MEC529: Introduction to Robotics (Theory and Applications)

(Fall 2024)

Amin Fakhari, Ph.D.

Department of Mechanical Engineering

Stony Brook University

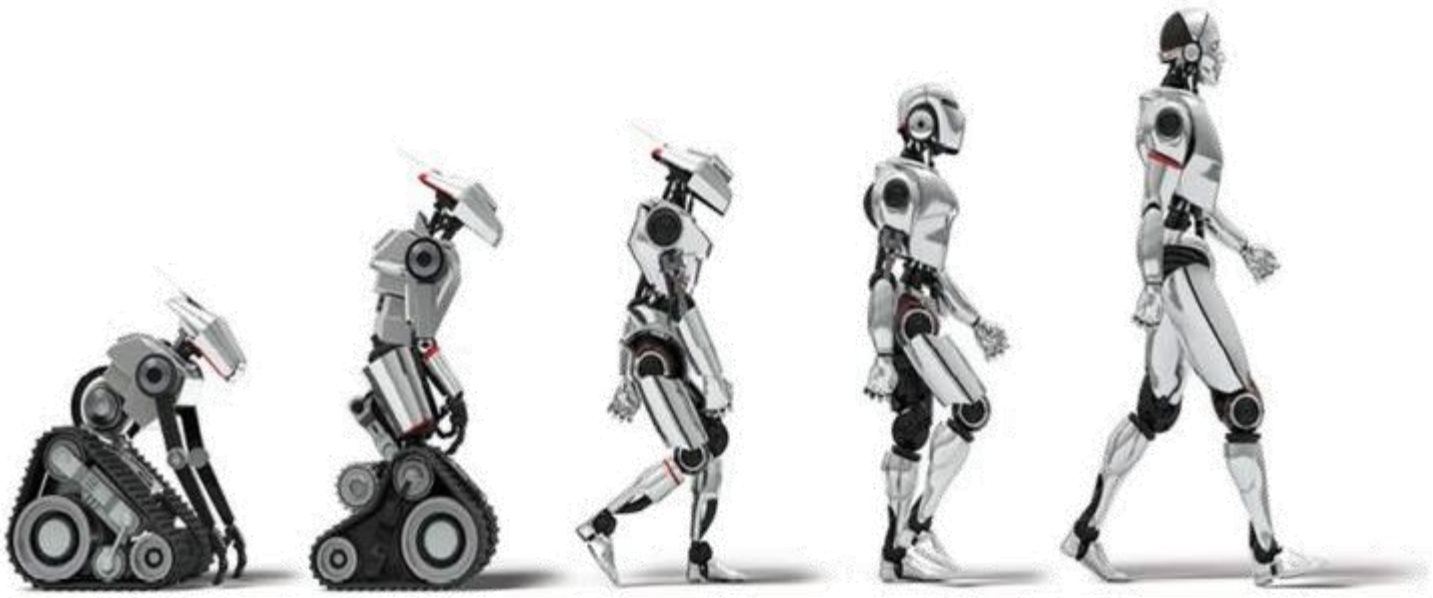
Ch1: A Review of Linear Algebra

Introduction

Robotics

Robotics is an interdisciplinary field which integrates computer science, mechanical engineering, electrical engineering, information engineering, bioengineering, computer engineering, control engineering, software engineering, mathematics, etc.

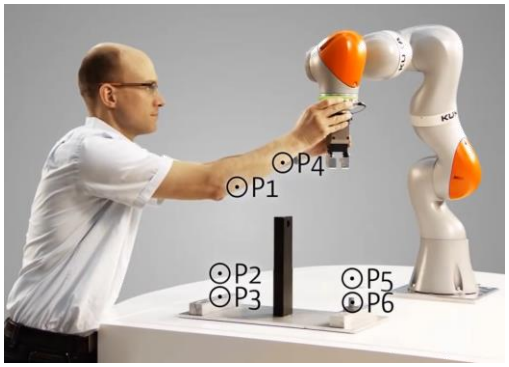
The goal of robotics is to design machines that can assist humans or replicate human actions.

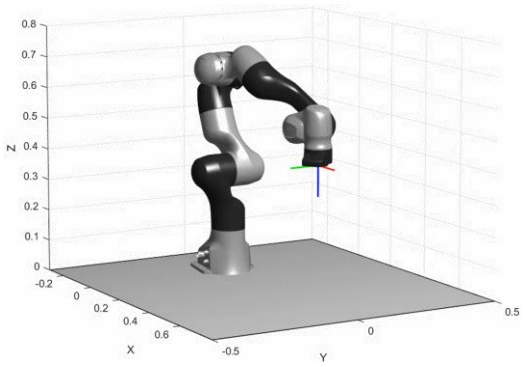


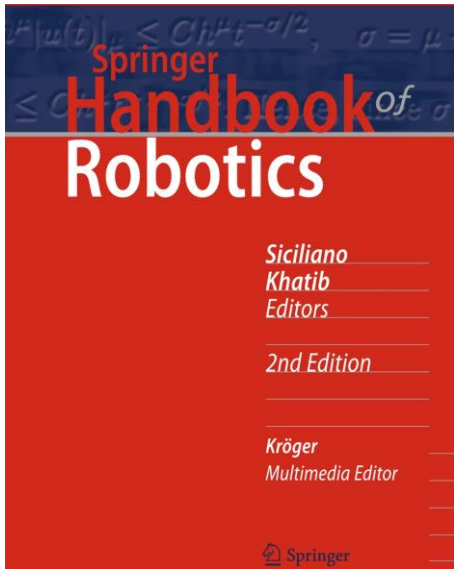
Robotics

- Classifications of robots based on structure:

- **Fixed-Base Robots** (e.g., Serial or Open-Chain Manipulators, Parallel Manipulators)
- **Mobile Robots**
  - **Ground Robots** (e.g., Wheeled Robots, Legged Robots)
  - **Submarine Robots**
  - **Aerial Robots**







https://link.springer.com/book/10.1007/978-3-319-32552-1

Vectors and Gradient

Basic Notation

$\forall$

$\exists$

$\in$

$\Longrightarrow$

$\Leftrightarrow$

$:=$

$\dot{x}$

$f: \mathcal{D} \to \mathcal{R}$

$\mathbb{R}$

$\mathbb{R}_+, \mathbb{R}_{++}$

$|x|$

Coordinate-Free Vector and Point

A **coordinate-free vector** v is a geometric quantity with a length and a direction.

Given a reference frame, vector v can be moved to a position such that the base of the arrow is at the origin without changing the orientation. Then, the vector v can be represented by its coordinates in the reference frame and denoted as $\boldsymbol{v}$.

