MA2001 LINEAR ALGEBRA

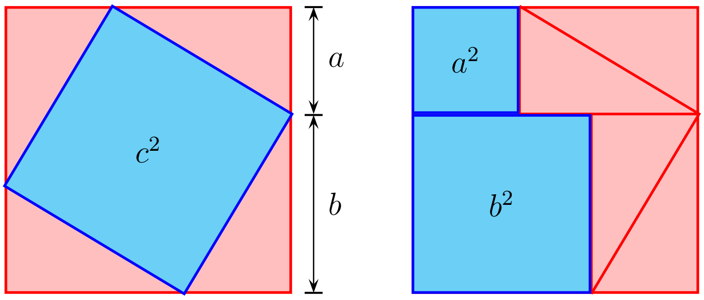
ORTHOGONALITY

National University of Singapore
Department of Mathematics

Pythagoras' Theorem

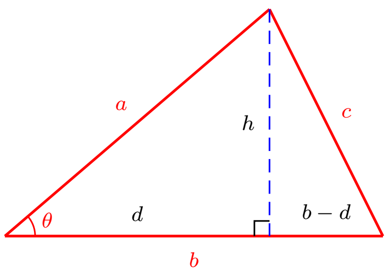
- **Pythagoras' Theorem**: In a right-angled triangle:
  - Let c be the length of the **hypotenuse**, and let a and b be the lengths of the other two sides.

  Then $a^2 + b^2 = c^2$.
- **Proof**. Consider the square of side $a + b$.



Pythagoras' Theorem

- **Cosine rule**: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

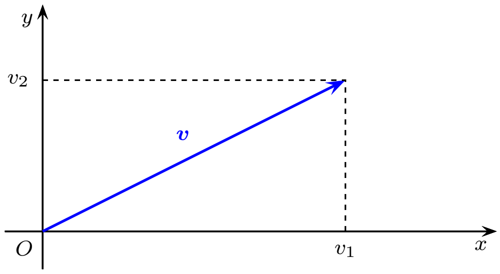


  - $a^2 = h^2 + d^2$ and $c^2 = h^2 + (b-d)^2$.

$$\begin{aligned} c^2 &= h^2 + (b-d)^2 \\ &= (a^2 - d^2) + (b-d)^2 \\ &= a^2 - d^2 + (b^2 - 2bd + d^2) \\ &= a^2 + b^2 - 2bd \\ &= a^2 + b^2 - 2b(a\cos\theta). \end{aligned}$$

Pythagoras' Theorem

- **Definition.** Let $\boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.



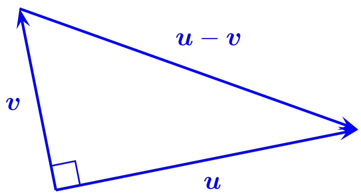
  - The **length** (or the **norm**) of $\boldsymbol{v}$ is $\|\boldsymbol{v}\| = \sqrt{v_1^2 + v_2^2}$.

  Let $\boldsymbol{u} = (u_1, u_2)$ and $\boldsymbol{v} = (v_1, v_2)$ be vectors in $\mathbb{R}^2$.

  - The **distance** between $\boldsymbol{u}$ and $\boldsymbol{v}$ is
    - $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$.

Angle between Vectors

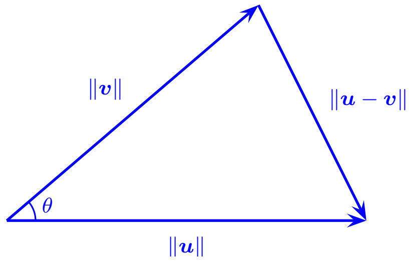
- When $\boldsymbol{u} = (u_1, u_2)$ and $\boldsymbol{v} = (v_1, v_2)$ are perpendicular?



$$
\begin{aligned}
&\boldsymbol{u} = (u_1, u_2) \text{ and } \boldsymbol{v} = (v_1, v_2) \text{ are perpendicular} \\
\Leftrightarrow\ & \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 = \|\boldsymbol{u} - \boldsymbol{v}\|^2 \\
\Leftrightarrow\ & (u_1^2 + u_2^2) + (v_1^2 + v_2^2) = (u_1 - v_1)^2 + (u_2 - v_2)^2 \\
\Leftrightarrow\ & u_1^2 + u_2^2 + v_1^2 + v_2^2 = u_1^2 + v_1^2 - 2u_1v_1 + u_2^2 + v_2^2 - 2u_2v_2 \\
\Leftrightarrow\ & 2u_1v_1 + 2u_2v_2 = 0 \\
\Leftrightarrow\ & u_1v_1 + u_2v_2 = 0.
\end{aligned}
$$

Angle between Vectors

- Let θ be the **angle** between $\boldsymbol{u} = (u_1, u_2)$ and $\boldsymbol{v} = (v_1, v_2)$.



Recall the **cosine rule**: $c^2 = a^2 + b^2 - 2ab\cos\theta$.

- ○ $\|\boldsymbol{u}-\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{u}\|\|\boldsymbol{v}\|\cos\theta$.

$$\begin{aligned}\cos\theta &= \frac{\|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - \|\boldsymbol{u}-\boldsymbol{v}\|^2}{2\|\boldsymbol{u}\|\|\boldsymbol{v}\|} \\ &= \frac{2(u_1v_1 + u_2v_2)}{2\|\boldsymbol{u}\|\|\boldsymbol{v}\|} = \frac{u_1v_1 + u_2v_2}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}.\end{aligned}$$

Definitions

- **Definition.** Let $\boldsymbol{u} = (u_1, u_2), \boldsymbol{v} = (v_1, v_2) \in \mathbb{R}^2$.
  - ○ Define the **dot product** (**inner product**) of $\boldsymbol{u}$ and $\boldsymbol{v}$:
    - $\boxed{\boldsymbol{u}\cdot\boldsymbol{v} = u_1v_1 + u_2v_2}$. $\quad (\boldsymbol{u}\cdot\boldsymbol{v} \in \mathbb{R})$ number, not a vector.

    Then the angle θ between $\boldsymbol{u}$ and $\boldsymbol{v}$ is given by
    - $\boxed{\theta = \cos^{-1}\left(\underbrace{\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}}_{\leq 1}\right)}$, $\quad \boldsymbol{u} \neq \boldsymbol{0}, \boldsymbol{v} \neq \boldsymbol{0}$.
- **Properties**:
  - ○ $\|\boldsymbol{u}\| = \sqrt{u_1^2 + u_2^2} = \sqrt{u_1u_1 + u_2u_2} = \sqrt{\boldsymbol{u}\cdot\boldsymbol{u}}$. $= \|u\|^2$
  - ○ $\boldsymbol{u}\cdot\boldsymbol{v} = 0 \Leftrightarrow \boldsymbol{u} \perp \boldsymbol{v}$ ($\boldsymbol{u}$ and $\boldsymbol{v}$ are perpendicular).
  - ○ $-1 \leq \dfrac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|} \leq 1 \Rightarrow |\boldsymbol{u}\cdot\boldsymbol{v}| \leq \|\boldsymbol{u}\|\|\boldsymbol{v}\|$.
  - ○ $\boldsymbol{u}\cdot\boldsymbol{v} = \underbrace{u_1v_1 + u_2v_2 = \begin{pmatrix} u_1 & u_2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}}_{\text{matrix multiplication.}} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$.

