

S V D

Linear mapping

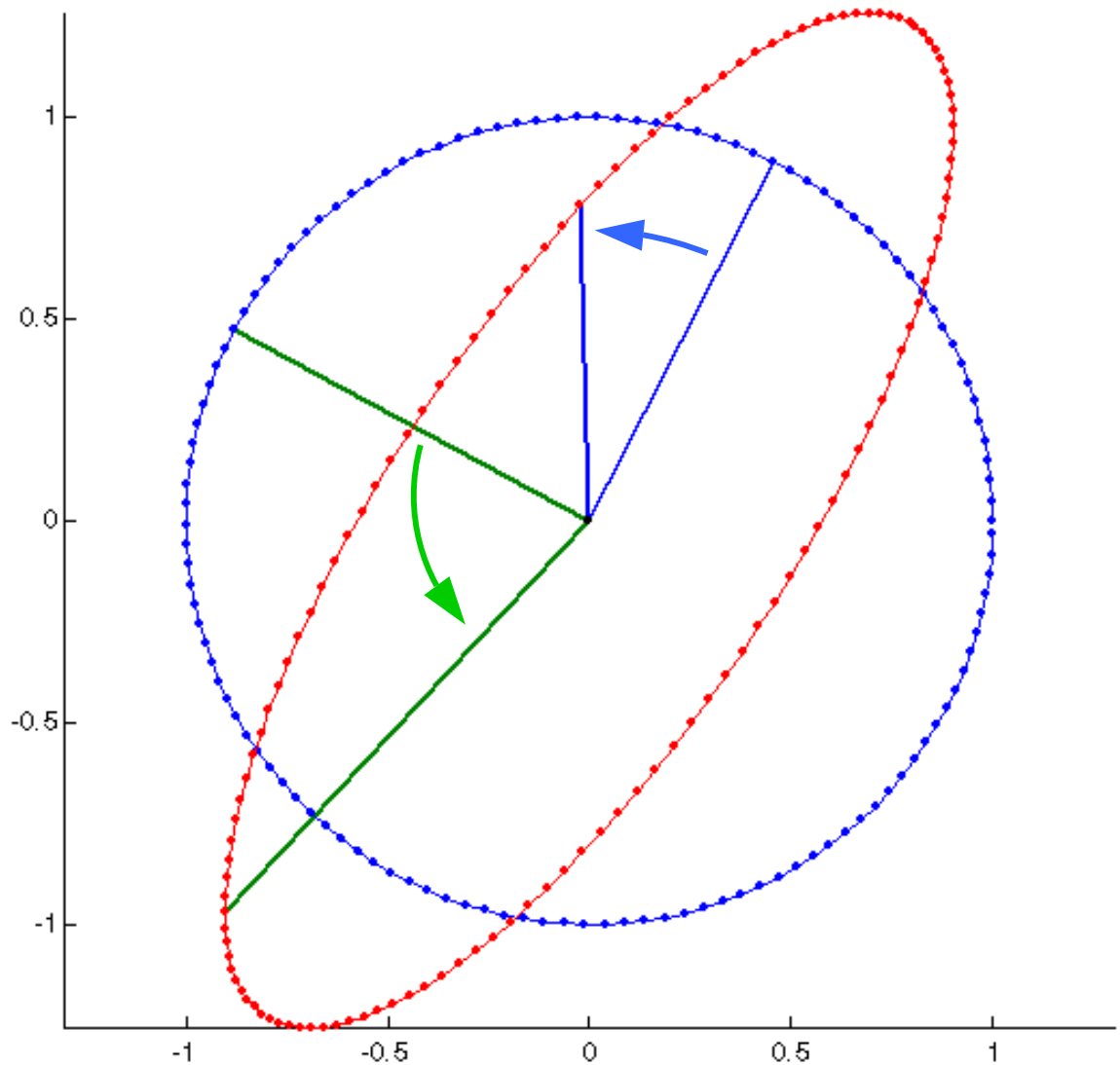
$A: \mathbb{R}^2 \rightarrow \mathbb{R}^2$... a linear mapping

