichtigt, d.h. $\tilde{A} \tilde{x} = \tilde{b}$ statt $Ax = b$ löst, bestimmt die Konditionszahl von A die Kondition dieses Problems !

- Man kann zeigen, dass in diesem Fall

$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \cdot \frac{\|\tilde{A} - A\|}{\|A\|}} \cdot \left(\frac{\|\tilde{b} - b\|}{\|b\|} + \frac{\|\tilde{A} - A\|}{\|A\|} \right)$$

- Folgerung

$$\frac{\|\tilde{x} - x\|}{\|x\|} \leq \frac{2\varepsilon \cdot \kappa(A)}{1 - \varepsilon \cdot \kappa(A)}, \text{ sofern } \frac{\|\tilde{b} - b\|}{\|b\|} \leq \varepsilon \text{ und } \frac{\|\tilde{A} - A\|}{\|A\|} \leq \varepsilon$$

Matrixnorm, Konditionszahl, EW der speziellen Matrizen

- U orthogonale Matrix:
 - ▶ Beispiel: Drehungen, (Householder-)Spiegelungen
 - ▶ $\|U\|_2 = 1$, $\kappa_2(U) = 1$,
 - ▶ $|\lambda_i| = 1$ für jeden Eigenwert λ_i von U

- A symmetrische Matrix:
 - ▶ alle Eigenwerte λ_i von A sind reell,
 - ▶ $\|A\|_2 = \max_i \{|\lambda_i|\}$
 - ▶ $\kappa_2(A) = \frac{\max_i \{|\lambda_i|\}}{\min_i \{|\lambda_i|\}}$

- Generell gilt $\kappa(AB) \leq \kappa(A) \cdot \kappa(B)$ i.A. “ $<$ ” !!!

Konditionszahl: Beispiel

$$M = \begin{bmatrix} 1 & 2 & 3 & 4 \\ -99 & -200 & 50 & 40 \\ 900 & 50 & -50 & -80 \\ 100 & -10 & 90 & 150 \end{bmatrix}$$

- LR-Zerlegung ohne Pivot:

$$M = L_1 R_1 :$$

	M	L_1	R_1
κ_1	2083.5	$9.642 \cdot 10^8$	$5.785 \cdot 10^5$
κ_2	1176.1	$1.108 \cdot 10^8$	$7.188 \cdot 10^5$
κ_∞	1236.4	$1.570 \cdot 10^8$	$1.384 \cdot 10^5$

- LR-Zerlegung mit Pivot :

$$M = L_2 R_2 :$$

	M	L_2	R_2
κ_1	2083.5	1.508	1704.7
κ_2	1176.1	1.183	1161.5
κ_∞	1236.4	1.429	1196.3

- QR-Zerlegung:

$$M = Q R_3 :$$

	M	Q	R_3
κ_1	2083.5	1.505	1726.3
κ_2	1176.1	1.0	1176.1
κ_∞	1236.4	1.505	1211.2

Eigenwerte und Diagonalisierbarkeit

- Gibt es eine Basis aus Eigenvektoren, d.h. $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ für $i = 1, 2, \dots, m$,
dann ist die Matrixdarstellung der Matrix A bzgl. der

Basis $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ die Diagonalmatrix $D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$

‣ d.h.

$$A \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_m \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \lambda_1 \mathbf{v}_1 & \cdots & \lambda_m \mathbf{v}_m \\ | & & | \end{bmatrix} = \underbrace{\begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_m \\ | & & | \end{bmatrix}}_V \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lambda_m \end{bmatrix}$$

‣ oder $A = V \cdot D \cdot V^{-1}$

Spektralsatz für symmetrische Matrizen

Theorem: Es sei $A \in \mathbb{R}^{m \times m}$ eine reelle **symmetrische** Matrix.

Dann gibt es eine **Orthonormalbasis aus Eigenvektoren**

- Gleichwertig dazu: es gibt eine **Faktorisierung** $A = VDV^T$

$$\text{oder } A = \sum_{i=1}^m \lambda_i \mathbf{v}_i \mathbf{v}_i^T \quad \text{oder} \quad A\mathbf{x} = \sum_{i=1}^m \lambda_i (\mathbf{v}_i \circ \mathbf{x}) \mathbf{v}_i \quad (*)$$

Dabei ist/sind

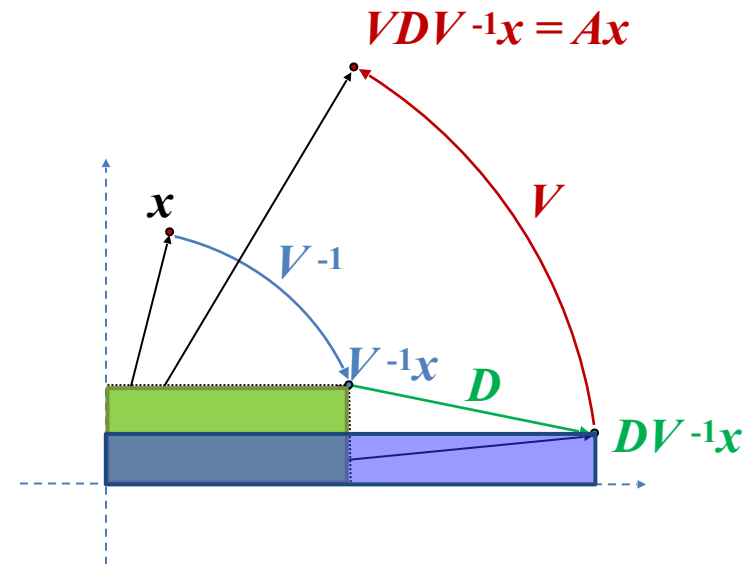
- ▶ D die Diagonalmatrix aller EW λ_i von A (inklusive Vielfachheit)
- ▶ die Spalten von V sind die normierten EV \mathbf{v}_i von A
- ▶ V ist orthogonal: (d.h. $V^{-1} = V^T$)
- ▶ im Fall einfacher EW, ist diese Zerlegung eindeutig bis auf die Reihenfolge der EWs und ± 1 -EVs.

(*) Hierbei ist " \circ " das Skalarprodukt von Vektoren: $\mathbf{v} \circ \mathbf{x} = \mathbf{v}^T \mathbf{x}$

- Veranschaulichung im Fall $n=m=2$

$$A = \frac{1}{5} \begin{bmatrix} 12 & -1 \\ 16 & \frac{3}{4} \end{bmatrix} = \underbrace{\frac{1}{5} \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}}_V \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}}_D \underbrace{\frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}}_{V^T}$$

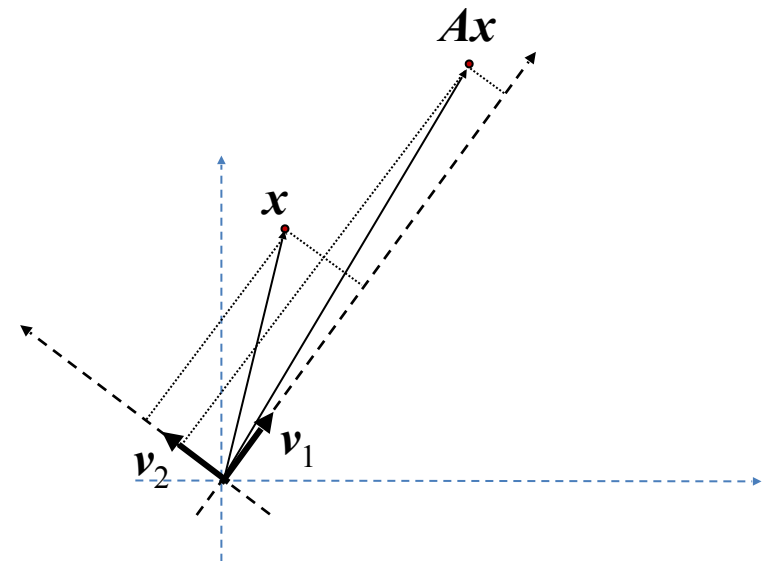
- V Drehung um $\text{atan}(4/3) \approx +53.1^\circ$
- D nicht isotrope Skalierung
- V^T Drehung um -53.13°



- Veranschaulichung im Fall $n=m=2$

$$A = \frac{1}{5} \begin{bmatrix} 12 & -1 \\ 16 & \frac{3}{4} \end{bmatrix} = \underbrace{\frac{1}{5} \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}}_V \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}}_D \underbrace{\frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}}_{V^T}$$

- V Drehung um $+53.13^\circ$
- D nicht isotrope Skalierung
- V^T Drehung um -53.13°



• Symmetrische Matrizen: Beispiel

► Bsp von oben : $\mathbf{S} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \lambda_1 = 3, \lambda_2 = 1$

► mit EVs $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ und $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ ergibt $\mathbf{V} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$

mit Inverser

$$\mathbf{V}^{-1} = \mathbf{V}^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

► somit $\mathbf{S} = \mathbf{V}\mathbf{D}\mathbf{V}^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$

- Folgerung des Theorems (A symmetrische Matrix)

