

Lecture 3: Analytic Geometry

Yi, Yung (이윤)

Mathematics for Machine Learning

<https://yung-web.github.io/home/courses/mathml.html>

KAIST EE

April 2, 2021

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

- A notion of the length of vectors
- **Definition.** A norm on a vector space V is a function $\|\cdot\| : V \mapsto \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ the following hold:
 - **Absolutely homogeneous:** $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$
 - **Triangle inequality:** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
 - **Positive definite:** $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| \iff \mathbf{x} = 0$

- **Manhattan Norm** (also called ℓ_1 norm) For $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$$

- **Euclidean Norm** (also called ℓ_2 norm) For $\mathbf{x} \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_2 := \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{\mathbf{x}^T \mathbf{x}}$$

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

- Need to talk about the length of a vector and the angle or distance between two vectors, where vectors are defined in abstract vector spaces
- To this end, we define **inner product** in an abstract manner.
- Dot product: A kind of inner product in vector space \mathbb{R}^n . $\mathbf{x}^T \mathbf{y} = \sum_{i=1}^n x_i y_i$
- **Question.** How can we generalize this and do a similar thing in some other vector spaces?

- An inner product is a mapping $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{R}$ that satisfies the following conditions for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $\lambda \in \mathbb{R}$:
 1. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
 2. $\langle \lambda \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$
 3. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$
 4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and equal iff $\mathbf{v} = \mathbf{0}$
- The pair $(V, \langle \cdot, \cdot \rangle)$ is called an inner product space.

- **Example.** $V = \mathbb{R}^n$ and the dot product $\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^T \mathbf{y}$
- **Example.** $V = \mathbb{R}^2$ and $\langle \mathbf{x}, \mathbf{y} \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2x_2 y_2$
- **Example.** $V = \{\text{continuous functions in } \mathbb{R} \text{ over } [a, b]\}$, $\langle u, v \rangle := \int_a^b u(x)v(x)dx$

- A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ that satisfies the following is called **symmetric, positive definite** (or just positive definite):

$$\forall \mathbf{x} \in V \setminus \{0\} : \mathbf{x}^T \mathbf{A} \mathbf{x} > 0.$$

If only \geq in the above holds, then \mathbf{A} is called **symmetric, positive semidefinite**.

- $\mathbf{A}_1 = \begin{pmatrix} 9 & 6 \\ 6 & 5 \end{pmatrix}$ is positive definite.
- $\mathbf{A}_2 = \begin{pmatrix} 9 & 6 \\ 6 & 3 \end{pmatrix}$ is not positive definite.

- Consider an n -dimensional vector space V with an inner product $\langle \cdot, \cdot \rangle$ and an ordered basis $B = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ of V .
- Any $\mathbf{x}, \mathbf{y} \in V$ can be represented as: $\mathbf{x} = \sum_{i=1}^n \psi_i \mathbf{b}_i$ and $\mathbf{y} = \sum_{j=1}^n \lambda_j \mathbf{b}_j$ for some ψ_i and λ_j , $i, j = 1, \dots, n$.

$$\langle \mathbf{x}, \mathbf{y} \rangle = \left\langle \sum_{i=1}^n \psi_i \mathbf{b}_i, \sum_{j=1}^n \lambda_j \mathbf{b}_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \psi_i \langle \mathbf{b}_i, \mathbf{b}_j \rangle \lambda_j = \hat{\mathbf{x}}^T \mathbf{A} \hat{\mathbf{y}},$$

where $\mathbf{A}_{ij} = \langle \mathbf{b}_i, \mathbf{b}_j \rangle$ and $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the coordinates w.r.t. B .

- Then, if $\forall \mathbf{x} \in V \setminus \{0\} : \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ (i.e., \mathbf{A} is symmetric, positive definite), $\hat{\mathbf{x}}^T \mathbf{A} \hat{\mathbf{y}}$ legitimately defines an inner product (w.r.t. B)
- Properties
 - The kernel of \mathbf{A} is only $\{0\}$, because $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for all $\mathbf{x} \neq 0 \implies \mathbf{A} \mathbf{x} \neq 0$ if $\mathbf{x} \neq 0$.
 - The diagonal elements a_{ii} of \mathbf{A} are all positive, because $a_{ii} = \mathbf{e}_i^T \mathbf{A} \mathbf{e}_i > 0$.

