

Applied Linear Algebra in Data Analysis

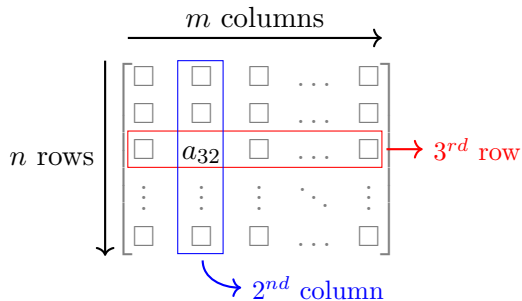
Matrices

Sivakumar Balasubramanian

Department of Bioengineering
Christian Medical College, Bagayam
Vellore 632002

Matrices

- **Matrices** are rectangular array of numbers. $\begin{bmatrix} 1.1 & -24 & \sqrt{2} \\ 0 & 1.12 & -5.24 \end{bmatrix}$



- Consider a matrix **A** with n rows and m columns. $\begin{cases} \text{Tall/Skinny} & n > m \\ \text{Square} & n = m \\ \text{Wide/Fat} & n < m \end{cases}$

Matrices

- ▶ n -vectors can be interpreted as $n \times 1$ matrices. These are called *column vectors*.
- ▶ A matrix with only one row is called a *row vector*, which can be referred to as n -row-vector. $\mathbf{x} = [1.45 \quad -3.1 \quad 12.4]$
- ▶ **Block matrices & Submatrices:** $\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix}$. What are the dimensions of the different matrices?

Matrices

- ▶ Matrices are also compact way to give a set of indexed column n -vectors, $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \dots \mathbf{x}_m$.

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \mathbf{x}_3 \quad \dots \quad \mathbf{x}_m]$$

- ▶ **Zero matrix** $= \mathbf{0}_{n \times m} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$

- ▶ **Identity matrix** is a square $n \times n$ matrix with all zero elements, except the diagonals where all elements are 1.

$$i_{mn} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3]$$

Matrices

- **Diagonal matrices** is a square matrix with non-zero elements on its diagonal.

$$\begin{bmatrix} 0.4 & 0 & 0 & 0 \\ 0 & -11 & 0 & 0 \\ 0 & 0 & 21 & 0 \\ 0 & 0 & 0 & 9.3 \end{bmatrix} = \mathbf{diag}(0.4, -11, 21, 9.3)$$

- **Triangular matrices:** Are square matrices. *Upper triangular* $a_{ij} = 0, \forall i > j$;
Lower triangular $a_{ij} = 0, \forall i < j$.

Matrix operations: Transpose

- **Transpose** switches the rows and columns of a matrix. \mathbf{A} is a $n \times m$ matrix, then its transpose is represented by \mathbf{A}^\top , which is a $m \times n$ matrix.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \implies \mathbf{A}^\top = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{13} & a_{23} \end{bmatrix}$$

Transpose converts between column and row vectors.

What is the transpose of a block matrix? $\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} \end{bmatrix}$

Matrix operations: Matrix Addition

- ▶ **Matrix addition** can only be carried out with matrices of same size. Like vectors we perform element wise addition.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

- ▶ **Properties of matrix addition:**

- ▶ *Commutative:* $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- ▶ *Associative:* $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- ▶ *Addition with zero matrix:* $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$
- ▶ *Transpose of sum:* $(\mathbf{A} + \mathbf{B})^{\top} = \mathbf{A}^{\top} + \mathbf{B}^{\top}$

Matrix operations: Scalar multiplication

- **Scalar multiplication** Each element of the matrix gets multiplied by the scalar.

$$\alpha \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} \\ \alpha a_{21} & \alpha a_{22} \end{bmatrix}$$

- We will mostly only deal with matrices with real entries. Such matrices are elements of the set $\mathbb{R}^{n \times m}$.
- Given the aforementioned matrix operations and their properties, is $\mathbb{R}^{n \times m}$ a vector space?

