SVD – Singular Value Decomposition

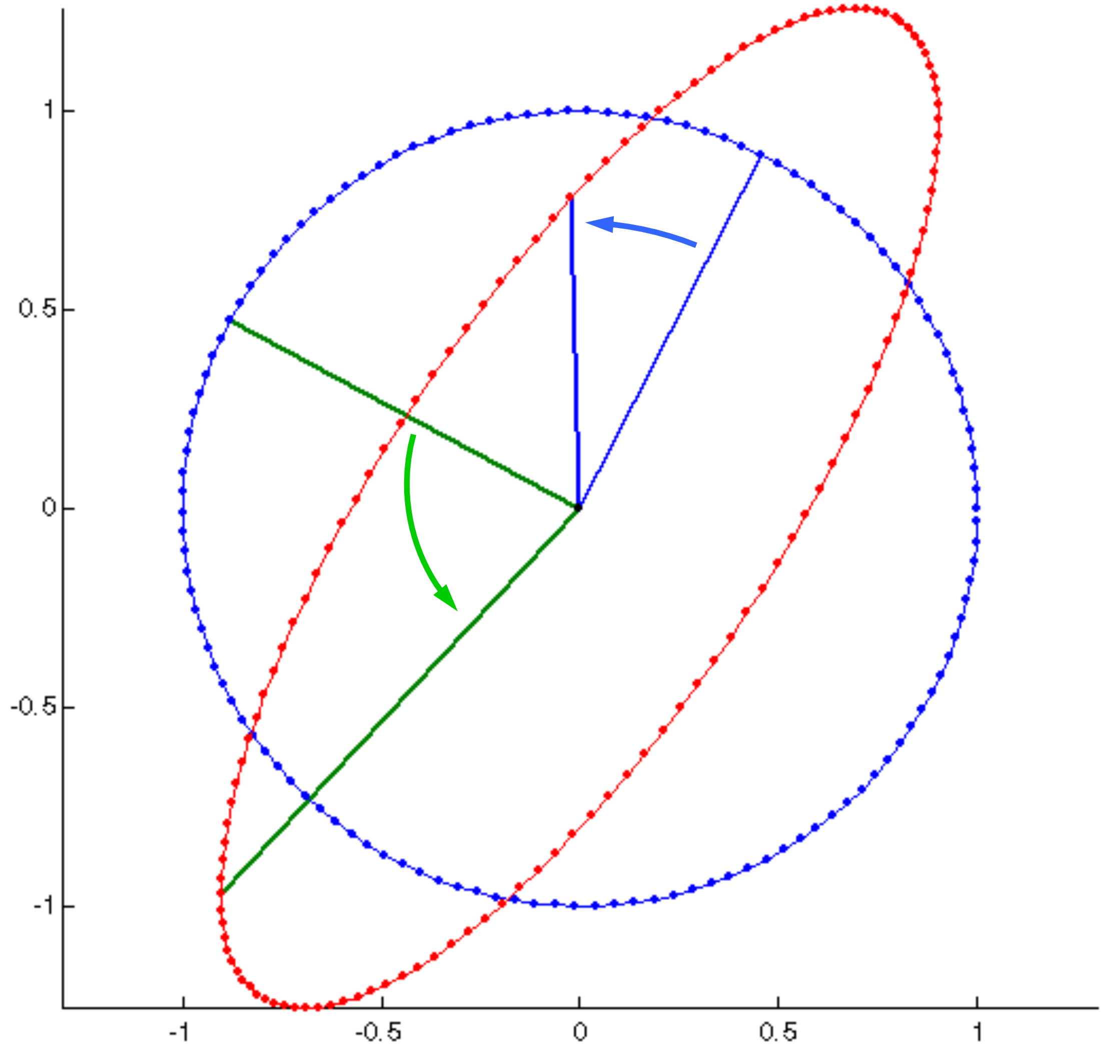
Tomas Pajdla

pajdla@cvut.cz

Linear mapping

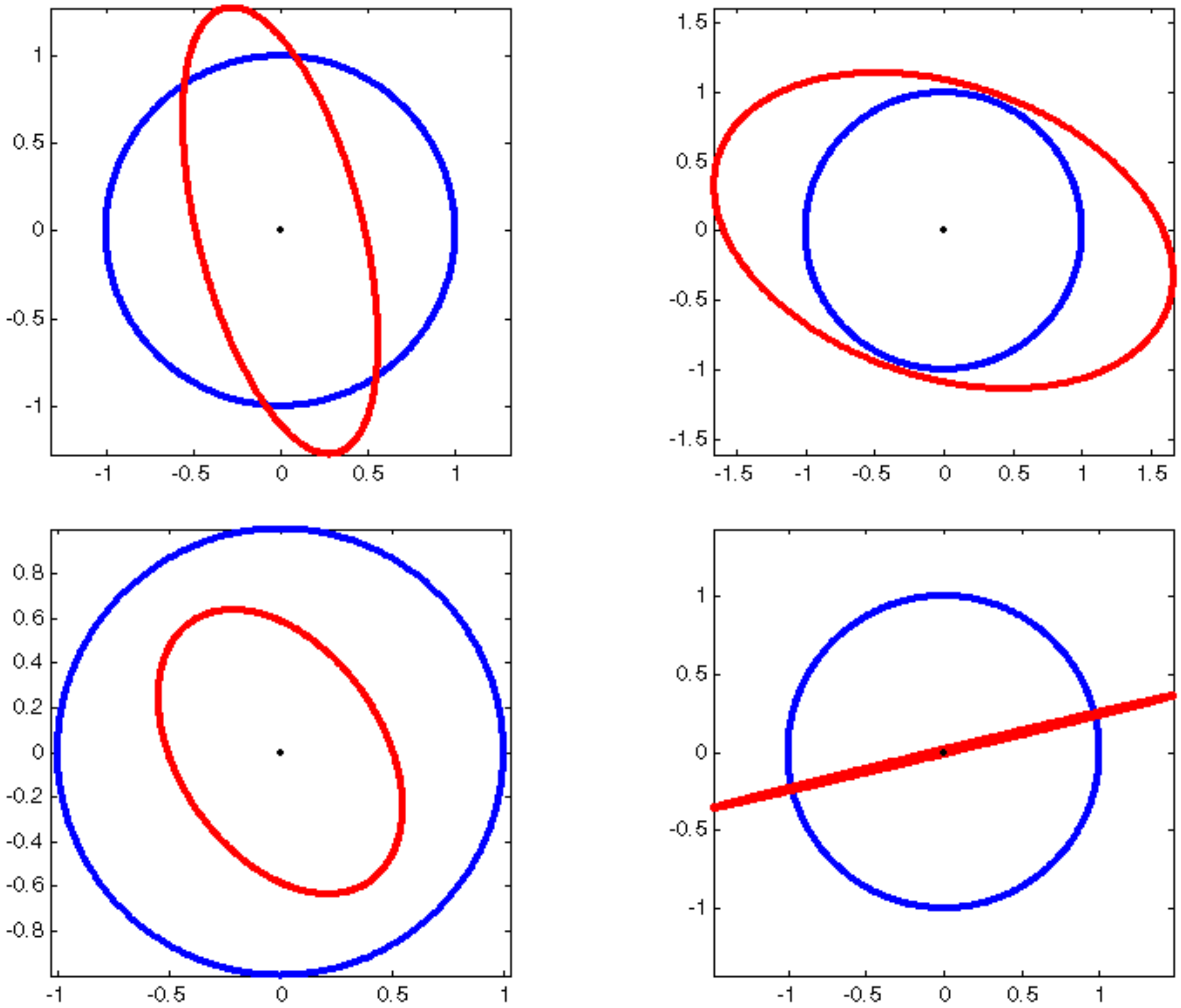
$\mathtt{A}:\mathbb{R}^2 \rightarrow \mathbb{R}^2$ ... a linear mapping

```
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
```