Definitions

- Let $\boldsymbol{u} = (u_1, \dots, u_n), \boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$.
  - The **dot product** (**inner product**) of $\boldsymbol{u}$ and $\boldsymbol{v}$ is
    - $\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + \cdots + u_n v_n$.
  - The **norm** (**length**) of $\boldsymbol{v}$ is
    - $\|\boldsymbol{v}\| = \sqrt{v_1^2 + \cdots + v_n^2}$.
  - $\boldsymbol{v}$ is called a **unit vector** if $\|\boldsymbol{v}\| = 1$.
  - The **distance** between $\boldsymbol{u}$ and $\boldsymbol{v}$ is
    - $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$.
  - The **angle** between $\boldsymbol{u}$ and $\boldsymbol{v}$ ($\boldsymbol{u} \neq \boldsymbol{0}$ and $\boldsymbol{v} \neq \boldsymbol{0}$) is
    - $\theta = \cos^{-1}\left(\dfrac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right), \qquad 0 \leq \theta \leq \pi.$

$\Downarrow$

$|u \cdot v| \leq \|u\| \, \|v\|$

Examples

- Let $\boldsymbol{u} = (1, -2, 2, -1)$ and $\boldsymbol{v} = (1, 0, 2, 0)$.
  - $\boldsymbol{u} \cdot \boldsymbol{v} = 1 \cdot 1 + (-2) \cdot 0 + 2 \cdot 2 + (-1) \cdot 0 = 5$.
  - $\|\boldsymbol{u}\| = \sqrt{1^2 + (-2)^2 + 2^2 + (-1)^2} = \sqrt{10}$.
  - $\|\boldsymbol{v}\| = \sqrt{1^2 + 0^2 + 2^2 + 0^2} = \sqrt{5}$.
  - $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\| = \|(0, -2, 0, -1)\|$.
    - $d(\boldsymbol{u}, \boldsymbol{v}) = \sqrt{0^2 + (-2)^2 + 0^2 + (-1)^2} = \sqrt{5}$.
  - Let θ be the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$.
    - $\cos\theta = \dfrac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|} = \dfrac{5}{\sqrt{10}\sqrt{5}} = \dfrac{1}{\sqrt{2}}$.
    - $\theta = \cos^{-1}\left(\dfrac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|}\right) = \cos^{-1}\left(\dfrac{1}{\sqrt{2}}\right) = \dfrac{\pi}{4}$.

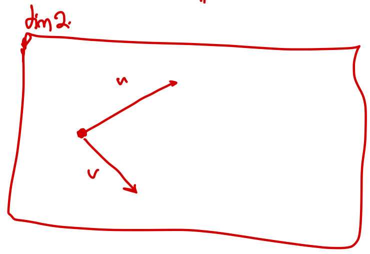
in $\mathbb{R}^4$?

$\Rightarrow$ translate to $\mathbb{R}^2$.

Vector space = span $\{u, v\}$

$u \nparallel v$.

dim 2



Dot Product and Matrix Multiplication

- Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^n$.
  - Suppose they are viewed as row vectors:
    - $\boldsymbol{u} = (u_1, \ldots, u_n)$, $\boldsymbol{v} = (v_1, \ldots, v_n)$.
      - $$\boldsymbol{u} \cdot \boldsymbol{v} = (u_1, \ldots, u_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \underset{1\times n}{\boldsymbol{u}}\,\underset{n\times 1}{\boldsymbol{v}^{\mathrm{T}}}. \in \mathbb{R}$$
  - Suppose they are viewed as column vectors:
    - $\boldsymbol{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$, $\boldsymbol{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$.
      - $$\boldsymbol{u} \cdot \boldsymbol{v} = (u_1, \ldots, u_n) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}.$$

Dot Product and Matrix Multiplication

- Let $\boldsymbol{A}$ be an $m \times n$ matrix and $\boldsymbol{B}$ an $n \times p$ matrix.
  - Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_m^{\mathrm{T}} \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_p \end{pmatrix}$,

    $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_p$ are column vectors in $\mathbb{R}^n$.
  - Recall that the (i, j)-entry of $\boldsymbol{AB}$ is $\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{b}_j$.
    - It is also given by $\boldsymbol{a}_i \cdot \boldsymbol{b}_j$.
  - $$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{b}_1 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{b}_j & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{b}_p \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ \boldsymbol{a}_i \cdot \boldsymbol{b}_1 & \cdots & \boxed{\boldsymbol{a}_i \cdot \boldsymbol{b}_j} & \cdots & \boldsymbol{a}_i \cdot \boldsymbol{b}_p \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_m \cdot \boldsymbol{b}_1 & \cdots & \boldsymbol{a}_m \cdot \boldsymbol{b}_j & \cdots & \boldsymbol{a}_m \cdot \boldsymbol{b}_p \end{pmatrix}$$

Properties

- **Theorem.** Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ and $c \in \mathbb{R}$.
  1. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$.
  2. $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$. Commutative,

     $\boldsymbol{w} \cdot (\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{w} \cdot \boldsymbol{u} + \boldsymbol{w} \cdot \boldsymbol{v}$.
  3. $(c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v}) = c(\boldsymbol{u} \cdot \boldsymbol{v})$.
  4. $\|c\boldsymbol{v}\| = |c| \, \|\boldsymbol{v}\|$.
  5. $\boldsymbol{v} \cdot \boldsymbol{v} \geq 0$ and $\boldsymbol{v} \cdot \boldsymbol{v} = 0 \Leftrightarrow \boldsymbol{v} = \boldsymbol{0}$.
- **Theorem.** Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$.
  1. $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$. (Cauchy-Schwarz inequality)
  2. $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$. (Triangle inequality)
  3. $d(\boldsymbol{u}, \boldsymbol{w}) \leq d(\boldsymbol{u}, \boldsymbol{v}) + d(\boldsymbol{v}, \boldsymbol{w})$. (Triangle inequality)

Properties

- **Proof.** We prove that $\boldsymbol{v} \cdot \boldsymbol{v} \geq 0$ & $\boldsymbol{v} \cdot \boldsymbol{v} = 0 \Leftrightarrow \boldsymbol{v} = \boldsymbol{0}$.
  - Let $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$, where $v_i \in \mathbb{R}$.
    - $\boldsymbol{v} \cdot \boldsymbol{v} = v_1^2 + v_2^2 + \cdots + v_n^2 \geq 0$.
  - For the second assertion:

$$\begin{aligned} \boldsymbol{v} \cdot \boldsymbol{v} = 0 &\Leftrightarrow v_1^2 + v_2^2 + \cdots + v_n^2 = 0 \\ &\Leftrightarrow v_1 = v_2 = \cdots = v_n = 0 \\ &\Leftrightarrow \boldsymbol{v} = (0, 0, \ldots, 0) = \boldsymbol{0}. \end{aligned}$$

- **Remark.** Note that $\|\boldsymbol{v}\| = \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$.
  - $\|\boldsymbol{v}\| \geq 0$ and $\|\boldsymbol{v}\| = 0 \Leftrightarrow \boldsymbol{v} = \boldsymbol{0}$.
- The proofs of other parts are left as exercises (Exercises 5.3 and 5.4).

Properties

- **Example** (Ex. 2.24g). If $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \mathbf{0}$, then $\boldsymbol{A} = \mathbf{0}$.

  Proof. Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_m \end{pmatrix}$, $\boldsymbol{a}_i$ is the ith row of $\boldsymbol{A}$.

  - $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} & \cdots & \boldsymbol{a}_m^{\mathrm{T}} \end{pmatrix}$. Then $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ has the form
    - $\begin{pmatrix} \boldsymbol{a}_1\boldsymbol{a}_1^{\mathrm{T}} & \cdots & \boldsymbol{a}_1\boldsymbol{a}_m^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_m\boldsymbol{a}_1^{\mathrm{T}} & \cdots & \boldsymbol{a}_m\boldsymbol{a}_m^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{a}_m \\ \vdots & \ddots & \vdots \\ \boldsymbol{a}_m \cdot \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_m \cdot \boldsymbol{a}_m \end{pmatrix}$

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \mathbf{0} &\Rightarrow \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 = \cdots = \boldsymbol{a}_m \cdot \boldsymbol{a}_m = 0 \\ &\Leftrightarrow \boldsymbol{a}_1 = \cdots = \boldsymbol{a}_m = \mathbf{0} \\ &\Leftrightarrow \boldsymbol{A} = \mathbf{0}. \end{aligned}$$

- **Exercise.** $\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = 0 \Leftrightarrow \boldsymbol{A} = \mathbf{0}$.

Orthogonal and Orthonormal Bases

Definitions

- Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^n$, and let θ (in radian) be the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$.
  - Suppose $\boldsymbol{u} \neq \mathbf{0}, \boldsymbol{v} \neq \mathbf{0}$. Then $\|\boldsymbol{u}\| \neq 0$, $\|\boldsymbol{v}\| \neq 0$.

$$\begin{aligned} \theta = \frac{\pi}{2} &\Leftrightarrow \cos\theta = 0 \\ &\Leftrightarrow \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|} = 0 \\ &\Leftrightarrow \boldsymbol{u} \cdot \boldsymbol{v} = 0. \end{aligned}$$

- **Definition**. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. They are said to be **orthogonal** if
  - $\boxed{\boldsymbol{u} \cdot \boldsymbol{v} = 0}$ denoted by $\boldsymbol{u} \perp \boldsymbol{v}$.