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

- Inner product naturally induces a norm by defining:

$$\|x\| := \sqrt{\langle x, x \rangle}$$

- Not every norm is induced by an inner product
- **Cachy-Schwarz inequality.** For the induced norm by the inner product,

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

- Now, we can introduce a notion of distance using a norm as:

Distance. $d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$

- If the dot product is used as an inner product in \mathbb{R}^n , it is **Euclidian distance**.
- **Note.** The distance between two vectors does **NOT** necessarily require the notion of norm. Norm is just sufficient.
- Generally, if the following is satisfied, it is a suitable notion of distance (also called **metric**).
 - **Positive definite.** $d(\mathbf{x}, \mathbf{y}) \geq 0$ for all \mathbf{x}, \mathbf{y} and $d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$
 - **Symmetric.** $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
 - **Triangle inequality.** $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$

- Using C-S inequality,

$$-1 \leq \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1$$

- Then, there exists a unique $\omega \in [0, \pi]$ with

$$\cos \omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

- We define ω as the angle between \mathbf{x} and \mathbf{y} .
- **Definition.** If $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, in other words their angle is $\pi/2$, we say that they are **orthogonal**, denoted by $\mathbf{x} \perp \mathbf{y}$. Additionally, if $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, they are **orthonormal**.

- Orthogonality is defined by a given inner product. Thus, different inner products may lead to different results about orthogonality.
- **Example.** Consider two vectors $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$
- Using the dot product as the inner product, they are orthogonal.
- However, using $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{y}$, they are not orthogonal.

$$\cos \omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} = -\frac{1}{3} \implies \omega \approx 1.91 \text{ rad} \approx 109.5^\circ$$

- **Definition.** A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is an **orthogonal matrix**, iff its columns (or rows) are **orthonormal** so that

$$\mathbf{A}\mathbf{A}^T = \mathbf{I} = \mathbf{A}^T\mathbf{A}, \text{ implying } \mathbf{A}^{-1} = \mathbf{A}^T.$$

- We can use $\mathbf{A}^{-1} = \mathbf{A}^T$ for the definition of orthogonal matrices.
- Fact 1. \mathbf{A}, \mathbf{B} : orthogonal $\implies \mathbf{AB}$: orthogonal
- Fact 2. \mathbf{A} : orthogonal $\implies \det(\mathbf{A}) = \pm 1$
- The linear mapping Φ by orthogonal matrices preserve **length** and **angle** (for the dot product)

$$\|\Phi(\mathbf{x})\| = \|\mathbf{Ax}\|^2 = (\mathbf{Ax})^T(\mathbf{Ax}) = \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$$

$$\cos \omega = \frac{(\mathbf{Ax})^T(\mathbf{Ay})}{\|\mathbf{Ax}\| \|\mathbf{Ay}\|} = \frac{\mathbf{x}^T\mathbf{A}^T\mathbf{Ay}}{\sqrt{\mathbf{x}^T\mathbf{A}^T\mathbf{Ax} \mathbf{y}^T\mathbf{A}^T\mathbf{Ay}}} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

- Basis that is orthonormal, i.e., they are all orthogonal to each other and their lengths are 1.
- Standard basis in \mathbb{R}^n , $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, is orthonormal.
- **Question.** How to obtain an orthonormal basis?

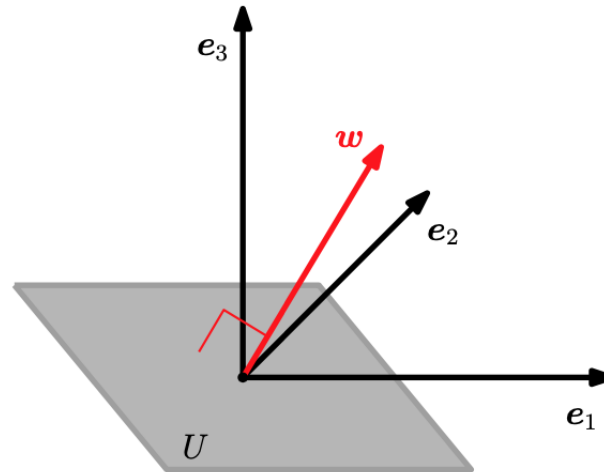
1. Use Gaussian elimination to find a basis for a vector space spanned by a set of vectors.
 - Given a set $\{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of unorthogonal and unnormalized basis vectors. Apply Gaussian elimination to the augmented matrix $(\mathbf{B}\mathbf{B}^T | \mathbf{B})$
2. Constructive way: Gram-Schmidt process (we will cover this later)

