Lecture 3: Analytic Geometry

Yi, Yung (이융)

Mathematics for Machine Learning
https://yung-web.github.io/home/courses/mathml.html
KAIST EE

April 6, 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: $\|\lambda \boldsymbol{x}\| = |\lambda| \, \|\boldsymbol{x}\|$
  - Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
  - Positive definite: $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| \iff \boldsymbol{x} = 0$

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

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

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

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

- 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 $\mathbb{R}^n$. $\boldsymbol{x}^\mathsf{T}\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$

- 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 $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and all scalars $\lambda \in \mathbb{R}$:

  1. $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$

  2. $\langle \lambda\boldsymbol{v}, \boldsymbol{w} \rangle = \lambda \langle \boldsymbol{v}, \boldsymbol{w} \rangle$

  3. $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$

  4. $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$ and equal iff $\boldsymbol{v} = 0$

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

- Example. $V = \mathbb{R}^n$ and the dot product $\langle x, y \rangle := x^\mathsf{T} y$

- Example. $V = \mathbb{R}^2$ and $\langle x, y \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2x_2 y_2$

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

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

If only $\geq$ in the above holds, then A is called symmetric, positive semidefinite.

- $A_1 = \begin{pmatrix} 9 & 6 \\ 6 & 5 \end{pmatrix}$ is positive definite.

- $A_2 = \begin{pmatrix} 9 & 6 \\ 6 & 3 \end{pmatrix}$ is not positive definite.

- Consider an n-dimensional vector space V with an inner product $\langle \cdot, \cdot \rangle$ and an ordered basis $B = (b_1, \ldots, b_n)$ of V.

- Any $x, y \in V$ can be represented as: $x = \sum_{i=1}^n \psi_i b_i$ and $y = \sum_{i=j}^n \lambda_j b_j$ for some ψ_i and λ_j, $i, j = 1, \ldots, n$.

$$\langle x, y \rangle = \left\langle \sum_{i=1}^n \psi_i b_i, \sum_{i=j}^n \lambda_j b_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \psi_i \langle b_i, b_j \rangle \lambda_j = \hat{x}^\mathsf{T} A \hat{y},$$

where $A_{ij} = \langle b_i, b_j \rangle$ and $\hat{x}$ and $\hat{y}$ are the coordinates w.r.t. B.

- Then, if $\forall x \in V \setminus \{0\} : x^\mathsf{T} A x > 0$ (i.e., A is symmetric, positive definite), $\hat{x}^\mathsf{T} A \hat{y}$ legitimately defines an inner product (w.r.t. B)

- Properties
  - The kernel of A is only $\{0\}$, because $x^\mathsf{T} A x > 0$ for all $x \neq 0 \implies Ax \neq 0$ if $x \neq 0$.
  - The diagonal elements a_{ii} of A are all positive, because $a_{ii} = e_i^\mathsf{T} A e_i > 0$.

- Inner product naturally induces a norm by defining:

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

- Not every norm is induced by an inner product

- Cachy-Schwarz inequality. For the induced norm by the inner product,

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \le \|\boldsymbol{x}\| \ \|\boldsymbol{y}\|$$

Distance

Angle, Orthogonal, and Orthonormal

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

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

- 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, called metric.
  - *Positive definite*. $d(\boldsymbol{x}, \boldsymbol{y}) \ge 0$ for all $\boldsymbol{x}, \boldsymbol{y}$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff \boldsymbol{x} = \boldsymbol{y}$
  - *Symmetric*. $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$
  - *Triangle inequality*. $d(\boldsymbol{x}, \boldsymbol{z}) \le d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$

- Using C-S inequality,

$$-1 \le \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \ \|\boldsymbol{y}\|} \le 1$$

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

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

- We define ω as the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

- Definition. If $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, in other words their angle is $\pi/2$, we say that they are orthogonal, denoted by $\boldsymbol{x} \perp \boldsymbol{y}$. Additionally, if $\|x\| = \|y\| = 1$, 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 $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $y = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

- Using the dot product as the inner product, they are orthogonal.

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

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

Orthogonal Matrix

- Definition. A square matrix $A \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, iff its columns (or rows) are orthonormal so that
$$AA^\top = I = A^\top A, \text{ implying } A^{-1} = A^\top.$$
  - We can use $A^{-1} = A^\top$ for the definition of orthogonal matrices.
  - Fact 1. A, B: orthogonal $\implies AB$: orthogonal
  - Fact 2. A: orthogonal $\implies \det(A) = \pm 1$

- The linear mapping Φ by orthogonal matrices preserve length and angle (for the dot product)
$$\|\Phi(A)\| = \|Ax\|^2 = (Ax)^\top(Ax) = x^\top A^\top A x = x^\top x = \|x\|^2$$
$$\cos \omega = \frac{(Ax)^\top(Ay)}{\|Ax\| \ \|Ay\|} = \frac{x^\top A^\top A y}{\sqrt{x^\top A^\top A x y^\top A^\top A y}} = \frac{x^\top y}{\|x\| \ \|y\|}$$

Roadmap

Orthonormal Basis

- Basis that is orthonormal, i.e., they are all orthogonal to each other and their lengths are 1.