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

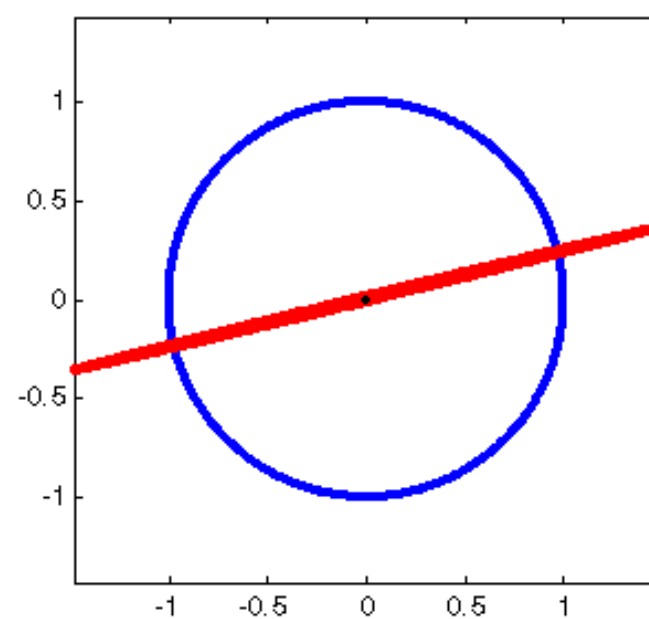
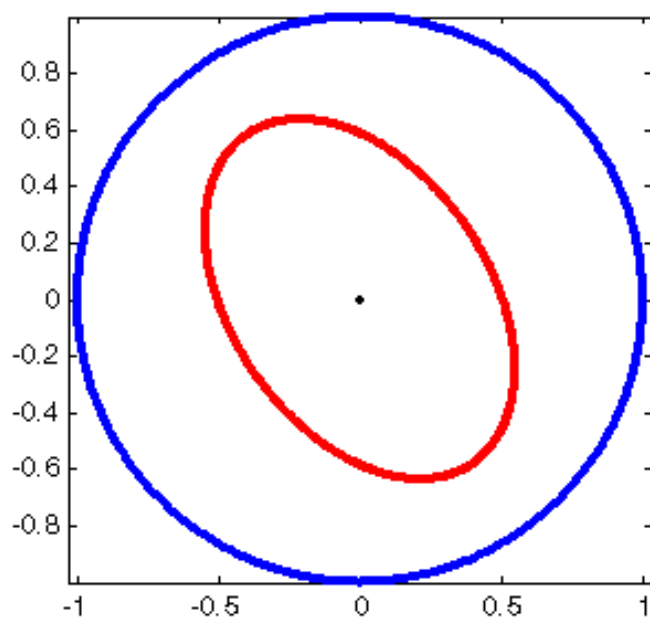
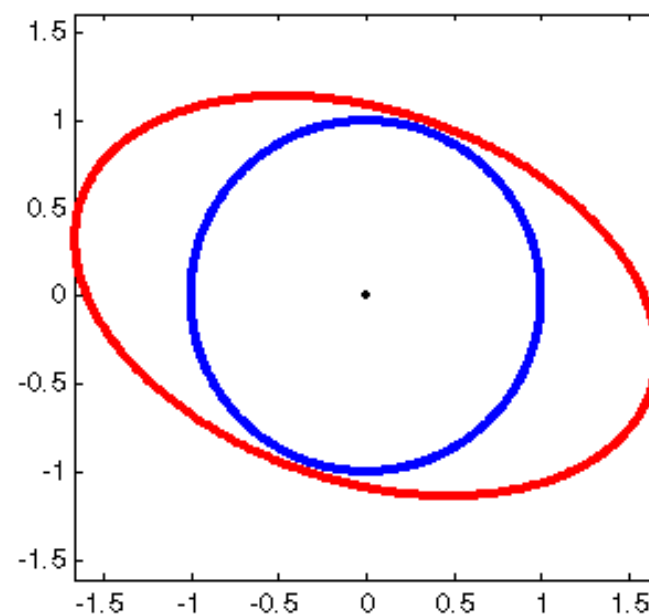
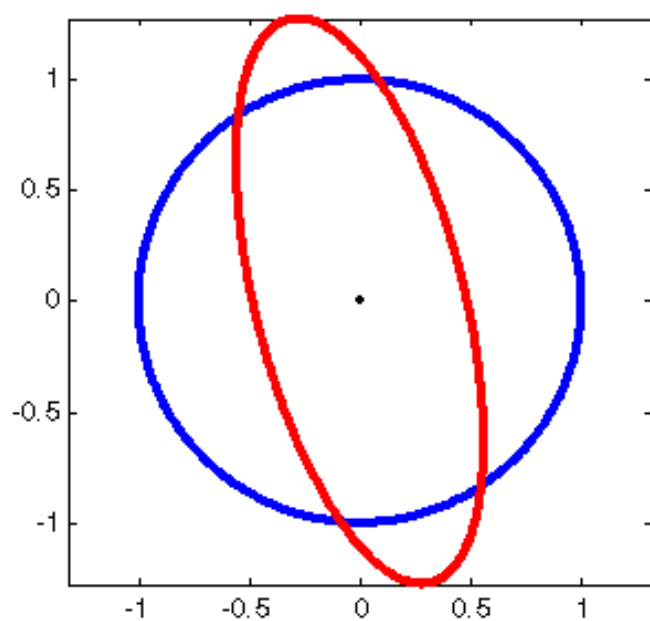
```
>> A
```

```
A =
```