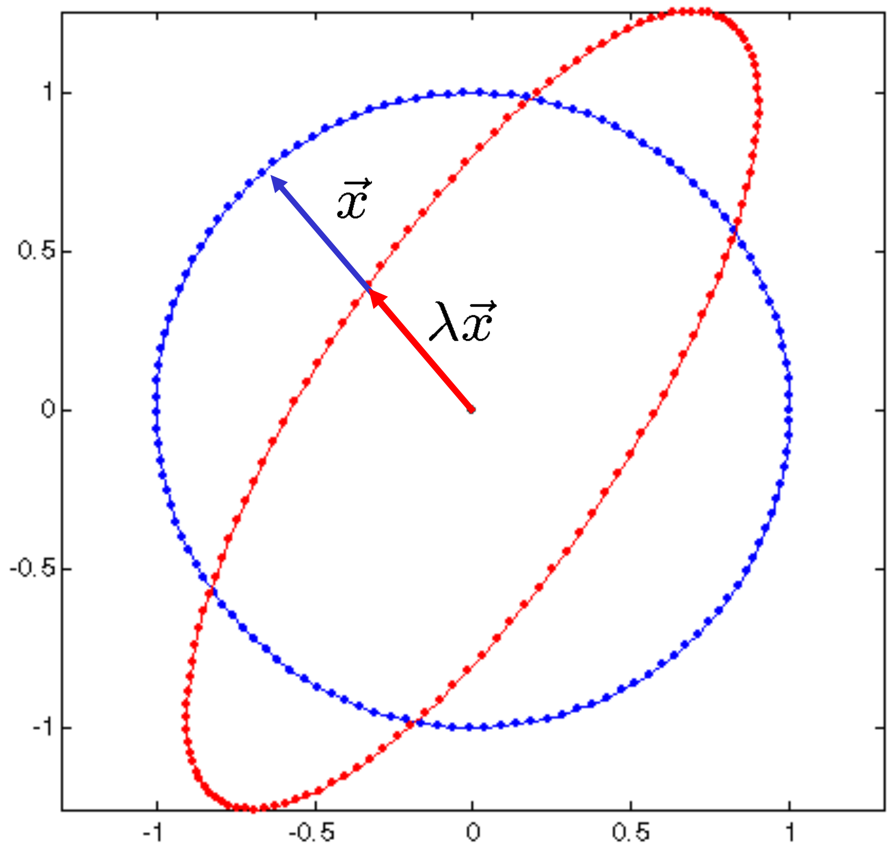
pajdla@cvut.cz

Observation: a linear mapping maps circles to ellipses or to line segments



pajdla@cvut.cz

Ellipse is a squashed circle



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

pajdla@cvut.cz

Theorem: A regular linear mapping maps circles to ellipses

Proof:

$\mathtt{A}:\mathbb{R}^n \rightarrow \mathbb{R}^n$ ... a regular linear mapping

$$\begin{aligned} \mathbf{x}^\top\mathbf{x} &= 1 && \dots && \mathbf{x} \text{ on a unit "circle"} \\ \mathbf{y} &= \mathtt{A}\,\mathbf{x} && \dots && \mathbf{x} \text{ is mapped to } \mathbf{y} \\ 1 = \mathbf{x}^\top\mathbf{x} &= (\mathtt{A}^{-1}\mathbf{y})^\top(\mathtt{A}^{-1}\mathbf{y}) \\ 1 &= \mathbf{y}^\top(\mathtt{A}^{-\top}\mathtt{A}^{-1})\,\mathbf{y} && \dots && \text{a conic} \end{aligned}$$

Let us show that the above conic is an ellipse.

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

Therefore $\|\mathtt{A}^{-1}\mathbf{z}\| > 0$ and $\frac{\mathbf{z}}{\|\mathtt{A}^{-1}\mathbf{z}\|}$ solves $1 = \frac{\mathbf{z}^\top}{\|\mathtt{A}^{-1}\mathbf{z}\|}(\mathtt{A}^{-\top}\mathtt{A}^{-1})\frac{\mathbf{z}}{\|\mathtt{A}^{-1}\mathbf{z}\|}$

S V D – Singular Value Decomposition

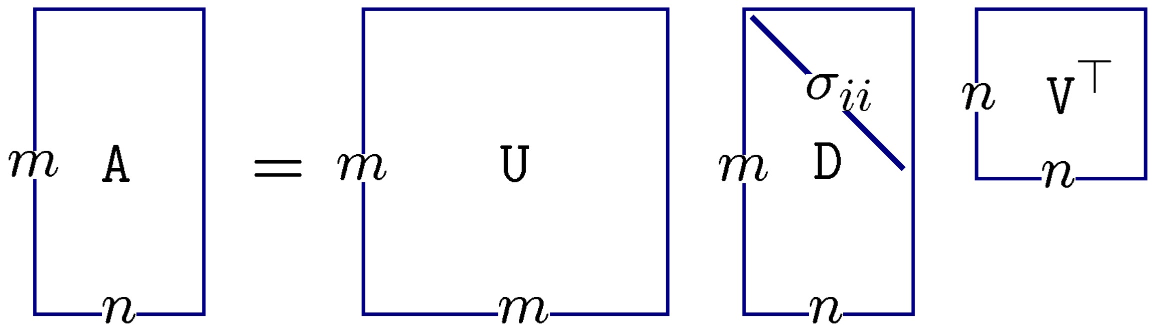
For every matrix $\mathtt{A} \in \mathbb{R}^{m \times n}$ exist matrices