► es gilt:
$$A = \lambda_1 \cdot \mathbf{v}_1 \cdot \mathbf{v}_1^T + \lambda_2 \cdot \mathbf{v}_2 \cdot \mathbf{v}_2^T + \cdots + \lambda_m \cdot \mathbf{v}_m \cdot \mathbf{v}_m^T$$

oder
$$A\mathbf{x} = \lambda_1 (\mathbf{v}_1 \circ \mathbf{x}) \mathbf{v}_1 + \lambda_2 (\mathbf{v}_2 \circ \mathbf{x}) \mathbf{v}_2 + \cdots + \lambda_m (\mathbf{v}_m \circ \mathbf{x}) \mathbf{v}_m$$

dabei ist $\{\mathbf{v}_i\}$ die ONB aus EVs der Matrix A und λ_i die EWs

► Anwendung

$$A^2 \mathbf{x} = \lambda_1^2 (\mathbf{v}_1 \circ \mathbf{x}) \mathbf{v}_1 + \lambda_2^2 (\mathbf{v}_2 \circ \mathbf{x}) \mathbf{v}_2 + \cdots + \lambda_m^2 (\mathbf{v}_m \circ \mathbf{x}) \mathbf{v}_m$$

$$f(A) = f(\lambda_1) \cdot \mathbf{v}_1 \cdot \mathbf{v}_1^T + f(\lambda_2) \cdot \mathbf{v}_2 \cdot \mathbf{v}_2^T + \cdots + f(\lambda_m) \cdot \mathbf{v}_m \cdot \mathbf{v}_m^T$$

$$\sqrt{A} \mathbf{x} = \sqrt{\lambda_1} (\mathbf{v}_1 \circ \mathbf{x}) \mathbf{v}_1 + \sqrt{\lambda_2} (\mathbf{v}_2 \circ \mathbf{x}) \mathbf{v}_2 + \cdots + \sqrt{\lambda_m} (\mathbf{v}_m \circ \mathbf{x}) \mathbf{v}_m$$

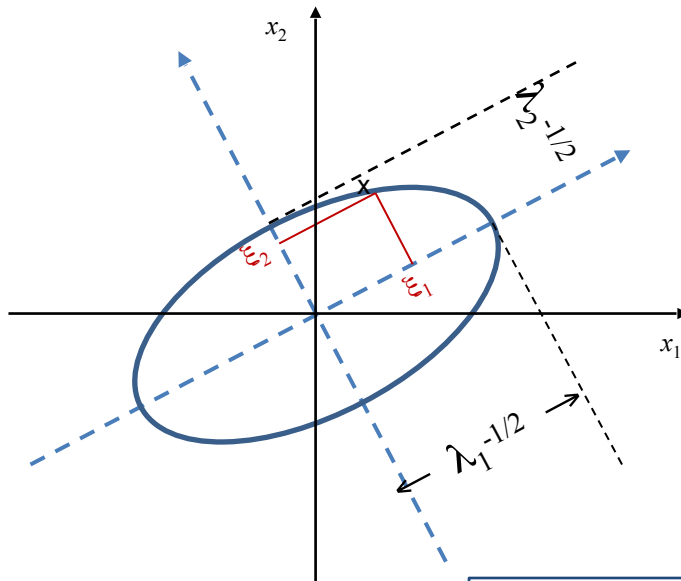
Geom. Interpretation: Hauptachsentransformation

- Gegeben **Quadratische Form** $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ (\mathbf{A} symmetrisch)
- Die gemäß Theorem existierende ONB $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ aus EV bilden die Hauptachsen der **Quadrik** $\{\mathbf{x} : Q(\mathbf{x}) = 1\}$
- Bezüglich dieser Achsen (in diesem Koordinatensystem) hat die Quadrik die Form $\left\{ \mathbf{x} = \sum_i \xi_i \mathbf{v}_i : \lambda_1 \xi_1^2 + \lambda_2 \xi_2^2 + \dots + \lambda_m \xi_m^2 = 1 \right\}$
- Dies ist ein verallgemeinertes Ellipsoid bzw. Hyperboloid
 - ▶ Hauptachsenrichtungen: \mathbf{v}_i ,
 - ▶ Hauptachsenabschnitten $|\lambda_i|^{-1/2}$

Geom. Interpretation: Hauptachsentransformation

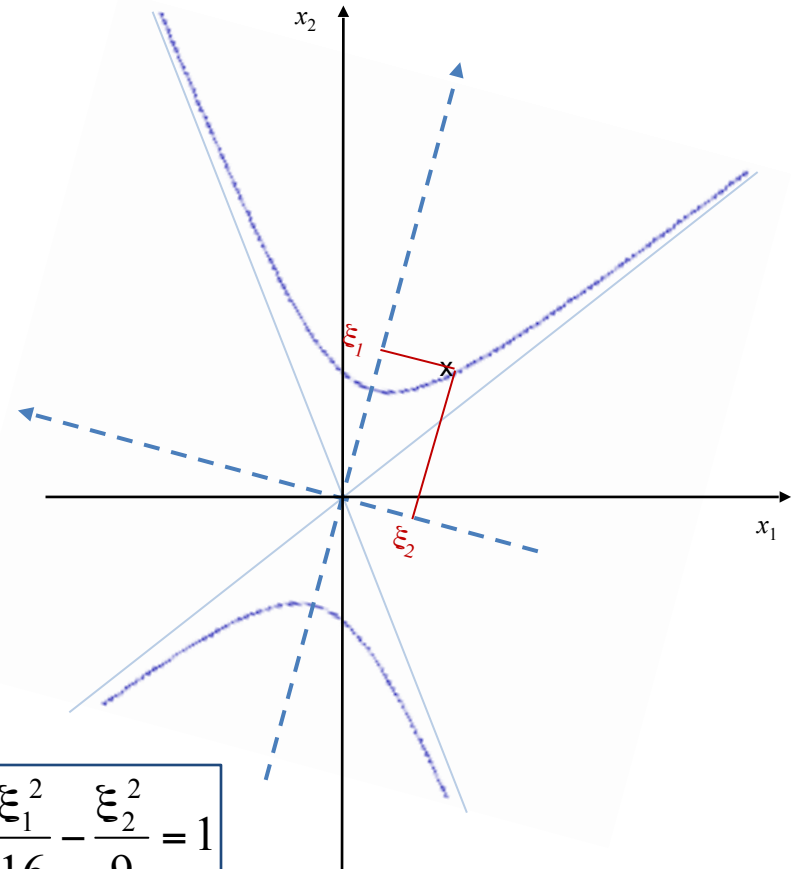
- nur positive EW \rightarrow Ellipse
- pos. & neg. EW \rightarrow Hyperbel

$$\lambda_1 > 0, \lambda_2 > 0$$



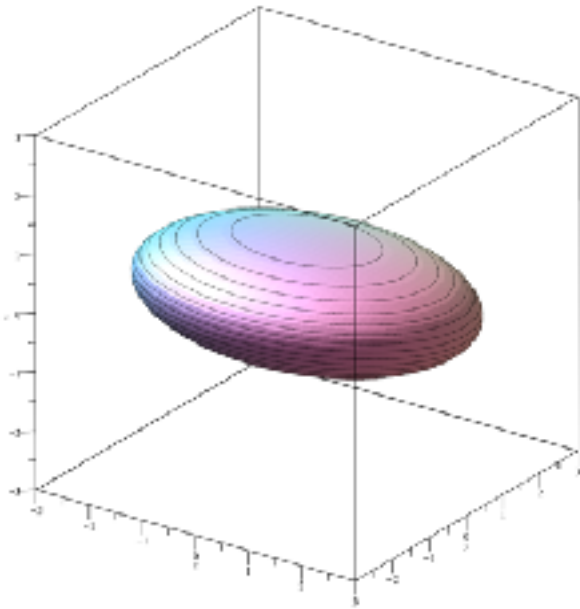
$$\frac{\xi_1^2}{25} + \frac{\xi_2^2}{4} = 1$$

$$\lambda_1 > 0, \lambda_2 < 0$$



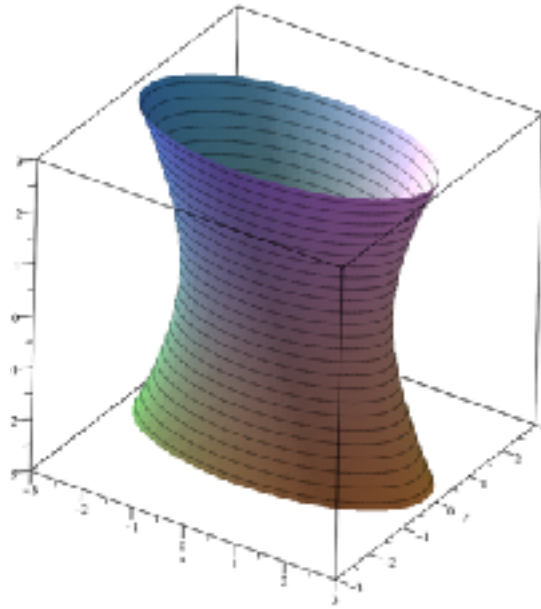
$$\frac{\xi_1^2}{16} - \frac{\xi_2^2}{9} = 1$$

- Quadriken im \mathbb{R}^3



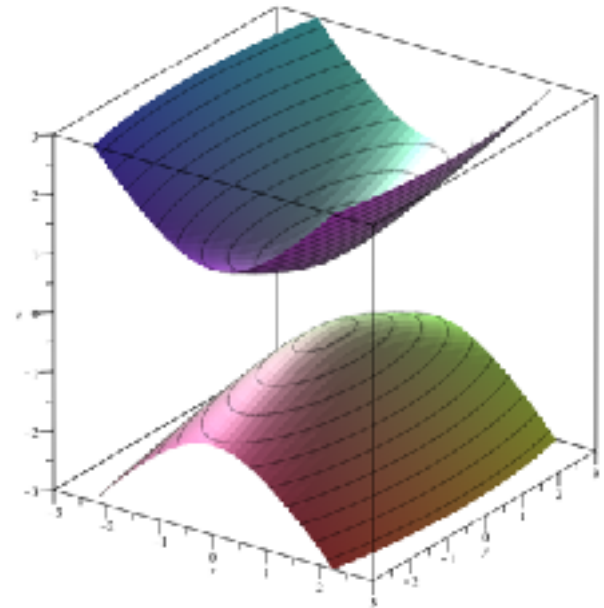
Ellipsoid

$$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0$$



Hyperboloid (1-schalig)

$$\lambda_1 > 0, \lambda_2 > 0, \lambda_3 < 0$$



Hyperboloid (2-schalig)

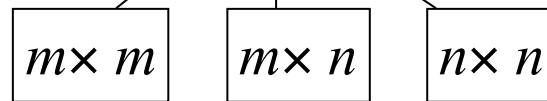
$$\lambda_1 < 0, \lambda_2 < 0, \lambda_3 > 0$$

Theorem: (Singulärwertzerlegung)

Es sei $A \in \mathbb{R}^{m \times n}$ eine beliebige Matrix:

Es gibt eine Faktorisierung

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



Dabei ist/sind

- Σ eine “Diagonalmatrix”, $\Sigma_{ij} = 0$ falls $i \neq j$ und $\Sigma_{11} \geq \Sigma_{22} \geq \dots \geq 0$
- U und V sind orthogonal
- die Spalten von U bzw. V sind EV von AA^T bzw. $A^T A$
- die Diagonalelemente $\sigma_i = \Sigma_{ii} \neq 0$ sind die **Singulärwerte** (singular values) von A , es gilt: $\sigma_i = \sqrt{\lambda_i}$ wobei λ_i EW von $A^T A$ bzw. $A A^T$

SVD (= singular value decomposition)

- Betrachte $A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$ (hier ist $m=3$, $n=2$).