- v refers to a physical quantity in the underlying space.
- $\boldsymbol{v}$ is a representation of v that depends on the choice of coordinate frame.

- A **point** p denotes a point in the physical space.
- A point p can be represented by as a **vector** from frame origin to p.
- $\boldsymbol{p}$ denotes the coordinate of a point p , which depends on the choice of reference frame.

Vector

$\boldsymbol{x} \in \mathbb{R}^n$: (an n-dimensional real vector in the column format)

$\mathbb{R}^n$: n-dimensional real space
(Euclidian Space)

$\boldsymbol{x}^T$:

Vector Norm

General Definition: Given $\boldsymbol{x} \in \mathbb{R}^n$, vector norm $\|\boldsymbol{x}\| \in \mathbb{R}_+$ is defined such that

- $\|\boldsymbol{x}\| > 0$ when $\boldsymbol{x} \neq \boldsymbol{0}$ and $\|\boldsymbol{x}\| = 0$ iff $\boldsymbol{x} = \boldsymbol{0}$.
- $\|k\boldsymbol{x}\| = |k|\|\boldsymbol{x}\|, \ \forall k \in \mathbb{R}$.
- $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|, \ \forall \boldsymbol{y} \in \mathbb{R}^n$.

❖ The p-norm (or ℓ_p-norm) of $\boldsymbol{x}$ for $p \in \mathbb{R}, p \geq 1$ is defined as $\|\boldsymbol{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$

e.g. $\|\boldsymbol{x}\|_2 = \|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}$ (Euclidean Norm)

Special case: $\|\boldsymbol{x}\|_\infty := \max_i |x_i|$

Schwartz Inequality: $|\boldsymbol{x}^T\boldsymbol{y}| \leq \|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

Unit Vector: $\hat{\boldsymbol{x}} = \boldsymbol{x}/\|\boldsymbol{x}\|_2 , \quad \|\hat{\boldsymbol{x}}\|_2 = \hat{\boldsymbol{x}}^T\hat{\boldsymbol{x}} = 1$

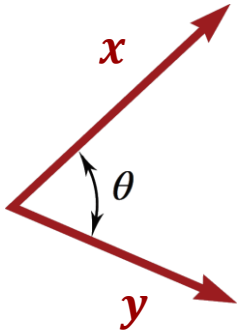
Dot Product or Scalar Product or Inner Product

Dot Product or Scalar Product or Inner Product of two vectors $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{y} \in \mathbb{R}^n$ is a <u>scalar</u> defined as

(Algebraic Definition) $$< \boldsymbol{x}, \boldsymbol{y} >= \boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{x}$$

(Geometric Definition) $$< \boldsymbol{x}, \boldsymbol{y} >= \boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \cos\theta$$

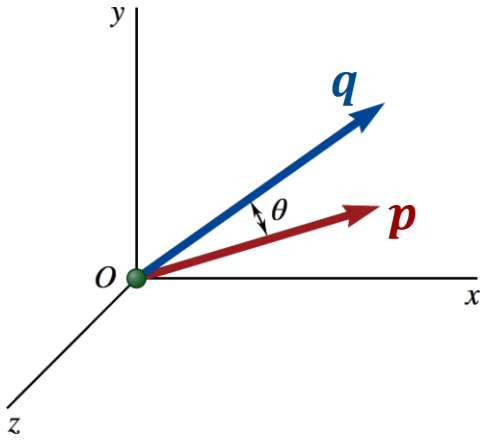
$$(0 \leq \theta \leq \pi)$$



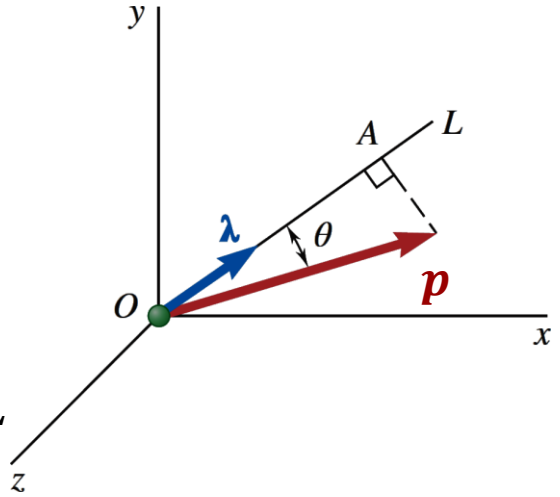
Orthogonal Vectors:

Applications of Dot Product

(1) Finding angle formed between two given vectors $\boldsymbol{p} \in \mathbb{R}^n$, $\boldsymbol{q} \in \mathbb{R}^n$ (or intersecting lines):



(2) Finding projection of a vector $\boldsymbol{p} \in \mathbb{R}^n$ on a given axis or directed line:

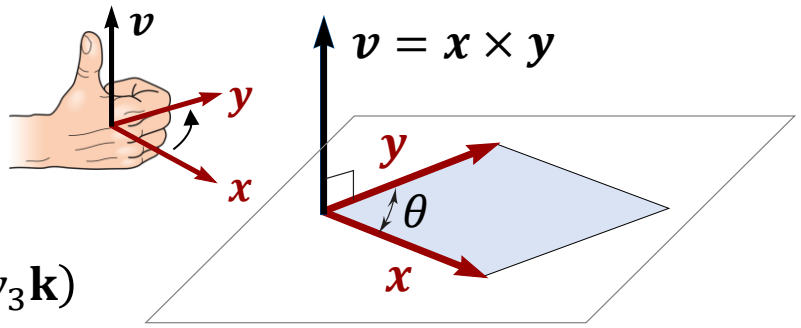


$\boldsymbol{\lambda}$: unit vector of line L

Cross Product or Vector Product

Cross product of $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$ (in the Euclidean space) is defined as a vector $\boldsymbol{v} = \boldsymbol{x} \times \boldsymbol{y} \in \mathbb{R}^3$ that is orthogonal to both $\boldsymbol{x}$ and $\boldsymbol{y}$ ($\boldsymbol{v} \perp \boldsymbol{x}$, $\boldsymbol{v} \perp \boldsymbol{y}$), with a direction given by the right-hand rule and a magnitude equal to the area of the parallelogram that the vectors span.

$$\|\boldsymbol{v}\|_2 = \|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2 \sin\theta \quad (0 \le \theta \le \pi)$$



$$\boldsymbol{v} = \boldsymbol{x} \times \boldsymbol{y} = (x_1\mathbf{i} + x_2\mathbf{j} + x_3\mathbf{k}) \times (y_1\mathbf{i} + y_2\mathbf{j} + y_3\mathbf{k})$$