- Consider D -dimensional vector space V and M -dimensional subspace $W \subset V$. The **orthogonal complement** U^\perp is a $(D - M)$ -dimensional subspace of V and contains all vectors in V that are orthogonal to every vector in U .
- $U \cap U^\perp = 0$
- Any vector $x \in V$ can be uniquely decomposed into:

$$\mathbf{x} = \sum_{m=1}^M \lambda_m \mathbf{b}_m + \sum_{j=1}^{D-M} \psi_j \mathbf{b}_j^\perp, \quad \lambda_m, \psi_j \in \mathbb{R},$$

where $(\mathbf{b}_1, \dots, \mathbf{b}_M)$ and $(\mathbf{b}_1^\perp, \dots, \mathbf{b}_{D-M}^\perp)$ are the **bases** of U and U^\perp , respectively.

Orthogonal Complement (2)



- The vector w with $\|w\| = 1$, which is orthogonal to U , is the basis of U^\perp .
- Such w is called **normal vector** to U .
- For a linear mapping represented by a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, the solution space of $\mathbf{A}\mathbf{x} = 0$ is $\text{row}(\mathbf{A})^\perp$, where $\text{row}(\mathbf{A})$ is the row space of \mathbf{A} (i.e., span of row vectors). In other words, $\text{row}(\mathbf{A})^\perp = \ker(\mathbf{A})$

- Earlier, $V = \{\text{continuous functions in } \mathbb{R} \text{ over } [a, b]\}$, the following is a proper inner product.

$$\langle u, v \rangle := \int_a^b u(x)v(x)dx$$

- **Example.** Choose $u(x) = \sin(x)$ and $v(x) = \cos(x)$, where we select $a = -\pi$ and $b = \pi$. Then, since $f(x) = u(x)v(x)$ is odd (i.e., $f(-x) = -f(x)$),

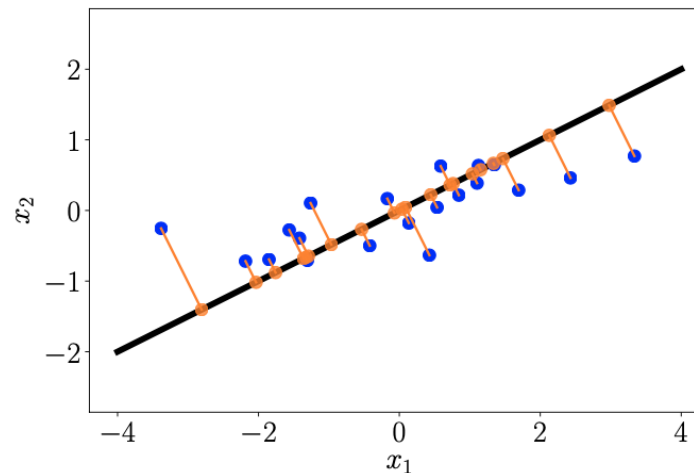
$$\int_{-\pi}^{\pi} u(x)v(x)dx = 0.$$

- Thus, u and v are orthogonal.
- Similarly, $\{1, \cos(x), \cos(2x), \cos(3x), \dots, \}$ is orthogonal over $[-\pi, \pi]$.

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

Projection: Motivation

- Big data: high dimensional
- However, most information is contained in a few dimensions
- **Projection**: A process of reducing the dimensions (hopefully) without loss of much information
- **Example**. Projection of 2D dataset onto 1D subspace