- Die SVD ergibt sich zu

$$\begin{bmatrix} 2/\sqrt{6} & 0 & 1/\sqrt{3} \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

- Die Singulärwerte sind absteigend angeordnet!

- $A = U \Sigma V^T \Rightarrow A^T = (U \Sigma V^T)^T = V \Sigma^T U^T$

- Illustration der SVD (Struktur, Dimension)

$$\underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_A = \underbrace{\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}}_U \underbrace{\begin{bmatrix} \bullet & & & & \\ & \bullet & & & \\ & & \bullet & & \\ & & & \bullet & \\ & & & & \bullet \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{V^T}$$

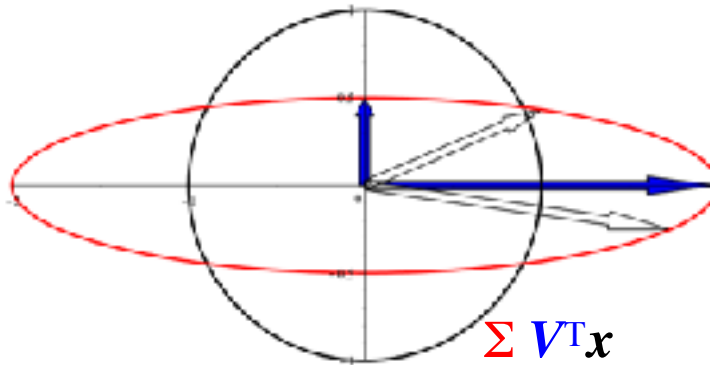
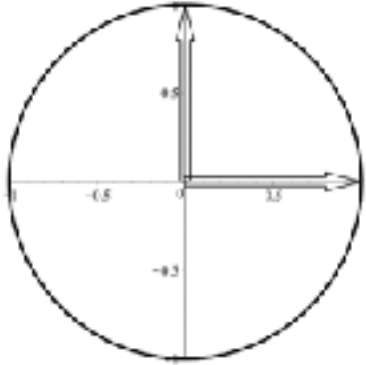
The diagram illustrates the SVD decomposition of a square matrix A (5x5) into three matrices: U (5x5), Σ (5x5), and V^T (5x3). The matrix U is shown with a yellow shaded 5x3 submatrix in its last three columns. The matrix Σ is shown with a yellow shaded 3x3 submatrix in its bottom-right corner. The matrix V^T is shown with a yellow shaded 3x3 submatrix in its bottom-right corner.

$$\underbrace{\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}}_A = \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_U \underbrace{\begin{bmatrix} \bullet & & & & \\ & \bullet & & & \\ & & \bullet & & \\ & & & \bullet & \\ & & & & \bullet \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}}_{V^T}$$

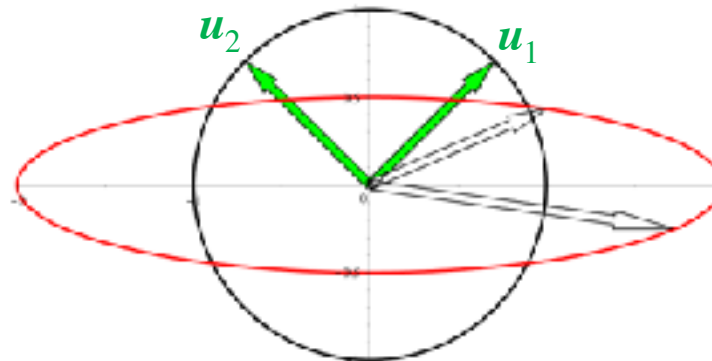
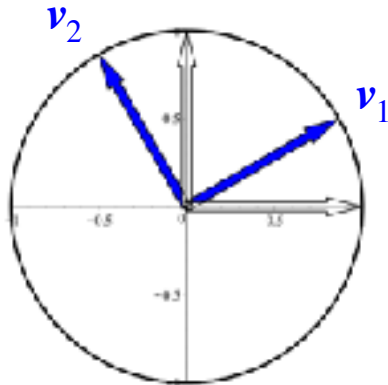
The diagram illustrates the SVD decomposition of a rectangular matrix A (3x5) into three matrices: U (3x3), Σ (3x5), and V^T (5x5). The matrix U is shown with a yellow shaded 3x3 submatrix in its bottom-right corner. The matrix Σ is shown with a yellow shaded 3x3 submatrix in its bottom-right corner. The matrix V^T is shown with a yellow shaded 3x3 submatrix in its bottom-right corner.

- A quadratisch: alles quadratisch

Veranschaulichung im Fall $n=m=2$



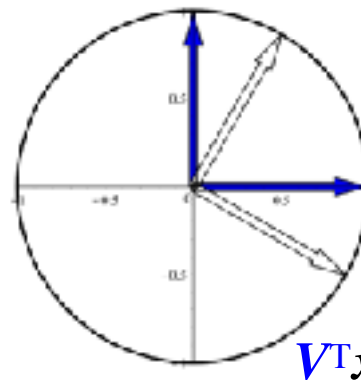
$$\Sigma V^T x$$



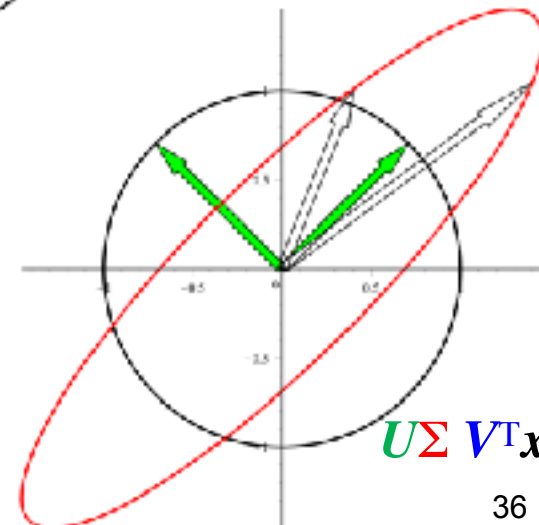
Beispiel

$$A = \begin{bmatrix} 1.402 & 0.401 \\ 1.048 & 1.013 \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}}_U \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}}_{V^T}$$



$$V^T x$$

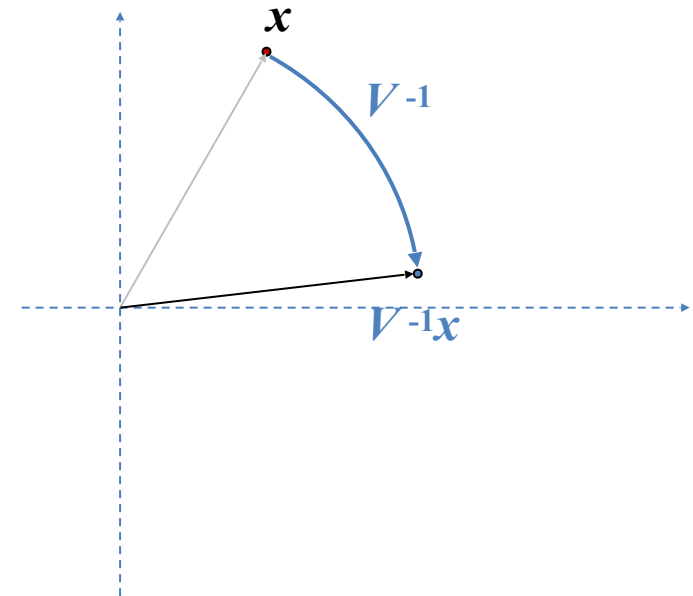


$$U \Sigma V^T x$$

- Veranschaulichung im Fall $n=m=2$

$$\underbrace{\frac{\sqrt{2}}{20} \begin{bmatrix} 1 & 18 \\ -17 & -6 \end{bmatrix}}_A = \underbrace{\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 1 \end{bmatrix}}_\Sigma \underbrace{\frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}}_{V^T}$$

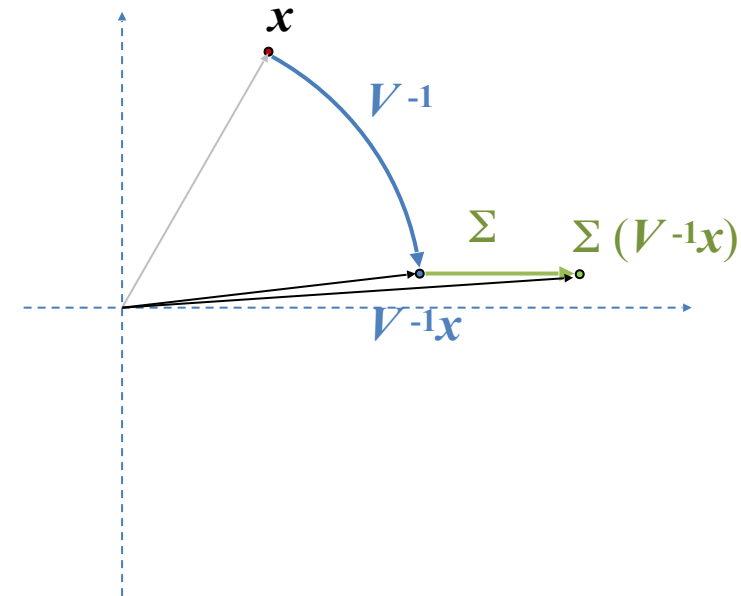
- V Drehung um $+53.13^\circ$
 V^T Drehung um -53.13°
- Σ nicht isotrope Skalierung
- U Drehung um -45°