- **Example.** Let $\mathbf{0} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^n$. Then $\mathbf{0} \cdot \boldsymbol{v} = 0$.
  - $\mathbf{0} \in \mathbb{R}^n$ is orthogonal to every vector $\boldsymbol{v} \in \mathbb{R}^n$.

Definitions

one vector set is orthogonal

- **Definitions**. Let $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ be a subset of $\mathbb{R}^n$.
  - S is called **orthogonal** if every pair of distinct vectors in S are orthogonal:
    - $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ for all $i \neq j$. $i<j$ assume check $\binom{k}{2}$ times for k vectors.
  - S is called **orthonormal** if S is orthogonal and every vector in S is a unit vector.
    - $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$ $\|S\|=1$.
- **Remarks**: $v_i \cdot v_i = \|v_i\|^2 = 1^2 = 1$.
  - If S is orthonormal, then S is orthogonal.
  - If S is orthogonal, then a subset of S is orthogonal.
  - If S is orthonormal, then a subset of S is orthonormal.
  - If S is orthogonal, then $S \cup \{\boldsymbol{0}\}$ is also orthogonal. orthogonal to every other vector.
  - If S is orthonormal, then $\boldsymbol{0} \notin S$.

every vector in set must be unit vectors

Normalizing

- Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$ ($\boldsymbol{u}_i \cdot \boldsymbol{u}_j = \boldsymbol{0}$ for all $i \neq j$).
  - Set $\boldsymbol{v}_1 = \dfrac{\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|}, \boldsymbol{v}_2 = \dfrac{\boldsymbol{u}_2}{\|\boldsymbol{u}_2\|}, \dots, \boldsymbol{v}_k = \dfrac{\boldsymbol{u}_k}{\|\boldsymbol{u}_k\|}$. divide each vector by length
  - $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \left(\dfrac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}\right) \cdot \left(\dfrac{\boldsymbol{u}_j}{\|\boldsymbol{u}_j\|}\right) = \dfrac{\boldsymbol{u}_i \cdot \boldsymbol{u}_j}{\|\boldsymbol{u}_i\| \, \|\boldsymbol{u}_j\|}$.
    - If $i \neq j$, $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \dfrac{\boldsymbol{u}_i \cdot \boldsymbol{u}_j}{\|\boldsymbol{u}_i\| \, \|\boldsymbol{u}_j\|} = 0$. $\perp$
    - If $i = j$, $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \dfrac{\boldsymbol{u}_i \cdot \boldsymbol{u}_i}{\|\boldsymbol{u}_i\| \, \|\boldsymbol{u}_i\|} = \dfrac{\|\boldsymbol{u}_i\|^2}{\|\boldsymbol{u}_i\|^2} = 1$.
  - Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ is an orthonormal set.
- The process of converting an orthogonal set of nonzero vectors to an orthonormal set of vectors, $\boldsymbol{u}_i \mapsto \boldsymbol{v}_i = \dfrac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$, is called **normalizing**.

Examples

- Let $\boldsymbol{u}_1 = (1, 2, 2, -1)$ and $\boldsymbol{u}_2 = (1, 1, -1, 1)$.
  - $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 1 \cdot \underset{1}{\cancel{2}} + 2 \cdot 1 + 2 \cdot (-1) + (-1) \cdot 1 = 0.$

    Then $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is an orthogonal set in $\mathbb{R}^4$.
  - Let
    - $\boldsymbol{v}_1 = \dfrac{\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|} = \dfrac{\boldsymbol{u}_1}{\sqrt{10}} = \left(\frac{1}{\sqrt{10}}, \frac{2}{\sqrt{10}}, \frac{2}{\sqrt{10}}, -\frac{1}{\sqrt{10}}\right).$
    - $\boldsymbol{v}_2 = \dfrac{\boldsymbol{u}_2}{\|\boldsymbol{u}_2\|} = \dfrac{\boldsymbol{u}_2}{\sqrt{4}} = \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right).$

    Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is an orthonormal set in $\mathbb{R}^4$.

Examples

- Let $\boldsymbol{u}_1 = (2, 0, 0)$, $\boldsymbol{u}_2 = (0, 1, 1)$, $\boldsymbol{u}_3 = (0, 1, -1)$.
  - $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 2 \cdot 0 + 0 \cdot 1 + 0 \cdot 1 = 0.$
  - $\boldsymbol{u}_1 \cdot \boldsymbol{u}_3 = 2 \cdot 0 + 0 \cdot 1 + 0 \cdot (-1) = 0.$
  - $\boldsymbol{u}_2 \cdot \boldsymbol{u}_3 = 0 \cdot 0 + 1 \cdot 1 + 1 \cdot (-1) = 0.$
    - Then $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthogonal set in $\mathbb{R}^3$.
  - Let
    - $\boldsymbol{v}_1 = \dfrac{\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|} = \dfrac{\boldsymbol{u}_1}{2} = (1, 0, 0).$
    - $\boldsymbol{v}_2 = \dfrac{\boldsymbol{u}_2}{\|\boldsymbol{u}_2\|} = \dfrac{\boldsymbol{u}_2}{\sqrt{2}} = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$
    - $\boldsymbol{v}_3 = \dfrac{\boldsymbol{u}_3}{\|\boldsymbol{u}_3\|} = \dfrac{\boldsymbol{u}_3}{\sqrt{2}} = \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$

    Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthonormal set in $\mathbb{R}^3$.

Examples

- Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ (column vectors) be a subset of $\mathbb{R}^n$.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_k^{\mathrm{T}} \end{pmatrix}$.
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_1 \cdot \boldsymbol{v}_k \\ \vdots & \ddots & \vdots \\ \boldsymbol{v}_k \cdot \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{v}_k \end{pmatrix} = (\boldsymbol{v}_i \cdot \boldsymbol{v}_j)_{k \times k}$.

all non-diagonal are 0.

$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \text{ is orthogonal} \Leftrightarrow \boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0 \text{ for all } i \neq j$$
$$\Leftrightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \text{ is diagonal.}$$

$$\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\} \text{ is orthonormal} \Leftrightarrow \boldsymbol{v}_i \cdot \boldsymbol{v}_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j, \end{cases}$$
$$\Leftrightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_k.$$

Examples

- Consider the standard basis $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ for $\mathbb{R}^n$,

  where $\boldsymbol{e}_i$ is the (column) vector of length n whose ith coordinate is 1 and 0 elsewhere.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{pmatrix}$. Then $\boldsymbol{A} = \boldsymbol{I}_n$.
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_n^{\mathrm{T}}\boldsymbol{I}_n = \boldsymbol{I}_n\boldsymbol{I}_n = \boldsymbol{I}_n$.

    Hence, $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is an orthonormal set.

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ be an orthonormal subset of $\mathbb{R}^n$.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{pmatrix}$. Then
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_n \Rightarrow \boldsymbol{A}$ is invertible.

  $\therefore$ $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ is a basis for $\mathbb{R}^n$.

Linear Independency

- **Theorem**. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be an orthogonal set of nonzero vectors in $\mathbb{R}^n$.
  - Then S is linearly independent.
- **Proof.** Suppose $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$. For any i,

$$\boldsymbol{v}_i \cdot (c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = \boldsymbol{v}_i \cdot \mathbf{0} = 0.$$

$$\begin{aligned} 0 &= \boldsymbol{v}_i \cdot (c_1\boldsymbol{v}_1 + \boldsymbol{c}_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) \\ &= \boldsymbol{v}_i \cdot (c_1\boldsymbol{v}_1) + \boldsymbol{v}_i \cdot (c_2\boldsymbol{v}_2) + \cdots + \boldsymbol{v}_i \cdot (c_k\boldsymbol{v}_k) \\ &= c_1(\boldsymbol{v}_i \cdot \boldsymbol{v}_1) + c_2(\boldsymbol{v}_i \cdot \boldsymbol{v}_2) + \cdots + c_k(\boldsymbol{v}_i \cdot \boldsymbol{v}_k). \end{aligned}$$

  - Recall that $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ if $i \neq j$. Then
    - the above equation is reduced to $c_i(\boldsymbol{v}_i \cdot \boldsymbol{v}_i) = 0$.
      - $\boldsymbol{v}_i \neq \mathbf{0} \Rightarrow \boldsymbol{v}_i \cdot \boldsymbol{v}_i > 0 \Rightarrow c_i = 0$.
  - Therefore, S is linearly independent.