Matrix operations: Matrix multiplication

- ▶ A useful multiplication operation can be defined for matrices.
- ▶ It is possible to *multiply* two matrices $\mathbf{A} \in \mathbb{R}^{n \times p}$ and $\mathbf{B} \in \mathbb{R}^{p \times m}$ through this *matrix multiplication* procedure.
- ▶ The product matrix $\mathbf{C} := \mathbf{AB} \in \mathbb{R}^{n \times m}$, if the number of columns of \mathbf{A} is equal to the number of rows of \mathbf{B} .

$$c_{ij} := \sum_{k=1}^p a_{ik} b_{kj} \quad \forall i \in \{1, \dots, n\}, j \in \{1 \dots m\}$$

Matrix multiplication

- *Inner product* is a special case of matrix multiplication between a *row vector* and a *column vector*.

$$\mathbf{x}^\top \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^\top \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n x_i y_i$$

Matrix multiplication: Post-multiplication by a column vector

- Consider a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ and a m -vector $\mathbf{x} \in \mathbb{R}^m$. We can multiply \mathbf{A} and \mathbf{x} to obtain $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^n$.

$$\mathbf{y} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m a_{1i}x_i \\ \sum_{i=1}^m a_{2i}x_i \\ \vdots \\ \sum_{i=1}^m a_{ni}x_i \end{bmatrix} = \sum_{i=1}^m x_i \begin{bmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{bmatrix} = \sum_{i=1}^m x_i \mathbf{a}_i$$

- Post-multiplying a matrix \mathbf{A} by a column vector \mathbf{x} results in a linear combination of the columns of matrix \mathbf{A} .
- \mathbf{x} provides the column mixture.

Matrix multiplication: Pre-multiplication by a row vector

- ▶ Let $\mathbf{x}^\top \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n \times m}$, then $\mathbf{y} = \mathbf{x}^\top \mathbf{A}$.

$$\mathbf{y} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \dots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nm} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i a_{i1} & \dots & \sum_{i=1}^n x_i a_{im} \end{bmatrix} = \sum_{i=1}^n x_i \tilde{\mathbf{a}}_i^\top$$

where, $\tilde{\mathbf{a}}_i^\top = \begin{bmatrix} a_{i1} & \dots & a_{im} \end{bmatrix}$

- ▶ Pre-multiplying a matrix \mathbf{A} by a row vector \mathbf{x} results in a linear combination of the rows of \mathbf{A} .
- ▶ \mathbf{x}^\top provides the row mixture.

Matrix multiplication

- Multiplying two matrices $\mathbf{A} \in \mathbb{R}^{n \times p}$ and $\mathbf{B} \in \mathbb{R}^{p \times m}$ produces $\mathbf{C} \in \mathbb{R}^{n \times m}$,

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pm} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{n2} & \dots & c_{nm} \end{bmatrix}$$

- **Four interpretations of matrix multiplication.**

1. Inner-Product interpretation
2. Column interpretation
3. Row interpretation
4. Outer product interpretation.

Matrix multiplication: Inner-product Interpretation

$$\mathbf{C} = \mathbf{AB}, \quad \mathbf{A} \in \mathbb{R}^{n \times p}, \mathbf{B} \in \mathbb{R}^{p \times m}, \mathbf{C} \in \mathbb{R}^{n \times m}$$

- ij^{th} element of \mathbf{C} is the inner product of the i^{th} row of \mathbf{A} and the j^{th} column of \mathbf{B} .

$$c_{ij} = \sum_{k=1}^p a_{ik}b_{kj} = \tilde{\mathbf{a}}_i^\top \mathbf{b}_j$$

where, $i \in \{1 \dots n\}, j \in \{1 \dots m\}$

Matrix multiplication: Column interpretation

$$\mathbf{C} = \mathbf{AB}, \quad \mathbf{A} \in \mathbb{R}^{n \times p}, \mathbf{B} \in \mathbb{R}^{p \times m}, \mathbf{C} \in \mathbb{R}^{n \times m}$$

- Columns of \mathbf{C} are the linear combinations of the columns of \mathbf{A} .

$$\mathbf{C} = \mathbf{A} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{b}_1 & \mathbf{A}\mathbf{b}_2 & \dots & \mathbf{A}\mathbf{b}_m \end{bmatrix}$$

- j^{th} column of \mathbf{C} is the linear combination of the columns of \mathbf{A}

$$\mathbf{c}_j = \sum_{k=1}^p b_{kj} \mathbf{a}_k$$

Matrix multiplication: Row interpretation

$$\mathbf{C} = \mathbf{AB}, \quad \mathbf{A} \in \mathbb{R}^{n \times p}, \quad \mathbf{B} \in \mathbb{R}^{p \times m}, \quad \mathbf{C} \in \mathbb{R}^{n \times m}$$

- Rows of \mathbf{C} are the linear combinations of the rows of \mathbf{B} .