- Veranschaulichung im Fall $n=m=2$

$$\underbrace{\frac{\sqrt{2}}{20} \begin{bmatrix} 1 & 18 \\ -17 & -6 \end{bmatrix}}_A = \underbrace{\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 1 \end{bmatrix}}_\Sigma \underbrace{\frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}}_{V^T}$$

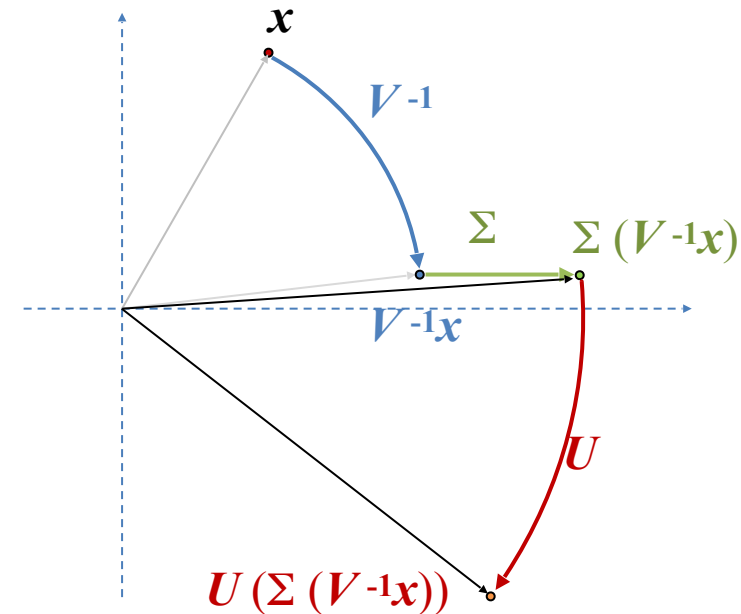
- V Drehung um $+53.13^\circ$
 V^T Drehung um -53.13°
- Σ nicht isotrope Skalierung
- U Drehung um -45°



- Veranschaulichung im Fall $n=m=2$

$$\underbrace{\frac{\sqrt{2}}{20} \begin{bmatrix} 1 & 18 \\ -17 & -6 \end{bmatrix}}_A = \underbrace{\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 1 \end{bmatrix}}_{\Sigma} \underbrace{\frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}}_{V^T}$$

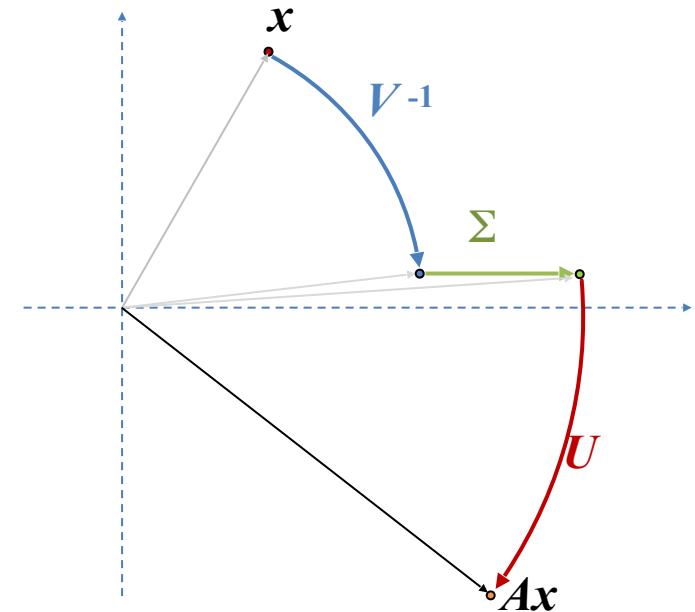
- V Drehung um $+53.13^\circ$
 V^T Drehung um -53.13°
- Σ nicht isotrope Skalierung
- U Drehung um -45°



- Veranschaulichung im Fall $n=m=2$

$$\underbrace{\frac{\sqrt{2}}{20} \begin{bmatrix} 1 & 18 \\ -17 & -6 \end{bmatrix}}_A = \underbrace{\frac{\sqrt{2}}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 1.5 & 0 \\ 0 & 1 \end{bmatrix}}_\Sigma \underbrace{\frac{1}{5} \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}}_{V^T}$$

- V Drehung um $+53.13^\circ$
 V^T Drehung um -53.13°
- Σ nicht isotrope Skalierung
- U Drehung um -45°



- Struktur, Dimension

$$\underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_A = \underbrace{\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}}_U \underbrace{\begin{bmatrix} \sigma_1 & & & & \\ & \sigma_r & & & \\ & & 0 & & \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \end{bmatrix}}_{V^T}$$

The diagram illustrates the SVD decomposition of a matrix A into three matrices: U , Σ , and V^T . Matrix A is a 5x3 matrix of asterisks. Matrix U is a 5x5 matrix of asterisks with a red shaded 3x3 submatrix in the bottom-right corner crossed out with a large 'X'. Matrix Σ is a 5x5 diagonal matrix with σ_1 and σ_r on the diagonal, and zeros elsewhere. Matrix V^T is a 3x3 matrix of asterisks with a red shaded 1x3 submatrix in the bottom row.

- Beachte: Einfluss haben nur
 - die $\sigma_i \neq 0$ (das sind genau die Singulärwerte)
 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$.
 - die Spalten $\{u_1, \dots, u_r\}$ von U
 - sowie die Spalten $\{v_1, \dots, v_r\}$ von V (= Zeilen von V^T)

- Folgerung des Theorems (A beliebige Matrix mit SVD)

$$A = \underbrace{\begin{bmatrix} | & & | & & | \\ | & & | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_r & \mathbf{u}_{r+1} & \cdots & \mathbf{u}_m \\ | & & | & & | \\ | & & | & & | \\ | & & | & & | \end{bmatrix}}_U \underbrace{\begin{bmatrix} \sigma_1 & \cdots & 0 & 0 \\ \vdots & \ddots & & \\ & & \sigma_r & \\ & & & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & 0 \end{bmatrix}}_\Sigma \underbrace{\begin{bmatrix} - & \mathbf{v}_1^T & - \\ & \vdots & \\ - & \mathbf{v}_r^T & - \\ & \vdots & \\ - & \mathbf{v}_n^T & - \end{bmatrix}}_{V^T}$$

- ▶ Rang: $\text{rank}(A) = r \ (\leq \min\{n, m\})$;
- ▶ Kern (Nullraum): $\ker(A) = \langle \mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_n \rangle$;
- ▶ Bild: $\text{im}(A) = \langle \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r \rangle$;
- ▶ Norm: $\|A\|_2 = \sigma_1 = \sigma_{\max}$, $\kappa_2(A) = \frac{\sigma_1}{\sigma_r} = \frac{\sigma_{\max}}{\sigma_{\min}}$ (falls $r = n$)

• Beispiel

$$\begin{aligned}
 A &= \frac{1}{25} \begin{bmatrix} 272 & 136 & -256 & 128 \\ 380 & 190 & 10 & -5 \\ 204 & 102 & -192 & 96 \end{bmatrix} = \\
 &= \frac{1}{25} \underbrace{\begin{bmatrix} 16 & -12 & 15 \\ 15 & 20 & 0 \\ 12 & -9 & -20 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 25 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_\Sigma \frac{1}{5} \underbrace{\begin{bmatrix} 4 & 2 & -2 & 1 \\ 2 & 1 & 4 & -2 \\ -2 & 4 & 1 & 2 \\ 1 & -2 & 2 & 4 \end{bmatrix}}_{V^T}
 \end{aligned}$$

Kern:

Bild

Norm

Rang

$$\ker(A) = \left\langle \begin{bmatrix} -2 \\ 4 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 2 \\ 4 \end{bmatrix} \right\rangle$$

$$\operatorname{im}(A) = \left\langle \begin{bmatrix} 16 \\ 15 \\ 12 \end{bmatrix}, \begin{bmatrix} -12 \\ 20 \\ -9 \end{bmatrix} \right\rangle$$

$$\|A\|_2 = 25$$

$$\operatorname{rank}(A) = 2$$

- Folgerung des Theorems A beliebige Matrix

► es gilt: $A = \sigma_1 \cdot \mathbf{u}_1 \cdot \mathbf{v}_1^T + \sigma_2 \cdot \mathbf{u}_2 \cdot \mathbf{v}_2^T + \cdots + \sigma_r \cdot \mathbf{u}_r \cdot \mathbf{v}_r^T$

oder $A\mathbf{x} = \sigma_1(\mathbf{v}_1 \circ \mathbf{x})\mathbf{u}_1 + \sigma_2(\mathbf{v}_2 \circ \mathbf{x})\mathbf{u}_2 + \cdots + \sigma_r(\mathbf{v}_r \circ \mathbf{x})\mathbf{u}_r$

dabei ist $\{\mathbf{v}_i\}$ die ONB aus EVs der Matrix $A^T A$,

$\{\mathbf{u}_i\}$ die ONB aus EVs der Matrix $A A^T$.