Definition

- **Corollary**. An orthonormal set is linearly independent.
- **Definition**. Let S be a basis for a vector space.
  - S is an **orthogonal basis** if it is orthogonal.
  - S is an **orthonormal basis** if it is orthonormal.
- **Remarks**.
  - Suppose S is a subset of a vector space V. To check if S is a basis for V, it suffices to check any two of the following three properties:
    - $|S| = \dim(V)$;
    - $\operatorname{span}(S) = V$;
    - S is linearly independent.
  - $\mathbf{0} \notin S \subseteq V$ is an orthogonal (orthonormal) basis:
    - $|S| = \dim V$ or $\operatorname{span}(S) = V$; and
    - S orthogonal (respectively, orthonormal).

Properties

- What are the advantages of orthogonal (orthonormal) basis?
- Let $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ be a basis for a vector space V.
  - For any $\boldsymbol{w} \in V$, there exist unique $c_1, \dots, c_k$ such that
    - $\boldsymbol{w} = c_1\boldsymbol{u}_1 + \cdots + c_k\boldsymbol{u}_k$.

    $(\boldsymbol{w})_S = (c_1, \dots, c_k)$, coordinate vector relative to S.
  - Solve the linear system $\begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k \end{pmatrix} [\boldsymbol{w}]_S = \boldsymbol{w}$.
  - Suppose that S is an orthogonal basis. For any i,

$$\begin{aligned} \boldsymbol{w} \cdot \boldsymbol{u}_i &= (c_1\boldsymbol{u}_1 + \cdots + c_k\boldsymbol{u}_k) \cdot \boldsymbol{u}_i \\ &= c_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_i) + \cdots + c_k(\boldsymbol{u}_k \cdot \boldsymbol{u}_i) \\ &= c_i(\boldsymbol{u}_i \cdot \boldsymbol{u}_i) \\ c_i &= \frac{\boldsymbol{w} \cdot \boldsymbol{u}_i}{\boldsymbol{u}_i \cdot \boldsymbol{u}_i} = \frac{\boldsymbol{w} \cdot \boldsymbol{u}_i}{\|\boldsymbol{u}_i\|^2}. \end{aligned}$$

Properties

- **Theorem**. Let $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ be an **orthogonal** basis for a vector space V. For any $\boldsymbol{w} \in V$,
  - $(\boldsymbol{w})_S = \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}, \dots, \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\right)$.
  - $\boldsymbol{w} = \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\right)\boldsymbol{u}_1 + \cdots + \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\right)\boldsymbol{u}_k$.
- If $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is an orthonormal basis, then
  - $\boldsymbol{v}_i \cdot \boldsymbol{v}_i = \|\boldsymbol{v}_i\|^2 = 1$ for all $i = 1, \dots, n$.
- **Theorem**. Let $S = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ be an **orthonormal** basis for a vector space V. For any $\boldsymbol{w} \in V$,
  - $(\boldsymbol{w})_S = (\boldsymbol{w} \cdot \boldsymbol{v}_1, \dots, \boldsymbol{w} \cdot \boldsymbol{v}_k)$,
  - $\boldsymbol{w} = (\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$.

Examples

- Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$, $\boldsymbol{v}_1 = \left(\frac{3}{5}, \frac{4}{5}\right)$, $\boldsymbol{v}_2 = \left(\frac{4}{5}, -\frac{3}{5}\right)$.
  - $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = \frac{3}{5} \cdot \frac{4}{5} + \frac{4}{5}(-\frac{3}{5}) = 0.$

    $\boldsymbol{v}_1 \cdot \boldsymbol{v}_1 = (\frac{3}{5})^2 + (\frac{4}{5})^2 = 1.$

    $\boldsymbol{v}_2 \cdot \boldsymbol{v}_2 = (\frac{4}{5})^2 + (-\frac{3}{5})^2 = 1.$
    - S is an orthonormal basis for $\mathbb{R}^2$.
  - For every $\boldsymbol{w} = (x, y) \in \mathbb{R}^2$.
    - $\boldsymbol{w} \cdot \boldsymbol{v}_1 = \dfrac{3x + 4y}{5}$; $\boldsymbol{w} \cdot \boldsymbol{v}_2 = \dfrac{4x - 3y}{5}$.
    - $(\boldsymbol{w})_S = \left(\dfrac{3x + 4y}{5}, \dfrac{4x - 3y}{5}\right)$.
    - $\boldsymbol{w} = \dfrac{3x + 4y}{5}\,\boldsymbol{v}_1 + \dfrac{4x - 3y}{5}\,\boldsymbol{v}_2$.

Examples

- Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, where
  $\boldsymbol{u}_1 = (1, 1, 1)$, $\boldsymbol{u}_2 = (1, 0, -1)$, $\boldsymbol{u}_3 = (1, -2, 1)$.
  - $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 1 \cdot 1 + 1 \cdot 0 + 1 \cdot (-1) = 0.$
    $\boldsymbol{u}_1 \cdot \boldsymbol{u}_3 = 1 \cdot 1 + 1 \cdot (-2) + 1 \cdot 1 = 0.$
    $\boldsymbol{u}_2 \cdot \boldsymbol{u}_3 = 1 \cdot 1 + 0 \cdot (-2) + (-1) \cdot 1 = 0.$
    - S is an orthogonal basis for $\mathbb{R}^3$.
  - Let $\boldsymbol{w} = (1, -1, 0) \in \mathbb{R}^3$. Then
    - $\dfrac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} = \dfrac{1 \cdot 1 + (-1) \cdot 1 + 0 \cdot 1}{1^2 + 1^2 + 1^2} = 0.$
    - $\dfrac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2} = \dfrac{1 \cdot 1 + (-1) \cdot 0 + 0 \cdot (-1)}{1^2 + 0^2 + (-1)^2} = \dfrac{1}{2}.$
    - $\dfrac{\boldsymbol{w} \cdot \boldsymbol{u}_3}{\boldsymbol{u}_3 \cdot \boldsymbol{u}_3} = \dfrac{1 \cdot 1 + (-1) \cdot (-2) + 0 \cdot 1}{1^2 + (-2)^2 + 1} = \dfrac{1}{2}.$

Examples

- Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, where
  $\boldsymbol{u}_1 = (1,1,1)$, $\boldsymbol{u}_2 = (1,0,-1)$, $\boldsymbol{u}_3 = (1,-2,1)$.
  - $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 1\cdot 1 + 1 \cdot 0 + 1 \cdot (-1) = 0.$
    $\boldsymbol{u}_1 \cdot \boldsymbol{u}_3 = 1\cdot 1 + 1 \cdot (-2) + 1 \cdot 1 = 0.$
    $\boldsymbol{u}_2 \cdot \boldsymbol{u}_3 = 1\cdot 1 + 0 \cdot (-2) + (-1) \cdot 1 = 0.$
    - S is an orthogonal basis for $\mathbb{R}^3$.
  - Let $\boldsymbol{w} = (1,-1,0) \in \mathbb{R}^3$. Then

$$\begin{aligned}(\boldsymbol{w})_S &= \left(\frac{\boldsymbol{w}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}, \frac{\boldsymbol{w}\cdot\boldsymbol{u}_2}{\boldsymbol{u}_2\cdot\boldsymbol{u}_2}, \frac{\boldsymbol{w}\cdot\boldsymbol{u}_3}{\boldsymbol{u}_3\cdot\boldsymbol{u}_3}\right) \\ &= \left(0, \frac{1}{2}, \frac{1}{2}\right).\end{aligned}$$

    - $\boldsymbol{w} = 0\boldsymbol{u}_1 + \frac{1}{2}\boldsymbol{u}_2 + \frac{1}{2}\boldsymbol{u}_3.$

Orthogonality

- **Definition**. Let V be a subspace of $\mathbb{R}^n$.
  - $\boldsymbol{u} \in \mathbb{R}^n$ is **orthogonal** (**perpendicular**) to V if $\boldsymbol{u}$ is orthogonal to every vector in V.
    - that is, $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ for all $\boldsymbol{v} \in V$.
- **Example**. Let $V = \{(x,y,z) \mid ax + by + cz = 0\}$,
  where a, b, c are not all zero.
  - Let $\boldsymbol{n} = (a,b,c)$. Then for any $\boldsymbol{v} = (x,y,z) \in V$,
    - $\boldsymbol{n} \cdot \boldsymbol{v} = (a,b,c) \cdot (x,y,z) = ax + by + cz = 0.$
  - $\boldsymbol{n} = (a,b,c)$ is a **normal vector** of the plane V.
  - $V = \{(x,y,z) \mid (a,b,c) \cdot (x,y,z) = 0\}$.
    - $V = \{\boldsymbol{u} \in \mathbb{R}^3 \mid \boldsymbol{n} \cdot \boldsymbol{u} = 0\}$.