$$\mathbf{C} = \begin{bmatrix} \tilde{\mathbf{a}}_1^\top \\ \tilde{\mathbf{a}}_2^\top \\ \vdots \\ \tilde{\mathbf{a}}_n^\top \end{bmatrix} \mathbf{B} = \begin{bmatrix} \tilde{\mathbf{a}}_1^\top \mathbf{B} \\ \tilde{\mathbf{a}}_2^\top \mathbf{B} \\ \vdots \\ \tilde{\mathbf{a}}_n^\top \mathbf{B} \end{bmatrix}$$

- i^{th} row of \mathbf{C} is the linear combination of the rows of \mathbf{B}

$$\tilde{\mathbf{c}}_i^\top = \sum_{k=1}^p a_{ik} \tilde{\mathbf{b}}_k^\top$$

Matrix multiplication: Outer product interpretation

- **Outer product:** Product between a column vector and a row vector. Let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. The *outer product* is defined as,

$$\mathbf{xy}^\top = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \dots & y_m \end{bmatrix} = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_m \\ x_2y_1 & x_2y_2 & \dots & x_2y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \dots & x_ny_m \end{bmatrix} \in \mathbb{R}^{n \times m}$$

Matrix multiplication: Outer product interpretation

$$\mathbf{C} = \mathbf{AB}, \quad \mathbf{A} \in \mathbb{R}^{n \times p}, \mathbf{B} \in \mathbb{R}^{p \times m}, \mathbf{C} \in \mathbb{R}^{n \times m}$$

- \mathbf{C} can be written as the sum of p outer products of columns of \mathbf{A} and rows of \mathbf{B} .

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \dots & \mathbf{a}_p \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{b}}_1^\top \\ \tilde{\mathbf{b}}_2^\top \\ \tilde{\mathbf{b}}_3^\top \\ \vdots \\ \tilde{\mathbf{b}}_p^\top \end{bmatrix} = \sum_{i=1}^p \mathbf{a}_i \tilde{\mathbf{b}}_i^\top$$

Properties of matrix multiplication

- ▶ **Not commutative:** $\mathbf{AB} \neq \mathbf{BA}$

The product of two matrices might not always be defined. When it is defined, \mathbf{AB} and \mathbf{BA} need not match.

- ▶ **Distributive:** $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ and $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$

- ▶ **Associative:** $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

- ▶ **Transpose:** $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$

- ▶ **Scalar product:** $\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B})$

Linear transformations

- ▶ Linear functions $f : \mathbb{R}^m \mapsto \mathbb{R}$,

$$y = f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}; \quad \mathbf{w}, \mathbf{x} \in \mathbb{R}^m, \quad y \in \mathbb{R}$$

- ▶ Generalization of the linear function is when its range \mathbb{R}^n :

$$\mathbf{y} = f(\mathbf{x}); \quad \mathbf{x} \in \mathbb{R}^m, \quad \mathbf{y} \in \mathbb{R}^n$$

- ▶ These can be represented as, $\mathbf{y} = \mathbf{A}\mathbf{x}$, $\mathbf{A} \in \mathbb{R}^{n \times m}$.

- ▶ Matrices can be thought of as representing a particular linear transformation.

Why does matrix multiplication have this strange definition?

Consider the following two functions,

$$\mathbf{y} = f(\mathbf{x}) = \mathbf{A}\mathbf{x} \longrightarrow \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{v} = g(\mathbf{u}) = \mathbf{B}\mathbf{u} \longrightarrow \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = g\left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) = \begin{bmatrix} \alpha u_1 + \beta u_2 \\ \gamma u_1 + \delta u_2 \end{bmatrix} = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{aligned} \mathbf{z} = h(\mathbf{u}) = f(g(\mathbf{u})) &= f\left(\begin{bmatrix} \alpha u_1 + \beta u_2 \\ \gamma u_1 + \delta u_2 \end{bmatrix}\right) = \begin{bmatrix} a\alpha u_1 + a\beta u_2 + b\gamma u_1 + b\delta u_2 \\ c\alpha u_1 + c\beta u_2 + d\gamma u_1 + d\delta u_2 \end{bmatrix} \\ &= \begin{bmatrix} (a\alpha + b\gamma)u_1 + (a\beta + b\delta)u_2 \\ (c\alpha + d\gamma)u_1 + (c\beta + d\delta)u_2 \end{bmatrix} = \begin{bmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \end{aligned}$$

$$\mathbf{z} = \mathbf{A}(\mathbf{B}\mathbf{u}) = (\mathbf{A}\mathbf{B})\mathbf{u} \implies \mathbf{A}\mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} a\alpha + b\gamma & a\beta + b\delta \\ c\alpha + d\gamma & c\beta + d\delta \end{bmatrix}$$

Matrix multiplication represents the composition of linear transformations.