Coordinate notation

$$\boldsymbol{x} = (x_1, x_2, x_3), \boldsymbol{y} = (y_1, y_2, y_3)$$

$$\boldsymbol{v} = \boldsymbol{x} \times \boldsymbol{y} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

Matrix notation

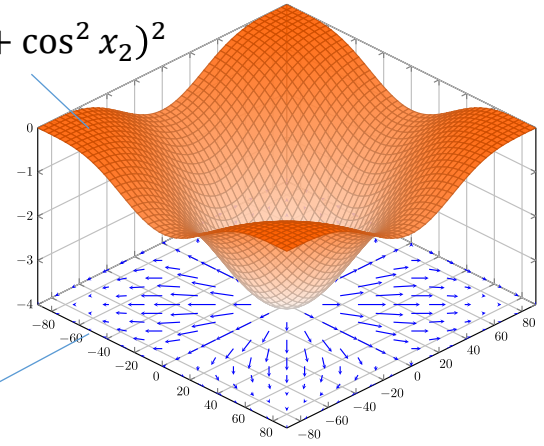
Gradient

For a **scalar function** $f: \mathbb{R}^n \to \mathbb{R}$ which is differentiable with respect to the elements x_i of $\boldsymbol{x} \in \mathbb{R}^n$, its **gradient** with respect to $\boldsymbol{x}$ is an n-dimensional column vector $\boldsymbol{\nabla}_{\boldsymbol{x}} f \in \mathbb{R}^n$ as:

(nabla symbol and pronounced "del")

$$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}) = \left(\frac{\partial f}{\partial \boldsymbol{x}}\right)^T = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}$$

$f(x_1, x_2) = -(\cos^2 x_1 + \cos^2 x_2)^2$



The gradient depicted as a projected vector field and shows the direction and rate at which a function increases the fastest at a given point.

- If $\boldsymbol{x}(t)$ is a differentiable function with respect to t:

$$\dot{f}(\boldsymbol{x}) = \frac{d}{dt} f\big(\boldsymbol{x}(t)\big) = \frac{\partial f}{\partial \boldsymbol{x}} \frac{d\boldsymbol{x}}{dt} = \frac{\partial f}{\partial \boldsymbol{x}} \dot{\boldsymbol{x}} = \boldsymbol{\nabla}_{\boldsymbol{x}}^T f(\boldsymbol{x}) \dot{\boldsymbol{x}} \qquad \text{(Chain Rule)}$$

Matrices and Jacobian

Matrix

$\boldsymbol{A} \in \mathbb{R}^{m \times n}$ (an m by n dimensional real matrix)

$\boldsymbol{A}^T \in \mathbb{R}^{n \times m}$

Tall, Wide (Fat), and Square Matrices:

Matrix-vector multiplication $\boldsymbol{Ax}$ as linear combination of columns of $\boldsymbol{A}$:

Particular Matrices

Square Matrix:

- Upper Triangular
- Lower Triangular
- Diagonal
    - Identity Matrix
- Null Matrix

Symmetric Matrix:

Skew-symmetric Matrix:

Partitioned Matrix: A matrix whose elements are matrices (blocks) of proper dimensions.

Matrix Operations

Trace of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$: $\mathrm{tr}(\boldsymbol{A})$

Sum of matrices: $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$

Symmetric and skew-symmetric part of a square matrix $\boldsymbol{A}$:

Product of matrices: $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$

Determinant of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$: $\det(\boldsymbol{A})$

Singular and Nonsingular Matrices:

k-order minors and k-order principal minors of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$:

Matrix Operations (cont.)

Linearly Independent Vectors $\boldsymbol{x}_i \in \mathbb{R}^m, i = 1, \dots, n$

Rank of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$: $\text{rank}(\boldsymbol{A})$

Inverse of a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$: $\boldsymbol{A}^{-1}$

Orthogonal Matrix:

Derivative of $\boldsymbol{A}(t) \in \mathbb{R}^{m \times n}$: $\dfrac{d}{dt}\boldsymbol{A}(t) = \dot{\boldsymbol{A}}(t)$

Derivative of $\boldsymbol{A}^{-1}(t) \in \mathbb{R}^{n \times n}$:

Cross Product as a Matrix-Vector Multiplication

Cross product $\boldsymbol{x} \times \boldsymbol{y}$ ($\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^3$) can be thought of as a multiplication of a vector by a 3×3 skew-symmetric matrix as

$$\boldsymbol{x} \times \boldsymbol{y} = \underbrace{\begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}}_{[\boldsymbol{x}]} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = [\boldsymbol{x}]\boldsymbol{y}$$

$$\boldsymbol{x} = (x_1, x_2, x_3),$$
$$\boldsymbol{y} = (y_1, y_2, y_3)$$

The matrix $[\boldsymbol{x}] \in \mathbb{R}^{3\times3}$ is a skew-symmetric matrix representation of $\boldsymbol{x}$. $[\boldsymbol{x}] = -[\boldsymbol{x}]^T$

- $[\boldsymbol{x}]\boldsymbol{y} = -[\boldsymbol{y}]\boldsymbol{x}$

Jacobian

For a **vector function** $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$ whose elements f_i are differentiable with respect to the elements x_i of $\boldsymbol{x} \in \mathbb{R}^n$, its **Jacobian** with respect to $\boldsymbol{x}$ is matrix $\boldsymbol{J_f} \in \mathbb{R}^{m \times n}$ as:

$$\boldsymbol{J_f}(\boldsymbol{x}) = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ \frac{\partial f_2(\boldsymbol{x})}{\partial \boldsymbol{x}} \\ \vdots \\ \frac{\partial f_m(\boldsymbol{x})}{\partial \boldsymbol{x}} \end{bmatrix}$$

- If $\boldsymbol{x}(t)$ is a differentiable function with respect to t:

$$\dot{\boldsymbol{f}}(\boldsymbol{x}) = \frac{d}{dt}\boldsymbol{f}\big(\boldsymbol{x}(t)\big) = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\frac{d\boldsymbol{x}}{dt} = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\dot{\boldsymbol{x}} = \boldsymbol{J_f}(\boldsymbol{x})\dot{\boldsymbol{x}} \qquad \text{(Chain Rule)}$$

Linear Transformation

Range/Column Space, Null Space, and Row Space

- The **linear transformation** (or linear map) between the vectors $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$ can be defined as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \qquad \boldsymbol{A} \in \mathbb{R}^{m \times n}$$

