

Mathematics for Machine Learning

— Linear Algebra: Projections & Gram-Schmidt Orthogonalization

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Fall 2023

Credits for the resource

- The slides are based on the textbooks:
 - *Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong: Mathematics for Machine Learning. Cambridge University Press. 2020.*
 - *Howard Anton, Chris Rorres, Anton Kaul: Elementary Linear Algebra. Wiley. 2019.*
- We could partially refer to the monograph:
Francesco Orabona: A Modern Introduction to Online Learning.
<https://arxiv.org/abs/1912.13213>

Outline

- 1 Orthogonal Projections
- 2 Gram-Schmidt Orthogonalization
- 3 Rotations

Outline

- 1 Orthogonal Projections
- 2 Gram-Schmidt Orthogonalization
- 3 Rotations

Motivations (1/2)

- In machine learning, we often need to deal with *high-dimensional* data.

Motivations (1/2)

- In machine learning, we often need to deal with *high-dimensional* data.
- High-dimensional data is often hard to analyze or visualize.

Motivations (1/2)

- In machine learning, we often need to deal with *high-dimensional* data.
- High-dimensional data is often hard to analyze or visualize.
- Sometimes, only a few dimensions contain most information.

Motivations (1/2)

- In machine learning, we often need to deal with *high-dimensional* data.
- High-dimensional data is often hard to analyze or visualize.
- Sometimes, only a few dimensions contain most information.
- We might try to project the original high-dimensional data onto a lower-dimensional space and work on it.

Motivations (1/2)

- In machine learning, we often need to deal with *high-dimensional* data.
- High-dimensional data is often hard to analyze or visualize.
- Sometimes, only a few dimensions contain most information.
- We might try to project the original high-dimensional data onto a lower-dimensional space and work on it.
- Note: When we compress or visualize high-dimensional data, we will lose information.

Motivations (2/2)

Examples (dimensionality reduction)

- Principal Component Analysis (PCA)

Motivations (2/2)

Examples (dimensionality reduction)

- Principal Component Analysis (PCA)
- Deep Neural Networks

Motivations (2/2)

Examples (dimensionality reduction)

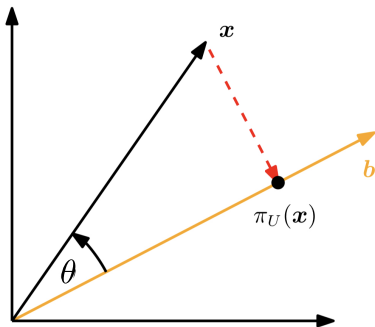
- Principal Component Analysis (PCA)
- Deep Neural Networks
- Classification

Motivations (2/2)

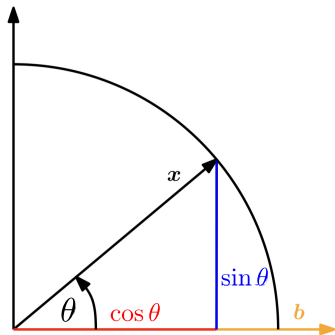
Examples (dimensionality reduction)

- Principal Component Analysis (PCA)
- Deep Neural Networks
- Classification
- Linear Regression

Projection from 2D to 1D



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



(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b .

Projection

Projection

Let V be a vector space and $U \subseteq V$ be a subspace of V . A linear mapping $\pi : V \mapsto U$ is called a **projection** if $\pi^2 = \pi \circ \pi = \pi$.

Projection

Projection

Let V be a vector space and $U \subseteq V$ be a subspace of V . A linear mapping $\pi : V \mapsto U$ is called a **projection** if $\pi^2 = \pi \circ \pi = \pi$.

- Recall that linear mappings can be expressed by transformation matrices.

