# Matrix Decomposition

矩阵分解

Liang Zheng
Australian National University
liang.zheng@anu.edu.au

# 4.1 Determinant 行列式

- We write the determinant as $\det(\boldsymbol{A})$ or sometimes as $|\boldsymbol{A}|$ so that

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

- The determinant of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is a function that maps $\boldsymbol{A}$ onto a real number.
- Example 4.1 (Testing for Matrix Invertibility)
- If $\boldsymbol{A}$ is a $1\times 1$ matrix, then $\boldsymbol{A} = a \Rightarrow \boldsymbol{A}^{-1} = \frac{1}{a}$. It holds if and only if $a \neq 0$.
- For $2\times 2$ matrices, if $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, recall that the inverse of $\boldsymbol{A}$ is

$$\boxed{\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}}$$

- Hence, $\boldsymbol{A}$ is invertible if and only if

$$a_{11}a_{22} - a_{12}a_{21} \neq 0$$

- This quantity is the determinant of $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$, i.e.,

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \underline{a_{11}a_{22} - a_{12}a_{21}}$$

- For a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ it holds that $\boldsymbol{A}$ is invertible if and only if $det(\boldsymbol{A}) \neq 0$.
- We have explicit (closed-form) expressions for determinants of small matrices in terms of the elements of the matrix. For $n = 1$,

$$\det(\boldsymbol{A}) = \det(a_{11}) = a_{11}$$

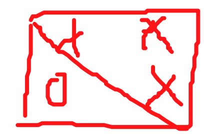
- For $n = 2$,

$$\det(\boldsymbol{A}) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

which we have observed in the preceding example.

- For $n = 3$ (known as Sarrus' rule),

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} \\ -a_{31}a_{22}a_{13} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33}$$



- We call a square matrix $\boldsymbol{T}$ an upper-triangular matrix if $T_{ij} = 0$ for $i > j$, i.e., the matrix is zero below its diagonal.
- Analogously, we define a lower-triangular matrix as a matrix with zeros above its diagonal.

- For a triangular matrix $\boldsymbol{T} \in \mathbb{R}^{n\times n}$ , the determinant is the product of the diagonal elements, i.e.,

$$\det(\boldsymbol{T}) = \prod_{i=1}^{n} T_{ii}$$

- How can we compute the determinant of an $n \times n$ $(n > 3)$ matrix?
- We reduce this problem to computing the determinant of $(n-1)\times(n-1)$ matrices. By recursively applying the Laplace expansion, we can compute determinants of an $n \times n$ matrix by ultimately computing determinants of $2\times 2$ matrices.
- **Theorem 4.2** (Laplace Expansion)**.**
- Consider a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. Then, for all $j = 1, \ldots, n$:
- 1. Expansion along column $j$

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{kj} \det(\boldsymbol{A}_{\boldsymbol{k},\boldsymbol{j}})$$

- *2.* Expansion along row $j$

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{\boldsymbol{j},\boldsymbol{k}})$$

- Here $\boldsymbol{A}_{k,j} \in \mathbb{R}^{(n-1)\times(n-1)}$ is the submatrix of $\boldsymbol{A}$ that we obtain when deleting row $k$ and column $j$.

2. Expansion along row $j$

$$\det(\boldsymbol{A}) = \sum_{k=1}^{n} (-1)^{k+j} a_{jk} \det(\boldsymbol{A}_{j,k})$$

- **Example 4.3 (Laplace Expansion)**
- Let us compute the determinant of

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

- Using the Laplace expansion **along the first row**, yielding

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = (-1)^{1+1} \cdot 1 \begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+2} \cdot 2 \begin{vmatrix} 3 & 2 \\ 0 & 1 \end{vmatrix} + (-1)^{1+3} \cdot 3 \begin{vmatrix} 3 & 1 \\ 0 & 0 \end{vmatrix}$$

- We compute the determinants of all the $2 \times 2$ matrices and obtain

$$\det(\boldsymbol{A}) = 1(1-0) - 2(3-0) + 3(0-0) = -5$$

- For completeness we can compare this result to computing the determinant using Sarrus' rule:

$$\det(\boldsymbol{A}) = 1 \cdot 1 \cdot 1 + 3 \cdot 0 \cdot 3 + 0 \cdot 2 \cdot 2 - 0 \cdot 1 \cdot 3 - 1 \cdot 0 \cdot 2 - 3 \cdot 2 \cdot 1 = 1 - 6 = -5.$$

# Properties of the determinant

- For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, we have the following properties
- The determinant of a matrix product is the product of the corresponding determinants $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$
- Determinants are invariant to transposition $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T)$
- If $\boldsymbol{A}$ is regular (invertible), then $\det(\boldsymbol{A}^{-1}) = \frac{1}{\det(\boldsymbol{A})}$

与高斯消元有关

- Adding a multiple of a column/row to another one does not change $\det(\boldsymbol{A})$
- Multiplication of a column/row with $\lambda \in \mathbb{R}$ scales $\det(\boldsymbol{A})$ by $\lambda$. In particular, $\det(\lambda\boldsymbol{A}) = \lambda^n \det(\boldsymbol{A})$
- Swapping two rows/columns changes the sign of $\det(\boldsymbol{A})$
- Because of the last three properties, we can use Gaussian elimination to compute $\det(\boldsymbol{A})$ by bringing $\boldsymbol{A}$ into row-echelon form. We can stop Gaussian elimination when we have $\boldsymbol{A}$ in a triangular form where the elements below the diagonal are all $0$. Recall: the determinant of a triangular matrix is the product of the diagonal elements.

