Roadmap

Roadmap

Norm

- A notion of the length of vectors
- **Definition.**A norm on a vector space V is a function $\|\cdot\| : V \to \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ and all $\mathbf{x}, \mathbf{y} \in V$ 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 $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$

Example for $V = \mathbb{R}^n$

- Manhattan Norm (also called ℓ_1 norm) For $\mathbf{x} = [x_1, \cdots, x_n] \in \mathbb{R}^n$,

$$\|\mathbf{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{n} |x_i|$$

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

$$\|\mathbf{x}\|_2 \stackrel{\text{def}}{=} \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\mathbf{x}^\top \mathbf{x}}$$

Roadmap

(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

Motivation

- 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 the notion of inner product in an abstract manner.
- Dot product: A kind of inner product in vector space $\mathbb{R}^n$. $\mathbf{x}^\top \mathbf{y} = \sum_{i=1}^{n} x_i y_i$
- **Question.**How can we generalize this and do a similar thing in some other vector spaces?

Formal Definition

- An inner product is a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ that satisfies the following conditions for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $\lambda \in \mathbb{R}$:

$$\begin{aligned}
&\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle \\
&\langle \lambda \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle \\
&\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle \\
&\langle \mathbf{v}, \mathbf{v} \rangle \geq 0 \text{ and } \langle \mathbf{v}, \mathbf{v} \rangle = 0 \text{ iff } \mathbf{v} = \mathbf{0}
\end{aligned}$$

- The pair $(V, \langle \cdot, \cdot \rangle)$ is called an inner product space.

Example

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

Symmetric, Positive Definite Matrix

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

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

  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.
- How to check? Compute $\det(\mathbf{A})$. For 2×2: $\mathbf{A}$ is PD iff $a_{11} > 0$ and $\det(\mathbf{A}) > 0$.
  - $\det(\mathbf{A}_1) = 9 \cdot 5 - 6^2 = 9 > 0$ ✓
  - $\det(\mathbf{A}_2) = 9 \cdot 3 - 6^2 = -9 < 0$ ×

Inner Product and Positive Definite Matrix (1)

- Consider an n-dimensional vector space V with an inner product $\langle \cdot, \cdot \rangle$ and an ordered basis $B = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$ 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, $i, j = 1, \ldots, n$.

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

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.

Inner Product and Positive Definite Matrix (2)

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

Roadmap

(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

Length

- Inner product naturally induces a norm by defining:

$$\|\mathbf{x}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

- Not every norm is induced by an inner product (e.g., ℓ_1 and ℓ_∞ norms are not)
- Cauchy-Schwarz inequality. For the induced norm by the inner product,

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

- Why it matters: This inequality guarantees $\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} \in [-1, 1]$, which makes the notion of angle between vectors well-defined.

Distance

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

  Distance. $d(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \|\mathbf{x} - \mathbf{y}\| = \sqrt{\langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle}$
- If the dot product is used as an inner product in $\mathbb{R}^n$, it is Euclidean 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, 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.

Example

- 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 $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{y}$, they are not orthogonal.

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

Orthogonal Matrix

- **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

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

  - We can use $\mathbf{A}^{-1} = \mathbf{A}^\top$ 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 preserves length and angle (for the dot product)

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

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

Roadmap

Orthonormal Basis

- Basis that is orthonormal, i.e., they are all orthogonal to each other and their lengths are 1.
- Standard basis in $\mathbb{R}^n$, $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, 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 $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ of non-orthogonal and unnormalized basis vectors. Apply Gaussian elimination to the augmented matrix $(\mathbf{B}\mathbf{B}^\top | \mathbf{B})$
2. Constructive way: Gram-Schmidt process (we will cover this later)

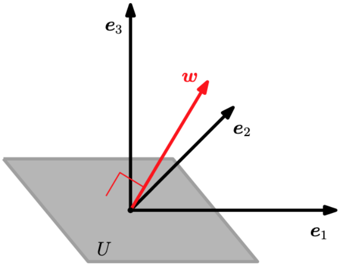
Orthogonal Complement (1)

- Consider D-dimensional vector space V and M-dimensional subspace $U \subset V$. 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.
- $U \cap U^{\perp} = \mathbf{0}$
- Any vector $\mathbf{x} \in V$ 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 $(\mathbf{b}_1 \ldots, \mathbf{b}_M)$ and $(\mathbf{b}_1^{\perp}, \ldots, \mathbf{b}_{D-M}^{\perp})$ are the bases of U and $U^{\perp}$, respectively.