$\mathtt{U} \in \mathbb{R}^{m \times m}, \mathtt{D} \in \mathbb{R}^{m \times n}, \mathtt{V} \in \mathbb{R}^{n \times n}$ such that

$\mathtt{U}^\top \mathtt{U} = \mathtt{I}$ and $\mathtt{V}^\top \mathtt{V} = \mathtt{I}$

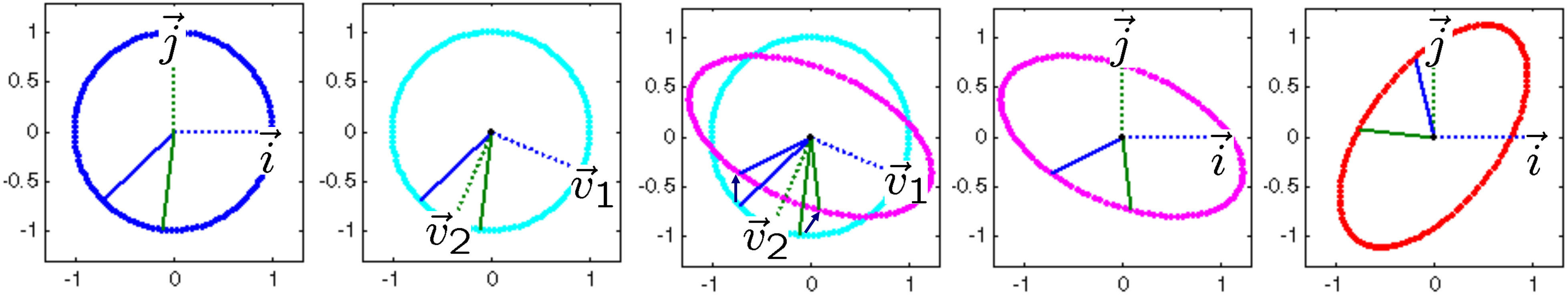
$\mathtt{D} = \mathrm{diag}([\sigma_{11}, \ldots, \sigma_{nn}]),\ \sigma_{11} \geq \ldots \geq \sigma_{nn} \geq 0$

$\mathtt{A} = \mathtt{U}\,\mathtt{D}\,\mathtt{V}^\top$



pajdla@cvut.cz

S V D – interpretation for regular 2×2 matrices



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

$$\mathbf{x}_\beta \xrightarrow{\mathtt{V}^{-1}(=\mathtt{V}^\top)} \mathbf{x}_{\beta'} \xrightarrow{\mathtt{D}} \mathbf{z}_{\beta'} \xrightarrow{\mathtt{V}} \mathbf{z}_\beta \xrightarrow{\mathtt{UV}^{-1}} \mathbf{y}_\beta$$

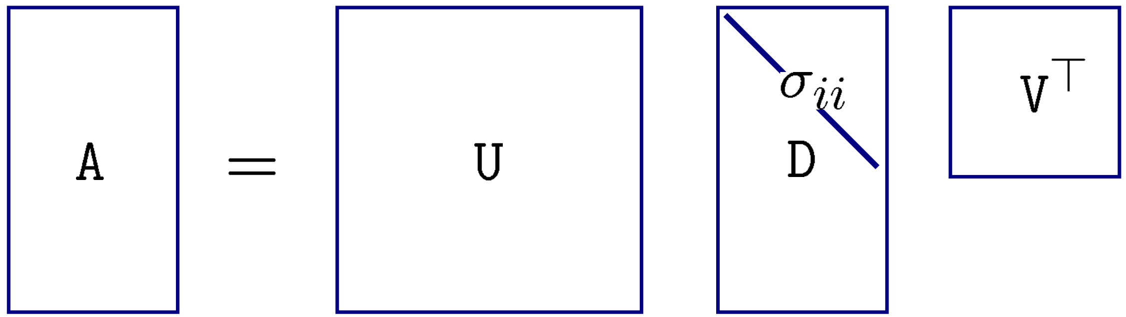
change of basis — "squashing" along coordinate axes — change of basis — rotation

$$\mathtt{A} = \mathtt{U\,D\,V}^\top = (\mathtt{U\,V}^{-1}) \quad \mathtt{V} \quad \mathtt{D} \quad \mathtt{V}^{-1}$$

