SVD
A: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ ... a linear mapping

```matlab
fi = 0:0.01:2*pi;
x = [cos(fi); sin(fi)];
A = randn(2,2);
y = A*x;
```

```
A =

0.7942   -0.4284
1.2336    0.2478
```
Observation: a linear mapping maps circles to ellipses or to line segments
A set $Y$ is an ellipse $\iff Y$ is a conic and $\forall \vec{x}$ on an unit circle $\exists \lambda \geq 0$ such that $\lambda \vec{x}$ is on $Y$
S V D – Singular Value Decomposition

For every matrix $A \in \mathbb{R}^{m \times n}$ exist matrices $U \in \mathbb{R}^{m \times m}, D \in \mathbb{R}^{m \times n}, V \in \mathbb{R}^{n \times n}$ such that

$U^T U = I$ and $V^T V = I$

$D = \text{diag}([\sigma_{11}, \ldots, \sigma_{nn}]), \sigma_{11} \geq \ldots \geq \sigma_{nn} \geq 0$

$A = U D V^T$
$S V D$ – interpretation for regular $2 \times 2$ matrices

\[ \mathbf{\beta} = (\vec{i}, \vec{j}) \quad \mathbf{\beta'} = (\vec{v}_1, \vec{v}_2) \]

\[
\begin{align*}
\mathbf{x}_\mathbf{\beta} & \xrightarrow{V^{-1}(=V^T)} \mathbf{x}_\mathbf{\beta'} \\
\text{change of basis} & \quad \text{"squashing"} & \quad \text{change of basis} & \quad \text{rotation}
\end{align*}
\]

along coordinate axes

\[
A = U D V^T = (U V^{-1}) V D V^{-1}
\]

289
S V D – interpretation in general

\[ A = U D V^\top \]

\[
\begin{array}{c}
A \\
\hline
U \\
D \\
V^\top
\end{array}
\]

\[ A = U D V^\top = (U V^{-1}) \quad V \quad D \quad V^{-1} \]

\[
\begin{array}{c}
A \\
\hline
U \\
V^{-1} \quad O \\
V^\top \\
D \\
V^{-1}
\end{array}
\]

\[
\begin{array}{c}
\sigma_{ii} \\
D \\
O \\
V^\top
\end{array}
\]
S V D – Low rank approximation

Let $A^{m \times n}$ be a real matrix of rank $r$.

We are looking for a real matrix $\bar{A}^{m \times n}$ of rank $k \leq r$ that best approximates $A$ in the sense that the largest difference between the matrices understood as linear mappings is minimized, i.e.

$$\bar{A} = \arg \min_{B \in \mathbb{R}^{m \times n}} \max_{\text{rank } B = k, \|y\| = 1} \|Ay - By\| = \arg \min_{B \in \mathbb{R}^{m \times n}, \text{rank } B = k} \|A - B\|$$

Interestingly, it is easy to find matrix $\bar{A}$ using SVD of $A$. 

S V D – Low rank approximation

Theorem:

Let \( A = U D V^\top \) be the singular value decomposition or a real matrix \( A^{m \times n} \). Then,

\[
A_k = \arg \min_{B \in \mathbb{R}^{m \times n}, \text{rank } B = k} \|A - B\|
\]

is obtained as

\[
A_k = U D_k V^\top
\]

where

\[
A = U D V^\top, \ D = \text{diag}([\sigma_{11}, \ldots, \sigma_{nn}])
\]

\[
D_k = \text{diag}([\sigma_{11}, \ldots, \sigma_{kk}, 0, 0, \ldots])
\]
Lemma: \( R^{m \times m} \) and \( R^\top R = I \), then \( ||RA|| = ||A|| \)

Proof:

\[
||RA|| = \max_{x \in \mathbb{R}^n} ||RAx|| = \max_{x \in \mathbb{R}^n} (x^\top A^\top R^\top RA x)^{\frac{1}{2}}
\]
\[
= \max_{x \in \mathbb{R}^n} (x^\top A^\top A x)^{\frac{1}{2}} = ||A||
\]

Lemma: \( R^{n \times n} \) and \( R^\top R = I \), then \( ||AR|| = ||A|| \)

Proof:

\[
||AR|| = \max_{x \in \mathbb{R}^n} ||ARx|| = \max_{y \in \mathbb{R}^n} ||Ay|| = ||A||
\]

since \( \{y \mid y = Rx, x \in \mathbb{R}^n, ||x|| = 1\} = \{x \mid x \in \mathbb{R}^n, ||x|| = 1\} \)
Lemma: \[ \|A - A_k\| = \sigma_{k+1,k+1} \]