# Example

- Let us use Gauss elimination in order to obtain the following determinant

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \; row2 - 3 \times row1$$

$$\rightsquigarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & -5 & -7 \\ 0 & 0 & 1 \end{bmatrix}$$

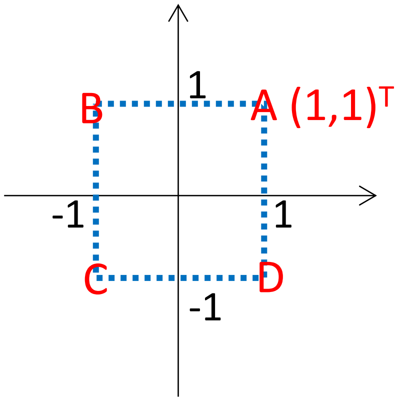
- Now we have the upper triangular form (row-echelon form).

$$\det(\boldsymbol{A}) = 1 \times (-5) \times 1 = -5$$

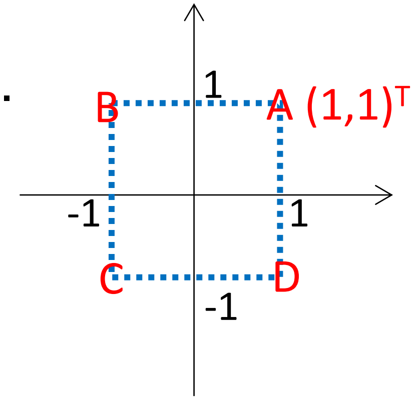
- We can verify this result with the previous example.

# Understanding of determinant

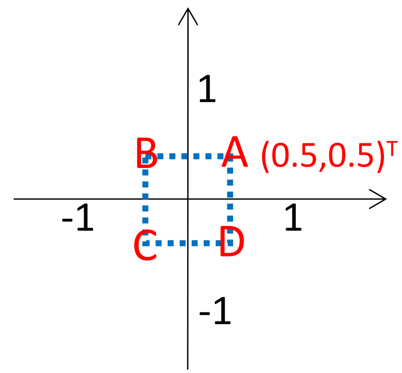
- Matrices characterize linear transformations.




Left multiply a matrix




$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$$

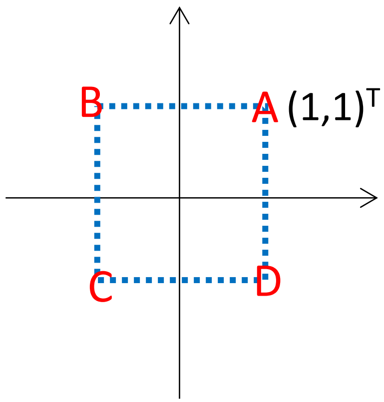



$$\begin{vmatrix} 0.5 & 0 \\ 0 & 0.5 \end{vmatrix} = 0.25$$

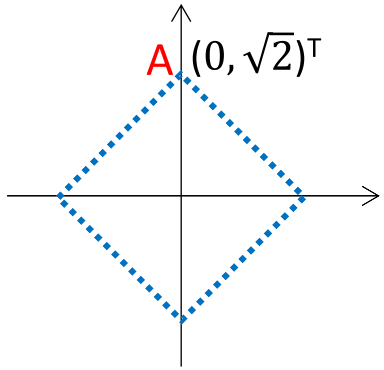
When determinant is greater than 1, it will enlarge a graph; otherwise it shrinks a graph

# Determinant and invertibility

$$\begin{vmatrix} \cos 45° & -\sin 45° \\ \sin 45° & \cos 45° \end{vmatrix} = 1$$




45° counterclock wise rotation

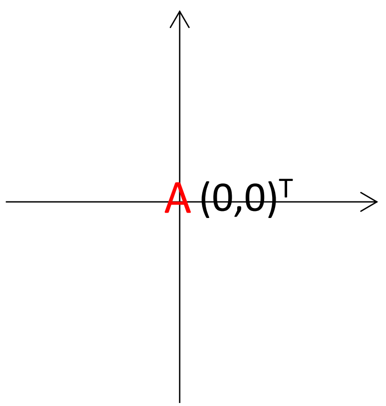



45° clockwise rotation

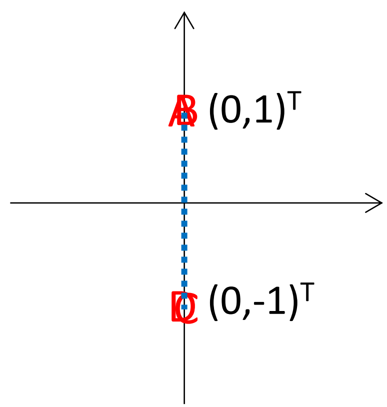
$$\begin{vmatrix} \cos(-45°) & -\sin(-45°) \\ \sin(-45°) & \cos(-45°) \end{vmatrix} = 1$$

$\begin{bmatrix} \cos 45° & -\sin 45° \\ \sin 45° & \cos 45° \end{bmatrix}$ and $\begin{bmatrix} \cos(-45°) & -\sin(-45°) \\ \sin(-45°) & \cos(-45°) \end{bmatrix}$ are inverses of each other

- Some linear transformations (matrices) are not invertible




$$\begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix} = 0$$




$$\begin{vmatrix} 0 & 0 \\ 0 & 1 \end{vmatrix} = 0$$

You cannot restore the original rectangle from these collapsed shapes.

# 4.2 Eigenvalues and Eigenvectors

特征值与特征向量

- For $\lambda \in \mathbb{R}$ and a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

$$p_{\boldsymbol{A}}(\lambda) := \det(\boldsymbol{A} - \lambda \boldsymbol{I})$$
$$= c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{n-1}\lambda^{n-1} + (-1)^n\lambda^n$$

$c_0, \cdots, c_{n-1} \in \mathbb{R}$, is the characteristic polynomial of $\boldsymbol{A}$.

- The characteristic polynomial $p_{\boldsymbol{A}}(\lambda) := \det(\boldsymbol{A} - \lambda \boldsymbol{I})$ will allow us to compute eigenvalues and eigenvectors.

- Example

- $\boldsymbol{A} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, we have,

$$p_{\boldsymbol{A}}(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} = (1-\lambda)(3-\lambda) - 8 = (\lambda - 5)(\lambda + 1)$$

- Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a square matrix. Then $\lambda \in \mathbb{R}$ is an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$ is the corresponding eigenvector of $\boldsymbol{A}$ if

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

- We call this equation the eigenvalue equation.