- Beachte, dass $A\mathbf{v}_i = \sigma_i \cdot \mathbf{u}_i$ und $A^T \mathbf{u}_i = \sigma_i \cdot \mathbf{v}_i$
- Die Singulärwerte σ_i sind die Quadratwurzeln aus den EWs von $A^T A$ oder $A A^T$
- SVD von A^T : alles transponieren

- **SVD folgt aus der Diagonalisierbarkeit symm. Matrizen:**
- $A^T A$ ist symmetrische, positiv semi-definite $n \times n$ -Matrix
- Nach dem Spektralsatz gibt es eine ONB $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ von $A^T A$ mit zugehörigen EWs λ_i , ggf. nach umordnen:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_n = 0$$
- ONB bedeutet $\mathbf{Id} = \sum_{i=1}^n \mathbf{v}_i \cdot \mathbf{v}_i^T$ folglich $\mathbf{x} = \sum_{i=1}^n (\mathbf{v}_i^T \mathbf{x}) \mathbf{v}_i$
und somit $\mathbf{Ax} = \sum_{i=1}^n (\mathbf{v}_i^T \mathbf{x}) \mathbf{Av}_i = \sum_{i=1}^r (\mathbf{v}_i^T \mathbf{x}) \mathbf{Av}_i$
- Setze $\sigma_i = \sqrt{\lambda_i}$, $\mathbf{u}_i = \frac{1}{\sigma_i} \mathbf{Av}_i$ für $1 \leq i \leq r$ dann gilt
 - ▶ $\mathbf{u}_i^T \mathbf{u}_j = \frac{1}{\sigma_i \sigma_j} (\mathbf{Av}_i)^T \mathbf{Av}_j = \frac{1}{\sigma_i \sigma_j} \mathbf{v}_i^T \mathbf{A}^T \mathbf{Av}_j = \frac{\lambda_j}{\sigma_i \sigma_j} \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}$
 - ▶ sowie $\mathbf{Ax} = \sum_{i=1}^r \sigma_i (\mathbf{v}_i^T \mathbf{x}) \mathbf{u}_i$

- Mit der SVD kann man eine Matrix durch Matrizen von niedrigem Rang approximieren (**low-rank approximation**)

- Problemstellung: Finde A_k vom Rang k so dass

$$A_k = \min_{X: \text{rank}(X)=k} \|A - X\| \longleftarrow \text{Frobenius-Norm (s.u.)}$$

- A_k und X sind ebenfalls $m \times n$ Matrizen!
- Typischerweise will man $k \ll r$.

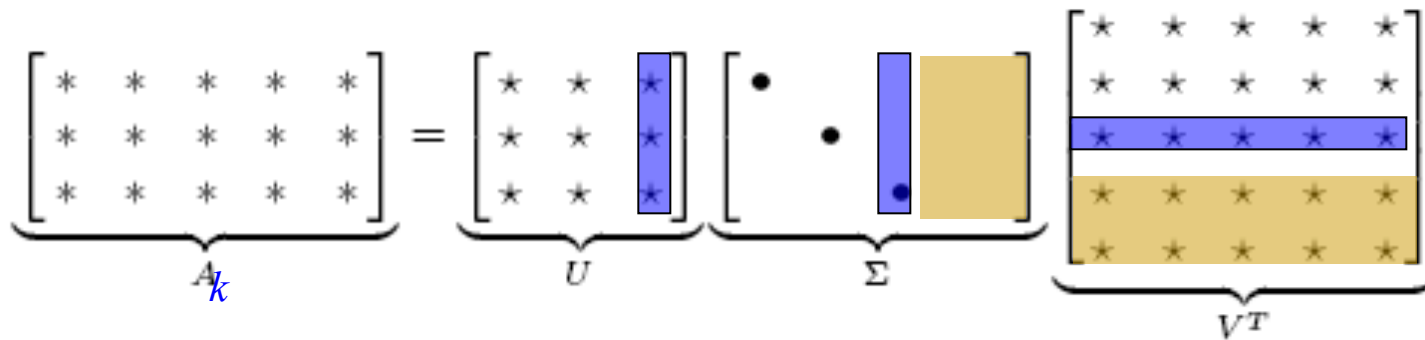
- Welche Norm? Meist Frobenius-Norm: $\|A\|_F = \left(\sum_i \sum_j a_{ij}^2 \right)^{1/2}$

Es gilt: $\|A\|_F = \left(\sum_{i=1}^r \sigma_i^2 \right)^{1/2}$ (σ_i die Singulärwerte von A !)

- Lösung mit Hilfe der SVD

$$A_k = U \operatorname{diag}(\sigma_1, \dots, \sigma_k, \underbrace{0, \dots, 0}) V^T$$

*setze die $r-k$ kleinsten
Singulärwerte Null*



$$A_k \mathbf{x} = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad \leftarrow \begin{array}{l} \text{Spaltenvektoren:} \\ \text{Summe von Rang 1 Matrizen} \end{array}$$

Was bringt die *low-rank-approximation* ?

$$A_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

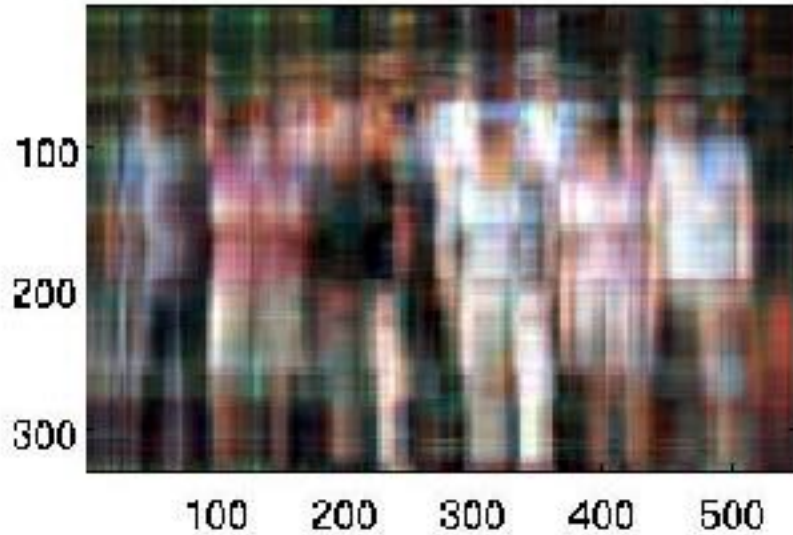
- Speichereffizienz: Nur $k \cdot (n+m)$ Werte (statt $n \cdot m$)
- Schnelle Berechnung von Matrix-Vektor-Multiplikation:

$$A_k \mathbf{x} = \sum_{i=1}^k \sigma_i (\mathbf{v}_i \circ \mathbf{x}) \mathbf{u}_i$$

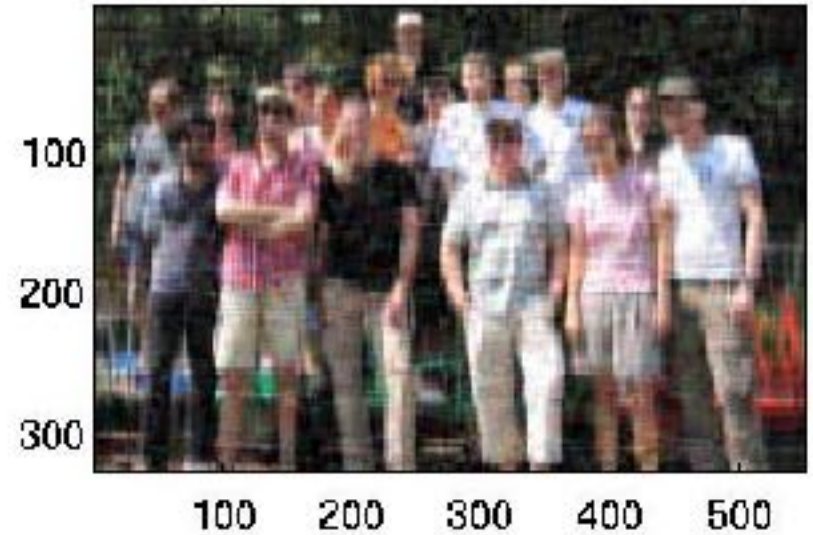
$k \cdot (2m+n)$ arithmetische Operationen (statt $2n \cdot m$)

- Berechnung von $\mathbf{y}^T A_k \mathbf{x} = \sum_{i=1}^k \sigma_i (\mathbf{u}_i \circ \mathbf{y}) \cdot (\mathbf{v}_i \circ \mathbf{x})$:
ca. $2k \cdot (n+m)$ arithm. Operationen (statt $2 \cdot n \cdot m$)