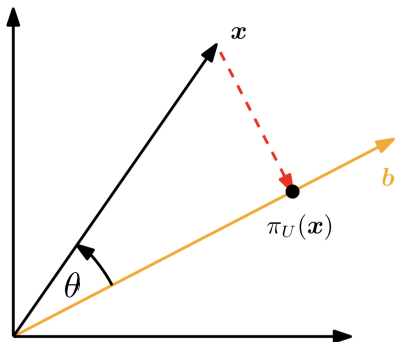
Projection

Projection

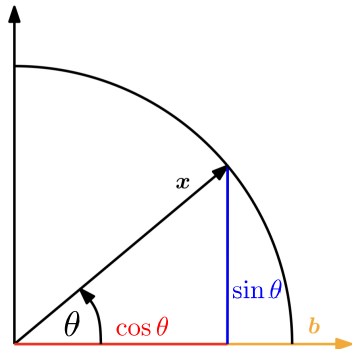
Let V be a vector space and $U \subseteq V$ be a subspace of V . A linear mapping $\pi : V \mapsto U$ is called a **projection** if $\pi^2 = \pi \circ \pi = \pi$.

- Recall that linear mappings can be expressed by transformation matrices.
- The projection matrices P_π exhibit the property that $P_\pi^2 = P_\pi$.

Illustration of projections onto 1-D



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



(b) Projection of a two-dimensional vector x with $\|x\| = 1$ onto a one-dimensional subspace spanned by b .

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to \mathbf{x} .
 - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to \mathbf{x} .
 - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.
 - $\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$.
- Projection $\pi_U(\mathbf{x})$ of \mathbf{x} onto U must be an element in U .
 - $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$.

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to \mathbf{x} .
 - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.
 - $\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$.
- Projection $\pi_U(\mathbf{x})$ of \mathbf{x} onto U must be an element in U .
 - $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$.
- Determining the coordinates:

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to \mathbf{x} .
 - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.
 - $\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$.
- Projection $\pi_U(\mathbf{x})$ of \mathbf{x} onto U must be an element in U .
 - $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$.
- Determining the coordinates:
Since $\pi_U(\mathbf{b}) = \lambda \mathbf{b}$:

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

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to \mathbf{x} .
 - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.
 - $\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$.
- Projection $\pi_U(\mathbf{x})$ of \mathbf{x} onto U must be an element in U .
 - $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$.
- Determining the coordinates:
Since $\pi_U(\mathbf{b}) = \lambda \mathbf{b}$:

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

$$\Leftrightarrow \langle \mathbf{x}, \mathbf{b} \rangle - \lambda \langle \mathbf{b}, \mathbf{b} \rangle = 0 \Leftrightarrow \lambda = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{\langle \mathbf{b}, \mathbf{x} \rangle}{\|\mathbf{b}\|^2}$$

Illustration of projections onto 1-D

- $\pi_U(\mathbf{x})$: closest to \mathbf{x} .
 - $\|\mathbf{x} - \pi_U(\mathbf{x})\|$ is minimal.
 - $\langle \pi_U(\mathbf{x}) - \mathbf{x}, \mathbf{b} \rangle = 0$.
- Projection $\pi_U(\mathbf{x})$ of \mathbf{x} onto U must be an element in U .
 - $\pi_U(\mathbf{x}) = \lambda \mathbf{b}$ for some $\lambda \in \mathbb{R}$.
- Determining the coordinates:
Since $\pi_U(\mathbf{b}) = \lambda \mathbf{b}$:

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

$$\Leftrightarrow \langle \mathbf{x}, \mathbf{b} \rangle - \lambda \langle \mathbf{b}, \mathbf{b} \rangle = 0 \Leftrightarrow \lambda = \frac{\langle \mathbf{x}, \mathbf{b} \rangle}{\langle \mathbf{b}, \mathbf{b} \rangle} = \frac{\langle \mathbf{b}, \mathbf{x} \rangle}{\|\mathbf{b}\|^2} = \frac{\mathbf{b}^\top \mathbf{x}}{\mathbf{b}^\top \mathbf{b}}.$$

- Finding the projection $\pi_U(\mathbf{x}) \in U$:

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

- Finding the projection $\pi_U(\mathbf{x}) \in U$:

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

Note that $\|\pi_U(\mathbf{x})\| = \|\lambda \mathbf{b}\| = |\lambda| \|\mathbf{b}\|$.

- Finding the projection $\pi_U(\mathbf{x}) \in U$:

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

Note that $\|\pi_U(\mathbf{x})\| = \|\lambda \mathbf{b}\| = |\lambda| \|\mathbf{b}\|$.