- The following statements are equivalent:
- $\lambda$ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$
- There exists an $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\}$ with $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ or equivalently $(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\boldsymbol{x} = \boldsymbol{0}$ can be solved non-trivially, i.e., $\boldsymbol{x} \neq \boldsymbol{0}$
- $\mathrm{rk}(\boldsymbol{A} - \lambda \boldsymbol{I}_n) < n$ rank
- $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$

非唯一性

- Non-uniqueness of eigenvectors
- If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ associated with eigenvalue $\lambda$, then for any $c \in \mathbb{R}\backslash\{0\}$ it holds that $c\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$ with the same eigenvalue since
$$\boldsymbol{A}(c\boldsymbol{x}) = c\boldsymbol{A}\boldsymbol{x} = c\lambda\boldsymbol{x} = \lambda(c\boldsymbol{x})$$
- Thus, all vectors that are collinear (point in the same or opporsite direction) to $\boldsymbol{x}$ are also eigenvectors of $\boldsymbol{A}$.

- **Theorem 4.8.** $\lambda \in \mathbb{R}$ is eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ if and only if $\lambda$ is a root of the characteristic polynomial $p_{\boldsymbol{A}}(\lambda)$ of $\boldsymbol{A}$. $p_{\boldsymbol{A}}(\lambda) := \det(\boldsymbol{A} - \lambda\boldsymbol{I})$
- $\boldsymbol{A} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, we have,

$$p_{\boldsymbol{A}}(\lambda) = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} = (1-\lambda)(3-\lambda) - 8 = (\lambda - 5)(\lambda + 1)$$

- Eigenvalues are $\lambda_1 = 5, \lambda_2 = -1$

- **Definition**. Let a square matrix $\boldsymbol{A}$ have an eigenvalue $\lambda_i$. The algebraic multiplicity of $\lambda_i$ is the number of times the root appears in the characteristic polynomial.
- $\boldsymbol{A} = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$, we have,

$$p_{\boldsymbol{A}}(\lambda) = \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \begin{vmatrix} 1-\lambda & 4 \\ 2 & 3-\lambda \end{vmatrix} = (1-\lambda)(3-\lambda) - 8 = (\lambda - 5)(\lambda + 1)$$

- Eigenvalues are $\lambda_1 = 5, \lambda_2 = -1$
- Hence it has two distinct eigenvalues and each occurs only once, so the algebraic multiplicity of both eigenvalues is one.
- $\boldsymbol{B} = \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}$, we have,

$$p_{\boldsymbol{B}}(\lambda) = \det(B - \lambda \boldsymbol{I}) = \begin{vmatrix} 5-\lambda & 0 \\ 0 & 5-\lambda \end{vmatrix} = (5-\lambda)^2$$

- Eigenvalues are $\lambda_1 = \lambda_2 = 5$
- The eigenvalue $5$ has algebraic multiplicity of $2$

- **Definition**. For $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, the union of the $\mathbf{0}$ vector and the set of all eigenvectors of $\boldsymbol{A}$ associated with an eigenvalue $\lambda$ is a subspace of $\mathbb{R}^n$, which is called the eigenspace of $\boldsymbol{A}$ with respect to $\lambda$ and is denoted by $E_\lambda$. The set of all eigenvalues of $\boldsymbol{A}$ is called the eigenspectrum, or just spectrum spectrum, of $\boldsymbol{A}$.
- If $\lambda$ is an eigenvalue of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, then the corresponding eigenspace $E_\lambda$ is the solution space of the homogeneous system of linear equations $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \mathbf{0}$
- **Example** (The Case of the Identity Matrix)
- The identity matrix $\boldsymbol{I} \in \mathbb{R}^{n \times n}$ has characteristic polynomial $p_{\boldsymbol{I}}(\lambda) = \det(\boldsymbol{I} - \lambda \boldsymbol{I}) = (1 - \lambda)^n = 0$. It has only one eigenvalue $\lambda = 1$ that occurs $n$ times.
- Moreover, $\boldsymbol{I}\boldsymbol{x} = \lambda \boldsymbol{x}$ holds for all vectors $\boldsymbol{x} \in \mathbb{R}^n \backslash \{\mathbf{0}\}$
- Therefore, the sole eigenspace $E_1$ of the identity matrix spans $n$ dimensions, and all $n$ standard basis vectors of $\mathbb{R}^n$ are eigenvectors of $\boldsymbol{I}$.

- Useful properties regarding eigenvalues and eigenvectors
- A matrix $\boldsymbol{A}$ and its transpose $\boldsymbol{A}^T$ possess the same eigenvalues, but not necessarily the same eigenvectors
- The eigenspace $E_\lambda$ is the null space of $\boldsymbol{A} - \lambda \boldsymbol{I}$ since

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \Leftrightarrow \boldsymbol{A}\boldsymbol{x} - \lambda \boldsymbol{x} = \boldsymbol{0}$$
$$\Leftrightarrow (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \Leftrightarrow \boldsymbol{x} \in \ker(\boldsymbol{A} - \lambda \boldsymbol{I})$$

- Symmetric, positive definite matrices always have positive, real eigenvalues.

$$\forall \boldsymbol{x} \in V \backslash \{\boldsymbol{0}\}: \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} > 0$$

- **Example (Computing Eigenvalues, Eigenvectors, and Eigenspaces)**
- Let us find the eigenvalues and eigenvectors of the $2 \times 2$ matrix

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$$

- **Step 1: Characteristic Polynomial**. We need to compute the roots of the characteristic polynomial $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$ to find the eigenvalues.