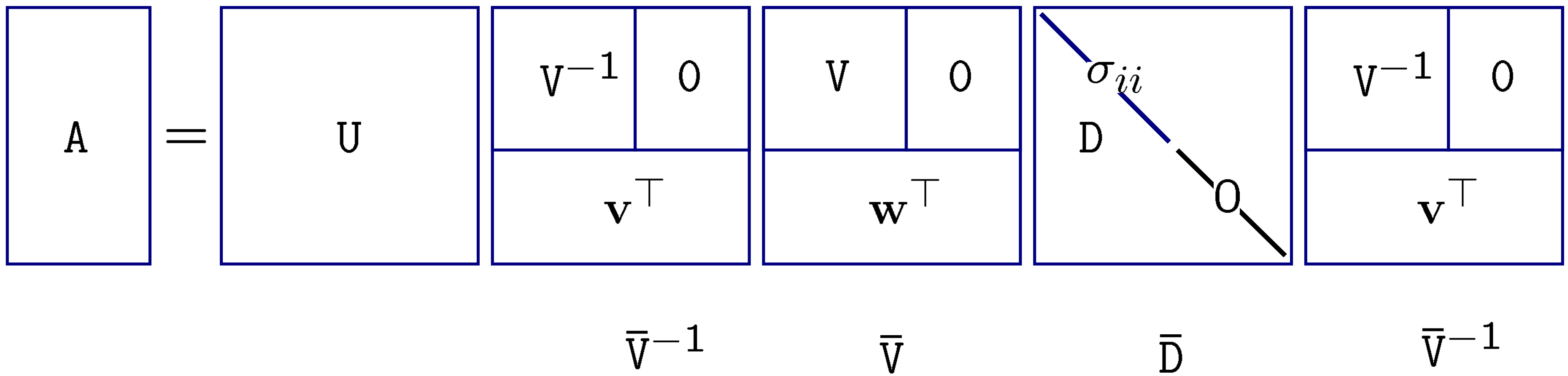
pajdla@cvut.cz

S V D – interpretation in general

$$A = U D V^\top$$



$$A = U D V^\top \quad = \quad (U V^{-1}) \quad \bar{V} \quad \bar{D} \quad \bar{V}^{-1}$$



pajdla@cvut.cz

S V D – Low rank approximation

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

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

$$\mathtt{A}_k = \arg \min_{\substack{\mathtt{B}\,\in\,\mathbb{R}^{m\times n} \\ rank\,\mathtt{B} = k}} \max_{\substack{\mathbf{y}\,\in\,\mathbb{R}^{n} \\ \|\mathbf{y}\| = 1}} \|\mathtt{A}\,\mathbf{y} - \mathtt{B}\,\mathbf{y}\| = \arg \min_{\substack{\mathtt{B}\,\in\,\mathbb{R}^{m\times n} \\ rank\,\mathtt{B} = k}} \|\mathtt{A} - \mathtt{B}\|$$

orm

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

S V D – Low rank approximation

Theorem:

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

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

is obtained as

$$\mathtt{A}_k = \mathtt{U}\,\mathtt{D}_k\,\mathtt{V}^\top$$

where

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

$$\mathtt{D}_k = \mathrm{diag}([\sigma_{11}, \ldots, \sigma_{kk}, 0, 0, \ldots])$$

pajdla@cvut.cz

S V D – Proof of the low rank approximation

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

Proof:

$$\begin{aligned}
\|\mathtt{R}\,\mathtt{A}\| &= \max_{\substack{\mathbf{x}\in\mathbb{R}^n\\ \|\mathbf{x}\|=1}} \|\mathtt{R}\,\mathtt{A}\,\mathbf{x}\| = \max_{\substack{\mathbf{x}\in\mathbb{R}^n\\ \|\mathbf{x}\|=1}} (\mathbf{x}^\top\mathtt{A}^\top\mathtt{R}^\top\mathtt{R}\,\mathtt{A}\,\mathbf{x})^{\frac{1}{2}} \\
&= \max_{\substack{\mathbf{x}\in\mathbb{R}^n\\ \|\mathbf{x}\|=1}} (\mathbf{x}^\top\mathtt{A}^\top\mathtt{A}\,\mathbf{x})^{\frac{1}{2}} = \|\mathtt{A}\|
\end{aligned}$$