- If we use the dot product as the inner product and let θ be the angle between \mathbf{x} and \mathbf{b} :

$$\|\pi_U(\mathbf{x})\| = \frac{|\mathbf{b}^\top \mathbf{x}|}{\|\mathbf{b}\|^2} \|\mathbf{b}\| = |\cos \theta| \|\mathbf{x}\| \|\mathbf{b}\| \frac{\|\mathbf{b}\|}{\|\mathbf{b}\|^2} = |\cos \theta| \|\mathbf{x}\|.$$

- Finding the projection matrix P_π :
 - Recall: projection is a linear mapping.
 - With the dot product as the inner product,

$$\|\pi_U(\mathbf{x})\| = \lambda \mathbf{b} = \mathbf{b} \lambda = \mathbf{b} \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} = \frac{\mathbf{b} \mathbf{b}^\top}{\|\mathbf{b}\|^2} \mathbf{x}.$$

- So,

$$P_\pi = \frac{\mathbf{b} \mathbf{b}^\top}{\|\mathbf{b}\|^2}.$$

Note: $\mathbf{b} \mathbf{b}^\top$ is a symmetric matrix.

- Finding the projection matrix P_π :
 - Recall: projection is a linear mapping.
 - With the dot product as the inner product,

$$\|\pi_U(\mathbf{x})\| = \lambda \mathbf{b} = \mathbf{b} \lambda = \mathbf{b} \frac{\mathbf{b}^\top \mathbf{x}}{\|\mathbf{b}\|^2} = \frac{\mathbf{b} \mathbf{b}^\top}{\|\mathbf{b}\|^2} \mathbf{x}.$$

- So,

$$P_\pi = \frac{\mathbf{b} \mathbf{b}^\top}{\|\mathbf{b}\|^2}.$$

Note: $\mathbf{b} \mathbf{b}^\top$ is a symmetric matrix.

Example

Find the projection matrix \mathbf{P}_π onto the line U through the origin spanned by $\mathbf{b} = [1 \ 2 \ 2]^\top$ and the projection of $\mathbf{x} = [1 \ 1 \ 1]^\top$.

Example

Find the projection matrix \mathbf{P}_π onto the line U through the origin spanned by $\mathbf{b} = [1 \ 2 \ 2]^\top$ and the projection of $\mathbf{x} = [1 \ 1 \ 1]^\top$.

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

Example

Find the projection matrix \mathbf{P}_π onto the line U through the origin spanned by $\mathbf{b} = [1 \ 2 \ 2]^\top$ and the projection of $\mathbf{x} = [1 \ 1 \ 1]^\top$.

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top\mathbf{b}} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} [1 \ 2 \ 2]$$

Example

Find the projection matrix \mathbf{P}_π onto the line U through the origin spanned by $\mathbf{b} = [1 \ 2 \ 2]^\top$ and the projection of $\mathbf{x} = [1 \ 1 \ 1]^\top$.

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top\mathbf{b}} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} [1 \ 2 \ 2] = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}.$$

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

Example

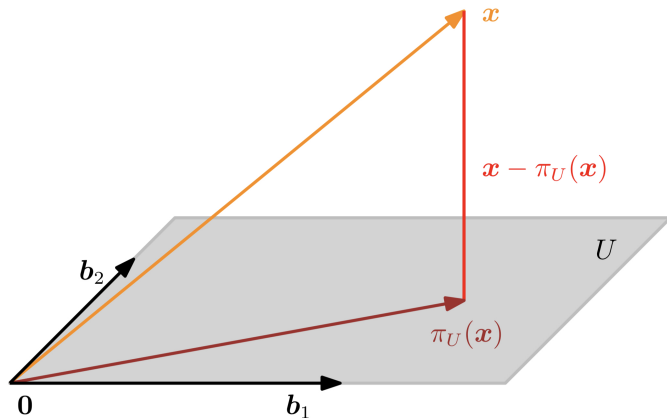
Find the projection matrix \mathbf{P}_π onto the line U through the origin spanned by $\mathbf{b} = [1 \ 2 \ 2]^\top$ and the projection of $\mathbf{x} = [1 \ 1 \ 1]^\top$.

$$\mathbf{P}_\pi = \frac{\mathbf{b}\mathbf{b}^\top}{\mathbf{b}^\top\mathbf{b}} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} [1 \ 2 \ 2] = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}.$$

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

Projection onto General Subspaces (1/4)

Orthogonal projections of $\mathbf{x} \in \mathbb{R}^n$ onto $U \subseteq \mathbb{R}^n$ with $\dim(U) = m \geq 1$.