- **Step 2: Eigenvalues.** The characteristic polynomial is

$$p_A(\lambda) = \det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \det\left(\begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) = \begin{vmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{vmatrix} = (4-\lambda)(3-\lambda) - 2 \cdot 1$$

- We factorize the characteristic polynomial and obtain
- $p_A(\lambda) = (4-\lambda)(3-\lambda) - 2 \cdot 1 = 10 - 7\lambda + \lambda^2 = (2-\lambda)(5-\lambda)$

  giving the roots $\lambda_1 = 2$ and $\lambda_2 = 5$.
- **Step 3: Eigenvectors and Eigenspaces.** From our definition of the eigenvector $\boldsymbol{x} \neq \boldsymbol{0}$, there will be a vector such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, i.e., $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. We find the eigenvectors that correspond to these eigenvalues by looking at vectors $\boldsymbol{x}$ such that

$$\begin{bmatrix} 4-\lambda & 2 \\ 1 & 3-\lambda \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

- For $\lambda = 5$ we obtain

$$\begin{bmatrix} 4-5 & 2 \\ 1 & 3-5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \boldsymbol{0}$$

- We solve this homogeneous system and obtain a solution space

$$E_5 = \text{span}[\begin{bmatrix} 2 \\ 1 \end{bmatrix}]$$

- This eigenspace is one-dimensional as it possesses a single basis vector.
- Analogously, we find the eigenvector for $\lambda = 2$ by solving

$$\begin{bmatrix} 4-2 & 2 \\ 1 & 3-2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

- The corresponding eigenspace is given as

$$E_2 = \text{span}[\begin{bmatrix} 1 \\ -1 \end{bmatrix}]$$

- **Definition**. Let $\lambda_i$ be an eigenvalue of a square matrix $\boldsymbol{A}$ . Then the geometric multiplicity of $\lambda_i$ is the number of linearly independent eigenvectors associated with $\lambda_i$. In other words, it is the dimensionality of the eigenspace spanned by the eigenvectors associated with $\lambda_i$.
- In our previous example, the geometric multiplicity of $\lambda = 5$ and $\lambda = 2$ is $1$.
- In another example, the matrix $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has two repeated eigenvalues $\lambda_1 = \lambda_2 = 2$. The algebraic multiplicity of $\lambda_1$ and $\lambda_2$ is $2$.
- The eigenvalue has only one distinct unit eigenvector $\boldsymbol{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and thus geometric multiplicity is $1$.
- **Theorem**. The eigenvectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ with $n$ distinct eigenvalues $\lambda_1, \dots, \lambda_n$ are linearly independent.
- Eigenvectors of a matrix with $n$ distinct eigenvalues form a basis of $\mathbb{R}^n$

- **Definition**. A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is defective if it possesses fewer than $n$ linearly independent eigenvectors
- Looking at the eigenspaces of a defective matrix, it follows that the sum of the dimensions of the eigenspaces is less than $n$.
- A defective matrix cannot have $n$ distinct eigenvalues, as distinct eigenvalues have linearly independent eigenvectors.
- **Theorem**. Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, we can always obtain a symmetric, positive semidefinite matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ by defining

$$\boldsymbol{S} := \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$$

- *Proof*. Symmetry: $\boldsymbol{S} := \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}^{\mathrm{T}}\right)^{\mathrm{T}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{\mathrm{T}} = \boldsymbol{S}^{\mathrm{T}}$
- positive semidefinite: $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{S}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{0}$

- If $\mathrm{rk}(\boldsymbol{A}) = n$, then $\boldsymbol{S} := \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is positive definite.

- **Theorem** (Spectral Theorem). If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, there exists an orthonormal basis of the corresponding vector space $V$ consisting of eigenvectors of $A$, and each eigenvalue is real

- **Example**

- Consider the matrix

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$$

- The characteristic polynomial of $\boldsymbol{A}$ is

$$p_A(\lambda) = (\lambda - 1)^2(\lambda - 7)$$

so that we obtain the eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 7$, where $\lambda_1$ is a repeated eigenvalue. Following our standard procedure for computing eigenvectors, we obtain the eigenspaces

$$E_1 = \text{span}[\underbrace{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}}_{=:\boldsymbol{x}_1}, \underbrace{\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}}_{=:\boldsymbol{x}_2}], E_7 = \text{span}[\underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{=:\boldsymbol{x}_3}]$$

- We see that $\boldsymbol{x}_3$ is orthogonal to both $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. However, since $\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 = 1 \neq 0$, they are not orthogonal. The spectral theorem states that there exists an orthogonal basis, but the one we have is not orthogonal.

- However, we can construct one.

- To construct such a basis, we exploit the fact that $\boldsymbol{x_1}, \boldsymbol{x_2}$ are eigenvectors associated with the same eigenvalue $\lambda$. Therefore, for any $\alpha, \beta \in \mathbb{R}$ it holds that

$$\boldsymbol{A}(\alpha\boldsymbol{x_1} + \beta\boldsymbol{x_2}) = \boldsymbol{A}\boldsymbol{x_1}\alpha + \boldsymbol{A}\boldsymbol{x_2}\beta = \lambda_1(\alpha\boldsymbol{x_1} + \beta\boldsymbol{x_2})$$

- i.e., any linear combination of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ is also an eigenvector of $\boldsymbol{A}$ associated with $\lambda_1$. The Gram-Schmidt algorithm is a method for iteratively constructing an orthogonal/orthonormal basis from a set of basis vectors using such linear combinations.

- Therefore, even if $\boldsymbol{x_1}$and $\boldsymbol{x_2}$ are not orthogonal, we can apply the Gram-Schmidt algorithm and find eigenvectors associated with $\lambda_1 = 1$ that are orthogonal to each other (and to $\boldsymbol{x_3}$). In our example, we will obtain

$$\boldsymbol{x_1'} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x_2'} = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

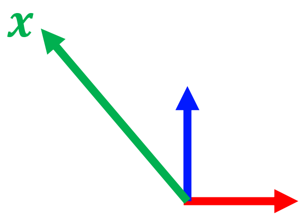
- which are orthogonal to each other, orthogonal to $\boldsymbol{x_3}$, and eigenvectors of $\boldsymbol{A}$ associated with $\lambda_1 = 1$.