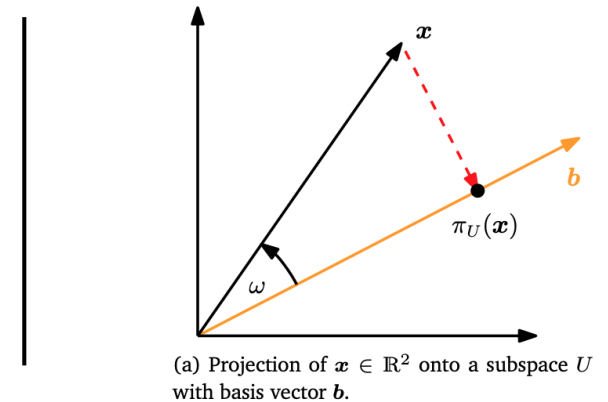
Projection onto Lines (1D Subspaces)

- Consider a 1D subspace $U \subset \mathbb{R}^n$ spanned by the basis \mathbf{b} .
- For $\mathbf{x} \in \mathbb{R}^n$, what is its projection $\pi_U(\mathbf{x})$ onto U (assume the dot product)?

$$\begin{aligned} \langle \mathbf{x} - \pi_U(\mathbf{x}), \mathbf{b} \rangle &= 0 \xleftrightarrow{\pi_U(\mathbf{x}) = \lambda \mathbf{b}} \langle \mathbf{x} - \lambda \mathbf{b}, \mathbf{b} \rangle = 0 \\ \implies \lambda &= \frac{\langle \mathbf{b}^\top, \mathbf{x} \rangle}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2}, \text{ and } \pi_U(\mathbf{x}) = \lambda \mathbf{b} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} \mathbf{b} \end{aligned}$$

- Projection matrix $\mathbf{P}_\pi \in \mathbb{R}^{n \times n}$ in $\pi_U(\mathbf{x}) = \mathbf{P}_\pi \mathbf{x}$

$$\pi_U(\mathbf{x}) = \lambda \mathbf{b} = \mathbf{b} \lambda = \frac{\mathbf{b} \mathbf{b}^\top}{\|\mathbf{b}\|^2} \mathbf{x}, \quad \mathbf{P}_\pi = \frac{\mathbf{b} \mathbf{b}^\top}{\|\mathbf{b}\|^2}$$

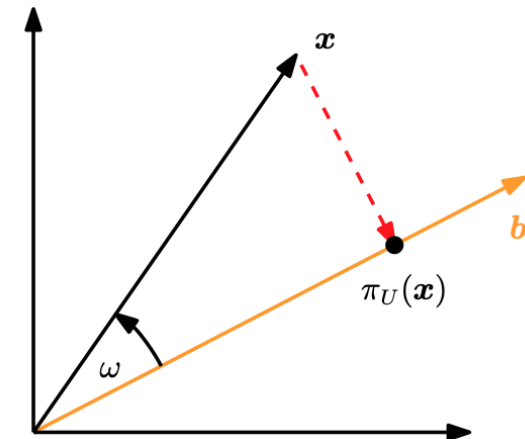


Inner Product and Projection

- We project \mathbf{x} onto \mathbf{b} , and let $\pi_{\mathbf{b}}(\mathbf{x})$ be the projected vector.
- **Question.** Understanding the inner product $\langle \mathbf{x}, \mathbf{b} \rangle$ from the projection perspective?

$$\langle \mathbf{x}, \mathbf{b} \rangle = \|\pi_{\mathbf{b}}(\mathbf{x})\| \times \|\mathbf{b}\|$$

- In other words, the inner product of \mathbf{x} and \mathbf{b} is the product of (length of the projection of \mathbf{x} onto \mathbf{b}) \times (length of \mathbf{b})



(a) Projection of $\mathbf{x} \in \mathbb{R}^2$ onto a subspace U with basis vector \mathbf{b} .

Example

- $\mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\|\mathbf{b}\|^2} = \frac{1}{9} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} (1 \ 2 \ 2) = \frac{1}{9} \begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix}$$

For $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$,

$$\pi_U(\mathbf{x}) = \mathbf{P}_\pi \mathbf{x} = \frac{1}{9} \begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{9} \begin{pmatrix} 5 \\ 10 \\ 10 \end{pmatrix} \in \text{span}\left[\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}\right]$$

Projection onto General Subspaces

- $\mathbb{R}^n \rightarrow 1\text{-Dim}$
- A basis vector \mathbf{b} in 1D subspace

$$\pi_U(\mathbf{x}) = \frac{\mathbf{b}\mathbf{b}^\top \mathbf{x}}{\mathbf{b}^\top \mathbf{b}}, \quad \lambda = \frac{\mathbf{b}^\top \mathbf{x}}{\mathbf{b}^\top \mathbf{b}}$$