- The **range space** or **column space** (or **range** or **image**) of the transformation/matrix $\boldsymbol{A}$ is the subspace generated by the linearly independent columns of matrix $\boldsymbol{A}$.

$$\mathcal{R}(\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}) = \{\boldsymbol{y}: \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$$

Note: $\dim \mathcal{R}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$

- The **null space** (or **null**) of the transformation/matrix $\boldsymbol{A}$ is the subspace

$$\mathcal{N}(\boldsymbol{A}) = \{\boldsymbol{x}: \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{x} \in \mathbb{R}^n\} \subseteq \mathbb{R}^n$$

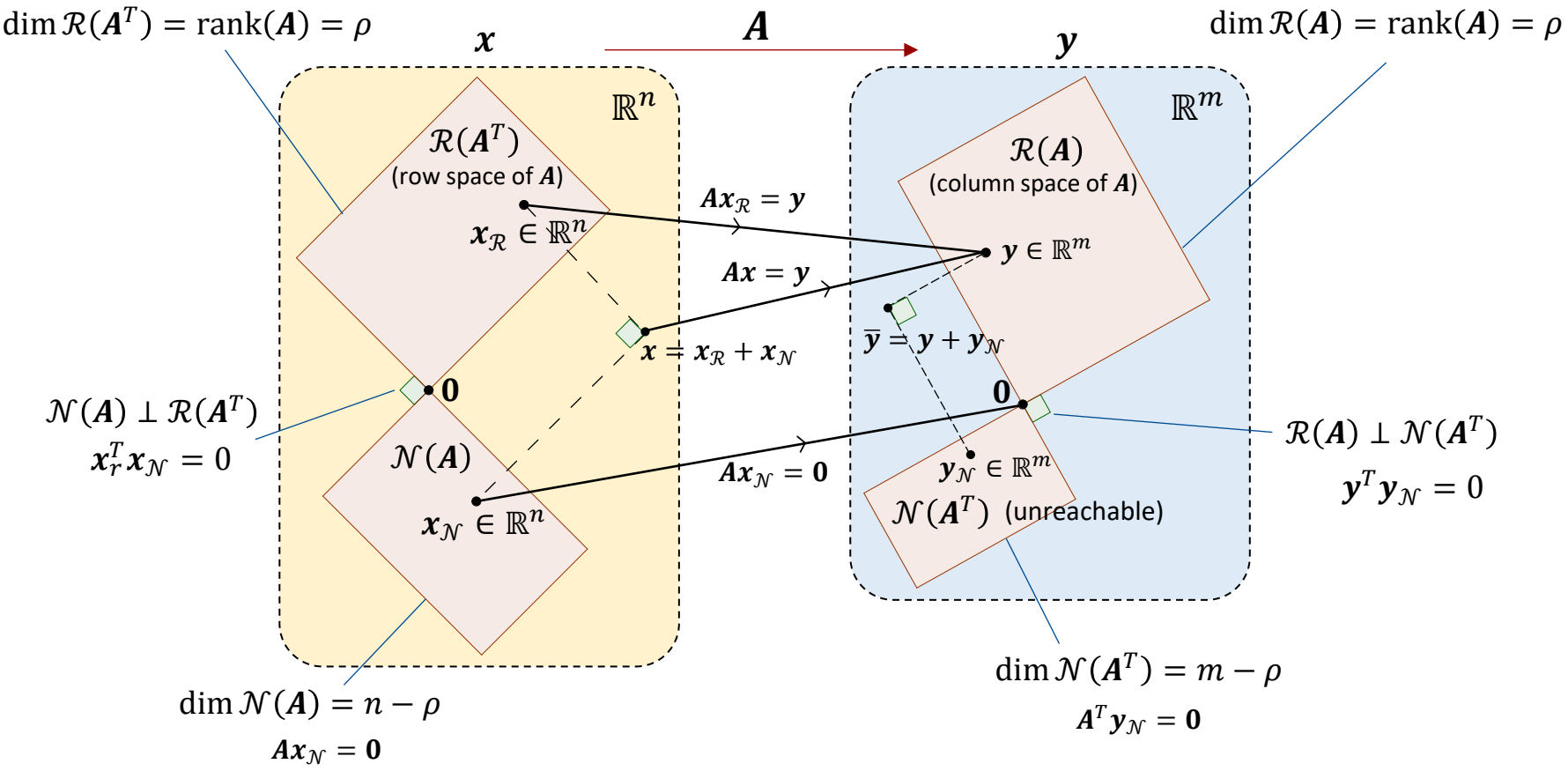
Note: $\dim \mathcal{R}(\boldsymbol{A}) + \dim \mathcal{N}(\boldsymbol{A}) = n$

Therefore, if $\text{rank}(\boldsymbol{A}) = \rho \leq \min\{m, n\}$, then $\dim \mathcal{R}(\boldsymbol{A}) = \rho$ and $\dim \mathcal{N}(\boldsymbol{A}) = n - \rho$.

- The **row space** of $\boldsymbol{A}$ is the same as the range/column space of $\boldsymbol{A}^T$, i.e., $\mathcal{R}(\boldsymbol{A}^T)$.

Range Space and Null Space of A and A^T

Consider the linear transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$.



Examples

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

$$\mathcal{R}(\boldsymbol{A}^T) = \text{Span}\left(\begin{pmatrix} -0.4287 \\ -0.5663 \\ -0.7039 \end{pmatrix}, \begin{pmatrix} 0.8060 \\ 0.1124 \\ -0.5812 \end{pmatrix}\right)$$

$$\mathcal{N}(\boldsymbol{A}) = \text{Span}\left(\begin{pmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{pmatrix}\right)$$

$$\dim \mathcal{N}(\boldsymbol{A}) = 1$$

$$\mathcal{R}(\boldsymbol{A}) = \text{Span}\left(\begin{pmatrix} -0.3863 \\ -0.9224 \end{pmatrix}, \begin{pmatrix} -0.9224 \\ 0.3863 \end{pmatrix}\right)$$

$$\dim \mathcal{R}(\boldsymbol{A}) = \dim \mathcal{R}(\boldsymbol{A}^T) = \text{rank}(\boldsymbol{A}) = 2$$

$$\mathcal{N}(\boldsymbol{A}^T) = \emptyset$$

$$\dim \mathcal{N}(\boldsymbol{A}^T) = 0$$

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

$$\mathcal{R}(\boldsymbol{A}^T) = \text{Span}\left(\begin{pmatrix} -0.6196 \\ -0.7849 \end{pmatrix}, \begin{pmatrix} -0.7849 \\ 0.6196 \end{pmatrix}\right)$$