Orthogonal Complement (2)



- The vector $\mathbf{w}$ with $\|\mathbf{w}\| = 1$, which is orthogonal to U, is the basis of $U^\perp$.
- Such $\mathbf{w}$ 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 $\mathbf{Ax} = 0$ 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, $\text{row}(\mathbf{A})^\perp = \ker(\mathbf{A})$

Inner Product of Functions

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

$$\langle u, v\rangle \stackrel{\text{def}}{=} \int_a^b u(x)v(x)dx$$

- **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., $f(-x) = -f(x)$),

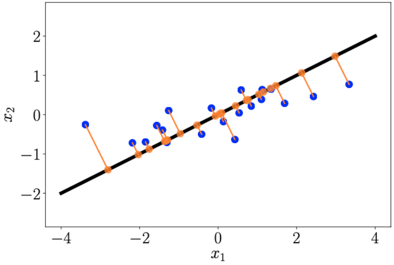
$$\int_{-\pi}^{\pi} u(x)v(x)dx = 0.$$

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

Roadmap

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[1]
- **Example.** Projection of 2D dataset onto 1D subspace



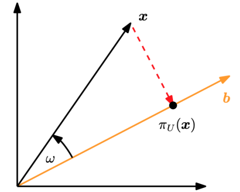
[1] In **L10**, we will formally study this with the topic of PCA (Principal Component Analysis).

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)?

$$\langle \mathbf{x} - \pi_U(\mathbf{x}), \mathbf{b} \rangle = 0 \xleftrightarrow{\pi_U(\mathbf{x}) = \lambda \mathbf{b}} \langle \mathbf{x} - \lambda \mathbf{b}, \mathbf{b} \rangle = 0$$

$$\implies \lambda = \frac{\langle \mathbf{b}, \mathbf{x} \rangle}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2}, \text{ and } \pi_U(\mathbf{x}) = \lambda \mathbf{b} = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} \mathbf{b}$$



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

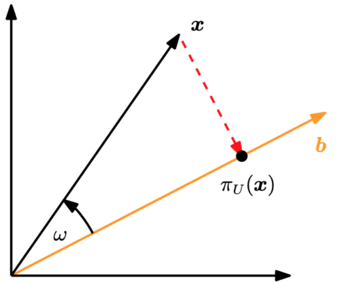
- Projection matrix $\mathbf{P}_\pi \in \mathbb{R}^{n \times n}$ in $\pi_U(\mathbf{x}) = \mathbf{P}_\pi \mathbf{x}$

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

- 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 $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.

Example

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

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

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

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

Projection onto General Subspaces

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

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

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

- $\mathbb{R}^n \to m$-Dim, $(m < n)$
- A basis matrix $B = \left(\mathbf{b}_1, \cdots, \mathbf{b}_m\right) \in \mathbb{R}^{n \times m}$

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

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

- $\lambda \in \mathbb{R}^1$ 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.
- The derivation is analogous to the case of 1-D lines (see pp. 71).

Example: Projection onto 2D Subspace

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

$$\pi_U(\mathbf{x}) = \frac{1}{6}\begin{pmatrix}5 & 2 & -1\\2 & 2 & 2\\-1 & 2 & 5\end{pmatrix}\begin{pmatrix}6\\0\\0\end{pmatrix} = \begin{pmatrix}5\\2\\-1\end{pmatrix}$$

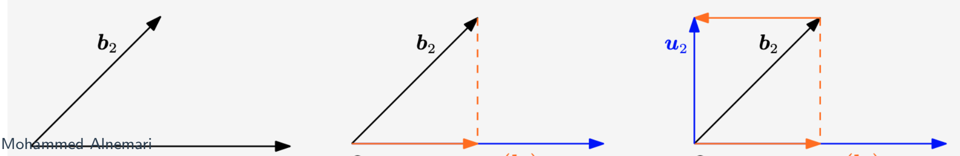
Gram-Schmidt Orthogonalization Method (G-S method)

- Constructively transform any basis $(\mathbf{b}_1, \ldots, \mathbf{b}_n)$ of n-dimensional vector space V into an orthogonal/orthonormal basis $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ of V
- Iteratively construct as follows

$$\mathbf{u}_1 \stackrel{\text{def}}{=} \mathbf{b}_1$$

$$\mathbf{u}_k \stackrel{\text{def}}{=} \mathbf{b}_k - \pi_{\text{span}[\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}]}(\mathbf{b}_k), \;\; k = 2, \ldots, n \qquad (*)$$