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

Proof:

$$\|\mathtt{A}\,\mathtt{R}\| = \max_{\substack{\mathbf{x}\in\mathbb{R}^n\\ \|\mathbf{x}\|=1}} \|\mathtt{A}\,\mathtt{R}\,\mathbf{x}\| = \max_{\substack{\mathbf{y}\in\mathbb{R}^n\\ \|\mathbf{y}\|=1}} \|\mathtt{A}\,\mathbf{y}\| = \|\mathtt{A}\|$$

since $\{\mathbf{y}\,|\,\mathbf{y} = \mathtt{R}\,\mathbf{x},\ \mathbf{x}\in\mathbb{R}^n,\ \|\mathbf{x}\| = 1\} = \{\mathbf{x}\,|\,\mathbf{x}\in\mathbb{R}^n,\ \|\mathbf{x}\| = 1\}$

pajdla@cvut.cz

S V D – Proof of the low rank approximation

Lemma: $\|\mathtt{A} - \mathtt{A}_k\| = \sigma_{k+1,k+1}$

Proof:

$$\begin{aligned}
\|\mathtt{A} - \mathtt{A}_k\| &= \|\mathtt{U}(\mathtt{D} - \mathtt{D}_k)\mathtt{V}^\top\| = \|\mathtt{D} - \mathtt{D}_k\| \\
&= \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \|\mathbf{x}\| = 1}} \left((\sigma_{11} - \sigma_{11})^2 x_1^2 + \ldots + (\sigma_{kk} - \sigma_{kk})^2 x_k^2 + \sigma_{k+1,k+1}^2 x_{k+1}^2 + \ldots\right)^{\frac{1}{2}} \\
&= \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \|\mathbf{x}\| = 1}} \left(0\, x_1^2 + \ldots + 0\, x_k^2 + \sigma_{k+1,k+1}^2 x_{k+1}^2 + \ldots + \sigma_{nn}^2 x_n^2\right)^{\frac{1}{2}} \\
&\leq \max_{\substack{\mathbf{x} \in \mathbb{R}^n \\ \|\mathbf{x}\| = 1}} \sigma_{k+1,k+1} \left(x_1^2 + \ldots + x_k^2 + x_{k+1}^2 + \ldots + x_n^2\right)^{\frac{1}{2}} = \sigma_{k+1,k+1}
\end{aligned}$$

Since $\|(\mathtt{D} - \mathtt{D}_k)\mathtt{V}^\top \mathbf{v}_{k+1,k+1}\| = \sigma_{k+1,k+1}$ we conclude that $\|\mathtt{A} - \mathtt{A}_k\| = \sigma_{k+1,k+1}$

S V D – Proof of the low rank approximation

Proof of the theorem: By contradiction. If $k = n$, then $\mathtt{A}_k = \mathtt{A}$. Assume that there is a matrix $\mathtt{B}$ with $\operatorname{rank} \mathtt{B} = k < \operatorname{rank} \mathtt{A}$ such that $\|\mathtt{A} - \mathtt{B}\| < \|\mathtt{A} - \mathtt{A}_k\| = \sigma_{k+1,k+1}$.

The null space N of $\mathtt{B}$ has dimension $n - k > 0$, and thus there is $\mathbf{x} \in N$ such that $\|\mathbf{x}\| = 1$. For every $\mathbf{x} \in N$, $\mathtt{B}\,\mathbf{x} = \mathbf{0}$. Take $\mathbf{x} \in N$ such that $\|\mathbf{x}\| = 1$.