Projection onto General Subspaces (2/4)

- Any projection can be represented as a linear combination of the basis vectors $\mathbf{b}_1, \dots, \mathbf{b}_m$ of U .
 - $\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i$.
- Find the coordinates $\lambda_1, \dots, \lambda_m$:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i = \mathbf{B}\boldsymbol{\lambda}$$

for $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}$, $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^\top \in \mathbb{R}^m$.

Projection onto General Subspaces (2/4)

- Any projection can be represented as a linear combination of the basis vectors $\mathbf{b}_1, \dots, \mathbf{b}_m$ of U .
 - $\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i$.
- Find the coordinates $\lambda_1, \dots, \lambda_m$:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i = \mathbf{B}\boldsymbol{\lambda} \quad (\text{closest to } \mathbf{x} \text{ on } U)$$

for $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}$, $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^\top \in \mathbb{R}^m$.

Note: $\mathbf{x} \perp \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ (\because minimum distance)

$$\begin{aligned} \langle \mathbf{b}_1, \mathbf{x} - \pi_U(\mathbf{x}) \rangle &= \mathbf{b}_1^\top (\mathbf{x} - \pi_U(\mathbf{x})) = 0 \\ &\vdots \\ \langle \mathbf{b}_m, \mathbf{x} - \pi_U(\mathbf{x}) \rangle &= \mathbf{b}_m^\top (\mathbf{x} - \pi_U(\mathbf{x})) = 0 \end{aligned}$$

Projection onto General Subspaces (3/4)

- Any projection can be represented as a linear combination of the basis vectors $\mathbf{b}_1, \dots, \mathbf{b}_m$ of U .
 - $\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i$.
- Find the coordinates $\lambda_1, \dots, \lambda_m$:

$$\pi_U(\mathbf{x}) = \sum_{i=1}^m \lambda_i \mathbf{b}_i = \mathbf{B}\boldsymbol{\lambda} \quad (\text{closest to } \mathbf{x} \text{ on } U)$$

for $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_m] \in \mathbb{R}^{n \times m}$, $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m]^\top \in \mathbb{R}^m$.

Note: $(\mathbf{x} - \pi_U(\mathbf{x})) \perp \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ (\because minimum distance)

$$\mathbf{b}_1^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = 0$$

$$\vdots$$

$$\mathbf{b}_m^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = 0$$

Projection onto General Subspaces (4/4)

Since

$$\mathbf{b}_1^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = 0$$

$$\vdots$$

$$\mathbf{b}_m^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = 0$$

We have

$$\begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} [\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}] = \mathbf{0} \Leftrightarrow \mathbf{B}^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = \mathbf{0}$$

$$\Leftrightarrow \mathbf{B}^\top \mathbf{B}\boldsymbol{\lambda} = \mathbf{B}^\top \mathbf{x}$$

Note: $\mathbf{B}^\top \mathbf{B}$ is invertible

Projection onto General Subspaces (4/4)

Since

$$\begin{aligned}\mathbf{b}_1^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) &= 0 \\ \vdots \\ \mathbf{b}_m^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) &= 0\end{aligned}$$

We have

$$\begin{aligned}\begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} [\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}] = \mathbf{0} &\Leftrightarrow \mathbf{B}^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = \mathbf{0} \\ &\Leftrightarrow \mathbf{B}^\top \mathbf{B}\boldsymbol{\lambda} = \mathbf{B}^\top \mathbf{x}\end{aligned}$$

Note: $\mathbf{B}^\top \mathbf{B}$ is invertible $\Rightarrow \boldsymbol{\lambda} = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}$.