Proof:

\[
\|A - A_k\| = \|U(D - D_k)V^\top\| = \|D - D_k\| \\
= \max_{\|x\| = 1} \left( (\sigma_{11} - \sigma_{11})^2 x_1^2 + \ldots + (\sigma_{kk} - \sigma_{kk})^2 x_k^2 + \sigma_{k+1,k+1}^2 x_{k+1}^2 + \ldots \right)^{\frac{1}{2}} \\
= \max_{\|x\| = 1} \left( 0 x_1^2 + \ldots + 0 x_k^2 + \sigma_{k+1,k+1}^2 x_{k+1}^2 + \ldots + \sigma_{nn}^2 x_n^2 \right)^{\frac{1}{2}} \\
\leq \max_{\|x\| = 1} \sigma_{k+1,k+1} \left( x_1^2 + \ldots + x_k^2 + x_{k+1}^2 + \ldots + x_n^2 \right)^{\frac{1}{2}} = \sigma_{k+1,k+1} \\
\]

Since \[ \|(D - D_k)V^\top v_{k+1,k+1}\| = \sigma_{k+1,k+1} \] we conclude that \[ \|A - A_k\| = \sigma_{k+1,k+1} \]
Proof of the theorem: By contradiction. If \( k = n \), then \( A_k = A \). Assume that there is a matrix \( B \) with rank \( \text{rank } B = k < \text{rank } A \) such that \( \|A - B\| < \|A - A_k\| = \sigma_{k+1,k+1} \).

The null space \( N \) of \( B \) has dimension \( n - k > 0 \), and thus there is \( x \in N \) such that \( \|x\| = 1 \). For every \( x \in N \), \( Bx = 0 \). Take \( x \in N \) such that \( \|x\| = 1 \).

Then

\[
\|Ax\| = \|(A - B)x\| \leq \|(A - B)\| \overset{\text{assumption}}{=} \sigma_{k+1,k+1}
\]

\[\forall x \in \mathbb{R}^n : \|A - B\| = \max_{y \in \mathbb{R}^n, \|y\| = 1} \|(A - B)y\| \geq \|(A - B)x\| \]

For every \( x \in M = \text{span}(v_1, \ldots, v_{k+1}) \), such that \( \|x\| = 1 \)

\[
\|Ax\| = \|D \left( \begin{array}{c} v_1^T \\ \vdots \\ v_n^T \end{array} \right) x\| = \|D \left( \begin{array}{c} v_1^T \\ \vdots \\ v_n^T \end{array} \right) \sum_{i=1}^{k+1} a_i v_i\| = \|D \left( \begin{array}{c} v_1^T \\ \vdots \\ v_n^T \end{array} \right) \sum_{i=1}^{k+1} a_i v_i\| = \]

S V D – Proof of the low rank approximation

\[
\begin{pmatrix}
a_1 \\
\vdots \\
a_{k+1} \\
0 \\
\vdots
\end{pmatrix}
\]

\[
= \|D\| \left( \begin{pmatrix}
a_1 \\
\vdots \\
a_{k+1} \\
0 \\
\vdots
\end{pmatrix} \right)
\]

\[
= \left( \sigma^2_{1,1} a_1^2 + \ldots + \sigma^2_{k+1,k+1} a_{k+1}^2 \right)^{\frac{1}{2}}
\]

\[
\geq \left( \sigma^2_{k+1,k+1} a_1^2 + \ldots + \sigma^2_{k+1,k+1} a_{k+1}^2 \right)^{\frac{1}{2}}
\]

\[
= \sigma_{k+1,k+1} \left( a_1^2 + \ldots + a_{k+1}^2 \right)^{\frac{1}{2}} = \sigma_{k+1,k+1}
\]

since \( 1 = \|x\| = (a_1^2 + \ldots + a_{k+1}^2)^{\frac{1}{2}}. \)

\( M \cap N \neq \{0\}, \) since \( \text{dim } M = k + 1, \text{dim } N = n - k \) and \( k + 1 + n - k = n + 1 > n, \) and therefore there is a unit vector \( x \in M \cap N \) such that \( \|Ax\| < \sigma_{k+1,k+1} \) and \( \|Ax\| \geq \sigma_{k+1,k+1}, \) which is absurd. Therefore, there is no such \( B. \)