    V is the set of all vectors orthogonal to $\boldsymbol{n} = (a,b,c)$.

Orthogonality

- **Theorem.** Let $V = \mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be a vector space.
  - $\boldsymbol{w}$ is orthogonal to V $\Leftrightarrow$ $\boldsymbol{w} \cdot \boldsymbol{v}_i = 0$ for all $i = 1, \ldots, k$.

  $(\Rightarrow)$ is trivial because $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in V$.

  $(\Leftarrow)$ Suppose $\boldsymbol{w} \cdot \boldsymbol{v}_i = 0$ for all $i = 1, \ldots, k$.
  - For any $\boldsymbol{v} \in V$, there exist $c_1, \ldots, c_k \in \mathbb{R}$ such that
    - $\boldsymbol{v} = c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k$.

$$\begin{aligned}\boldsymbol{w} \cdot \boldsymbol{v} &= \boldsymbol{w} \cdot (c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k)\\ &= c_1(\boldsymbol{w} \cdot \boldsymbol{v}_1) + \cdots + c_k(\boldsymbol{w} \cdot \boldsymbol{v}_k)\\ &= c_1 0 + \cdots + c_k 0 = 0.\end{aligned}$$

  - $\boldsymbol{w}$ is orthogonal to all $\boldsymbol{v} \in V$; so $\boldsymbol{w}$ is orthogonal to V.
- **Exercise.** Let W be a subspace of $\mathbb{R}^n$.
  - Prove that $W^\perp = \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v}$ is orthogonal to $W\}$ is a subspace of $\mathbb{R}^n$.

*

Examples

- **Example.** Let $V = \mathrm{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.

  $\boldsymbol{v}_1 = (1, 1, 1, 0)$ and $\boldsymbol{v}_2 = (0, -1, -1, 1)$.

  Let $\boldsymbol{w} = (w, x, y, z) \in \mathbb{R}^4$. Then

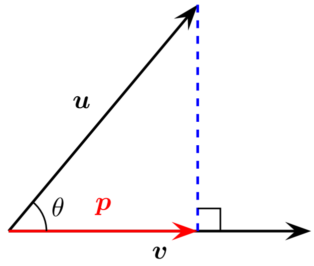
$$\begin{aligned}&\boldsymbol{w} \text{ is orthogonal to } V\\ \Leftrightarrow\ &\boldsymbol{w} \text{ is orthogonal to } \boldsymbol{v}_1 \text{ and } \boldsymbol{v}_2\\ \Leftrightarrow\ &\boldsymbol{w} \cdot \boldsymbol{v}_1 = \boldsymbol{w} \cdot \boldsymbol{v}_2 = 0\\ \Leftrightarrow\ &\begin{cases} w + x + y \quad = 0,\\ -x - y + z = 0.\end{cases}\\ \Leftrightarrow\ &(w, x, y, z) = (-t, -s + t, s, t).\end{aligned}$$

  - $\boldsymbol{w}$ is orthogonal to V

    $\Leftrightarrow \boldsymbol{w} = (-t, -s + t, s, t)$ for some $s, t \in \mathbb{R}$
    $\Leftrightarrow \boldsymbol{w} \in \mathrm{span}\{(0, -1, 1, 0), (-1, 1, 0, 1)\}$.

Projection

- Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^n$, $\boldsymbol{v} \neq \boldsymbol{0}$.

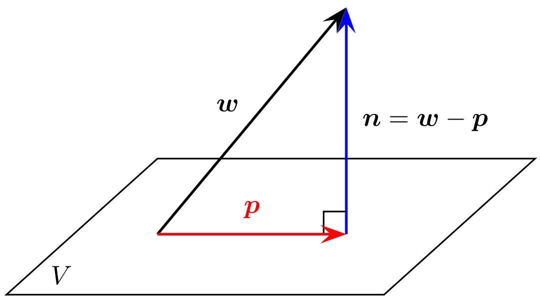


  - Let $\boldsymbol{p}$ be the projection of $\boldsymbol{u}$ onto $\boldsymbol{v}$. Then
    - $\boldsymbol{p} = \|\boldsymbol{p}\| \dfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$ and $\|\boldsymbol{p}\| = \|\boldsymbol{u}\| \cos\theta$.

    Then $\boldsymbol{p} = \|\boldsymbol{u}\| \dfrac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|} \dfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \left(\dfrac{\boldsymbol{u}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v}$.
    - If $\boldsymbol{v}$ is a unit vector, then $\boldsymbol{p} = (\boldsymbol{u}\cdot\boldsymbol{v})\boldsymbol{v}$.

Projection

- Let V be a vector subspace of $\mathbb{R}^n$ and $\boldsymbol{w} \in \mathbb{R}^n$.



  - Can we find a vector $\boldsymbol{p} \in V$ such that $\boldsymbol{n} = \boldsymbol{w} - \boldsymbol{p}$ is orthogonal to V?
    - Exercise 5.18 states that such $\boldsymbol{p}$ exists and unique.
    - $\boldsymbol{p}$ is called the **projection** of $\boldsymbol{w}$ onto V.

Projection

- Let V be a vector subspace of $\mathbb{R}^n$ and $\boldsymbol{w} \in \mathbb{R}^n$.
  - Assume that $\boldsymbol{w} = \boldsymbol{p} + \boldsymbol{n}$, where
    - $\boldsymbol{p} \in V$ and $\boldsymbol{n}$ is orthogonal to V.
  - Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be an orthonormal basis for V.
    - $\boldsymbol{n} = \boldsymbol{w} - \boldsymbol{p}$ is orthogonal to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$.
    - $(\boldsymbol{w} - \boldsymbol{p}) \cdot \boldsymbol{v}_i = 0 \Leftrightarrow \boldsymbol{w} \cdot \boldsymbol{v}_i = \boldsymbol{p} \cdot \boldsymbol{v}_i$ for all i.
  - Recall that $\boldsymbol{p} \in V$ can be written as
    - $\boldsymbol{p} = (\boldsymbol{p} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{p} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$.

  $\therefore$ $\boldsymbol{p} = (\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$.
  - Conversely, if $\boldsymbol{p} = (\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$,

$$(\boldsymbol{w} - \boldsymbol{p}) \cdot \boldsymbol{v}_i = \boldsymbol{w} \cdot \boldsymbol{v}_i - \boldsymbol{p} \cdot \boldsymbol{v}_i = \boldsymbol{w} \cdot \boldsymbol{v}_i - \boldsymbol{w} \cdot \boldsymbol{v}_i = 0.$$

  $\therefore$ $\boldsymbol{n} = \boldsymbol{w} - \boldsymbol{p}$ is orthogonal to V.

Projection

- **Theorem**. Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be an **orthonormal basis** for a vector space V. The projection of $\boldsymbol{w}$ onto V is
  - $(\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{w} \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$.
- Suppose $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ is an orthogonal basis for V.
  - Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an orthonormal basis for V,
    - where $\boldsymbol{v}_i = \dfrac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$, $i = 1, 2, \ldots, k$.
  - The projection of $\boldsymbol{w}$ onto V is

$$\begin{aligned} &(\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k \\ &= \left(\boldsymbol{w} \cdot \frac{\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|}\right)\frac{\boldsymbol{u}_1}{\|\boldsymbol{u}_1\|} + \cdots + \left(\boldsymbol{w} \cdot \frac{\boldsymbol{u}_k}{\|\boldsymbol{u}_k\|}\right)\frac{\boldsymbol{u}_k}{\|\boldsymbol{u}_k\|} \\ &= \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\right)\boldsymbol{u}_1 + \cdots + \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\right)\boldsymbol{u}_k. \end{aligned}$$

Projection

- **Theorem.** Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be an **orthogonal basis** for a vector space V. The projection of $\boldsymbol{w}$ onto V is
  - $\left(\dfrac{\boldsymbol{w}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\right)\boldsymbol{u}_1 + \left(\dfrac{\boldsymbol{w}\cdot\boldsymbol{u}_2}{\boldsymbol{u}_2\cdot\boldsymbol{u}_2}\right)\boldsymbol{u}_2 + \cdots + \left(\dfrac{\boldsymbol{w}\cdot\boldsymbol{u}_k}{\boldsymbol{u}_k\cdot\boldsymbol{u}_k}\right)\boldsymbol{u}_k.$

  It is the sum of projections of $\boldsymbol{w}$ onto $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$.