- **Theorem.** The determinant of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the product of its eigenvalues,

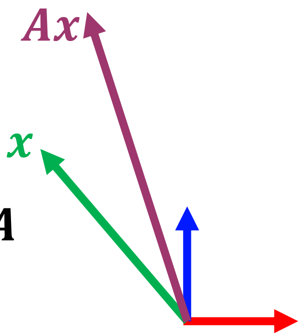
$$\det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i$$

where $\lambda_i$ are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

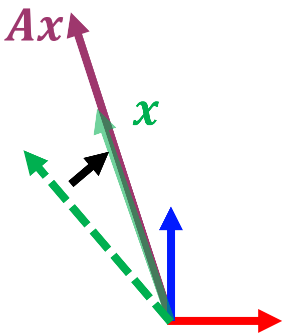
# Some understandings




Left multiplied by a matrix $\boldsymbol{A}$

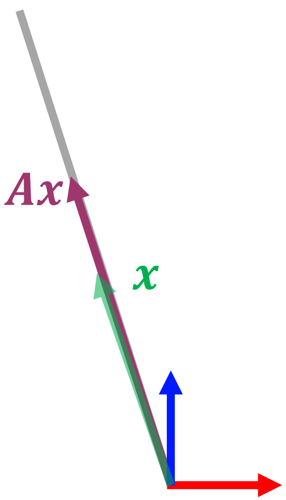



Everything is normal




Now, $\boldsymbol{x}$ and $\boldsymbol{Ax}$ are of the same line. The length of $\boldsymbol{Ax}$ is greater than $\boldsymbol{x}$

$$\boldsymbol{Ax} = \lambda \boldsymbol{x}$$




The grey line is the eigenspace of $\boldsymbol{A}$ with respect to $\lambda$
Every vector on this grey line is an eigenvector of $\boldsymbol{A}$, and they all correspond to the eigenvalue $\lambda$

# 4.4 Eigendecomposition and Diagonalization

特征分解与对角化

- A diagonal matrix is a matrix that has value zero on all off-diagonal elements,

$$\boldsymbol{D} = \begin{bmatrix} c_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & c_n \end{bmatrix}$$

- Diagonal matrices allow fast computation of determinants, powers, and inverses.
- The determinant is the product of its diagonal entries.
- a matrix power $\boldsymbol{D}^k$ is given by each diagonal element raised to the power $k$.
- The inverse $\boldsymbol{D}^{-1}$ is the reciprocal of its diagonal elements if all of them are nonzero.

- Two matrices $\boldsymbol{A}$, $\boldsymbol{D}$ are similar if there exists an invertible matrix $\boldsymbol{P}$, such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

- **Definition**. A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable if it is similar to a diagonal matrix, i.e., if there exists an invertible matrix $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ such that $\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$.

- Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, and let $\lambda_1, \dots, \lambda_n$ be a set of scalars, and let $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$ be a set of vectors in $\mathbb{R}^n$. We define $\boldsymbol{P} := [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n]$ and let $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ be a diagonal matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then we can show that

$$\boldsymbol{AP} = \boldsymbol{PD}$$

$D = P^{-1}AP$

if and only if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$ are corresponding eigenvectors of $\boldsymbol{A}$.

- We can see that this statement holds because

$$\boldsymbol{AP} = \boldsymbol{A}[\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n] = [\boldsymbol{Ap}_1, \cdots, \boldsymbol{Ap}_n]$$

$$\boxed{\boldsymbol{PD} = [\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n] \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = [\lambda_1 \boldsymbol{p}_1, \cdots, \lambda_n \boldsymbol{p}_n]}$$

- Thus $\boldsymbol{AP} = \boldsymbol{PD}$ implies that

$$\boxed{\begin{aligned} \boldsymbol{Ap}_1 &= \lambda_1 \boldsymbol{p}_1 \\ &\vdots \\ \boldsymbol{Ap}_n &= \lambda_n \boldsymbol{p}_n \end{aligned}}$$

- Therefore, the columns of $\boldsymbol{P}$ must be eigenvectors of $\boldsymbol{A}$.
- Our definition of diagonalization requires that $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is invertible, i.e., $\boldsymbol{P}$ has full rank. This requires us to have $n$ linearly independent eigenvectors $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_n$, i.e., the $\boldsymbol{p}_i$ form a basis of $\mathbb{R}^n$.

- **Theorem** (Eigendecomposition)**.**
- A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ can be factored into

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$$

where $\boldsymbol{P} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{A}$, if and only if the eigenvectors of $\boldsymbol{A}$ form a basis of $\mathbb{R}^n$

- **Theorem**. A symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ can always be diagonalized.

- **Theorem** (Spectral Theorem). If $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is symmetric, there exists an orthonormal basis of the corresponding vector space $V$ consisting of eigenvectors of $\boldsymbol{A}$, and each eigenvalue is real.

- The spectral theorem states that we can find an ONB of eigenvectors of $\mathbb{R}^n$. This makes $\boldsymbol{P}$ an orthogonal matrix ($\boldsymbol{P}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{I}$) so that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}}$ or equivalently $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$

# Example

- Let us compute the eigendecomposition of $\boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}}$
- **Step 1: Compute eigenvalues and eigenvectors.** The characteristic polynomial of $\boldsymbol{A}$ is

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \det\left(\begin{bmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{bmatrix}\right) = (2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda - 3)(\lambda - 1)$$

- Therefore, the eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 1$ and $\lambda_2 = 3$, and the associated (normalized) eigenvectors are obtained via

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \boldsymbol{p}_1 = 1\boldsymbol{p}_1, \qquad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \boldsymbol{p}_2 = 3\boldsymbol{p}_2$$

- This yields

$$\boldsymbol{p}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \boldsymbol{p}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- **Step 2: Check for existence**. The eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2$ form a basis of $\mathbb{R}^2$. Therefore, $\boldsymbol{A}$ can be diagonalized.
- **Step 3: Construct the matrix $\boldsymbol{P}$ to diagonalize $\boldsymbol{A}$**. We collect the eigenvectors of $\boldsymbol{A}$ in $\boldsymbol{P}$ so that

$$\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2] = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

- We then obtain

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} = \boldsymbol{D}$$

- Equivalently, we get

$$\underbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}_{\boldsymbol{A}} = \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{\boldsymbol{P}} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}}_{\boldsymbol{D}} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{\boldsymbol{P}^{\mathrm{T}}}$$

- Diagonal matrices $\boldsymbol{D}$ can efficiently be raised to a power. Therefore, we can find a matrix power for a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ via the eigenvalue decomposition (if it exists) so that

$$\boldsymbol{A}^k = (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1})^k = \boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1}$$

- Computing $\boldsymbol{D}^k$ is efficient because we apply this operation individually to any diagonal element.