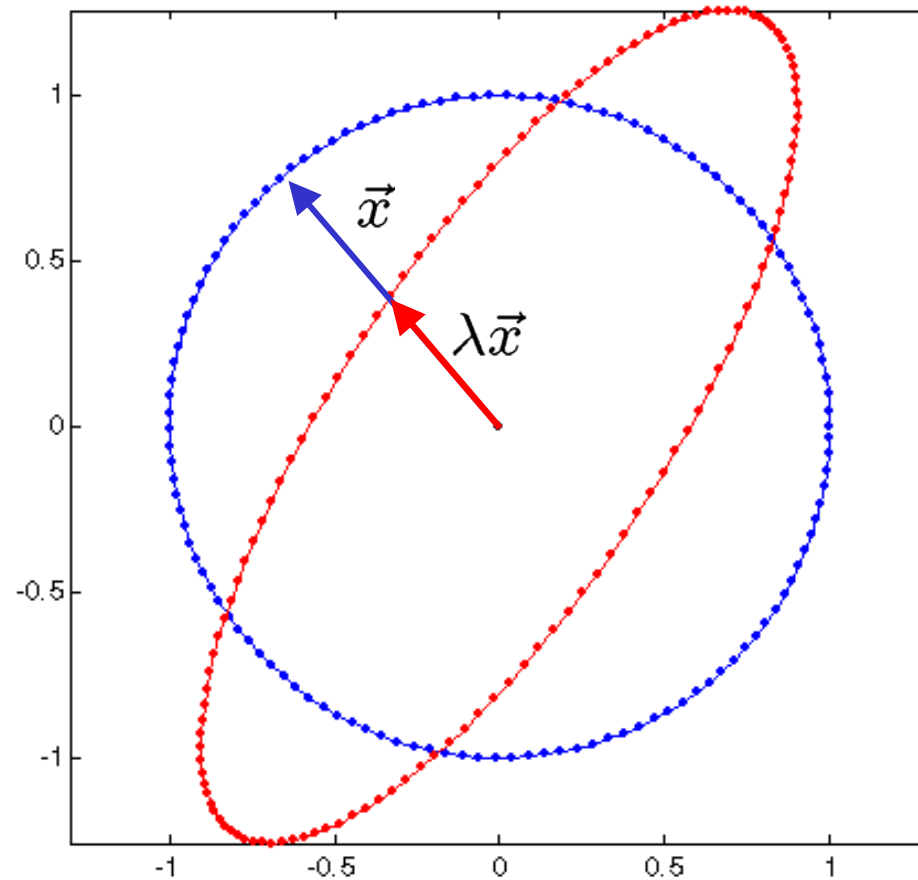
```
0.7942   -0.4284  
1.2336    0.2478
```