- **Example.** Let $V = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$, where

  $\boldsymbol{u}_1 = (1, 0, 1)$ and $\boldsymbol{u}_2 = (1, 0, -1)$.

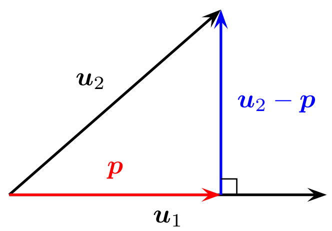
  - $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 1\cdot 1 + 0\cdot 0 + 1\cdot(-1) = 0.$

  The projection of $\boldsymbol{w} = (1, 1, 0)$ onto V is

$$
\begin{aligned}
&\frac{\boldsymbol{w}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_2}\boldsymbol{u}_1 + \frac{\boldsymbol{w}\cdot\boldsymbol{u}_2}{\boldsymbol{u}_2\cdot\boldsymbol{u}_2}\boldsymbol{u}_2 \\
&= \frac{1}{2}(1, 0, 1) + \frac{1}{2}(1, 0, -1) = (1, 0, 0).
\end{aligned}
$$

Gram-Schmidt Process

- How to find an orthogonal basis for a given vector space?
- $\dim V = 1$: Any basis is orthogonal.
- Suppose $\dim V = 2$. Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be a basis for V.



  - The projection of $\boldsymbol{u}_2$ onto $\boldsymbol{u}_1$: $\boldsymbol{p} = \dfrac{\boldsymbol{u}_2\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1.$
    - $\boldsymbol{u}_2 - \boldsymbol{p} \neq \boldsymbol{0}$ and it is orthogonal to $\boldsymbol{u}_1$.
  - $\{\boldsymbol{u}_1, \boldsymbol{u}_2 - \boldsymbol{p}\}$ is an orthogonal basis for V.

Gram-Schmidt Process

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be a basis for a vector space V.
  - We obtain an orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ for V:

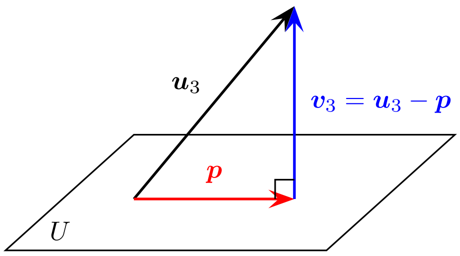
$$\begin{aligned}\boldsymbol{v}_1 &= \boldsymbol{u}_1\\ \boldsymbol{v}_2 &= \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1.\end{aligned}$$

- **Example**. Let $V = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$.
  - $\boldsymbol{u}_1 = (1, -1, 2)$ and $\boldsymbol{u}_2 = (2, 1, 0)$.

$$\begin{aligned}\boldsymbol{v}_1 &= \boldsymbol{u}_1 = (1, -1, 2)\\ \boldsymbol{v}_2 &= \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1\\ &= (2, 1, 0) - \frac{1}{6}(1, -1, 2) = \left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right).\end{aligned}$$

Gram-Schmidt Process

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ be basis for a vector space V.
  - Let $U = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$. Then $\dim(U) = 2$ and
    - U has an orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.



  - $\boldsymbol{v}_3 = \boldsymbol{u}_3 - \left(\frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 + \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2\right)$.

Gram-Schmidt Process

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ be basis for a vector space V, where
  - $\boldsymbol{u}_1 = (1, -1, 2)$, $\boldsymbol{u}_2 = (2, 1, 0)$ and $\boldsymbol{u}_3 = (0, 0, 1)$.

  $U = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ has an orthogonal basis:

$$\boldsymbol{v}_1 = \boldsymbol{u}_1 = (1, -1, 2)$$
$$\boldsymbol{v}_2 = \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 = \left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right).$$

  Use $\boldsymbol{v}_3 = \boldsymbol{u}_3 - \boldsymbol{p}$, where $\boldsymbol{p}$ is the projection of $\boldsymbol{u}_3$ onto U

$$\begin{aligned}\boldsymbol{v}_3 &= \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 \\ &= (0, 0, 1) - \frac{2}{6}(1, -1, 2) - \frac{-1/3}{29/6}\left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right) \\ &= \left(-\frac{6}{29}, \frac{12}{29}, \frac{9}{29}\right).\end{aligned}$$

Gram-Schmidt Process

- **(Gram-Schmidt Process).** Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for a vector space V. Define

$$\begin{aligned}\boldsymbol{v}_1 &= \boldsymbol{u}_1 \\ \boldsymbol{v}_2 &= \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 \\ \boldsymbol{v}_3 &= \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 \\ \vdots \quad & \quad \vdots \\ \boldsymbol{v}_k &= \boldsymbol{u}_k - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 - \cdots - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_{k-1}}{\boldsymbol{v}_{k-1} \cdot \boldsymbol{v}_{k-1}}\boldsymbol{v}_{k-1}\end{aligned}$$

  - Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an orthogonal basis for V.

  Define $\boldsymbol{w}_1 = \dfrac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}, \boldsymbol{w}_2 = \dfrac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}, \ldots, \boldsymbol{w}_k = \dfrac{\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|}$.

  - Then $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ is an orthonormal basis for V.

Example

- Let $V = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, where
  - $\boldsymbol{u}_1 = (1,1,1,1)$, $\boldsymbol{u}_2 = (1,2,2,1)$, $\boldsymbol{u}_3 = (2,3,1,6)$.

$$\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{u}_1 = (1,1,1,1) \\
\boldsymbol{v}_2 &= \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 \\
&= (1,2,2,1) - \frac{6}{4}(1,1,1,1) = \left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right) \\
\boldsymbol{v}_3 &= \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 \\
&= (2,3,1,6) - \frac{12}{4}(1,2,2,1) - \frac{-2}{1}\left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right) \\
&= (-2,1,-1,2).
\end{aligned}$$

Example

- Let $V = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, where
  - $\boldsymbol{u}_1 = (1,1,1,1)$, $\boldsymbol{u}_2 = (1,2,2,1)$, $\boldsymbol{u}_3 = (2,3,1,6)$.
    - Orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$.
      - $\boldsymbol{v}_1 = (1,1,1,1)$,
      - $\boldsymbol{v}_2 = \left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$.
      - $\boldsymbol{v}_3 = (-2,1,-1,2)$.
    - Orthonormal basis $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$.
      - $\boldsymbol{w}_1 = \dfrac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} = \dfrac{1}{2}(1,1,1,1)$.
      - $\boldsymbol{w}_2 = \dfrac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} = \dfrac{1}{2}(-1,1,1,-1)$.
      - $\boldsymbol{w}_3 = \dfrac{\boldsymbol{v}_3}{\|\boldsymbol{v}_3\|} = \dfrac{1}{\sqrt{10}}(-2,1,-1,2)$.

Decomposition

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for V.
  - Orthonormal basis: $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ such that
    - $\mathrm{span}\{\boldsymbol{w}_1\} = \mathrm{span}\{\boldsymbol{u}_1\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$.
    - ....
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$.

    Therefore,

    - $\boldsymbol{u}_1 = b_{11}\boldsymbol{w}_1$;
    - $\boldsymbol{u}_2 = b_{12}\boldsymbol{w}_1 + b_{22}\boldsymbol{w}_2$;
    - $\boldsymbol{u}_3 = b_{13}\boldsymbol{w}_1 + b_{23}\boldsymbol{w}_2 + b_{33}\boldsymbol{w}_3$;
    - ....
    - $\boldsymbol{u}_k = b_{1k}\boldsymbol{w}_1 + b_{2k}\boldsymbol{w}_2 + b_{23}\boldsymbol{w}_3 + \cdots + c_{kk}\boldsymbol{w}_k$.

Decomposition

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for V.
  - Orthonormal basis: $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ such that
    - $\mathrm{span}\{\boldsymbol{w}_1\} = \mathrm{span}\{\boldsymbol{u}_1\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$.
    - ....
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$.

    Therefore,

    - $\boldsymbol{u}_1 = \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k \end{pmatrix} \begin{pmatrix} b_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

Decomposition

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for V.
  - Orthonormal basis: $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ such that
    - $\mathrm{span}\{\boldsymbol{w}_1\} = \mathrm{span}\{\boldsymbol{u}_1\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$.
    - ....
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$.