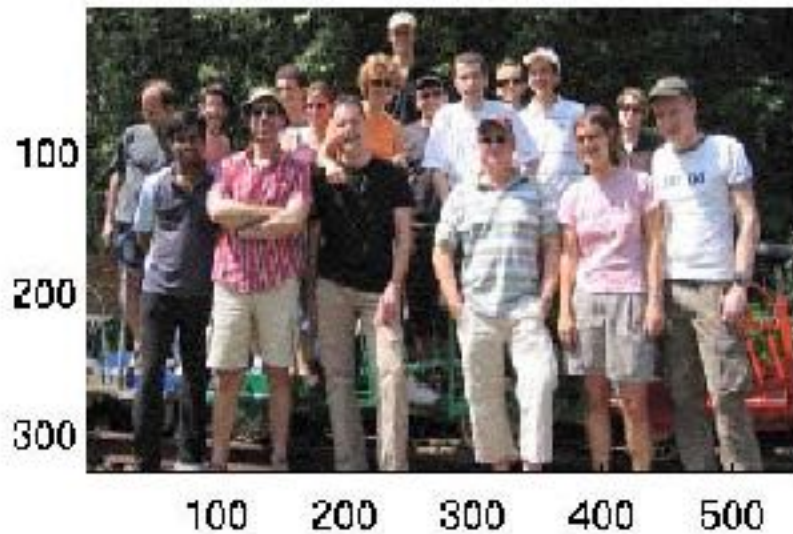
5 Singulaerwerte



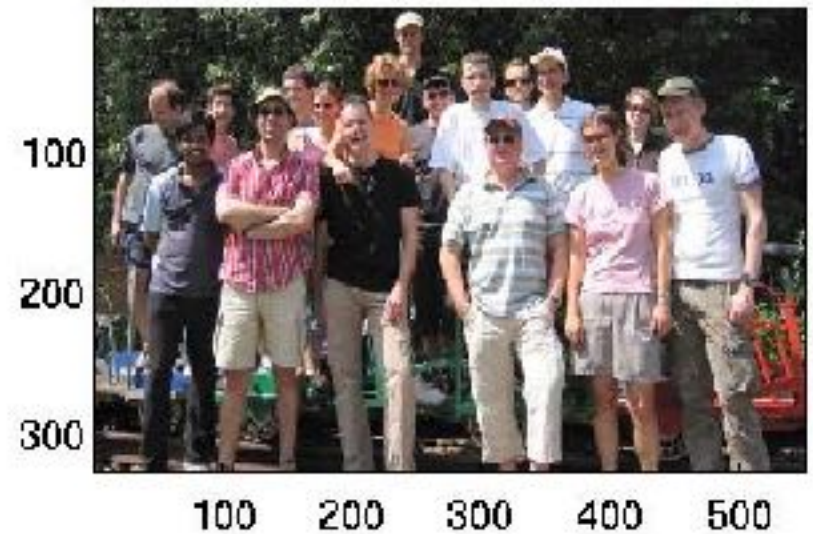
20 Singulaerwerte



80 Singulaerwerte



320 Singulaerwerte



Quelle: Tobias Damm: TU Kaiserslautern

- **Approximationsfehler (wie gut ist diese Approximation?)**

- ▶ Es ist die (eindeutige) beste Approximierende bezüglich der Frobenius-Norm $\| \cdot \|_F$

- ▶ Es gilt:
$$\min_{X: \text{rank}(X)=k} \|A - X\|_F^2 = \|A - A_k\|_F^2 = \sum_{j>k} \sigma_j^2$$

- ▶ sowie
$$\min_{X: \text{rank}(X)=k} \|A - X\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

sofern die Singulärwerte der Größe nach absteigend geordnet sind
 $\sigma_i \geq \sigma_{i+1}$

- ▶ Aufwand bei Standardfall $\text{rank}(A) = \min\{m, n\}$:
 - ★ Speicherbedarf $k(m+n)$ im Gegensatz zu mn
 - ★ #(Multiplikationen) bei Matrix-Vektor-Multiplikation: $k(m+n)$ im Gegensatz zu mn

- $||| A |||_{\infty}$ (Matrixnorm bzgl. Maximums-Norm)
- $||| A |||_2$ (Matrixnorm bzgl. Euklidischer Norm)
- $||| A |||_1$ (Matrixnorm bzgl. Summen-Norm)

- $\| A \|_F$ (Frobeniusnorm $\| A \|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$)

(letztere ist keine „echte“ Matrixnorm/Operatornorm !)

- Es gilt:
 - $||| A |||_2 = \sigma_1 (= \max \{ \sigma_i \})$
 - $\| A \|_F = \left(\sum_i \sigma_i^2 \right)^{1/2}$

- Lösungstheorie für $A\mathbf{x} = \mathbf{b}$, wobei
 - ▶ $n = m$ und $\det(A) \neq 0$: eindeutige Lösung
 - ▶ $n = m$ und $\det(A) = 0$ oder $n \neq m$
 - ★ nur lösbar falls $\mathbf{b} \in \text{im}(A)$
 - ★ alle Lösungen: $\mathbf{x}_0 + \ker(A)$ wobei \mathbf{x}_0 eine spezielle Lösung
 - ▶ $\dim(\ker(A)) + \dim(\text{im}(A)) = \#(\text{Spalten})$
 $\text{rank}(A) = \dim(\text{im}(A))$;
- SVD von A : $A = U \Sigma V^T$
 - ▶ $\text{im}(A) = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\} = \{\mathbf{u}_{r+1}, \dots, \mathbf{u}_m\}^\perp$
 - ▶ $\ker(A) = \text{span}\{\mathbf{v}_{r+1}, \dots, \mathbf{v}_n\} = \{\mathbf{v}_1, \dots, \mathbf{v}_r\}^\perp$

- **Pseudo-Inverse**

$$A^{\sim 1} = V \Sigma^{\sim 1} U^T$$

wobei

$$\Sigma^{\sim 1} = \begin{bmatrix} 1/\sigma_1 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 1/\sigma_r & & 0 \\ \vdots & & \vdots & 0 & 0 \\ 0 & \dots & 0 & 0 & \ddots \end{bmatrix}$$

- **Wichtig:** $A^{\sim 1}$ und $\Sigma^{\sim 1}$ sind $n \times m$ -Matrizen (wie Σ^T bzw. A^T !)
- $$\mathbf{x} = A^{\sim 1} \mathbf{b} = \sum_{i=1}^r \frac{1}{\sigma_i} (\mathbf{u}_i \circ \mathbf{b}) \mathbf{v}_i$$
- Beachte, dass $A^{\sim 1} = V \Sigma^{\sim 1} U^T$ nicht ganz genau die SVD von $A^{\sim 1}$ ist (falsche Reihenfolge der Singulärwerte!)

- 1. Fall: A hat **maximalen Rang**,
d.h. $\text{rank}(A) = \min\{n, m\}$:

- ▶ überbestimmtes System ($n < m$):

$$\mathbf{x} = A^{\sim 1} \mathbf{b} \text{ löst } \|\mathbf{Ax} - \mathbf{b}\|_2 = \min$$

Dies zeigt, dass $\mathbf{x} = A^{\sim 1} \mathbf{b}$ die Lösung des linearen Ausgleichsproblems ist

- ▶ unterbestimmtes System ($n > m$):

$$\mathbf{x} = A^{\sim 1} \mathbf{b} \text{ löst } \mathbf{Ax} = \mathbf{b} \text{ und erfüllt } \|\mathbf{x}\|_2 = \min$$

- 2. Fall: $\text{rank}(A) < \min\{n, m\}$:
Die Lösung $\mathbf{x} = A^{\sim 1} \mathbf{b}$

- ▶ minimiert das Residuum $\|\mathbf{Ax} - \mathbf{b}\|_2 = \min$ **und**

- ▶ ist unter allen diesen Lösungen diejenige mit kleinster Norm $\|\mathbf{x}\|_2 = \min$

- Theorie: $r = \text{rank}(A) = \dim(\text{im}(A))$ (Theorie)

Praxis: **numerischer Rang** : $\#(\{ i : \sigma_i > \varepsilon \})$

- Zu gegebenen (hoffentlich) lin. unabhängigen Vektoren $\{b_1, b_2, \dots, b_n\}$ soll Orthonormalbasis bestimmt werden:
 - ▶ Gram-Schmidt: numerisch sehr problematisch
 - ▶ modifizierter Gram-Schmidt (besser, sehr gut aber nur mit weiterer „Nachorthogonalisierung“)
 - ▶ QR-Zerlegung von $A = [b_1, b_2, \dots, b_n]$
 - ▶ SVD dieser Matrix $A = U \Sigma V^T$
 - ★ ONB : $u_1, u_2, \dots, u_r // [b_1, b_2, \dots, b_n]$ lin. unabh.
sofern $r = n$

Wikipedia:

- **Latente Semantische Analyse (LSA)** bzw. ***Latent Semantic Indexing*** (kurz **LSI**), ist ein (nicht mehr patentgeschütztes) Verfahren des Information Retrieval, das 1990 von Deerwester et al.[1] beschrieben wurde. Verfahren wie das *LSI* sind insbesondere für die Suche auf großen Datenmengen wie dem Internet von Interesse (google, yahoo,). Das Ziel von *LSI* ist es, „Hauptkomponenten“ von Dokumenten zu finden.
- http://de.wikipedia.org/wiki/Latente_Semantische_Analyse

[1] S Deerwester, S Dumais, G Furnas, T Landauer, R Harshman: [Indexing by Latent Semantic Analysis](#). In: *Journal of the American society for information science*. 1990

- Basiert auf dem sog. *vector space model* ein wichtiges Modell im Bereich des *information retrieval*
- Zentrales Objekt ist dabei die *term-document-matrix* :

← documents →

	C	d_1	d_2	d_3	d_4	d_5	d_6
↑ terms ↓	ship	1	0	1	0	0	0
	boat	0	1	0	0	0	0
	ocean	1	1	0	0	0	0
	wood	1	0	0	1	1	0
	tree	0	0	0	1	0	1

- Größen in der Praxis:
 - ▶ #(documents): bis zu mehreren Mio
 - ▶ #(terms): 10.000 – 200.000

- Der eigentliche *retrieval process* basiert (u.a.) auf Vektor-Matrix-Multiplikationen und inneren Produkten
- Die TD-Matrix wird durch das LSI zusätzlich bearbeitet, um sie zu verkleinern. Dies ist gerade für größere Dokumentensammlungen sinnvoll, da hier die TD-Matrizen im Allgemeinen sehr groß werden.
- Dazu wird die TD-Matrix über die [Singularwertzerlegung](#) zerlegt. Danach werden „unwichtige“ Teile der TD-Matrix abgeschnitten. Diese Verkleinerung hilft [Komplexität](#) und damit Rechenzeit beim **Retrievalprozess** (Vergleich der Dokumente oder Anfragen) zu sparen.
- Am Ende des Algorithmus steht eine neue, kleinere TD-Matrix, in der die Terme der originalen TD-Matrix zu Konzepten generalisiert sind.