- Intuition: In $(*)$, $\pi_{\text{span}[\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}]}(\mathbf{b}_k)$ is the projection of $\mathbf{b}_k$ onto the subspace spanned by $[\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}]$. By subtracting this projection, $\mathbf{u}_k$ becomes orthogonal



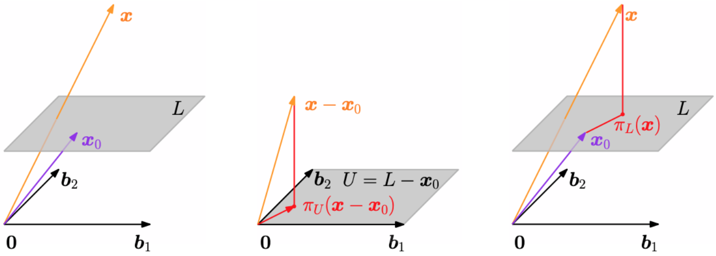
Example: G-S method

- A basis $(\mathbf{b}_1, \mathbf{b}_2)$ of $\mathbb{R}^2$, $\mathbf{b}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and $\mathbf{b}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
- $\mathbf{u}_1 = \mathbf{b}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and

$$\mathbf{u}_2 = \mathbf{b}_2 - \pi_{\mathrm{span}[\mathbf{u}_1]}(\mathbf{b}_2) = \mathbf{b}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2}\mathbf{b}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

- $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. If we want them to be orthonormal, then just normalization would do the job.

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

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

Roadmap

Rotation

- Length and angle preservation: two properties of linear mappings with orthogonal matrices. Let's look at some of their special cases.
- A linear mapping that rotates the given coordinate system by an angle θ.
- Basis change
- $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$
- 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

Key Takeaways

- Concept chain: Inner Product $\to$ Norm (length) $\to$ Distance $\to$ Angle $\to$ Orthogonality $\to$ Projection
- Each concept builds on the previous one. Once we have an inner product, we get all the others for free.
- Projections are the most applied concept in this lecture:
  - Dimensionality reduction (PCA)
  - Least-squares fitting and linear regression
  - Signal processing and data compression
- Key formulas to remember:
  - Cauchy-Schwarz: $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$
  - Projection matrix: $\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1}\mathbf{B}^\top$
  - Gram-Schmidt: $\mathbf{u}_k = \mathbf{b}_k - \sum_{i=1}^{k-1} \frac{\langle \mathbf{b}_k, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$

Why This Matters for AI/ML

- Norms $\to$ Regularization (ℓ_1 for sparsity, ℓ_2 for weight decay)
- Inner products $\to$ Cosine similarity in NLP (comparing word/sentence embeddings)
- Orthogonal projections $\to$ Principal Component Analysis (PCA) for dimensionality reduction
- Least squares via projections $\to$ Linear regression (fitting $\mathbf{y} \approx \mathbf{X}\theta$)
- Orthogonal matrices $\to$ Preserving gradient norms during backpropagation (orthogonal initialization)
- Gram-Schmidt $\to$ QR decomposition, used in numerical linear algebra for solving systems and computing eigenvalues

More on Norms: ℓ_p Family and Geometry

- In $\mathbb{R}^n$, for $p \geq 1$, the ℓ_p norm is

$$\|\mathbf{x}\|_p \stackrel{\text{def}}{=} \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}.$$

- Special cases: $p = 1$ (diamond-shaped unit ball in 2D), $p = 2$ (circle), $p = \infty$ (square).
- ℓ_∞ norm:

$$\|\mathbf{x}\|_\infty \stackrel{\text{def}}{=} \max_i |x_i|.$$

Question.(Geometry) Sketch the unit balls of $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ in $\mathbb{R}^2$. How does the shape relate to what the norm "penalizes"?

Induced Norms from Inner Products

- Any inner product $\langle \cdot, \cdot \rangle$ induces a norm:

$$\|\mathbf{x}\| \stackrel{\text{def}}{=} \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}.$$

- In $\mathbb{R}^n$, if $\langle \mathbf{x}, \mathbf{y} \rangle \stackrel{\text{def}}{=} \mathbf{x}^\top \mathbf{A} \mathbf{y}$ with $\mathbf{A} \succ 0$, then

$$\|\mathbf{x}\|_{\mathbf{A}} = \sqrt{\mathbf{x}^\top \mathbf{A} \mathbf{x}}.$$

- This norm "stretches" space according to $\mathbf{A}$ (elliptical unit ball).