    Therefore,

    - $\boldsymbol{u}_2 = \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k \end{pmatrix} \begin{pmatrix} b_{12} \\ b_{22} \\ \vdots \\ 0 \end{pmatrix}$

Decomposition

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for V.
  - Orthonormal basis: $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ such that
    - $\mathrm{span}\{\boldsymbol{w}_1\} = \mathrm{span}\{\boldsymbol{u}_1\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$;
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$.
    - ....
    - $\mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$.

    Therefore,

    - $\boldsymbol{u}_k = \begin{pmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k \end{pmatrix} \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{kk} \end{pmatrix}$

Decomposition

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for V.
  - Orthonormal basis: $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ such that
    - $\operatorname{span}\{\boldsymbol{w}_1\} = \operatorname{span}\{\boldsymbol{u}_1\}$;
    - $\operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\} = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$;
    - $\operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\} = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$.
    - ....
    - $\operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$.

    Therefore, we can write $\begin{pmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_k\end{pmatrix}$ as

    - $\begin{pmatrix}\boldsymbol{w}_1 & \cdots & \boldsymbol{w}_k\end{pmatrix}\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ 0 & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & b_{kk}\end{pmatrix}$

      $=$ orthonormal columns $\times$ upper triangular.

Decomposition

- **Theorem**. Let $\boldsymbol{A}$ be an $m \times n$ matrix whose columns are **linearly independent**. Then there exist
  - An $m \times n$ matrix $\boldsymbol{Q}$ whose columns form an **orthonormal** set, and
  - An invertible $n \times n$ **upper triangular** matrix $\boldsymbol{R}$

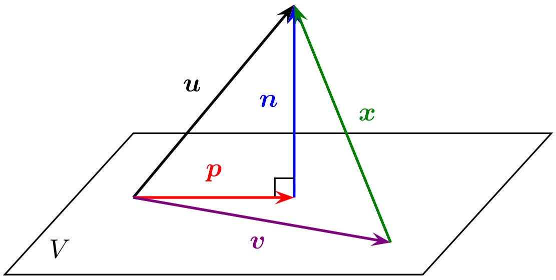
  such that $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$.

- **Application**: Solve linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.
  1. $(\boldsymbol{Q}\boldsymbol{R})\boldsymbol{x} = \boldsymbol{b}$.
  2. $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} \Rightarrow \boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$.
  3. Solve $\boldsymbol{x}$ by back-substitution.

- **Remark**. One may choose $\boldsymbol{R}$ so that the diagonal entries are all positive. Can you prove it?

Projection

- Recall the projection of $\boldsymbol{u}$ onto a vector space V:

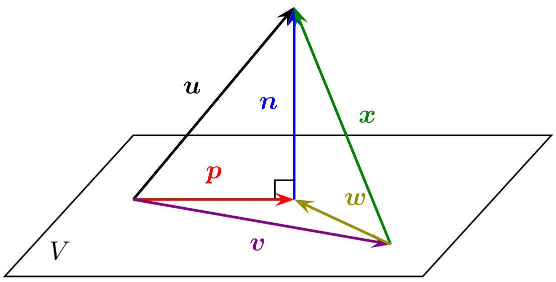


  - Among all the vectors $\boldsymbol{v} \in V$, the one with the **shortest** distance to $\boldsymbol{u}$ is $\boldsymbol{p}$, the **projection** of $\boldsymbol{u}$ onto V.
    - $d(\boldsymbol{u}, \boldsymbol{p}) \le d(\boldsymbol{u}, \boldsymbol{v})$ for all $\boldsymbol{v} \in V$.

    $\boldsymbol{p}$ is the **best approximation** of $\boldsymbol{u}$ in V.

Best Approximation

- Take $\boldsymbol{v} \in V$. Set $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v}$.
  - Need to show that $\|\boldsymbol{n}\| \le \|\boldsymbol{x}\|$.



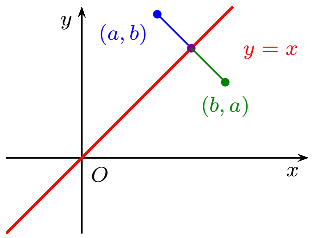
  - Let $\boldsymbol{w} = \boldsymbol{p} - \boldsymbol{v}$. Note that $\boldsymbol{n}$ is orthogonal to $\boldsymbol{w}$.

$$\begin{aligned}\|\boldsymbol{x}\|^2 &= \boldsymbol{x} \cdot \boldsymbol{x} = (\boldsymbol{n} + \boldsymbol{w}) \cdot (\boldsymbol{n} + \boldsymbol{w}) \\ &= \boldsymbol{n} \cdot \boldsymbol{n} + 2(\boldsymbol{n} \cdot \boldsymbol{w}) + \boldsymbol{w} \cdot \boldsymbol{w} \\ &= \|\boldsymbol{n}\|^2 + \|\boldsymbol{w}\|^2 \ge \|\boldsymbol{n}\|^2.\end{aligned}$$

Best Approximation

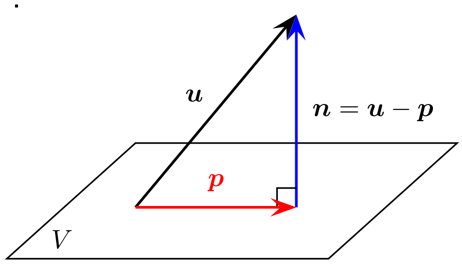
- **Theorem**. Let V be a subspace of $\mathbb{R}^n$.
  - For $\boldsymbol{u} \in \mathbb{R}^n$, let $\boldsymbol{p}$ be the projection of $\boldsymbol{u}$ onto V.
    - Then $\boldsymbol{p}$ is the **best approximation** of $\boldsymbol{u}$ in V.
      - $d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{v})$ for all $\boldsymbol{v} \in V$.

    Moreover, $d(\boldsymbol{u}, \boldsymbol{p}) = d(\boldsymbol{u}, \boldsymbol{v}) \Leftrightarrow \boldsymbol{v} = \boldsymbol{p}$.
- **Example**. Best approximation of (a, b) in $\operatorname{span}\{(1, 1)\}$.
  - $\boldsymbol{p} = \dfrac{(a, b) \cdot (1, 1)}{(1, 1) \cdot (1, 1)}(1, 1) = \dfrac{a + b}{2}(1, 1)$.



Examples

- Consider the plane $V = \{(x, y, z) \mid ax + by + cz = 0\}$.
  - Normal vector (a, b, c) is orthogonal to V.
    - Let $\boldsymbol{u} = (x_0, y_0, z_0) \in \mathbb{R}^3$.



- $\boldsymbol{n} = \boldsymbol{u} - \boldsymbol{p}$ is the projection of $\boldsymbol{u}$ onto (a, b, c):

$$\frac{(x_0, y_0, z_0) \cdot (a, b, c)}{\|(a, b, c)\|^2}(a, b, c) = \frac{ax_0 + by_0 + cz_0}{\|(a, b, c)\|} \frac{(a, b, c)}{\|(a, b, c)\|}.$$

$$\therefore \quad \|\boldsymbol{n}\| = \frac{|ax_0 + by_0 + cz_0|}{\sqrt{a^2 + b^2 + c^2}}.$$

Examples

- Let $V = \text{span}\{(1,0,1),(1,1,1)\}$.
  - Find the shortest distance from $\boldsymbol{u} = (1,2,3)$ to V.

1. Find an orthogonal basis:
   - $(1,0,1)$ and
   - $(1,1,1) - \dfrac{(1,1,1)\cdot(1,0,1)}{(1,0,1)\cdot(1,0,1)}(1,1,1) = (0,1,0)$.
2. Find the projection of $(1,2,3)$ onto V:
   - $\dfrac{(1,2,3)\cdot(1,0,1)}{(1,0,1)\cdot(1,0,1)} = 2$
   - $\dfrac{(1,2,3)\cdot(0,1,0)}{(0,1,0)\cdot(0,1,0)} = 2.$
   - $\boldsymbol{p} = 2(1,0,1) + 2(0,1,0) = (2,2,2)$.
3. Find the distance:
   - $d(\boldsymbol{u},\boldsymbol{p}) = \|\boldsymbol{u}-\boldsymbol{p}\| = \|(-1,0,1)\| = \sqrt{2}$.