Projection onto General Subspaces (4/4)

Since

$$\begin{aligned} \mathbf{b}_1^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) &= 0 \\ &\vdots \\ \mathbf{b}_m^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) &= 0 \end{aligned}$$

We have

$$\begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} [\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}] = \mathbf{0} \Leftrightarrow \mathbf{B}^\top (\mathbf{x} - \mathbf{B}\boldsymbol{\lambda}) = \mathbf{0}$$

$$\Leftrightarrow \mathbf{B}^\top \mathbf{B}\boldsymbol{\lambda} = \mathbf{B}^\top \mathbf{x}$$

Note: $\mathbf{B}^\top \mathbf{B}$ is invertible $\Rightarrow \boldsymbol{\lambda} = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}$.

- $\pi_U(\mathbf{x}) = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}$

Projection onto General Subspaces (4/4)

Since

$$\begin{aligned}\mathbf{b}_1^\top (\mathbf{x} - \mathbf{B}\lambda) &= 0 \\ \vdots \\ \mathbf{b}_m^\top (\mathbf{x} - \mathbf{B}\lambda) &= 0\end{aligned}$$

We have

$$\begin{aligned}\begin{bmatrix} \mathbf{b}_1^\top \\ \vdots \\ \mathbf{b}_m^\top \end{bmatrix} [\mathbf{x} - \mathbf{B}\lambda] = \mathbf{0} &\Leftrightarrow \mathbf{B}^\top (\mathbf{x} - \mathbf{B}\lambda) = \mathbf{0} \\ &\Leftrightarrow \mathbf{B}^\top \mathbf{B}\lambda = \mathbf{B}^\top \mathbf{x}\end{aligned}$$

Note: $\mathbf{B}^\top \mathbf{B}$ is invertible $\Rightarrow \lambda = (\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x}$.

- $\pi_U(\mathbf{x}) = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top \mathbf{x} \Rightarrow$ Projection matrix $\mathbf{P}_\pi = \mathbf{B}(\mathbf{B}^\top \mathbf{B})^{-1} \mathbf{B}^\top$.

But wait a minute ...

Why $B^T B$ is invertible?

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{A}\mathbf{x} = \mathbf{0}$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \implies$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \implies \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0} \Rightarrow$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax}$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} = (\mathbf{Ax})^\top (\mathbf{Ax})$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \implies \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0} \implies \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} = (\mathbf{Ax})^\top (\mathbf{Ax}) = \|\mathbf{Ax}\|^2 = 0$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

• **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} = (\mathbf{Ax})^\top (\mathbf{Ax}) = \|\mathbf{Ax}\|^2 = 0 \Rightarrow \mathbf{Ax} = \mathbf{0}$

But wait a minute ...

Why $\mathbf{B}^\top \mathbf{B}$ is invertible?

Fact

$\text{rank}(\mathbf{A}^\top \mathbf{A}) = \text{rank}(\mathbf{A})$ for any $\mathbf{A} \in \mathbb{R}^{n \times m}$.

- **Claim:** $\text{null}(\mathbf{A}) = \text{null}(\mathbf{A}^\top \mathbf{A})$.

(\Rightarrow): $\mathbf{Ax} = \mathbf{0} \implies \mathbf{A}^\top \mathbf{Ax} = \mathbf{0}$.

(\Leftarrow): $\mathbf{A}^\top \mathbf{Ax} = \mathbf{0} \implies \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} = (\mathbf{Ax})^\top (\mathbf{Ax}) = \|\mathbf{Ax}\|^2 = 0 \implies \mathbf{Ax} = \mathbf{0}$

- $\text{rank}(\mathbf{A}) = m = \text{rank}(\mathbf{A}^\top \mathbf{A})$ (\because the Dimension Theorem).

Example

Example

For a subspace $U = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \right\} \subseteq \mathbb{R}^3$ and $\mathbf{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \in \mathbb{R}^3$.

Find

- the coordinates λ of \mathbf{x} in terms of U
- the projection point $\pi_U(\mathbf{x})$
- the projection matrix \mathbf{P}_π .

- First, we find that the spanning set of U is a basis (check its linear independence!).

- First, we find that the spanning set of U is a basis (check its linear independence!).
- Derive $\mathbf{B} =$

- First, we find that the spanning set of U is a basis (check its linear independence!).

- Derive $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$.

- First, we find that the spanning set of U is a basis (check its linear independence!).
- Derive $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$.
- Compute $\mathbf{B}^\top \mathbf{B}$ and $\mathbf{B}^\top \mathbf{x}$:

$$\mathbf{B}^\top \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 5 & 5 \end{bmatrix},$$

- First, we find that the spanning set of U is a basis (check its linear independence!).