Rank of a matrix \mathbf{A}

- **Rank of a matrix \mathbf{A} :** dimension of the subspace spanned by the columns of \mathbf{A} or the rows of $\mathbf{A} \in \mathbb{R}^{n \times m}$.

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \dim \text{span}(\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}) \rightarrow \text{Column rank} \\ &= \dim \text{span}\left(\left\{\tilde{\mathbf{a}}_1^\top, \tilde{\mathbf{a}}_2^\top, \dots, \tilde{\mathbf{a}}_n^\top\right\}\right) \rightarrow \text{Row rank} \end{aligned}$$

- Column Rank is always equal to the row rank.
- Rank tells us the number of independent columns/rows in the matrix.
- **Full rank matrix \mathbf{A} :** $\text{rank}(\mathbf{A}) = \min(n, m)$
Rank deficient matrix \mathbf{A} : $\text{rank}(\mathbf{A}) < \min(n, m)$

Matrix Inverse

- ▶ Consider the square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$. $\mathbf{B} \in \mathbb{R}^{n \times n}$ is the inverse of \mathbf{A} , if $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$, and \mathbf{B} is represented as \mathbf{A}^{-1} .
- ▶ Not all matrices have inverses. A matrix with an inverse is called **non-singular**, otherwise it is called **singular**.
- ▶ For a non-singular matrix \mathbf{A} , \mathbf{A}^{-1} is unique. \mathbf{A}^{-1} is both the left and right inverse.
- ▶ A matrix \mathbf{A} has an inverse, if and only if \mathbf{A} is full rank, i.e. $\text{rank}(\mathbf{A}) = n$
- ▶ $\mathbf{Ax} = \mathbf{b}$ can be solved as follows, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. *It is never solved like this in practice.*
- ▶ Inverse of product of matrices, $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- ▶ $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$ and $(\mathbf{A}^{-1})^\top = (\mathbf{A}^\top)^{-1}$

Complex Vectors and Matrices

- ▶ Similar to \mathbb{R}^n , we can have \mathbb{C}^n . $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{r1} + jx_{i1} \\ x_{r2} + jx_{i2} \\ \vdots \\ x_{rn} + jx_{in} \end{bmatrix}$
- ▶ Vector addition and scalar multiplication are the same. The scalar is a complex number.
- ▶ Additive identity, and scalar multiplication identity are the same. So is the **standard basis** $\{\mathbf{e}_i\}_{i=1}^n$
- ▶ **Linear independence:** The set $\{\mathbf{v}_i\}_{i=1}^n$ with $\mathbf{v}_i \in \mathbb{C}^n$ is linearly independent, if $\sum_{i=1}^n c_i \mathbf{v}_i = 0, \implies c_i = 0, \forall 1 \leq i \leq n, c_i \in \mathbb{C}$
- ▶ **Inner product:** $\mathbf{x}^* \mathbf{y} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^n \bar{x}_i y_i$

Complex Vectors and Matrices

- ▶ **Length:** $\|\mathbf{x}\|_2^2 = \mathbf{x}^* \mathbf{x} = \sum_{i=1}^n \bar{x}_i x_i = \sum_{i=1}^n |x_i|^2$
- ▶ **Orthogonality:** $\mathbf{x}^* \mathbf{y} = 0$
- ▶ Complex matrices have complex entries. $\mathbf{A} \in \mathbb{C}^{m \times n}$ such that $a_{ij} \in \mathbb{C}, \forall 1 \leq i \leq m, 1 \leq j \leq n$
- ▶ The transpose operation is generalized to conjugate transpose known as the *Hermitian*. $\mathbf{A}^* = \overline{\mathbf{A}}^\top$.
- ▶ The idea of symmetric matrices $\mathbb{R}^{n \times n}$ are now generalized to $\mathbb{C}^{n \times n}$ as $\mathbf{A} = \mathbf{A}^*$. Such matrices are called **Hermitian** matrices.
- ▶ Orthogonal matrices in the complex case are called **Unitary** matrices, $\mathbf{U}^* \mathbf{U} = \mathbf{I} \implies \mathbf{U}^{-1} = \mathbf{U}^*$.