$$\dim \mathcal{R}(\boldsymbol{A}) = \dim \mathcal{R}(\boldsymbol{A}^T) = \text{rank}(\boldsymbol{A}) = 2$$

$$\mathcal{N}(\boldsymbol{A}) = \emptyset$$

$$\dim \mathcal{N}(\boldsymbol{A}) = 0$$

$$\mathcal{R}(\boldsymbol{A}) = \text{Span}\left(\begin{pmatrix} -0.2298 \\ -0.5247 \\ -0.8196 \end{pmatrix}, \begin{pmatrix} 0.8835 \\ 0.2408 \\ -0.4019 \end{pmatrix}\right)$$

$$\mathcal{N}(\boldsymbol{A}^T) = \text{Span}\left(\begin{pmatrix} 0.4082 \\ -0.8165 \\ 0.4082 \end{pmatrix}\right)$$

$$\dim \mathcal{N}(\boldsymbol{A}^T) = 1$$

Examples

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 8 & 12 \end{pmatrix}$$

$$\mathcal{R}(\boldsymbol{A}^T) = \text{Span}\left(\begin{pmatrix} -0.2673 \\ -0.5345 \\ -0.8018 \end{pmatrix}\right)$$

$$\mathcal{N}(\boldsymbol{A}) = \text{Span}\left(\begin{pmatrix} 0.9636 \\ -0.1482 \\ -0.2224 \end{pmatrix}, \begin{pmatrix} 0 \\ -0.8321 \\ 0.5547 \end{pmatrix}\right)$$

$$\dim \mathcal{N}(\boldsymbol{A}) = 2$$

$$\mathcal{R}(\boldsymbol{A}) = \text{Span}\left(\begin{pmatrix} -0.2425 \\ -0.9701 \end{pmatrix}\right)$$

$$\dim \mathcal{R}(\boldsymbol{A}) = \dim \mathcal{R}(\boldsymbol{A}^T) = \text{rank}(\boldsymbol{A}) = 1$$

$$\mathcal{N}(\boldsymbol{A}^T) = \text{Span}\left(\begin{pmatrix} -0.9701 \\ 0.2425 \end{pmatrix}\right)$$

$$\dim \mathcal{N}(\boldsymbol{A}^T) = 1$$

Note: According to a geometric interpretation, the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ transforms the unit sphere in $\mathbb{R}^n$ defined by $\|\boldsymbol{x}\|_2 = 1$ into the set of vectors $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \in \mathbb{R}^m$ which define an ellipsoid of dimension $\rho = \text{rank}(\boldsymbol{A})$ in $\boldsymbol{y} \in \mathbb{R}^m$.

Pseudo-Inverse

Pseudo-Inverse

The inverse of a matrix $\boldsymbol{A}$ can be defined as $\boldsymbol{A}^{-1}$ only when the matrix is square and nonsingular (full rank). However, the inverse operation can be extended to the case of non-square (and singular) matrices as **Pseudo-Inverse** or **Moore–Penrose Inverse** $\boldsymbol{A}^{+}$.

For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, a pseudoinverse is defined as a matrix $\boldsymbol{A}^{+} \in \mathbb{R}^{n\times m}$ satisfying all of the following four criteria (Moore–Penrose conditions):

- $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{A}$
- $\boldsymbol{A}^{+}\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{A}^{+}$
- $(\boldsymbol{A}\boldsymbol{A}^{+})^{T} = \boldsymbol{A}\boldsymbol{A}^{+}$
- $(\boldsymbol{A}^{+}\boldsymbol{A})^{T} = \boldsymbol{A}^{+}\boldsymbol{A}$

Note: The pseudoinverse $\boldsymbol{A}^{+}$ exists for <u>any matrix</u> $\boldsymbol{A}$.

❖ If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is not full rank, i.e., $\mathrm{rank}(\boldsymbol{A}) < \min\{m, n\}$, a computationally simple and accurate way to compute pseudoinverse $\boldsymbol{A}^{+}$ is by using the **Singular Value Decomposition**.

❖ If $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is full rank, i.e., $\mathrm{rank}(\boldsymbol{A}) = \min\{m, n\}$, then $\boldsymbol{A}^{+} \in \mathbb{R}^{n\times m}$ can be given a particularly simple algebraic expression (see Special Cases in the next slides).

Special Case 1: Right Pseudo-Inverse

- If $m < n$ (fat matrix) and $\text{rank}(\boldsymbol{A}) = m$ (equivalently, $\boldsymbol{A}$ has linearly independent rows and $\boldsymbol{A}\boldsymbol{A}^T$ is invertible):

$$\boldsymbol{A}^+ = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1} \qquad \text{(right pseudo-inverse as } \boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}_m\text{)}$$

Weighted Right Pseudo-inverse:

If $\boldsymbol{W}_r \in \mathbb{R}^{n\times n}$ is a positive definite matrix[1], a weighted right pseudo-inverse is given by

$$\boldsymbol{A}^+ = \boldsymbol{W}_r^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{W}_r^{-1}\boldsymbol{A}^T)^{-1}$$

[1] A square not necessarily symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is Positive Definite (PD or $\boldsymbol{A} > 0$) if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for all nonzero $\boldsymbol{x} \in \mathbb{R}^n$.

Special Case 2: Left Pseudo-Inverse

- If $m > n$ (tall matrix) and $\text{rank}(\boldsymbol{A}) = n$ (equivalently, $\boldsymbol{A}$ has linearly independent columns and $\boldsymbol{A}^T\boldsymbol{A}$ is invertible):

$$\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T \qquad \text{(left pseudo-inverse as } \boldsymbol{A}^+\boldsymbol{A} = \boldsymbol{I}_n\text{)}$$

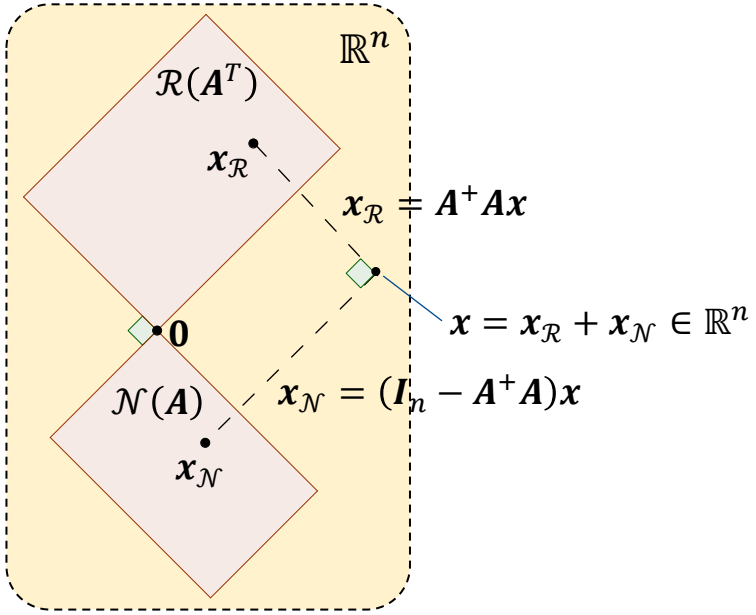
Weighted Left Pseudo-inverse:

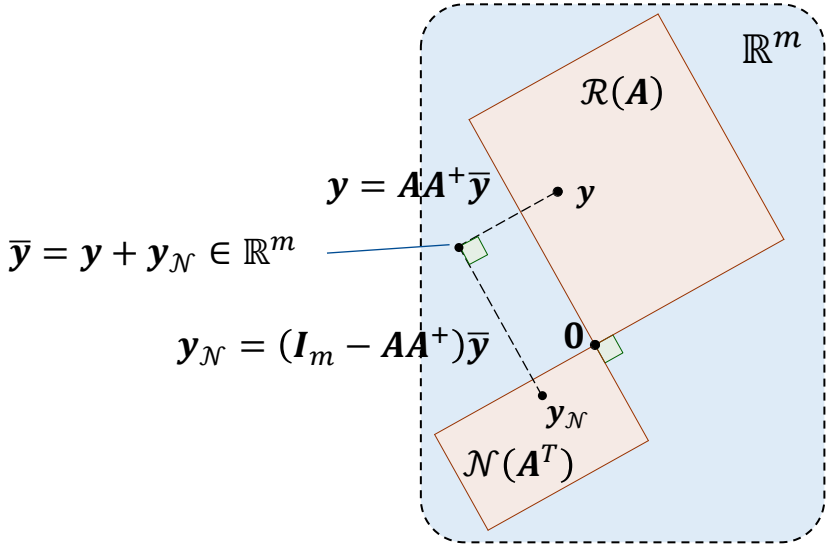
If $\boldsymbol{W}_l \in \mathbb{R}^{m\times m}$ is a positive definite matrix, a weighted right pseudo-inverse is given by

$$\boldsymbol{A}^+ = (\boldsymbol{A}^T\boldsymbol{W}_l\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{W}_l$$

Remarks

- If $\boldsymbol{A}$ is a square full-rank matrix, then $\boldsymbol{A}^{+} = \boldsymbol{A}^{-1}$.
- $\boldsymbol{A}^{+}\boldsymbol{A}$ projects any vector onto the row space of $\boldsymbol{A}$ or column space of $\boldsymbol{A}^{T}$, i.e., $\mathcal{R}(\boldsymbol{A}^{T})$.
- $\boldsymbol{I}_n - \boldsymbol{A}^{+}\boldsymbol{A}$ projects any vector onto the null space of $\boldsymbol{A}$, i.e., $\mathcal{N}(\boldsymbol{A})$.
- $\boldsymbol{A}\boldsymbol{A}^{+}$ projects any vector onto the column space of $\boldsymbol{A}$, i.e., $\mathcal{R}(\boldsymbol{A})$.
- $\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^{+}$ projects any vector onto the null space of $\boldsymbol{A}^{T}$, i.e., $\mathcal{N}(\boldsymbol{A}^{T})$.
- The pseudo-inverse is very useful to invert a linear transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.





Eigenvalue, Eigenvector, and SVD

Eigenvalues and Eigenvectors

If the vector resulting from the linear transformation $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ on a vector $\boldsymbol{u}$ has the same direction of $\boldsymbol{u}$ (with $\boldsymbol{u} \neq \boldsymbol{0}$), then $\boldsymbol{Au} = \lambda \boldsymbol{u}$.

For each square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ there exist n **eigenvalues** (in general, complex numbers) denoted by $\lambda_i(\boldsymbol{A})$, $i = 1, \dots, n$ that satisfy

$$\det(\boldsymbol{A} - \lambda_i(\boldsymbol{A})\boldsymbol{I}_n) = 0 \qquad \text{(characteristic equation)}$$

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

Eigenvectors $\boldsymbol{u}_i$ associated with the eigenvalues λ_i satisfy $(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{u}_i = \boldsymbol{0}$, $\quad i = 1, \dots, n$.

- If $\boldsymbol{A} = \boldsymbol{A}^T$, then $\lambda_i(\boldsymbol{A}) \in \mathbb{R}$, $i = 1, \dots, n$ (i.e., all eigenvalues are real) and the eigenvectors are always orthogonal, regardless of whether $\boldsymbol{A}$ is full rank or not.
- For any $\boldsymbol{A}$, if the eigenvectors $\boldsymbol{u}_i$ are linearly independent, matrix $\boldsymbol{U}$ formed by the column vectors $\boldsymbol{u}_i$ is invertible and $\boldsymbol{\Lambda} = \boldsymbol{U}^{-1}\boldsymbol{AU}$ where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. If $\boldsymbol{A}$ is symmetric, $\boldsymbol{U}$ is orthogonal ($\boldsymbol{UU}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$) and $\boldsymbol{\Lambda} = \boldsymbol{U}^T\boldsymbol{AU}$.

  $\Rightarrow$ Eigendecomposition: $\boldsymbol{A} = \boldsymbol{U\Lambda U}^{-1}$ and if $\boldsymbol{A}$ is symmetric $\boldsymbol{A} = \boldsymbol{U\Lambda U}^T$.

Singular Values and Condition Number

For a nonsquare matrix it is not possible to define eigenvalues and eigenvectors. An extension of the eigenvalue concept can be obtained by **Singular Values**.