Then $\quad \|\mathtt{A}\,\mathbf{x}\| = \|(\mathtt{A} - \mathtt{B})\,\mathbf{x}\| \leq \|(\mathtt{A} - \mathtt{B})\| \overset{\text{assumption}}{<} \sigma_{k+1,k+1}$

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

For every $\mathbf{x} \in M = span(\mathbf{v}_1, \ldots, \mathbf{v}_{k+1})$, such that $\|\mathbf{x}\| = 1$

$$\|\mathtt{A}\,\mathbf{x}\| = \|\mathtt{D} \begin{pmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_n^\top \end{pmatrix} \mathbf{x}\| = \begin{matrix} \mathbf{x} \in M \\ \Downarrow \\ \mathbf{x} = \sum_{i=1}^{k+1} a_i \mathbf{v}_i \end{matrix} = \|\mathtt{D} \begin{pmatrix} \mathbf{v}_1^\top \\ \vdots \\ \mathbf{v}_n^\top \end{pmatrix} \sum_{i=1}^{k+1} a_i \mathbf{v}_i\| =$$

pajdla@cvut.cz

S V D – Proof of the low rank approximation

$$
\begin{aligned}
&= \|\mathtt{D} \begin{pmatrix} a_1 \\ \vdots \\ a_{k+1} \\ 0 \\ \vdots \end{pmatrix} \| \\
&= (\sigma_{11}^2 a_1^2 + \ldots + \sigma_{k+1,k+1}^2 a_{k+1}^2)^{\frac{1}{2}} \\
&\geq (\sigma_{k+1,k+1}^2 a_1^2 + \ldots + \sigma_{k+1,k+1}^2 a_{k+1}^2)^{\frac{1}{2}} \\
&= \sigma_{k+1,k+1} (a_1^2 + \ldots + a_{k+1}^2)^{\frac{1}{2}} = \sigma_{k+1,k+1}
\end{aligned}
$$

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

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

S V D – Low rank approximation in Frobenius norm

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

We are looking for a real matrix $\mathtt{A}_k^{m\times n}$ of rank $k \leq r$ that best approximates $\mathtt{A}$ in the sense that the largest difference between the matrices understood as vectors from $\mathbb{R}^{mn}$ is minimized, i.e.

$$\mathtt{A}_k = \arg \min_{\substack{\mathtt{B} \in \mathbb{R}^{m\times n} \\ rank\,\mathtt{B} = k}} \|\mathtt{A}-\mathtt{B}\|_F = \arg \min_{\substack{\mathtt{B} \in \mathbb{R}^{m\times n} \\ rank\,\mathtt{B} = k}} \sum_{i,j=1}^{i=m,j=n} (\mathtt{A}_{i,j}-\mathtt{B}_{i,j})^2$$

Again, it is easy to find matrix $\mathtt{A}_k$ using SVD of $\mathtt{A}$.

pajdla@cvut.cz

S V D – Low rank approximation in Frobenius norm

Theorem:

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

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

is obtained as

$$\mathtt{A}_k = \mathtt{U}\,\mathtt{D}_k\,\mathtt{V}^\top$$

where

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

$$\mathtt{D}_k = \mathrm{diag}([\sigma_{11}, \ldots, \sigma_{kk}, 0, 0, \ldots])$$

pajdla@cvut.cz

S V D – Low rank approximation in Frobenius norm

Lemma: $(\mathtt{U}^\top \mathtt{U} = \mathtt{I}) \Rightarrow (\|\mathtt{U}\,\mathtt{A}\|_F = \|\mathtt{A}\|_F)$

Proof:

$$\|\mathtt{U}\,\mathtt{A}\|_F \;=\; trace((\mathtt{U}\,\mathtt{A})^\top(\mathtt{U}\,\mathtt{A})) = trace(\mathtt{A}^\top\mathtt{U}^\top\mathtt{U}\,\mathtt{A}) = trace(\mathtt{A}^\top\mathtt{A}) = \|\mathtt{A}\|_F$$

Lemma: $(\mathtt{V}\,\mathtt{V}^\top = \mathtt{I}) \Rightarrow (\|\mathtt{A}\,\mathtt{V}^\top\|_F = \|\mathtt{A}\|_F)$

Proof:

$$\|\mathtt{A}\,\mathtt{V}^\top\|_F \;=\; trace((\mathtt{A}\,\mathtt{V}^\top)(\mathtt{A}\,\mathtt{V}^\top)^\top) = trace(\mathtt{A}\,\mathtt{V}^\top\mathtt{V}\,\mathtt{A}^\top) = trace(\mathtt{A}\,\mathtt{A}^\top) = \|\mathtt{A}\|_F$$