- Assume that the eigendecomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ exists. Then,

$$\det(\boldsymbol{A}) = \det(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}) = \det(\boldsymbol{P})\det(\boldsymbol{D})\det(\boldsymbol{P}^{-1})$$
$$= \det(\boldsymbol{D}) = \prod_i d_{ii}$$

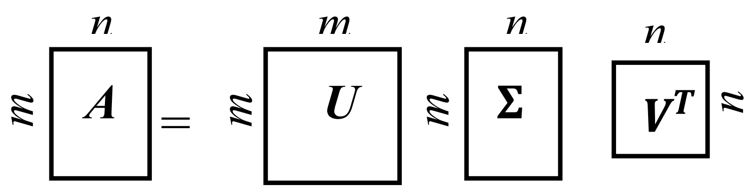
allows for an efficient computation of the determinant of $\boldsymbol{A}$.

- Eigendecomposition requires square matrices.
- We introduce a more general matrix decomposition technique, the singular value decomposition.

# 4.5 Singular Value Decomposition SVD

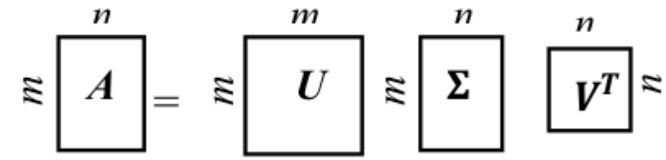
Can take any shape of inputs

- **Theorem** (SVD Theorem). Let $A^{m\times n}$ be a rectangular matrix of rank $\mathrm{r} \in [0, \min(\mathrm{m}, \mathrm{n})]$ .The SVD of $A$ is a decomposition of the form




with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ with column vectors $\boldsymbol{u}_i, i = 1,\ldots,m$, and an orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ with column vectors $\boldsymbol{v}_j, j = 1,\ldots,n$. Moreover, $\boldsymbol{\Sigma}$ is an $m\times n$ matrix with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij} = 0, i \neq j$ .

- The diagonal entries $\sigma_i, i = 1, \ldots, r$ of $\boldsymbol{\Sigma}$ are called the singular values
- $\boldsymbol{u}_i$ are called the left-singular vectors
- $\boldsymbol{v}_j$ are called the right-singular vectors
- By convention, the singular values are ordered $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \geq 0$




- The singular value matrix $\boldsymbol{\Sigma}$ is unique.
- $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times n}$ is rectangular and of the same size as $\boldsymbol{A}$. This means that $\boldsymbol{\Sigma}$ has a diagonal submatrix that contains the singular values and needs additional zero padding.
- If $m > n$, $\boldsymbol{\Sigma}$ has diagonal structure up to row $n$ and consists of $\mathbf{0}^{\mathrm{T}}$ row vectors from $n+1$ to $m$,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ \hline 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

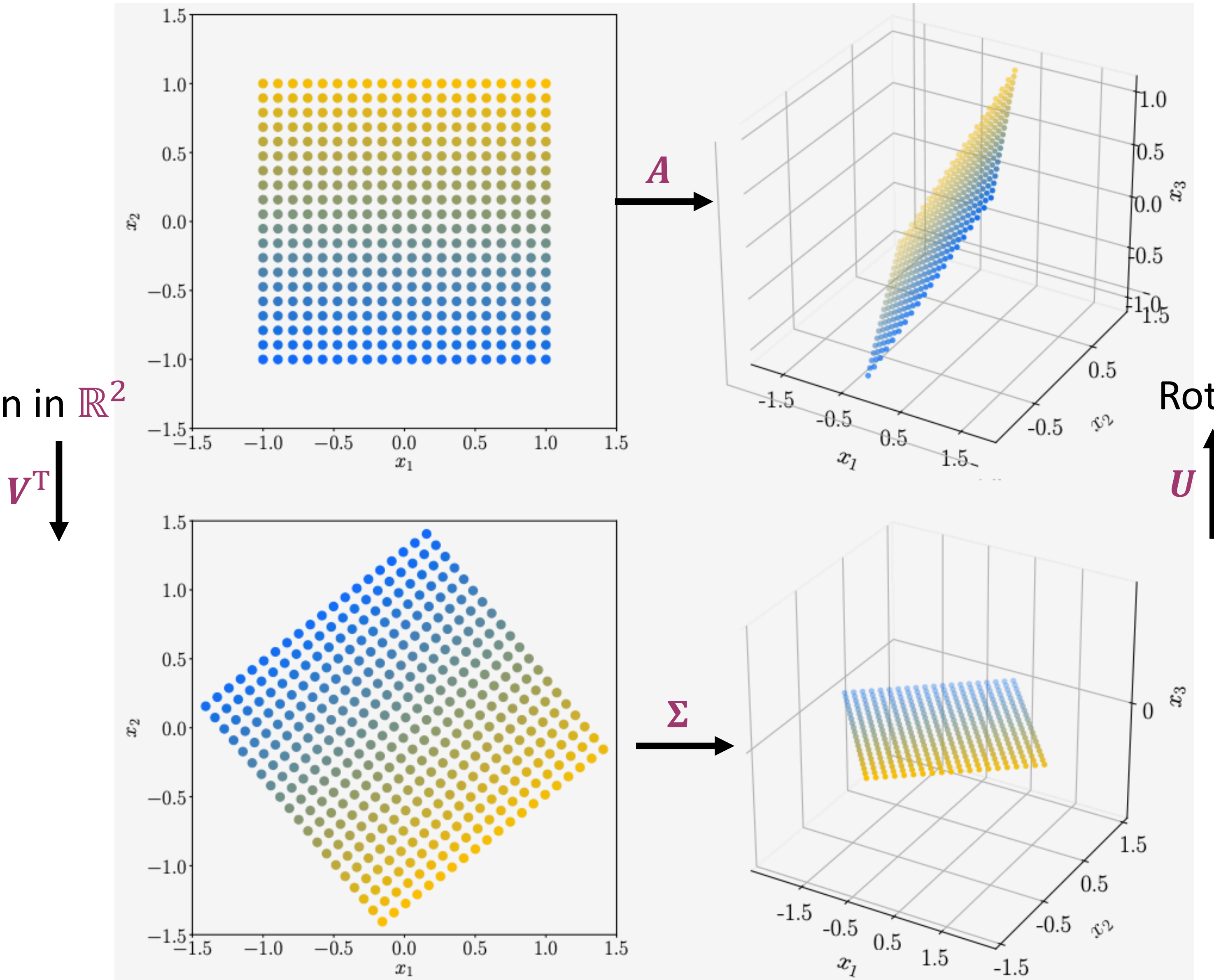
- If $m < n$, $\boldsymbol{\Sigma}$ has a diagonal structure up to column $m$ and columns that consist of $\mathbf{0}$ from $m+1$ to $n$:

$$\boldsymbol{\Sigma} = \left[\begin{array}{ccc|ccc} \sigma_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \cdots & 0 \end{array}\right]$$

- The SVD exists for any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ .
- **Example 4.12 (Vectors and the SVD)**
- Consider a mapping of a square grid of vectors $\mathcal{X} \in \mathbb{R}^2$ that fit in a box of size $2\times 2$ entered at the origin. Using the standard basis, we map these vectors using

$$\boldsymbol{A} = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}$$

- $\boldsymbol{A} = \begin{bmatrix} 1 & -0.8 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = \begin{bmatrix} -0.79 & 0 & -0.62 \\ 0.38 & -0.78 & -0.49 \\ -0.48 & -0.62 & 0.62 \end{bmatrix} \begin{bmatrix} 1.62 & 0 \\ 0 & 1.0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -0.78 & 0.62 \\ -0.62 & -0.78 \end{bmatrix}$




Rotation in $\mathbb{R}^2$

Rotation in $\mathbb{R}^3$

all vectors lie in the $x_1 - x_2$ plane.

# 4.5.2 Construction of the SVD

- The SVD of a general matrix shares some similarities with the eigendecomposition of a square matrix
- Compare the eigendecomposition of an SPD (Symmetric, positive definite) matrix

$$\boldsymbol{S} = \boldsymbol{S}^{\mathrm{T}} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}}$$

with the corresponding SVD

$$\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

- If we set

$$\boldsymbol{U} = \boldsymbol{P} = \boldsymbol{V}, \qquad \boldsymbol{D} = \boldsymbol{\Sigma}$$

we see that the SVD of SPD matrices is their eigendecomposition.

Given any matrix A:

- We can always diagonalize $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and obtain

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}\begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}\boldsymbol{P}^{\mathrm{T}} \quad (1)$$

where $\boldsymbol{P}$ is an orthogonal matrix, which is composed of the orthonormal eigenbasis. The $\lambda_i \geq 0$ are the eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

- Let us assume the SVD of $\boldsymbol{A}$ exists and takes the form of $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$

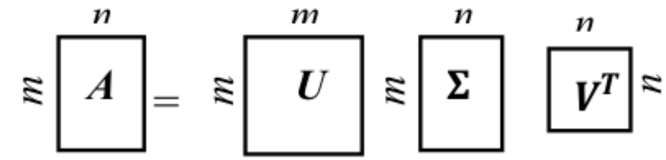
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\right)^{\mathrm{T}}\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}\right) = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$$

where $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal matrices. Therefore, with $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I}$ we obtain

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Sigma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{bmatrix}\boldsymbol{V}^{\mathrm{T}} \quad (\mathbf{2})$$

- Comparing now (1) and (2), we identify

$$\boxed{\begin{aligned} \boldsymbol{V} &= \boldsymbol{P} \\ \sigma_i^2 &= \lambda_i \end{aligned}}$$




- To obtain the left-singular vectors $\boldsymbol{U}$.

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} \Leftrightarrow \boldsymbol{AV} = \boldsymbol{U\Sigma V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{U\Sigma}$$

- We have,

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i, i = 1, \dots, r$$

where $r$ is the rank of $\boldsymbol{A}$. So, we can calculate

$$\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i,\ i = 1, \dots, r \quad (1)$$

- We look at matrices with full rank, i.e., $r = \min(m, n)$. Remember that $\boldsymbol{U}$ is an $m \times m$ matrix.
- If $m \leq n$, $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m]$; All the $\boldsymbol{u}_i$ have been calculated through (1)
- If $m > n$, $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n, \dots, \boldsymbol{u}_m]$;
  - $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ have been calculate through (1)
  - In order to calculate $\boldsymbol{u}_{n+1}, \dots, \boldsymbol{u}_m$, you use the fact that $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n, \dots, \boldsymbol{u}_m$ are orthonormal vectors.