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



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

Theorem: A regular linear mapping maps circles to ellipses

Proof:

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

$\mathbf{x}^\top \mathbf{x} = 1$... \mathbf{x} on a unit "circle"

$\mathbf{y} = A\mathbf{x}$... \mathbf{x} is mapped to \mathbf{y}

$$1 = \mathbf{x}^\top \mathbf{x} = (A^{-1}\mathbf{y})^\top (A^{-1}\mathbf{y})$$

$$1 = \mathbf{y}^\top (A^{-\top} A^{-1}) \mathbf{y} \quad \dots \quad \text{a conic}$$

Let us show that the above conic is an ellipse.

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

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

S V D – Singular Value Decomposition

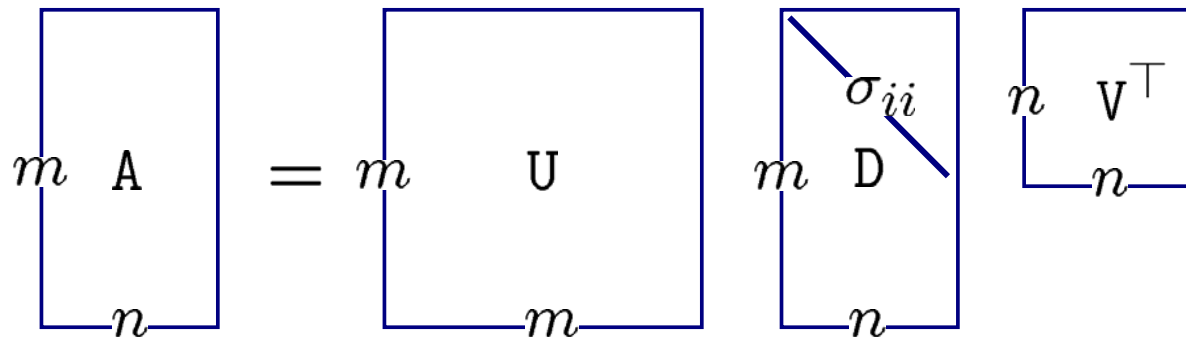
For every matrix $A \in \mathbb{R}^{m \times n}$ exist matrices