For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, matrices $\boldsymbol{A}\boldsymbol{A}^T \in \mathbb{R}^{m\times m}$ and $\boldsymbol{A}^T\boldsymbol{A} \in \mathbb{R}^{n\times n}$ are symmetric and positive semi-definite (PSD), and, therefore, have (m for $\boldsymbol{A}\boldsymbol{A}^T$ and n for $\boldsymbol{A}^T\boldsymbol{A}$) real and nonnegative eigenvalues. Moreover, they have always the same non-zero eigenvalues. The **Singular Values** for matrix $\boldsymbol{A}$ are given by the square roots of the eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$ (if $m < n$) or $\boldsymbol{A}^T\boldsymbol{A}$ (if $n < m$):

$$\sigma_i = \sqrt{\lambda_i}\,, \qquad i = 1, \dots, \min(m, n)$$

Notes:

- The number of singular values σ_i is always $\min(m, n)$ and the number of non-zero singular values σ_i is equal to $\rho = \text{rank}(\boldsymbol{A})$ ($\sigma_1 \geq \cdots \geq \sigma_\rho > 0$).
- According to a geometric interpretation, the matrix $\boldsymbol{A}$ transforms the unit sphere in $\mathbb{R}^n$ defined by $\|\boldsymbol{x}\| = 1$ into the set of vectors $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ which define an ellipsoid of dimension $\rho = \text{rank}(\boldsymbol{A})$ in $\mathbb{R}^m$. The singular values σ_i are the lengths of the various axes of the ellipsoid.
- The **Condition Number** of matrix $\boldsymbol{A}$ is defined as $\kappa = \sigma_1/\sigma_\rho$ (the eccentricity of the ellipsoid).

Singular Value Decomposition (SVD)

The **Singular Value Decomposition** (**SVD**) of the matrix $\boldsymbol{A}$ is then given by $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$

- $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_m] \in \mathbb{R}^{m\times m}$ is an orthogonal matrix ($\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m$) that $\boldsymbol{u}_i$ (called left singular vectors) are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ ($\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{u}_i = \lambda_i\boldsymbol{u}_i = \sigma_i^2\boldsymbol{u}_i$).
- $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n] \in \mathbb{R}^{n\times n}$ is an orthogonal matrix ($\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n$) that $\boldsymbol{v}_i$ (called right singular vectors) are eigenvectors of $\boldsymbol{A}^T\boldsymbol{A}$ ($\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i = \sigma_i^2\boldsymbol{v}_i$).
- $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{m\times n}$ is a matrix where $\boldsymbol{D} = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_r\}$ and $\sigma_1 \geq \cdots \geq \sigma_\rho > 0$.

❖ SVD is used to compute the (right or left) **pseudoinverse** $\boldsymbol{A}^+$, even if $\boldsymbol{A}$ is not full rank.

$$\boldsymbol{A}^+ = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T$$

$$\boldsymbol{\Sigma}^+ = \begin{bmatrix} \boldsymbol{D}^+ & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{D}^+ = \text{diag}\left\{\frac{1}{\sigma_1}, \frac{1}{\sigma_2}, \dots, \frac{1}{\sigma_r}\right\}$$

Solving $\boldsymbol{Ax=y}$

Solving $\boldsymbol{Ax = y}$

Consider $\boldsymbol{Ax = y}$ (where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{y} \in \mathbb{R}^{m}$ are given and $\boldsymbol{x} \in \mathbb{R}^{n}$ is unknown).

- If $\boldsymbol{A}$ is invertible, i.e., square and full rank, (i.e., $\mathrm{rank}(\boldsymbol{A}) = n = m$), then there is a unique solution $\boldsymbol{x}$ as $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$.
- If $\boldsymbol{A}$ is not invertible, i.e., (1) $\boldsymbol{A}$ is not square ($n \neq m$) but full rank or (2) rank deficient $\mathrm{rank}(\boldsymbol{A}) < \min(m, n)$ (square or nonsquare), then $\boldsymbol{Ax = y}$ can still be solved (or approximately solved) for $\boldsymbol{x}$ with the Moore–Penrose pseudoinverse $\boldsymbol{A}^{+}$ as $\boldsymbol{x}^{*} = \boldsymbol{A}^{+}\boldsymbol{y}$ which minimizes the norm $\|\boldsymbol{Ax} - \boldsymbol{y}\|_2$.
- Exact solution(s) exist if and only if $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ or $\boldsymbol{AA}^{+}\boldsymbol{y} = \boldsymbol{y}$, otherwise, the solution will be approximate with minimum error.

Solving $\boldsymbol{Ax = y}$

(Case 1)

- If $m < n$ (fat matrix) and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is full (row) rank, i.e., $\text{rank}(\boldsymbol{A}) = m$ or $\mathcal{N}(\boldsymbol{A}^T) = \emptyset$ (this satisfies the solution existence condition $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$), infinite exact solutions exist. The solutions are in the form

$$\boldsymbol{x} = \underbrace{\boldsymbol{x}^*}_{\text{particular solution}} + \underbrace{\boldsymbol{P}\boldsymbol{x}_0}_{\text{homogeneous solution}} = \boldsymbol{A}^+\boldsymbol{y} + (\boldsymbol{I}_n - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}_0 \quad \text{for arbitrary vector } \boldsymbol{x}_0 \in \mathbb{R}^n \qquad (1)$$

  where $\boldsymbol{A}^+ = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$ and it is the right pseudo-inverse as $\boldsymbol{A}\boldsymbol{A}^+ = \boldsymbol{I}_m$.

- The solution (1) is derived from this constrained linear optimization problem:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{y} \end{aligned}$$

- Among all these solutions, $\boldsymbol{x}^* = \boldsymbol{A}^+\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A}^T)$ minimizes $\|\boldsymbol{x}\|_2$ (you can see this by setting $\boldsymbol{x}_0 = \boldsymbol{0}$ is the above optimization problem).
- Matrix $\boldsymbol{P} = (\boldsymbol{I}_n - \boldsymbol{A}^+\boldsymbol{A}) \in \mathbb{R}^{n \times n}$ is null-space projection matrix where $\mathcal{R}(\boldsymbol{P}) = \mathcal{N}(\boldsymbol{A})$.
- The term $\boldsymbol{x}_{\mathcal{N}} = (\boldsymbol{I}_n - \boldsymbol{A}^+\boldsymbol{A})\boldsymbol{x}_0$ is the projection of $\boldsymbol{x}_0$ in $\mathcal{N}(\boldsymbol{A})$ where $\boldsymbol{A}\boldsymbol{x}_{\mathcal{N}} = \boldsymbol{0}$.
- We can also use the weighted right pseudo-inverse as $\boldsymbol{A}^+ = \boldsymbol{W}_r^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{W}_r^{-1}\boldsymbol{A}^T)^{-1}$. $\boldsymbol{W}_r \in \mathbb{R}^{n \times n}$ is a positive definite matrix.

Solving $\boldsymbol{Ax = y}$

(Case 2)

- If $m > n$ (tall matrix) and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is full (column) rank, i.e., $\text{rank}(\boldsymbol{A}) = n$ or $\mathcal{N}(\boldsymbol{A}) = \emptyset$, when $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$, a unique exact solution exists, and when $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{A})$, no exact solutions but an approximate solutions exist. In both cases, the solutions are in the form

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y} \qquad (1)$$

  where $\boldsymbol{A}^{+} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ and it is the left pseudo-inverse as $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{I}_n$.