Question. If $\mathbf{A} = \operatorname{diag}(4, 1)$ in $\mathbb{R}^2$, what does the unit ball $\{\mathbf{x} : \mathbf{x}^\top \mathbf{A} \mathbf{x} \leq 1\}$ look like?

Norm Equivalence in Finite Dimensions (Key Fact)

- In finite-dimensional spaces, all norms are equivalent: for any two norms $\|\cdot\|_a, \|\cdot\|_b$ on $\mathbb{R}^n$,

$$\exists\, c, C > 0 \text{ s.t. } c\,\|\mathbf{x}\|_a \le \|\mathbf{x}\|_b \le C\,\|\mathbf{x}\|_a\,,\ \ \forall \mathbf{x}.$$

- Example inequalities (useful to remember):

$$\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \sqrt{n}\,\|\mathbf{x}\|_\infty\,, \qquad \|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1 \le \sqrt{n}\,\|\mathbf{x}\|_2\,.$$

Question. Prove $\|\mathbf{x}\|_2 \le \|\mathbf{x}\|_1$ using $|x_i| \ge 0$ and compare squares.

Cauchy–Schwarz Inequality: Two Proof Ideas

- Statement: $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$.
- Proof idea (quadratic trick): for any $t \in \mathbb{R}$,

$$0 \leq \|\mathbf{x} - t\mathbf{y}\|^2 = \langle \mathbf{x} - t\mathbf{y}, \mathbf{x} - t\mathbf{y} \rangle = \|\mathbf{x}\|^2 - 2t\langle \mathbf{x}, \mathbf{y} \rangle + t^2 \|\mathbf{y}\|^2 .$$

- Discriminant ≤ 0 yields

$$\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \|\mathbf{x}\|^2 \, \|\mathbf{y}\|^2 .$$

Question. When does equality hold in Cauchy–Schwarz? (Answer should be a geometric condition on $\mathbf{x}, \mathbf{y}$.)

Angles, Correlation, and Orthogonality

- For nonzero vectors, define

$$\cos\omega = \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\ \|\mathbf{y}\|} \in [-1, 1].$$

- Interpretation:
  - $\cos\omega \approx 1$: vectors are aligned (strong positive correlation).
  - $\cos\omega \approx -1$: opposite directions (strong negative correlation).
  - $\cos\omega = 0$: orthogonal (uncorrelated in Euclidean geometry).

Question. In $\mathbb{R}^n$ with dot product, show that $\mathbf{x} \perp \mathbf{y}$ implies $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ (Pythagorean theorem).

Orthogonal Complement and the Fundamental Subspaces

- For $\mathbf{A} \in \mathbb{R}^{m\times n}$ (dot product):

$$\ker(\mathbf{A}) = \operatorname{row}(\mathbf{A})^{\perp}, \qquad \ker(\mathbf{A}^{\top}) = \operatorname{col}(\mathbf{A})^{\perp}.$$

- Intuition: rows impose linear constraints; nullspace is all vectors orthogonal to every row.
- Dimensions (Rank–Nullity):

$$\dim(\ker(\mathbf{A})) = n - \operatorname{rank}(\mathbf{A}).$$

Question. If $\mathbf{A}$ has rank r, what is $\dim(\operatorname{row}(\mathbf{A})^{\perp})$? Explain why this matches rank–nullity.

Projection Matrices: Properties You Must Know

- For a subspace $U = \mathrm{col}(\mathbf{B})$ with full column rank, the Euclidean projection matrix is

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

- Key properties (Euclidean projection):

$$\mathbf{P}^2 = \mathbf{P} \quad \text{(idempotent)}, \qquad \mathbf{P}^\top = \mathbf{P} \quad \text{(symmetric)}.$$

- Decomposition: $\mathbf{x} = \underbrace{\mathbf{P}\mathbf{x}}_{\in U} + \underbrace{(\mathbf{I} - \mathbf{P})\mathbf{x}}_{\in U^\perp}$.

Question. Prove $\mathbf{P}^2 = \mathbf{P}$ directly from the formula $\mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1}\mathbf{B}^\top$.

