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

\begin{verbatim}
fi = 0:0.01:2*pi;
x = [cos(fi); sin(fi)];
A = randn(2,2);
y = A*x;

>> A

A =

0.7942  -0.4284
1.2336   0.2478
\end{verbatim}
Observation: a linear mapping maps circles to ellipses or to line segments.
A set $Y$ is an ellipse $\Leftrightarrow$ $Y$ is a conic and $\forall \vec{x}$ on an unit circle
$\exists \lambda \geq 0$ such that $\lambda \vec{x}$ is on $Y$
Theorem: A regular linear mapping maps circles to ellipses

Proof:

\[ A : \mathbb{R}^n \to \mathbb{R}^n \quad \ldots \quad \text{a regular linear mapping} \]

\[ x^\top x = 1 \quad \ldots \quad \text{x on a unit "circle"} \]

\[ y = A x \quad \ldots \quad \text{x is mapped to y} \]

\[ 1 = x^\top x = (A^{-1}y)^\top (A^{-1}y) \]

\[ 1 = y^\top (A^{-\top}A^{-1}) y \quad \ldots \quad \text{a conic} \]

Let us show that the above conic is an ellipse.

Take \( z \) on the unit circle. Then \( z^\top (A^{-\top}A^{-1}) z = (A^{-1}z)^\top (A^{-1}z) = \|A^{-1}z\|^2 > 0 \) since \( \|z\| = 1 \) and for a regular \( A \), \( A^{-1}x = 0 \Rightarrow x = 0 \).

Therefore \( \|A^{-1}z\| > 0 \) and \( \frac{z}{\|A^{-1}z\|} \) solves \( 1 = \frac{z^\top (A^{-\top}A^{-1})}{\|A^{-1}z\|} \frac{z}{\|A^{-1}z\|} \).
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 $V^T V = I$

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

$A = UDV^T$
SVD – interpretation for regular $2 \times 2$ matrices

$\beta = (\vec{i}, \vec{j})$ \hspace{1cm} $\beta' = (\vec{v}_1, \vec{v}_2)$

\[
\begin{align*}
\mathbf{x}_{\beta} & \xrightarrow{V^{-1}(=V^\top)} \mathbf{x}_{\beta'} \\
\text{change of basis} & \hspace{1cm} \text{"squashing"} \\
& \xrightarrow{D} \mathbf{z}_{\beta'} \\
& \xrightarrow{V} \mathbf{z}_{\beta} \\
& \xrightarrow{V^{-1}U} \mathbf{y}_{\beta}
\end{align*}
\]

along coordinate axes

\[A = UDV^\top = (UV^{-1}) V D V^{-1}\]
S V D – interpretation in general

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

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

\[ A = U D V^\top = (U V^{-1}) \begin{array}{c}
V^{-1} \\
0 \\
V \\
0 \\
\end{array} \begin{array}{c}
D \\
0 \\
\end{array} \begin{array}{c}
V^{-1} \\
0 \\
\end{array} \]

\[ \begin{array}{c}
A \\
U \\
V^{-1} \\
0 \\
V \\
0 \\
\end{array} \begin{array}{c}
V^{-1} \\
0 \\
\end{array} \begin{array}{c}
D \\
0 \\
\end{array} \begin{array}{c}
V^{-1} \\
0 \\
\end{array} \]

\[ \begin{array}{c}
\sigma_{ii} \\
V^{-1} \\
V \\
\bar{D} \\
\bar{V}^{-1} \\
\end{array} \]
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.

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

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

Let $A = UDV^\top$ be the singular value decomposition of a real matrix $A^{m \times n}$. Then,

$$A_k = \arg \min_{\substack{B \in \mathbb{R}^{m \times n} \
 rank B = k}} \|A - B\|$$

is obtained as

$$A_k = UD_k V^\top$$

where

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

$$D_k = \text{diag}([\sigma_{11}, \ldots, \sigma_{kk}, 0, 0, \ldots])$$
S V D – Proof of the low rank approximation

Lemma: $R^{m \times m}$ and $R^\top R = I$, then $\|RA\| = \|A\|$

Proof:

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

$$= \max_{x \in \mathbb{R}^n, \|x\| = 1} (x^\top A^\top A x)^{\frac{1}{2}} = \|Ax\|$$

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

Proof:

$$\|AR\| = \max_{x \in \mathbb{R}^n, \|x\| = 1} \|ARx\| = \max_{y \in \mathbb{R}^n, \|y\| = 1} \|Ay\| = \|A\|$$

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:

$$\|A - A_k\| = \|U(D - D_k)V^\top\| = \|D - D_k\|$$

$$= \max_{\|x\| = 1} ((\sigma_{11} - \sigma_{11})^2 x_1^2 + \ldots + (\sigma_{kk} - \sigma_{kk})^2 x_k^2 + \sigma_{k+1,k+1}^2 x_{k+1}^2 + \ldots)^{\frac{1}{2}}$$

$$= \max_{\|x\| = 1} (0 x_1^2 + \ldots + 0 x_k^2 + \sigma_{k+1,k+1}^2 x_{k+1}^2 + \ldots + \sigma_{nn}^2 x_n^2)^{\frac{1}{2}}$$

$$\leq \max_{\|x\| = 1} \sigma_{k+1,k+1} (x_1^2 + \ldots + x_k^2 + x_{k+1}^2 + \ldots + x_n^2)^{\frac{1}{2}} = \sigma_{k+1,k+1}$$

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 $\text{rank } B = k < \text{rank } A$ 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 $\|x\| = 1$

$$\|Ax\| = \|D \begin{pmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{pmatrix} x\| = \|D \begin{pmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{pmatrix} \sum_{i=1}^{k+1} a_i v_i\| = \|D \begin{pmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{pmatrix} \sum_{i=1}^{k+1} a_i v_i\| = \|D \begin{pmatrix} v_1^\top \\ \vdots \\ v_n^\top \end{pmatrix} \sum_{i=1}^{k+1} a_i v_i\| =$$
S V D – Proof of the low rank approximation

\[
\begin{align*}
&= \|D \begin{pmatrix} a_1 \\ \vdots \\ a_{k+1} \\ 0 \\ \vdots \end{pmatrix} \|
\\
&= (\sigma_{11}^2 a_1^2 + \ldots + \sigma_{k+1,k+1}^2 a_{k+1}^2)^{\frac{1}{2}}
\\
&\geq (\sigma_{k+1,k+1}^2 a_1^2 + \ldots + \sigma_{k+1,k+1}^2 a_{k+1}^2)^{\frac{1}{2}}
\\
&= \sigma_{k+1,k+1} (a_1^2 + \ldots + a_{k+1}^2)^{\frac{1}{2}} = \sigma_{k+1,k+1}
\\
\text{since } 1 = \|x\| = (a_1^2 + \ldots + a_{k+1}^2)^{\frac{1}{2}}.
\end{align*}
\]

\( M \cap N \neq \{0\}, \text{ since } \dim M = k + 1, \dim N = n - k \text{ and } k + 1 + n - k = n + 1 > n, \) 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 \( B. \)