- Derive $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$.

- Compute $\mathbf{B}^\top \mathbf{B}$ and $\mathbf{B}^\top \mathbf{x}$:

$$\mathbf{B}^\top \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 5 & 5 \end{bmatrix},$$

$$\mathbf{B}^\top \mathbf{x} =$$

- First, we find that the spanning set of U is a basis (check its linear independence!).

- Derive $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$.

- Compute $\mathbf{B}^\top \mathbf{B}$ and $\mathbf{B}^\top \mathbf{x}$:

$$\mathbf{B}^\top \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 5 & 5 \end{bmatrix},$$

$$\mathbf{B}^\top \mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

- First, we find that the spanning set of U is a basis (check its linear independence!).
- Derive $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$.
- Compute $\mathbf{B}^\top \mathbf{B}$ and $\mathbf{B}^\top \mathbf{x}$:

$$\mathbf{B}^\top \mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 5 & 5 \end{bmatrix},$$

$$\mathbf{B}^\top \mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}.$$

- Then, solve $\mathbf{B}^\top \mathbf{B} \boldsymbol{\lambda} = \mathbf{B}^\top \mathbf{x}$ to find $\boldsymbol{\lambda}$:

$$\begin{bmatrix} 3 & 3 \\ 5 & 5 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$\text{So } \boldsymbol{\lambda} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}.$$

- Finally, the projection:

$$\pi_U(\mathbf{x}) = \mathbf{B} \boldsymbol{\lambda} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}.$$

What if $B = (\mathbf{b}_1, \dots, \mathbf{b}_m)$ is orthonormal?

- $\pi_U(\mathbf{x}) = B(B^\top B)^{-1}B^\top \mathbf{x}$

What if $B = (\mathbf{b}_1, \dots, \mathbf{b}_m)$ is orthonormal?

- $\pi_U(\mathbf{x}) = B(\mathbf{B}^\top B)^{-1} B^\top \mathbf{x} \Rightarrow \pi_U(\mathbf{x}) = BB^\top \mathbf{x}.$
 - $\because B^\top B = I.$
- Coordinates: $\lambda = (B^\top B)^{-1} B^\top \mathbf{x} = B^\top \mathbf{x}.$

Outline

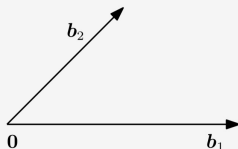
- 1 Orthogonal Projections
- 2 Gram-Schmidt Orthogonalization
- 3 Rotations

Illustration of Gram-Schmidt Orthogonalization

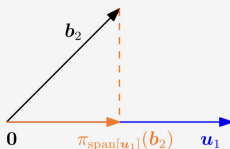
- Goal:** Transform any basis $(\mathbf{b}_1, \dots, \mathbf{b}_n)$ of an n -dimensional vector space V into an orthogonal/orthonormal basis of V .

$$\mathbf{u}_1 := \mathbf{b}_1$$

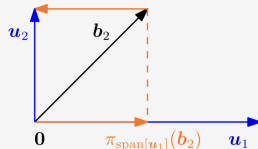
$$\mathbf{u}_k := \mathbf{b}_k - \pi_{\text{span}(\{\mathbf{u}_1, \dots, \mathbf{u}_{k-1}\})}(\mathbf{b}_k), \quad k = 2, \dots, n.$$



(a) Original non-orthogonal basis vectors $\mathbf{b}_1, \mathbf{b}_2$.



(b) First new basis vector $\mathbf{u}_1 = \mathbf{b}_1$ and projection of \mathbf{b}_2 onto the subspace spanned by \mathbf{u}_1 .



(c) Orthogonal basis vectors \mathbf{u}_1 and $\mathbf{u}_2 = \mathbf{b}_2 - \pi_{\text{span}[\mathbf{u}_1]}(\mathbf{b}_2)$.