Least Squares = Projection onto Column Space

- Overdetermined system $\mathbf{A}\mathbf{x} \approx \mathbf{b}$ (usually $m > n$). We solve:

$$\min_{\mathbf{x}} \ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 .$$

- Optimality condition (normal equations):

$$\mathbf{A}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0 \quad \Longleftrightarrow \quad \mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{b}.$$

- Geometric meaning: $\mathbf{A}\mathbf{x}^\star$ is the projection of $\mathbf{b}$ onto $\mathrm{col}(\mathbf{A})$.

Question. Why is the residual $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}^\star$ orthogonal to $\mathrm{col}(\mathbf{A})$?

Gram–Schmidt: One Clean Formula (and Orthonormalization)

- Given linearly independent $(\mathbf{b}_1, \ldots, \mathbf{b}_k)$, define:

$$\mathbf{u}_1 = \mathbf{b}_1, \qquad \mathbf{u}_j = \mathbf{b}_j - \sum_{i=1}^{j-1} \frac{\langle \mathbf{b}_j, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i, \quad j \geq 2.$$

- Then $(\mathbf{u}_1, \ldots, \mathbf{u}_k)$ is orthogonal. Orthonormalize via

$$\mathbf{e}_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}.$$

Question. Apply Gram–Schmidt to $\mathbf{b}_1 = (1, 1, 0)$ and $\mathbf{b}_2 = (1, 0, 1)$ in $\mathbb{R}^3$. Compute an orthonormal basis.

Rotations and Orthogonal Matrices: Eigen-Geometry

- In $\mathbb{R}^2$, a rotation $\mathbf{R}(\theta)$ is orthogonal with $\det(\mathbf{R}) = +1$.
- In general, orthogonal matrices satisfy $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$, so they preserve dot products:

$$\langle \mathbf{Qx}, \mathbf{Qy} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle.$$

- Eigen-geometry:
  - Pure rotations in 2D have complex eigenvalues $e^{\pm i\theta}$ unless $\theta \in \{0, \pi\}$.
  - Reflections (also orthogonal) have $\det = -1$ and real eigenvalues ± 1.

Question. Show that if $\mathbf{Q}$ is orthogonal, then $\|\mathbf{Qx}\|_2 = \|\mathbf{x}\|_2$. Then argue why angles are preserved.

Concepts at a Glance

Concept	**What It Means**	**Why It Matters**
Norm	Length of a vector	Regularization in ML
Inner Product	Generalized dot product	Defines geometry on vector spaces
Cauchy-Schwarz	$\lvert\langle \mathbf{x}, \mathbf{y} \rangle\rvert \leq \lVert\mathbf{x}\rVert \, \lVert\mathbf{y}\rVert$	Makes angles well-defined
Orthogonality	$\langle \mathbf{x}, \mathbf{y} \rangle = 0$	Uncorrelated features
Orthogonal Matrix	$\mathbf{A}^{-1} = \mathbf{A}^{\top}$	Preserves lengths and angles
Projection	Closest point in subspace	PCA, least squares
Gram-Schmidt	Builds orthonormal basis	QR decomposition, numerical methods
Rotation	Orthogonal + det $= +1$	Image transformations, robotics

Pattern: Inner product $\rightarrow$ norm $\rightarrow$ distance $\rightarrow$ angle $\rightarrow$ orthogonality $\rightarrow$ projection. Each concept builds on the previous one.

Common Mistakes to Avoid

(1) Not every norm comes from an inner product (ℓ_1 and ℓ_∞ do not)
Only ℓ_2-type norms are induced by inner products

(2) Orthogonality depends on the choice of inner product
Two vectors can be orthogonal under one inner product but not another

(3) $\mathbf{P}_\pi \neq \mathbf{P}_\pi^{-1}$ — projection matrices are **not** invertible (except $\mathbf{P} = \mathbf{I}$)
They satisfy $\mathbf{P}^2 = \mathbf{P}$ (idempotent), not $\mathbf{P}\mathbf{P}^{-1} = \mathbf{I}$

(4) Gram-Schmidt requires linearly independent input
If input vectors are dependent, you get $\mathbf{u}_k = \mathbf{0}$ at some step

(5) Orthogonal matrix $\neq$ matrix with orthogonal columns
Columns must be orthogonal AND unit length (orthonormal)

Review Question 1: Norms

Question. Let $\mathbf{x} = (3, -4) \in \mathbb{R}^2$.