- Der eigentliche *retrieval process* basiert (u.a.) auf Vektor-Matrix-Multiplikationen und inneren Produkten
- Beispiel: 10 Mio Dokumente, 100.000 terms:
- A ist eine $100.000 \times 10.000.000$ -Matrix
- Exakt (vollbesetzt):
 - $y^T A x$: ca. 10^{12} Multiplikationen
- Approximation durch Rang $k = 100$:
 - $y^T A_k x$: ca. 10^9 Multiplikationen

mehr Info:

- Diplomarbeit von *Régis Newo Kenmogne*
Univ. Saarbrücken 2005

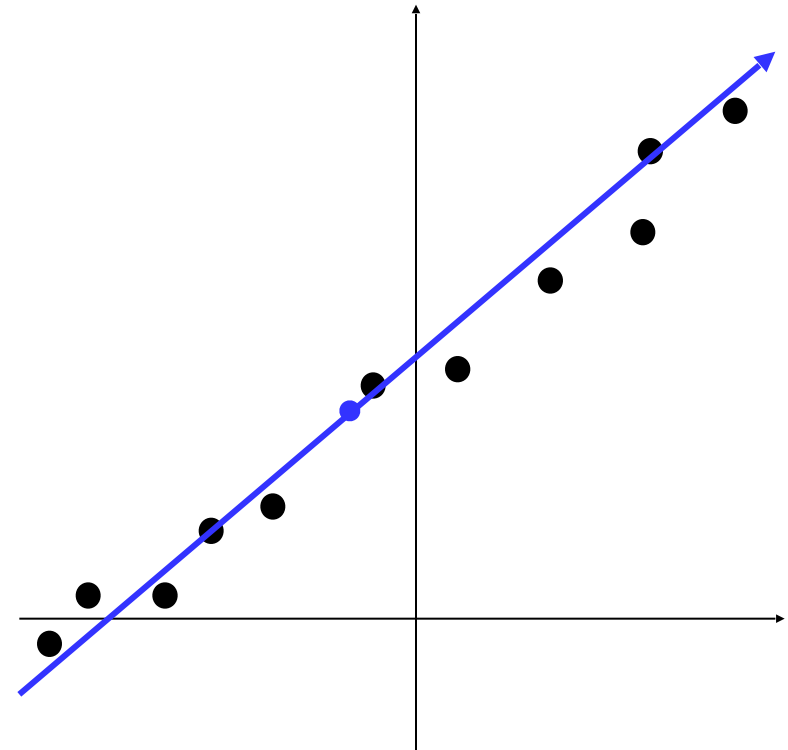
http://www.uni-hildesheim.de/media/fb4/informatik/AG_IIS/Dokumente-Allgemein/Mitglieder/newo/dipl_newo.pdf

- Stanford: *CS276 Information Retrieval and Web Search*

<http://web.stanford.edu/class/cs276/handouts/lecture13-lsi.ppt>

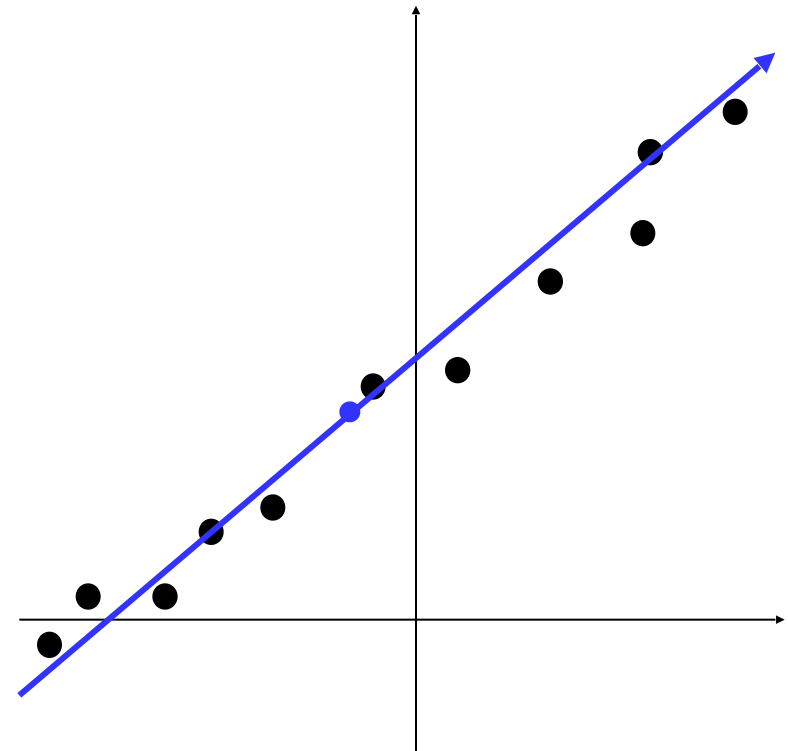
Motivation – wozu/warum

- Analyse (vieler) n -dimensionaler Daten (n groß)
z.B. Merkmalsvektoren (Bitmuster)
[Karhunen-Loeve-Transformation (KLT)]
- Hoch-dimensionales Problem
auf niedrig-dimensionales
reduzieren unter
Beibehaltung der
wesentlichen Merkmale
(Reduziere/projiziere
auf wenige Dimensionen)



Motivation – wozu/warum

- Bestimmung der Basis eines Unterraums, in dem die Varianz ziemlich groß ist, die sog. **principal components** oder
- **Hauptachsen**
- Nützlich für Datenkompression



Gegeben : Daten $\{X_i\}_{1 \leq i \leq N} \in \mathbb{R}^n$

Schritte zur Bestimmung der PCA

- Balanciere die Daten um Mittelwert

$$\bar{X} = \frac{1}{N} \sum_i X_i, \quad \bar{X}_i = X_i - \bar{X}$$

- Bilde Kovarianz-Matrix (symm. und pos. semi-def.)

$$C = \text{cov}(X, X) = \frac{1}{N-1} \sum_i \bar{X}_i \cdot \bar{X}_i^T$$

Gegeben : Daten $\{\mathbf{X}_i\}_{1 \leq i \leq N} \in \mathbb{R}^n$

- Diagonalisierung der Kovarianzmatrix :

$$\mathbf{C} = \mathbf{Q} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & \lambda_n \end{bmatrix} \mathbf{Q}^T$$

wobei \mathbf{Q} orthogonal ist und $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$

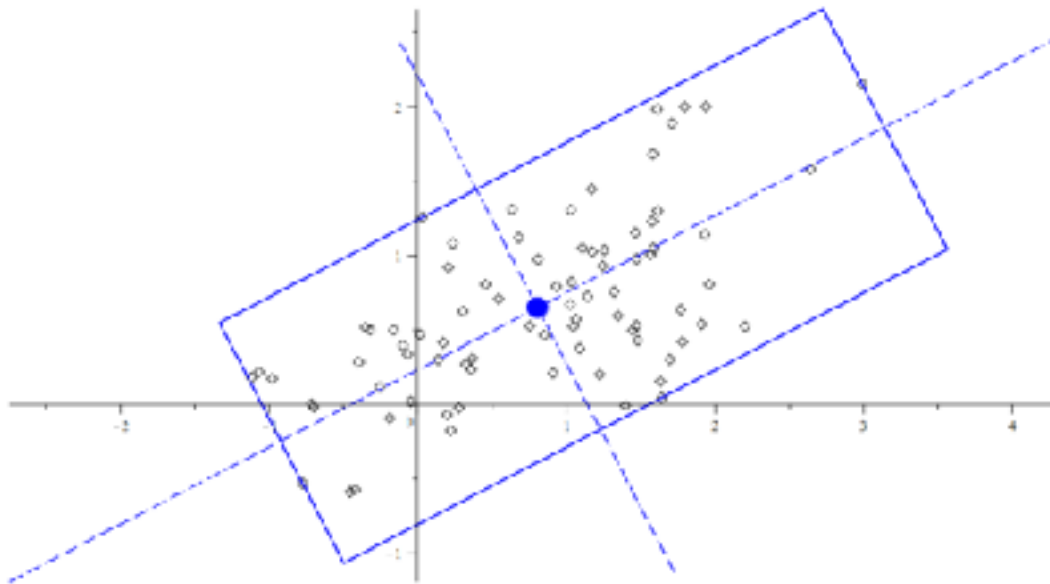
- Die Spalten \mathbf{q}_j von \mathbf{Q} sind die Eigenvektoren
diese bezeichnet man als **principal components**
oder als **Hauptachsen**

Beispiel : Daten $\{P_i\}_{1 \leq i \leq 80} \in \mathbb{R}^2$

- Balancieren,
- Kovarianzmatrix,
- Diagonalisieren

$$C = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

dabei sind $\{u_1, u_2\}$ orthonormal und $\lambda_1 \geq \lambda_2$.