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top \mathbf{b}}$$

- $\mathbb{R}^n \rightarrow m\text{-Dim}, (m < n)$

- A basis matrix
 $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_m) \in \mathbb{R}^{n \times m}$

$$\pi_U(\mathbf{x}) = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}, \quad \boldsymbol{\lambda} = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}$$

$$\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top$$

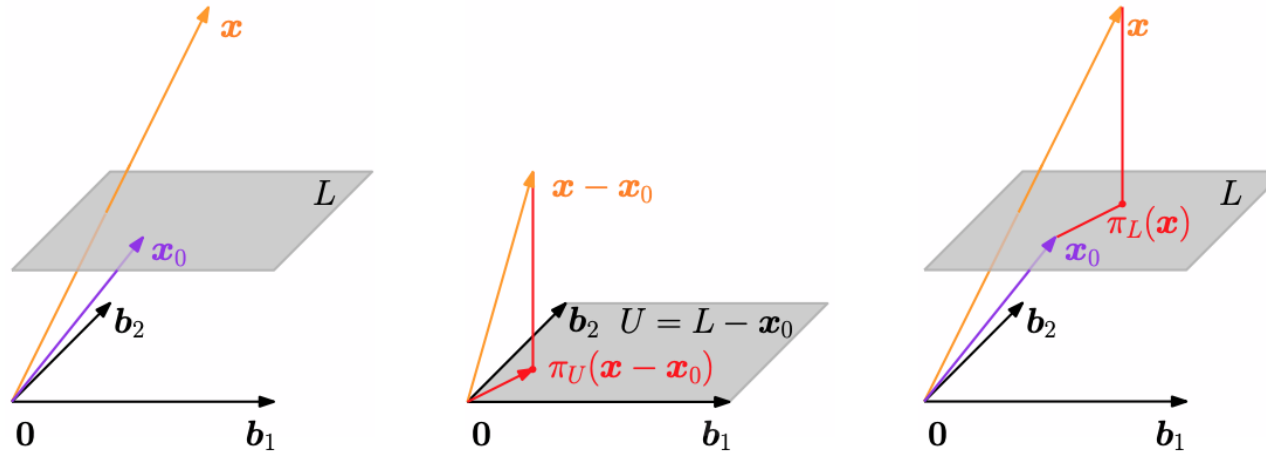
- $\lambda \in \mathbb{R}^1$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$ are the coordinates in the projected spaces, respectively.
- $(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top$ is called **pseudo-inverse**.
- How to derive is analogous to that of lines (see pp. 71).

Example: Projection onto 2D Subspace

- $U = \text{span}\left[\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}\right] \subset \mathbb{R}^3$ and $\mathbf{x} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$. Check that $\{(1 \ 1 \ 1)^T, (0 \ 1 \ 2)^T\}$ is a basis.
- Let $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$. Then, $\mathbf{B}^T \mathbf{B} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$
- Can see that $\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T = \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$, and
$$\pi_U(\mathbf{x}) = \frac{1}{6} \begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

-

Projection onto Affine Subspaces



- Affine space: $L = \mathbf{x}_0 + U$
- Affine subspaces are not vector spaces
- Idea: (i) move \mathbf{x} to a point in U , (ii) do the projection, (iii) move back to L

$$\pi_L(\mathbf{x}) = \mathbf{x}_0 + \pi_U(\mathbf{x} - \mathbf{x}_0)$$

- (1) Norms
- (2) Inner Products
- (3) Lengths and Distances
- (4) Angles and Orthogonality
- (5) Orthonormal Basis
- (6) Orthogonal Complement
- (7) Inner Product of Functions
- (8) Orthogonal Projections
- (9) Rotations

- A linear mapping that rotates the given coordinate system by an angle θ .
- Basis change
- $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}$ and $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \rightarrow \begin{pmatrix} -\sin \theta \\ \cos \theta \end{pmatrix}$
- Rotation matrix $\mathbf{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$
- Properties
 - Preserves distance: $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{R}_\theta(\mathbf{x}) - \mathbf{R}_\theta(\mathbf{y})\|$
 - Preserves angle

Questions?

1)