(a) Compute $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, and $\|\mathbf{x}\|_\infty$.

(b) Which three properties must a function satisfy to be a valid norm?

(c) Is $f(\mathbf{x}) = |x_1|$ a norm on $\mathbb{R}^2$? Why or why not?

Hint: $\|\mathbf{x}\|_1 = \sum |x_i|$, $\|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}$, $\|\mathbf{x}\|_\infty = \max |x_i|$. For (c), check positive definiteness: can $f(\mathbf{x}) = 0$ with $\mathbf{x} \neq \mathbf{0}$?

Review Question 2: Inner Products

Question. Consider $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + 2x_2 y_2$ on $\mathbb{R}^2$.

(a) Verify this is a valid inner product (check all four axioms).

(b) What is the matrix $\mathbf{A}$ such that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^\top \mathbf{A} \mathbf{y}$?

(c) Is $\mathbf{A}$ symmetric and positive definite?

Hint: Write $\mathbf{A} = \operatorname{diag}(1, 2)$. Check: symmetric? All eigenvalues positive? Then $\mathbf{x}^\top \mathbf{A} \mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$.

Review Question 3: Cauchy-Schwarz and Lengths

Question. Let $\mathbf{x} = (1, 2, 2)$ and $\mathbf{y} = (2, 1, 0)$ in $\mathbb{R}^3$ with the dot product.

(a) Compute $\langle \mathbf{x}, \mathbf{y} \rangle$, $\|\mathbf{x}\|$, and $\|\mathbf{y}\|$.

(b) Verify the Cauchy-Schwarz inequality $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$.

(c) Compute the distance $d(\mathbf{x}, \mathbf{y})$.

Hint: $\langle \mathbf{x}, \mathbf{y} \rangle = 1 \cdot 2 + 2 \cdot 1 + 2 \cdot 0 = 4$. Distance: $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$.

Review Question 4: Angles and Orthogonality

Question. In $\mathbb{R}^3$ with the dot product:

(a) Find the angle between $\mathbf{x} = (1, 0, 1)$ and $\mathbf{y} = (0, 1, 0)$.

(b) Are they orthogonal? How do you know?

(c) Find a vector $\mathbf{z} \neq \mathbf{0}$ that is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$.

Hint: $\cos\omega = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|\|\mathbf{y}\|}$. Two vectors are orthogonal iff $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. For (c), solve $\langle \mathbf{z}, \mathbf{x} \rangle = 0$ and $\langle \mathbf{z}, \mathbf{y} \rangle = 0$ simultaneously.

Review Question 5: Orthogonal Matrices

Question. Let $\mathbf{A} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$.

(a) Verify that $\mathbf{A}$ is an orthogonal matrix.

(b) What is $\mathbf{A}^{-1}$?

(c) What is $\det(\mathbf{A})$? Is $\mathbf{A}$ a rotation or a reflection?

Hint: Check $\mathbf{A}^\top\mathbf{A} = \mathbf{I}$. For orthogonal matrices, $\mathbf{A}^{-1} = \mathbf{A}^\top$. $\det = +1$ means rotation, $\det = -1$ means reflection.

Question. Apply the Gram-Schmidt process to find an orthonormal basis from $\mathbf{b}_1 = (1, 1, 0)$ and $\mathbf{b}_2 = (1, 0, 1)$.

(a) Compute $\mathbf{u}_1 = \mathbf{b}_1$.

(b) Compute $\mathbf{u}_2 = \mathbf{b}_2 - \frac{\langle \mathbf{b}_2, \mathbf{u}_1 \rangle}{\langle \mathbf{u}_1, \mathbf{u}_1 \rangle}\mathbf{u}_1$.

(c) Normalize $\mathbf{u}_1$ and $\mathbf{u}_2$ to get an orthonormal basis.

Hint: $\langle \mathbf{b}_2, \mathbf{u}_1 \rangle = 1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0 = 1$, $\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = 2$, so $\mathbf{u}_2 = (1, 0, 1) - \frac{1}{2}(1, 1, 0) = (\frac{1}{2}, -\frac{1}{2}, 1)$.

Review Question 7: Orthogonal Complement

Question. Let $U = \text{span}\{(1,1,0),(0,1,1)\} \subset \mathbb{R}^3$.

(a) What is $\dim(U^\perp)$?