- Standard basis in $\mathbb{R}^n$, $\{e_1, \ldots, e_n\}$, 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 $\{b_1, \ldots, b_n\}$ of unorthogonal and unnormalized basis vectors. Apply Gaussian elimination to the augmented matrix $(BB^\top | B)$

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.
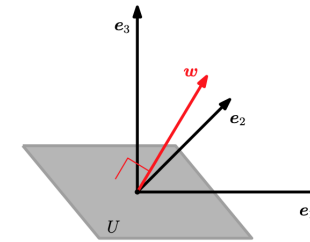
- $U \cap U^\perp = 0$

- Any vector $x \in V$ can be uniquely decomposed into:

$$x = \sum_{m=1}^{M} \lambda_m b_m + \sum_{j=1}^{D-M} \psi_j b_j^\perp, \quad \lambda_m, \psi_j \in \mathbb{R},$$

where $(b_1 \ldots, b_M)$ and $(b_1^\perp, \ldots, b_{D-M}^\perp)$ are the bases of U and $U^\perp$, respectively.

- 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 $A \in \mathbb{R}^{m \times n}$, the solution space of $Ax = 0$ is $\text{row}(A)^\perp$, where $\text{row}(A)$ is the row space of A (i.e., span of row vectors). In other words, $\text{row}(A)^\perp = \ker(A)$

- Remind: $V = \{$continuous functions in $\mathbb{R}$ 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, $\{1, \cos(x), \cos(2x), \cos(3x), \ldots, \}$ is orthogonal over $[-\pi, \pi]$.
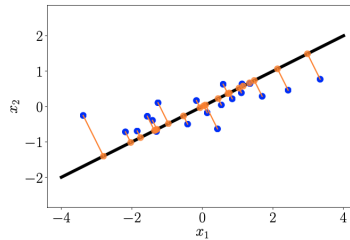
- Big data: high dimensional
- However, most information is contained in a few dimensions
- Projection: A process of reducing the dimensions (hopefully) without loss of much information[1]
- Example. Projection of 2D dataset onto 1D subspace

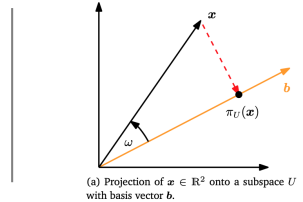[1] In L10, we will formally study this with the topic of PCA (Principal Component Analysis).

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

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

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

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

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

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

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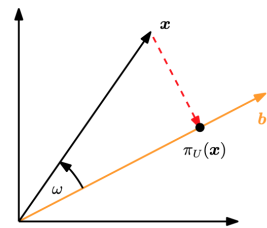
$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle = \|\pi_{\boldsymbol{b}}(\boldsymbol{x})\| \times \|\boldsymbol{b}\|$$

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

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

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

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

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

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

- $\mathbb{R}^n \to$ 1-Dim

- A basis vector $\boldsymbol{b}$ in 1D subspace

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

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

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

- A basis matrix
$$B = (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m) \in \mathbb{R}^{n \times m}$$
$$\pi_U(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{x}, \quad \boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top \boldsymbol{x}$$
$$\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top$$

- $\lambda \in \mathbb{R}^1$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$ are the coordinates in the projected spaces, respectively.

- $(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top$ is called pseudo-inverse.

- How to derive is analogous to the case of 1-D lines (see pp. 71).

- $U = \mathrm{span}[\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}] \subset \mathbb{R}^3$ and $\boldsymbol{x} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}$. Check that $\{(1\ 1\ 1)^\top, (0\ 1\ 2)^\top\}$ is a basis.

- Let $\boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$. Then, $\boldsymbol{B}^\top \boldsymbol{B} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$

- Can see that $\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1}\boldsymbol{B}^\top = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}$, and

$$\pi_U(\boldsymbol{x}) = \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix}\begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$
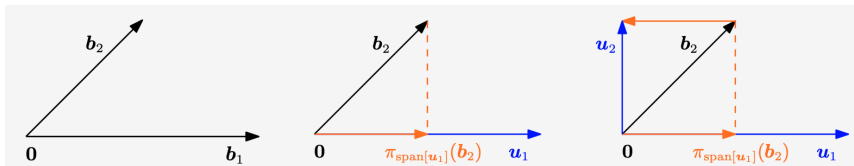
- Constructively transform any basis $(\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ of n-dimensional vector space V into an orthogonal/orthonormal basis $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ of V

- Iteratively construct as follows

$$\boldsymbol{u}_1 := \boldsymbol{b}_1$$
$$\boldsymbol{u}_k := \boldsymbol{b}_k - \pi_{\mathrm{span}[\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{k-1}]}(\boldsymbol{b}_k), \ \ k = 2, \ldots, n \qquad (*)$$

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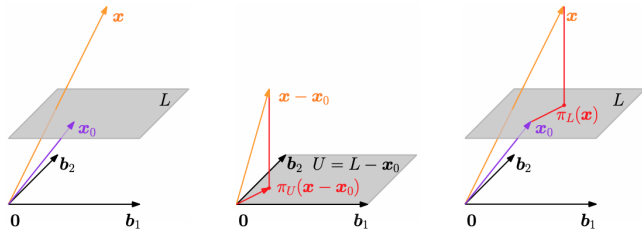
- $\boldsymbol{u}_1 = \boldsymbol{b}_1 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ and

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

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

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

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

(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

- 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
- $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and $\boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$
- Rotation matrix $\boldsymbol{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
- Properties
  - Preserves distance: $\|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{R}_\theta(\boldsymbol{x}) - \boldsymbol{R}_\theta(\boldsymbol{y})\|$
  - Preserves angle

Questions?

1)