Lemma: $\|\mathtt{A} - \mathtt{A}_k\|_F = \sum_{i=k+1}^{n} \sigma_{i,i}^2$

Proof:

$$\|\mathtt{A} - \mathtt{A}_k\|_F \;=\; \|\mathtt{U}(\mathtt{D} - \mathtt{D}_k)\mathtt{V}^\top\|_F = \|\mathtt{D} - \mathtt{D}_k\|_F = \sigma_{k+1,k+1}^2 + \sigma_{k+2,k+2}^2 + \cdots + \sigma_{n,n}^2$$

pajdla@cvut.cz

S V D – Low rank approximation in Frobenius norm

Let us make first make quite a general observation:

$$\begin{aligned}
\mathtt{A}_k &= \arg \min_{\substack{\mathtt{B} \in \mathbb{R}^{m\times n} \\ rank\, \mathtt{B} = k}} \|\mathtt{A} - \mathtt{B}\|_F = \arg \min_{\substack{\bar{\mathtt{U}}, \bar{\mathtt{D}}, \bar{\mathtt{V}} \\ rank\, \bar{\mathtt{D}} = k \\ \bar{\mathtt{U}}^\top \bar{\mathtt{U}} = \mathtt{I} \\ \bar{\mathtt{V}}\, \bar{\mathtt{V}}^\top = \mathtt{I}}} \|\mathtt{U}\,\mathtt{D}\,\mathtt{V}^\top - \bar{\mathtt{U}}\,\bar{\mathtt{D}}\,\bar{\mathtt{V}}^\top\|_F \\
&= \arg \min_{\substack{\bar{\mathtt{U}}, \bar{\mathtt{D}}, \bar{\mathtt{V}} \\ rank\, \bar{\mathtt{D}} = k \\ \bar{\mathtt{U}}^\top \bar{\mathtt{U}} = \mathtt{I} \\ \bar{\mathtt{V}}\, \bar{\mathtt{V}}^\top = \mathtt{I}}} \|\mathtt{U}^\top \bar{\mathtt{U}}\,\bar{\mathtt{D}}\,\bar{\mathtt{V}}^\top \mathtt{V} - \mathtt{D}\|_F = \arg \min_{\substack{\tilde{\mathtt{U}}, \bar{\mathtt{D}}, \tilde{\mathtt{V}} \\ rank\, \bar{\mathtt{D}} = k \\ \tilde{\mathtt{U}}^\top \tilde{\mathtt{U}} = \mathtt{I} \\ \tilde{\mathtt{V}}\, \tilde{\mathtt{V}}^\top = \mathtt{I}}} \|\tilde{\mathtt{U}}^\top \bar{\mathtt{D}}\,\tilde{\mathtt{V}} - \mathtt{D}\|_F
\end{aligned}$$

And then see the proof for $m = n = 2$ and $k = 1$

pajdla@cvut.cz

S V D – Low rank approximation in Frobenius norm

Proof for $m = n = 2$ and $k = 1$

$$\left\| \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_{11} & v_{21} \\ v_{12} & v_{22} \end{bmatrix} - \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \right\|_F =$$

$$= \left\| \begin{bmatrix} s\, v_{11}\, u_{11} & s\, v_{21}\, u_{11} \\ s\, v_{11}\, u_{21} & s\, v_{21}\, u_{21} \end{bmatrix} - \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \right\|_F$$

$$= \left\| s\, v_{11} \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} - \begin{bmatrix} \sigma_{11} \\ 0 \end{bmatrix} \right\|_F + \left\| s\, v_{21} \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma_{22} \end{bmatrix} \right\|_F$$

$$\geq \left\| a \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} - \begin{bmatrix} \sigma_{11} \\ 0 \end{bmatrix} \right\|_F + \left\| b \begin{bmatrix} u_{11} \\ u_{21} \end{bmatrix} - \begin{bmatrix} 0 \\ \sigma_{22} \end{bmatrix} \right\|_F$$

$$\geq (\sigma_{11}\, u_{21})^2 + (\sigma_{22}\, u_{11})^2 \geq \sigma_{22}^2\, u_{21}^2 + \sigma_{22}^2\, u_{11}^2 = \sigma_{22}^2\,(u_{21}^2 + u_{11}^2) = \sigma_{22}^2$$

pajdla@cvut.cz