  - If $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ or $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{y} = \boldsymbol{y}$, then $\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y}$ gives the unique exact solution.
  - If $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{A})$ or $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{y} \neq \boldsymbol{y}$, then $\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y}$ minimizes the norm of the error $\|\boldsymbol{Ax} - \boldsymbol{y}\|_2$ and gives an approximate solution $\boldsymbol{x}^*$.

- The solution (1) is derived from this optimization problem: $\min_{\boldsymbol{x}} \frac{1}{2}(\boldsymbol{Ax} - \boldsymbol{y})^T(\boldsymbol{Ax} - \boldsymbol{y})$

- We can also use the weighted left pseudo-inverse as $\boldsymbol{A}^{+} = (\boldsymbol{A}^T\boldsymbol{W}_l\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{W}_l$. $\boldsymbol{W}_l \in \mathbb{R}^{m \times m}$ is a positive definite matrix.

Solving $\boldsymbol{Ax = y}$

(Case 3: General Case) - (Method 1)

- If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ (square or non-square) is rank deficient (i.e., $\text{rank}(\boldsymbol{A}) < \min(m, n)$, $\mathcal{N}(\boldsymbol{A}) \neq \emptyset$, and $\mathcal{N}(\boldsymbol{A}^T) \neq \emptyset$), when $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$, infinite exact solutions exist, and when $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{A})$, no exact solutions but infinite approximate solutions exist. In both cases, the solutions are in the form

$$\boldsymbol{x} = \boldsymbol{A}^{+}\boldsymbol{y} + (\boldsymbol{I}_n - \boldsymbol{A}^{+}\boldsymbol{A})\boldsymbol{x}_0 \quad \text{for arbitrary vector } \boldsymbol{x}_0 \in \mathbb{R}^n$$

  and $\boldsymbol{A}^{+}$ is pseudo-inverse which is computed using the **Singular Value Decomposition (SVD)**.

  - If $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ or $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{y} = \boldsymbol{y}$, exact solution $\boldsymbol{A}^{+}\boldsymbol{y}$ minimizes $\|\boldsymbol{x}\|_2$ and $(\boldsymbol{I}_n - \boldsymbol{A}^{+}\boldsymbol{A})\boldsymbol{x}_0$ represents all vectors in $\mathcal{N}(\boldsymbol{A})$.
  - If $\boldsymbol{y} \notin \mathcal{R}(\boldsymbol{A})$ or $\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{y} \neq \boldsymbol{y}$, approximate solution $\boldsymbol{A}^{+}\boldsymbol{y}$ minimizes both $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2$ and $\|\boldsymbol{x}\|_2$ while all solutions $\boldsymbol{A}^{+}\boldsymbol{y} + (\boldsymbol{I}_n - \boldsymbol{A}^{+}\boldsymbol{A})\boldsymbol{x}_0$ minimize only $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2$.

Solving $\boldsymbol{Ax = y}$

(Case 3: General Case) - (Method 2)

The **Damped Least Squares (DLS)** can be used to compute an approximate inverse or pseudoinverse of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, especially when $\boldsymbol{A}$ is singular or near-singular:

$$\boldsymbol{A}_D^+ \approx (\boldsymbol{A}^T\boldsymbol{A} + \lambda \boldsymbol{I}_n)^{-1}\boldsymbol{A}^T \quad \text{(right inverse)} \qquad (1)$$

$$\boldsymbol{A}_D^+ \approx \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T + \lambda \boldsymbol{I}_m)^{-1} \quad \text{(left inverse)} \qquad (2)$$

$\lambda \in \mathbb{R}_+$ is the **damping (regularization) parameter** which introduces a trade-off between accuracy of the solution and numerical stability. A larger λ improves stability for very ill-conditioned matrices but may reduce accuracy (i.e., away from true least squares solution). In general, choosing the optimal damping factor λ can be challenging.

- These two forms above (i.e., right and left inverses) are mathematically equivalent. For computational efficiency, use (1) when $n \leq m$ and use (2) when $m \leq n$ (which relates to which matrix $\boldsymbol{A}^T\boldsymbol{A}$ or $\boldsymbol{A}\,\boldsymbol{A}^T$ is smaller).
- The addition of $\lambda \boldsymbol{I}_n$ or $\lambda \boldsymbol{I}_m$ ensures that the matrix is invertible even when $\boldsymbol{A}^T\boldsymbol{A}$ or $\boldsymbol{A}\,\boldsymbol{A}^T$ are singular or nearly singular.

Solving $\boldsymbol{Ax = y}$

(Case 3: General Case) - (Method 2)

Using Damped Least Squares (DLS), we can find approximate solutions $\boldsymbol{x}$ to $\boldsymbol{Ax = y}$ in the form

$$\boldsymbol{x} \approx \boldsymbol{A}_D^+ \boldsymbol{y} + (\boldsymbol{I}_n - \boldsymbol{A}_D^+ \boldsymbol{A})\boldsymbol{x}_0 \quad \text{for arbitrary vector } \boldsymbol{x}_0 \in \mathbb{R}^n \qquad (1)$$

The approximate particular solution $\boldsymbol{A}_D^+ \boldsymbol{y}$ minimizes $\|\boldsymbol{Ax} - \boldsymbol{y}\|_2 + \lambda \|\boldsymbol{x}\|_2$ and the approximate homogeneous solution $(\boldsymbol{I}_n - \boldsymbol{A}^+ \boldsymbol{A})\boldsymbol{x}_0$ represents all vectors in $\mathcal{N}(\boldsymbol{A})$.

- The solution (1) is derived from this optimization problem:

$$\min_{\boldsymbol{x}} \quad \frac{1}{2}(\boldsymbol{Ax} - \boldsymbol{y})^T(\boldsymbol{Ax} - \boldsymbol{y}) + \frac{1}{2}\lambda \dot{\boldsymbol{x}}^T \dot{\boldsymbol{x}}$$

DLS Method vs SVD Method:

- DLS is simpler and computationally less expensive, making it ideal for real-time applications like robotics requiring quick, stable solutions, but the solutions are approximate, and it introduces a uniform bias.
- SVD provides a more comprehensive and flexible approach, providing a complete picture of the matrix's properties and allowing for selective treatment of singular values, leading to potentially more accurate solutions, but at the cost of higher computational effort. Moreover, SVD is more sensitive to small changes in the matrix and to ill-conditioning.