(b) Find a basis for $U^\perp$ (find a vector orthogonal to both spanning vectors).

(c) Verify that $U \cap U^\perp = \{\mathbf{0}\}$.

Hint: $\dim(U^\perp) = \dim(\mathbb{R}^3) - \dim(U) = 3 - 2 = 1$. Solve $\langle \mathbf{z}, (1,1,0)\rangle = 0$ and $\langle \mathbf{z}, (0,1,1)\rangle = 0$, i.e., $z_1 + z_2 = 0$ and $z_2 + z_3 = 0$.

Review Question 8: Projection onto Lines

Question. Project $\mathbf{x} = (3, 1)$ onto the line spanned by $\mathbf{b} = (1, 2)$ in $\mathbb{R}^2$.

(a) Compute the scalar $\lambda = \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2}$.

(b) Compute the projection $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$.

(c) Compute the projection matrix $\mathbf{P}_\pi$ and verify $\mathbf{P}_\pi^2 = \mathbf{P}_\pi$.

Hint: $\lambda = \frac{3 \cdot 1 + 1 \cdot 2}{1^2 + 2^2} = \frac{5}{5} = 1$, so $\pi_U(\mathbf{x}) = 1 \cdot (1, 2) = (1, 2)$. This means $\mathbf{x} - \pi_U(\mathbf{x}) = (2, -1)$, which should be $\perp$ $\mathbf{b}$.

Question. Let $U = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$ and $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

(a) Set up $\mathbf{B}$ and compute $\mathbf{B}^\top \mathbf{B}$.

(b) Compute the projection $\pi_U(\mathbf{x}) = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1}\mathbf{B}^\top \mathbf{x}$.

(c) Verify that $\mathbf{x} - \pi_U(\mathbf{x})$ is orthogonal to both columns of $\mathbf{B}$.

Hint: $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{B}^\top \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. Use $(\mathbf{B}^\top \mathbf{B})^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$.

Review Question 10: Rotations

Question. Consider the rotation matrix $\mathbf{R}(\theta)$ in $\mathbb{R}^2$.

(a) Write out $\mathbf{R}(\pi/4)$ (rotation by 45 degrees).

(b) Apply $\mathbf{R}(\pi/4)$ to the vector $\mathbf{x} = (1, 0)$. What is the result?

(c) Show that $\mathbf{R}(\theta)$ is orthogonal by verifying $\mathbf{R}^\top \mathbf{R} = \mathbf{I}$.

Hint: $\mathbf{R}(\pi/4) = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Apply to $(1, 0)$: you get $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ — rotated 45° counterclockwise.

Theory Connection: Analytic Geometry in Machine Learning

- Regularization: ℓ_1 norm (Lasso) promotes sparsity, ℓ_2 norm (Ridge) prevents large weights
- Cosine similarity: $\cos\omega = \frac{\mathbf{x}^\top \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$ measures how similar two feature vectors are
- Least squares: The optimal fit $\hat{\mathbf{x}} = (\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{A}^\top \mathbf{b}$ is a projection!
- Feature orthogonality: Uncorrelated features are orthogonal in data space — ideal for learning

Every ML model uses norms for regularization and inner products for similarity.

Theory Connection: Analytic Geometry in NLP

- Word embeddings: Words are vectors in $\mathbb{R}^{300}$; similarity $=$ cosine of angle between them
- Word analogies: "king $-$ man $+$ woman $\approx$ queen" works because of vector arithmetic
- Document similarity: TF-IDF vectors compared via cosine similarity for search engines
- Sentence embeddings: Transformers output vectors; inner products score relevance

NLP lives in high-dimensional inner product spaces — analytic geometry is the language of language models.

Theory Connection: Analytic Geometry in Deep Learning

- Orthogonal initialization: Weight matrices initialized as orthogonal preserve gradient norms during backpropagation
- Batch normalization: Projects activations to have unit norm — uses projection concepts
- Attention mechanism: Query-key dot products $\mathbf{q}^\top \mathbf{k}$ measure relevance via inner products
- Gradient orthogonality: When gradients from different tasks are orthogonal, multi-task learning works well

Deep learning architectures are designed using inner products, norms, and projections.