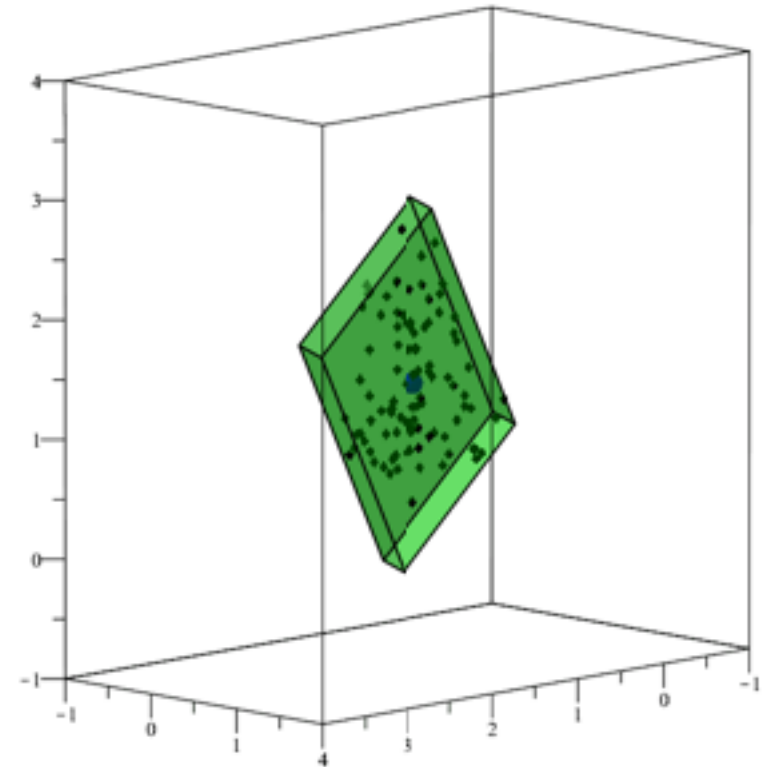
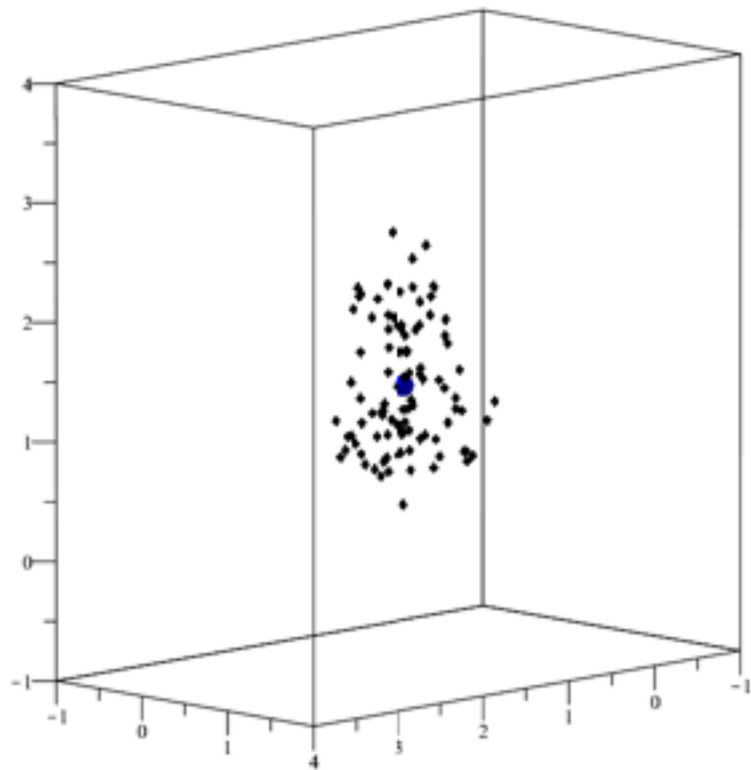


OBB = object oriented bounding box

Beispiel : Daten $\{P_i\}_{1 \leq i \leq 100} \in \mathbb{R}^3$

- Balancieren,
- Kovarianzmatrix,
- Diagonalisieren

$$C = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$



Reduktion auf wenige Hauptachsen:

$$C = Q \Lambda Q^T \text{ oder } C = \sum_{j=1}^n \lambda_j \mathbf{q}_j \cdot \mathbf{q}_j^T$$

- Berücksichtige nur die k größten EWs !

$$C \approx \sum_{j=1}^k \lambda_j \mathbf{q}_j \cdot \mathbf{q}_j^T$$

- Die zugehörigen EVs \mathbf{q}_j (Spalten in Q) sind die Basisvektoren des niedrig dimensionalen Unterraums (mit Nullpunkt \bar{X})
- Die Koeffizienten des Vektors X_i bzgl. dieser Basis ergeben sich als Skalarprodukt $\bar{X}_i \circ \mathbf{q}_j \quad (j = 1, \dots, k)$.

$$X_i \approx \bar{X} + \sum_{j=1}^k \lambda_j (\bar{X}_i \circ \mathbf{q}_j) \mathbf{q}_j$$

PCA und SVD

- Betrachte die Matrix der balancierten Daten:
$$A = \begin{bmatrix} | & | & & | \\ \bar{X}_1 & \bar{X}_2 & \dots & \bar{X}_N \\ | & | & & | \end{bmatrix}$$
- dann ist die Kovarianz-Matrix $C = \frac{1}{N-1} A A^T$
- SVD von A : $A = U \Sigma V^T$
- $C = \frac{1}{N-1} A A^T = \frac{1}{N-1} U \Sigma V^T V \Sigma^T U^T = \frac{1}{N-1} U \Sigma \Sigma^T U^T$
 - EW: $\lambda_i = \frac{1}{(N-1)} \sigma_i^2$
 - die Hauptachsen u_i

- Weiteres Beispiel: **morphable model**
(ein statistisches Geometrie Modell)

[V. Blanz, T. Vetter:](#)

[A Morphable Model For The Synthesis of 3D Faces.](#)
[SIGGRAPH'99](#)

[video \(youtube\)](#)

- Demo
 - Grundlage ca 300 Datensätze (3D Scans von Gesichtern)
 - Geometriemodell mit 53.500 3D-Punkten (106.500 Dreiecken)

→ PCA für 300 „Punkte“ in einem 162.000-dimensionalen Raum

- Vorweg: Exakte Bestimmung der SVD ist nicht möglich!
- Im Wesentlichen sind EWs und EVs von einer (pos. semi-definiten) Matrix zu bestimmen,
z.B. von $A^T A$ oder von $A A^T$ oder von
$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$
- (EW sind Nullstellen von Polynomen, letztere lassen sich bekanntlich für $n > 5$ nicht „algebraisch lösen“ !
- Schwieriges Problem – nur iterativ näherungsweise lösbar !
- Lösung in „zwei“ Schritten
 1. (Ähnlichkeits-)Transformation auf Tridiagonalgestalt
 2. EWs und EVs einer Tridiagonalmatrix

1. Schritt: Ähnlichkeits-Transformation auf Tridiagonalgestalt für eine pos. semi-definite Matrix B

- **Ähnlichkeitstransformation:**
 $B_1 = S^{-1} B S$ und B haben dieselben EWs,
 und S transformiert die EVs.
- Ansatz: S als Produkt von $n-2$ Householder-Spiegelungen $Q_i = Id - 2 \mathbf{n}_i \cdot \mathbf{n}_i^T$
 (also eine orthogonale Transformation)

1. Schritt: Ähnlichkeits-Transformation auf Tridiagonalgestalt für eine pos. semi-definite Matrix B (cont'd)

- Bestimmung der Householder-Sp.: $Q_i = Id - 2 n_i \cdot n_i^T = Q_i^T$

$\triangleright Q_1 : \text{Ziel}$

$$\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \\ \vdots \\ b_{n1} \end{bmatrix} \mapsto \begin{bmatrix} b_{11} \\ * \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
 erfolgt durch $n_1 = \gamma [0, b_{21} - *, b_{31}, \dots, b_{n1}]^T$

\triangleright Dann gilt: $Q_1 B Q_1 =$

$$\begin{bmatrix} * & * & 0 & \dots & 0 \\ * & * & * & \dots & * \\ 0 & * & * & & * \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & * & * & \dots & * \end{bmatrix}$$

$\triangleright Q_2, \dots, Q_{n-2}$ analog

2. Schritt: EWs und EVs einer tridiagonalen, pos. semi-definiten Matrix B

- **QR-Algorithmus zur EW Bestimmung**

- ein iteratives Verfahren:

- ▶ $B_0 := B ;$

- ▶ QR-Zerlegung $B_0 = Q_0 R_0 \rightarrow B_1 := R_0 Q_0 ;$

(dann gilt: $Q_0 B_1 Q_0^T = B_0$, d.h. B_1 und sind B_0 ähnlich!)

- ▶

- ▶ QR-Zerlegung $B_i = Q_i R_i \rightarrow B_{i+1} := R_i Q_i ;$

2. Schritt: EWs und EVs einer tridiagonalen, pos. semi-definiten Matrix B (cont'd)

- Eigenschaften der oben konstruierten Folge $\{ B_i \}$:

- ▶ $Q_i B_{i+1} Q_i^T = B_i$

- ▶ B_i sind symmetrisch, tridiagonal und positiv semi-definit

- ▶ $\lim_{i \rightarrow \infty} B_i = \begin{bmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \lambda_n \end{bmatrix} \quad (\lambda_j \text{ die EWs von } B)$

- ▶ Abbruch sofern alle Nicht-Diagonalelemente $< \varepsilon$

- Aufwand für die Bestimmung der QR-Zerlegung einer tridiagonalen Matrix $TB = QR$

- Q ist das Produkt von nur $n-1$ Jacobi-Rotationen

- $Q_1 = \begin{bmatrix} \cos & -\sin & 0 & \dots & 0 \\ \sin & \cos & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}, Q_{n-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & & 0 & \cos & -\sin \\ 0 & \dots & 0 & \sin & \cos \end{bmatrix}$

- $R = \begin{bmatrix} * & * & 0 & \dots & 0 \\ 0 & * & * & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & * & * \\ 0 & 0 & \dots & 0 & * \end{bmatrix}$

Aufwand nur: $O(n)$

- Lineare Algebra (Erinnerung)
 - Vektor- und Matrix-Normen, Konditionszahl
 - EW, EV, Diagonalisierbarkeit symmetrischer Matrizen (Spektralsatz)
- SVD – was ist das?
- Was kann man anhand der SVD bestimmen
 - Rang, $\ker(A)$, $\text{im}(A)$, $\|A\|_2$, Konditionszahl, Rang- k -Appr.
- Anwendungen:
 - Datenkompression
 - Lösen allgemeiner Gleichungen mit der Pseudo-Inversen
 - LSA / LSI
 - PCA
- Anmerkungen zur (numerischen) Bestimmung der SVD