# Summary of the SVD

- Given $\in \mathbb{R}^{m\times n}$, $\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$
- $\boldsymbol{V}$: eigendecomposition of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ =PDP^T
- $\boldsymbol{\Sigma}$: nonzero elements are $\sigma_i$ obtained from eigendecomposition of $\boldsymbol{A}^T\boldsymbol{A}$
- $\boldsymbol{U}$: calculate $\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i$
  - If $m \le n$, $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m]$;
  - If $m > n$, $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_n, \dots, \boldsymbol{u}_m]$;
    - For $i > n$, the $\boldsymbol{u}_i$ are orthonormal vectors that satisfy

$$[\boldsymbol{u}_1^{\mathrm{T}}, \boldsymbol{u}_2^{\mathrm{T}}, \dots, \boldsymbol{u}_n^{\mathrm{T}}]\boldsymbol{u}_i = \mathbf{0}$$

# 4.5.2 Construction of the SVD

- **Example (Computing the SVD)**
- Let us find the singular value decomposition of

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}$$

- The SVD requires us to compute the right-singular vectors $\boldsymbol{v}_j$, the singular values $\sigma_k$, and the left-singular vectors $\boldsymbol{u}_i$.
- **Step 1: Right-singular vectors as the eigenbasis of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ .**
- We start by computing

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

- We compute the singular values $\sigma_k$ and right-singular vectors $\boldsymbol{v}_j$ through the eigenvalue decomposition of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, which is given as

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}}$$

and we obtain the right-singular vectors as the columns of $\boldsymbol{P}$ so that

$$V = P = \begin{bmatrix} \frac{5}{\sqrt{30}} & 0 & \frac{-1}{\sqrt{6}} \\ \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{30}} & \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

- **Step 2: Singular-value matrix.**
- As the singular values $\sigma_i$ are the square roots of the eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ we obtain them straight from $\begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. Since $\mathrm{rk}(\boldsymbol{A}) = 2$, there are only two non-zero singular values: $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$. The singular value matrix must be the same size as $\boldsymbol{A}$, and we obtain

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

- **Step 3: Right-singular vectors are calculated using $\boldsymbol{u}_i = \frac{1}{\sigma_i}\boldsymbol{A}\boldsymbol{v}_i$**

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\boldsymbol{v}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{\sqrt{30}} \\ \frac{-2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{bmatrix}$$

$$\boldsymbol{u}_2 = \frac{1}{\sigma_2}\boldsymbol{A}\boldsymbol{v}_2 = \frac{1}{\sqrt{1}}\begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix}$$

- We obtain the left-singular vectors as the columns of $\boldsymbol{S}$ so that

$$\boldsymbol{U} = \boldsymbol{S} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

- Now we have computed $\boldsymbol{U}$, $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$.
- You can verify that
- $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ -2 & 1 & 0 \end{bmatrix} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{6} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} \frac{5}{\sqrt{30}} & \frac{-2}{\sqrt{30}} & \frac{1}{\sqrt{30}} \\ 0 & \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$

# Another example

- Calculate the SVD of $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} = \boldsymbol{U\Sigma V}^{\mathrm{T}}$
- We first calculate $\boldsymbol{V}$ as the eigenbasis of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$\boldsymbol{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

- The singular value matrix is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

- We finally calculate $\boldsymbol{U}$

$$\boldsymbol{u}_1 = \frac{1}{\sigma_1}\boldsymbol{A}\boldsymbol{v}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

$$\boldsymbol{u}_2 = \frac{1}{\sigma_2}\boldsymbol{A}\boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

- Now we have calculated $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$; we want to calculate $\boldsymbol{u}_3$
- We make use of the fact that $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, and $\boldsymbol{u}_3$ are an orthonormal basis.

$$\begin{cases} \boldsymbol{u}_1^{\mathrm{T}}\boldsymbol{u}_3 = 0 \\ \boldsymbol{u}_2^{\mathrm{T}}\boldsymbol{u}_3 = 0 \\ \|\boldsymbol{u}_3\|_2 = 1 \end{cases},$$

- We can obtain
- $\boldsymbol{u}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$
- In all, the SVD of $\boldsymbol{A}$ is written as

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{-1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

# 4.5.3 Eigenvalue Decomposition vs. Singular Value Decomposition

- Let us consider the eigendecomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ and the SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{T}}$.
- The SVD always exists for any matrix $\mathbb{R}^{m\times n}$. The eigendecomposition is only defined for square matrices $\mathbb{R}^{n\times n}$ and only exists if we can find a basis of eigenvectors of $\mathbb{R}^{n}$
- The vectors in the eigendecomposition matrix $\boldsymbol{P}$ are not necessarily orthogonal. On the other hand, the vectors in the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ in the SVD are orthonormal, so they represent rotations.
- Both the eigendecomposition and the SVD are compositions of three linear mappings:
  1. Change of basis in the domain
  2. Independent scaling of each new basis vector and mapping from domain to codomain
  3. Change of basis in the codomain

A key difference between the eigendecomposition and the SVD is that in the SVD, domain and codomain can be vector spaces of different dimensions

# 4.5.3 Eigenvalue Decomposition vs. Singular Value Decomposition

- In the SVD, the left- and right-singular vector matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are generally not inverse of each other (they perform basis changes in different vector spaces). In the eigendecomposition, the basis change matrices $\boldsymbol{P}$ and $\boldsymbol{P}^{-1}$ are inverses of each other.
- In the SVD, the entries in the diagonal matrix $\boldsymbol{\Sigma}$ are all real and nonnegative, which is not generally true for the diagonal matrix in the eigendecomposition.
- The SVD and the eigendecomposition are closely related through their projections
  - The right-singular vectors of $\boldsymbol{A}$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.
  - The nonzero singular values of $\boldsymbol{A}$ are the square roots of the nonzero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.
- For symmetric matrices $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, the eigenvalue decomposition and the SVD are one and the same, which follows from the spectral theorem.

# Further understanding

$PDP^T$

- A symmetric matrix represents a combination of rotation and scaling
- Through matrix decomposition, we can explain the effect of linear transformation defined by this matrix.
- Eigenvalues quantify the scaling effect.
- Eigenvectors quantify the direction of the scaling $P$
- The application of eigendecomposition is limited.
- SVD is a universal one by finding an orthonormal basis.