Theory Connection: Analytic Geometry in Computer Vision

- Image similarity: Compare images as vectors; distance $= \|\text{img}_1 - \text{img}_2\|_2$
- Rotation matrices: 3D object pose estimation uses $\mathbf{R}(\theta)$ to model rotations
- Projective geometry: Camera models project 3D scenes onto 2D images — projection matrices!
- Face recognition: FaceNet maps faces to unit sphere; recognition = nearest neighbor by angle

Computer vision is geometry — every transformation, comparison, and detection uses these tools.

Theory Connection: Analytic Geometry in Data Science

- PCA: Projects data onto directions of maximum variance — orthogonal projection onto top eigenvectors
- Dimensionality reduction: From $\mathbb{R}^D$ to $\mathbb{R}^d$ ($d \ll D$) using projection matrices
- Anomaly detection: Points far from projected subspace (large $\|\mathbf{x} - \mathbf{P}\mathbf{x}\|$) are anomalies
- Recommendation systems: User and item vectors; rating $\approx \langle \text{user}, \text{item} \rangle$

Data science reduces high-dimensional data to lower dimensions via projections — the core of this lecture.

PhD View: Hilbert Spaces and Functional Analysis

- Hilbert spaces: Complete inner product spaces — generalize $\mathbb{R}^n$ to infinite dimensions
- L^2 space: Functions f with $\int |f(x)|^2 dx < \infty$ form a Hilbert space with $\langle f, g \rangle = \int f(x)g(x)dx$
- Fourier series: Projection of a function onto orthonormal basis $\{e^{inx}\}$ — Gram-Schmidt in function space
- Riesz representation: Every continuous linear functional on a Hilbert space is an inner product — deep duality result
- Connection to ML: Kernel methods implicitly work in infinite-dimensional Hilbert spaces

PhD View: Reproducing Kernel Hilbert Spaces (RKHS)

- Kernel trick: Map data $\mathbf{x} \mapsto \phi(\mathbf{x})$ into high-dim space; compute $\langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = k(\mathbf{x}, \mathbf{y})$ without explicit mapping
- RKHS: A Hilbert space of functions where evaluation is a continuous linear functional
- Reproducing property: $f(\mathbf{x}) = \langle f, k(\cdot, \mathbf{x}) \rangle$ — function value = inner product with kernel
- Representer theorem: Optimal solution in RKHS is $f^* = \sum_{i=1}^{n} \alpha_i k(\cdot, \mathbf{x}_i)$ — finite despite infinite-dim space
- Applications: Support Vector Machines, Gaussian Processes, kernel PCA

PhD View: Compressed Sensing and Sparse Recovery

- Problem: Recover $\mathbf{x} \in \mathbb{R}^n$ from $m \ll n$ measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$ — underdetermined!
- Key insight: If $\mathbf{x}$ is s-sparse, solve $\min \|\mathbf{x}\|_1$ s.t. $\mathbf{A}\mathbf{x} = \mathbf{y}$ (convex relaxation of ℓ_0)
- Restricted Isometry Property (RIP): Random $\mathbf{A}$ preserves norms of sparse vectors with high probability
- Geometry: ℓ_1 ball has corners aligned with axes — promotes sparse solutions
- Applications: MRI acceleration, single-pixel cameras, radar, genomics

PhD View: Riemannian Geometry and Manifold Learning

- Riemannian manifolds: Curved spaces with a smoothly varying inner product (metric tensor g_{ij}) at each point
- Geodesics: Shortest paths on manifolds — generalize straight lines in Euclidean space
- Manifold hypothesis: High-dimensional data often lies on a low-dimensional manifold
- Algorithms: Isomap, LLE, t-SNE, UMAP — discover manifold structure via local distances and angles
- Natural gradient descent: Uses Riemannian metric (Fisher information) instead of Euclidean — faster convergence

PhD View: Random Projections and Johnson-Lindenstrauss

- JL Lemma: n points in $\mathbb{R}^D$ can be projected to $\mathbb{R}^d$ with $d = O(\frac{\log n}{\epsilon^2})$ preserving all pairwise distances within $(1 \pm \epsilon)$
- Random projection: Multiply by random Gaussian matrix $\mathbf{A} \in \mathbb{R}^{d \times D}$ — fast, data-independent
- Concentration of measure: In high dimensions, random vectors are nearly orthogonal — inner products concentrate
- Locality-Sensitive Hashing (LSH): Random projections $\rightarrow$ hash functions that preserve cosine similarity
- Applications: Approximate nearest neighbors, streaming algorithms, sketching for massive datasets

Thank You!

Questions?