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

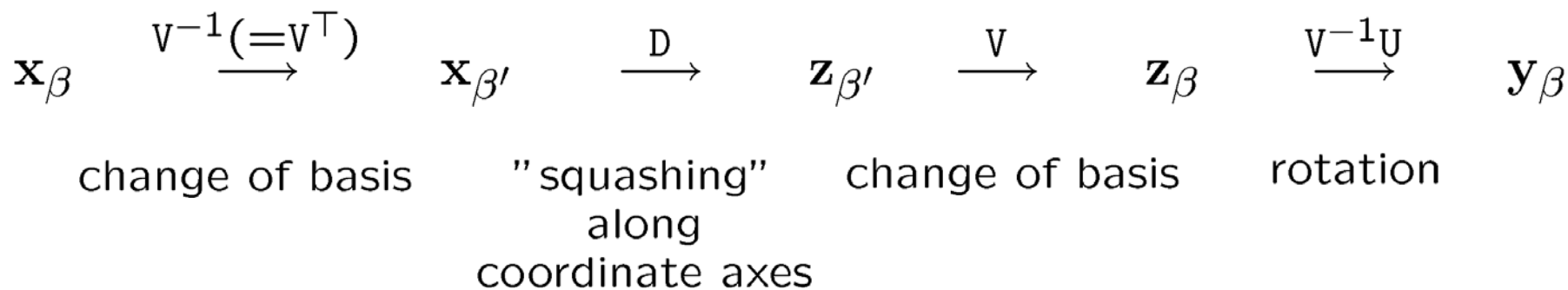
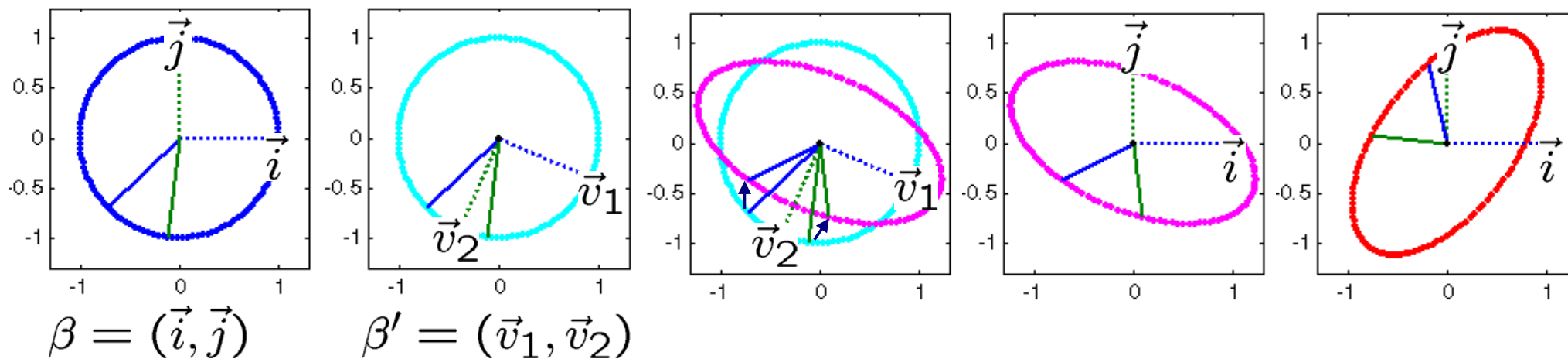
$$U^T U = I \text{ and } V^T V = I$$

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

$$A = U D V^T$$



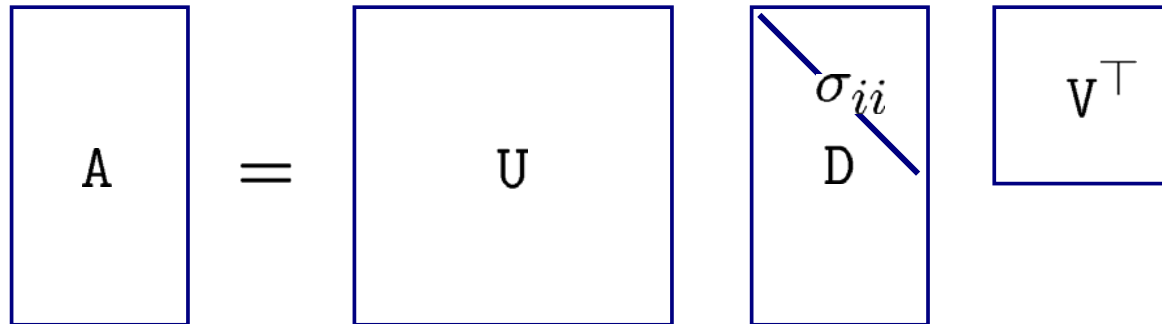
SVD – interpretation for regular 2×2 matrices