Example

Example

Consider a basis $(\mathbf{b}_1, \mathbf{b}_2)$ of \mathbb{R}^2 , where $\mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Apply the Gram-Schmidt method to construct an orthonormal basis $(\mathbf{u}_1, \mathbf{u}_2)$ of \mathbb{R}^2 (assuming the dot product as the inner product).

$$\mathbf{u}_1 := \mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

$$\mathbf{u}_2 := \mathbf{b}_2 - \pi_{\text{span}(\mathbf{u}_1)}(\mathbf{b}_2)$$

$$\mathbf{u}_1 := \mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

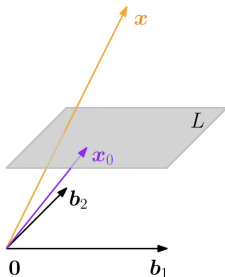
$$\begin{aligned} \mathbf{u}_2 &:= \mathbf{b}_2 - \pi_{\text{span}(\mathbf{u}_1)}(\mathbf{b}_2) = \mathbf{b}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2} \mathbf{b}_2 \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

$$\mathbf{u}_1 := \mathbf{b}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix},$$

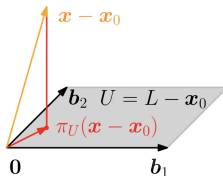
$$\begin{aligned} \mathbf{u}_2 &:= \mathbf{b}_2 - \pi_{\text{span}(\mathbf{u}_1)}(\mathbf{b}_2) = \mathbf{b}_2 - \frac{\mathbf{u}_1 \mathbf{u}_1^\top}{\|\mathbf{u}_1\|^2} \mathbf{b}_2 \\ &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

Projection onto Affine Spaces

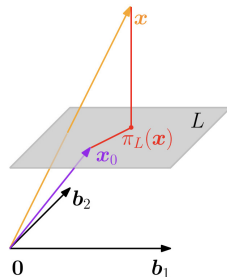
- Given an affine space $L = \mathbf{x}_0 + U$.
 - U is a low-dimensional subspace of V .
- $\pi_L(\mathbf{x}) = \mathbf{x}_0 + \pi_U(\mathbf{x} - \mathbf{x}_0)$



(a) Setting.



(b) Reduce problem to projection π_U onto vector subspace.

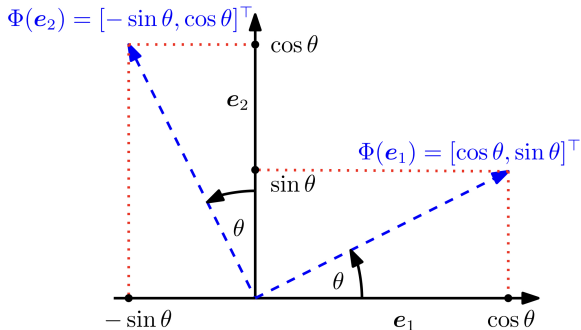


(c) Add support point back in to get affine projection π_L .

Outline

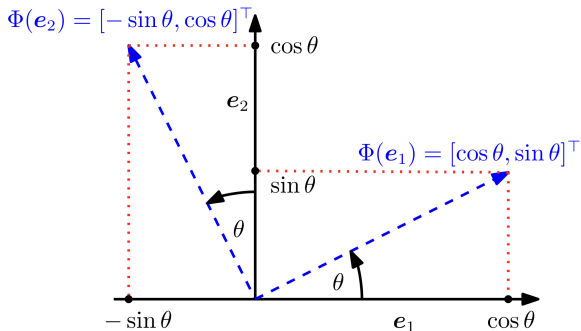
- 1 Orthogonal Projections
- 2 Gram-Schmidt Orthogonalization
- 3 Rotations**

Rotations in \mathbb{R}^2 as An Example



- Standard basis $\mathbf{e} = \{\mathbf{e}_1 = [1 \ 0]^\top, \ \mathbf{e}_2 = [0 \ 1]^\top\}$.
- $\mathbf{R}(\theta) = [\Phi(\mathbf{e}_1) \ \Phi(\mathbf{e}_2)]$

Rotations in \mathbb{R}^2 as An Example



- Standard basis $\mathbf{e} = \{\mathbf{e}_1 = [1 \ 0]^\top, \ \mathbf{e}_2 = [0 \ 1]^\top\}$.
- $\mathbf{R}(\theta) = [\Phi(\mathbf{e}_1) \ \Phi(\mathbf{e}_2)] = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$.

Discussions