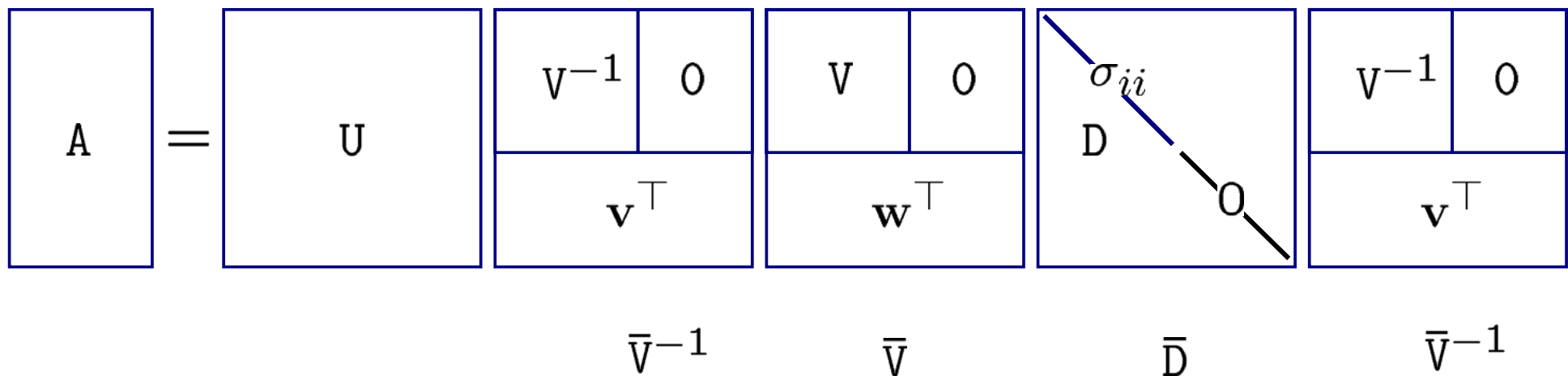
$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T = (\mathbf{U}\mathbf{V}^{-1}) \mathbf{V} \mathbf{D} \mathbf{V}^{-1}$$

SVD – interpretation in general

$$A = UDV^T$$



$$A = UDV^T = (UV^{-1}) \bar{V} \bar{D} \bar{V}^{-1}$$



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.

$$\bar{A} = \arg \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank } B = k}} \max_{\substack{\mathbf{y} \in \mathbb{R}^n \\ \|\mathbf{y}\| = 1}} \underbrace{\|A\mathbf{y} - B\mathbf{y}\|}_{\text{vector norm}} = \arg \min_{\substack{B \in \mathbb{R}^{m \times n} \\ \text{rank } B = k}} \underbrace{\|A - B\|}_{\text{matrix norm}}$$

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

S V D – Low rank approximation

Theorem:

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

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

is obtained as

$$A_k = U D_k V^T$$

where

$$A = UDV^T, D = \text{diag}([\sigma_{11}, \dots, \sigma_{nn}])$$

$$D_k = \text{diag}([\sigma_{11}, \dots, \sigma_{kk}, 0, 0, \dots])$$

S V D – Proof of the low rank approximation

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

Proof:

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

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

Proof:

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

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

SVD – Proof of the low rank approximation

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

Proof:

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

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

SVD – Proof of the low rank approximation

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 $\mathbf{x} \in N$ such that $\|\mathbf{x}\| = 1$. For every $\mathbf{x} \in N$, $B\mathbf{x} = \mathbf{0}$. Take $\mathbf{x} \in N$ such that $\|\mathbf{x}\| = 1$.

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

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

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

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

S V D – Proof of the low rank approximation

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

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

$M \cap N \neq \{\mathbf{0}\}$, 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 $\|A\mathbf{x}\| < \sigma_{k+1,k+1}$ and $\|A\mathbf{x}\| \geq \sigma_{k+1,k+1}$, which is absurd. Therefore, there is no such B.