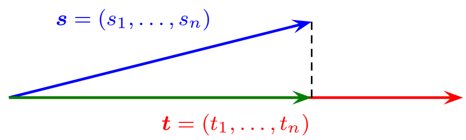
Least Squares Solution

- Determine the speed of an object:

time	t_1	t_2	$\cdots$	t_n
distance	s_1	s_2	$\cdots$	s_n

  - Due to the experimental error, there is no v so that
    - $s_1 = vt_1, s_2 = vt_2, \ldots, s_n = vt_n$.
    - $(s_1, s_2, \ldots, s_n) = v(t_1, t_2, \ldots, t_n)$.

    What is the **best** choice of v?



- Find v s.t. $v\boldsymbol{t}$ is the projection of $\boldsymbol{s}$ onto $\text{span}\{\boldsymbol{t}\}$.
  - $v = \dfrac{\boldsymbol{s}\cdot\boldsymbol{t}}{\boldsymbol{t}\cdot\boldsymbol{t}} = \dfrac{s_1t_1 + \cdots + s_nt_n}{t_1^2 + \cdots + t_n^2}$.

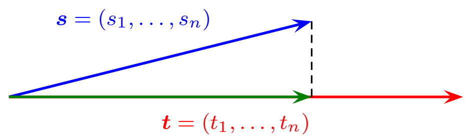
Least Squares Solution

- Determine the speed of an object:

time	t_1	t_2	$\cdots$	t_n
distance	s_1	s_2	$\cdots$	s_n

  - Due to the experimental error, there is no v so that
    - $s_1 = vt_1, s_2 = vt_2, \dots, s_n = vt_n$.
    - $(s_1, s_2, \dots, s_n) = v(t_1, t_2, \dots, t_n)$.

    What is the **best** choice of v?



    - Find v so that $v\boldsymbol{t}$ is closest to $\boldsymbol{s}$, i.e., minimize
      - $\|\boldsymbol{s} - v\boldsymbol{t}\| = \sqrt{(s_1 - vt_1)^2 + \cdots + (s_n - vt_n)^2}$.

Least Squares Solution

- Let $V = \{\boldsymbol{Ax} \mid \boldsymbol{x} \in \mathbb{R}^n\}$ be the column space of $\boldsymbol{A}$.
  - $\boldsymbol{Ax} = \boldsymbol{b}$ is consistent $\Leftrightarrow \boldsymbol{b} \in V$.

  Suppose $\boldsymbol{b} \notin V$. Then for all $\boldsymbol{x}$, $\boldsymbol{Ax} \neq \boldsymbol{b}$.
  - Although $\boldsymbol{Ax} = \boldsymbol{b}$ is not solvable, we may seek for $\boldsymbol{x}$ so that $\boldsymbol{Ax}$ is **closest** to $\boldsymbol{b}$.
    - Find $\boldsymbol{x}$ so that $\boldsymbol{Ax}$ is the projection of $\boldsymbol{b}$ onto V, i.e.,
      - $\|\boldsymbol{b} - \boldsymbol{Ax}\|$ is minimized.
- **Definition**. Let $\boldsymbol{A}$ be an $m \times n$ matrix, $\boldsymbol{b} \in \mathbb{R}^m$.
  - $\boldsymbol{u} \in \mathbb{R}^n$ is a **least squares solution** to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ if
    - $\|\boldsymbol{b} - \boldsymbol{Au}\| \leq \|\boldsymbol{b} - \boldsymbol{Av}\|$ for all $\boldsymbol{v} \in \mathbb{R}^n$.

Least Squares Solution

- **Theorem**. Let $\boldsymbol{A}$ be an $m \times n$ matrix, $\boldsymbol{b} \in \mathbb{R}^m$.
  - Let $\boldsymbol{p}$ be the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$.
    - Then $\|\boldsymbol{b} - \boldsymbol{p}\| \leq \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{v}\|$ for all $\boldsymbol{v} \in \mathbb{R}^n$,

    i.e., $\boldsymbol{u}$ is a **least squares solution** to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
    $\Leftrightarrow$ $\boldsymbol{u}$ is a solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p}$.

- **Proof**. Recall that among all the vectors in V, $\boldsymbol{p}$, the projection of $\boldsymbol{b}$ onto V, has the shortest distance to $\boldsymbol{b}$:
  - $d(\boldsymbol{b}, \boldsymbol{p}) \leq d(\boldsymbol{b}, \boldsymbol{w})$ for all $w \in V$.

  On the other hand, $V = \{\boldsymbol{A}\boldsymbol{v} \mid \boldsymbol{v} \in \mathbb{R}^n\}$. So
  - $d(\boldsymbol{b}, \boldsymbol{p}) \leq d(\boldsymbol{b}, \boldsymbol{A}\boldsymbol{v})$ for all $\boldsymbol{v} \in \mathbb{R}^n$,
    i.e., $\|\boldsymbol{b} - \boldsymbol{p}\| \leq \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{v}\|$ for all $\boldsymbol{v} \in \mathbb{R}^n$.

  $\|\boldsymbol{b} - \boldsymbol{p}\| = \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{v}\| \Leftrightarrow \boldsymbol{p} = \boldsymbol{A}\boldsymbol{v}$.

Examples

- Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

  1. Find the projection of $\boldsymbol{b}$ onto V:
     - V = coln space of $\boldsymbol{A} = \operatorname{span}\left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}$.
     - The projection is (by (5.3.3)) $\boldsymbol{p} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$.
  2. Solve the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{p}$.
     - $\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \Leftrightarrow \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.

Examples

- Consider the following data:

x	1	0	1
y	1	2	3

  Assume that the data satisfies $y = ax + b$.
  - What are the best choices of a and b?
    - $y = ax + b = \begin{pmatrix} x & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$.
  - The least squares solution to the system:
    - $\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$: $\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.
  - The best linear function which fits the data is
    - $y = 0x + 2$.

Methodology

- Find a **least squares solution** to $\boldsymbol{Ax} = \boldsymbol{b}$:
  1. Find an orthogonal (orthonormal) basis for V, the column space of $\boldsymbol{A}$.
  2. Find the projection $\boldsymbol{p}$ of $\boldsymbol{b}$ onto V.
  3. Solve the linear system $\boldsymbol{Ax} = \boldsymbol{p}$.

  Then a solution to $\boldsymbol{Ax} = \boldsymbol{p}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$.
- **Questions**.
  - Is the system $\boldsymbol{Ax} = \boldsymbol{p}$ solvable?
    - Yes! Because $\boldsymbol{p}$ lies in the column space of $\boldsymbol{A}$.
  - If $\boldsymbol{Ax} = \boldsymbol{b}$ is already consistent, what is the least squares solution?
    - $\boldsymbol{b} = \boldsymbol{p} \in V$. Solution $=$ Least squares solution.

Methodology

- Find a **least squares solution** to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.
  - Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$.
  - $V = \mathrm{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n\}$ = column space of $\boldsymbol{A}$.

$\boldsymbol{u}$ is a least squares solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$

$\Leftrightarrow$ $\boldsymbol{A}\boldsymbol{u}$ = projection of $\boldsymbol{b}$ onto V

$\Leftrightarrow$ $\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}$ is orthogonal to V

$\Leftrightarrow$ $\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}$ is orthogonal to $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$

$\Leftrightarrow$ $\boldsymbol{a}_i \cdot (\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}) = 0$ for all $i = 1, \dots, n$

$\Leftrightarrow$ $\boldsymbol{a}_i^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}) = 0$ for all $i = 1, \dots, n$

$$\Leftrightarrow \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{pmatrix} (\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{u} - \boldsymbol{b}) = \boldsymbol{0} \Leftrightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

Methodology

- **Theorem**. (Find the least squares solutions)
  - $\boldsymbol{u}$ is a **least squares solution** to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$

    $\Leftrightarrow$ $\boldsymbol{u}$ is a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.
- **Example**. Recall the system
  - $\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

  Its least square solutions are precisely the solutions to
  - $\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.
    - $\begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix} \Rightarrow \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.

Examples

- Suppose r, s and t are parameters satisfying
  - $t = cr^2 + ds + e.$

ith experiment	1	2	3	4	5	6
r_i	0	0	1	1	2	2
s_i	0	1	2	0	1	2
t_i	0.5	1.6	2.8	0.8	5.1	5.9

  - $$\begin{cases} cr_1^2 + ds_1 + e = t_1 \\ cr_2^2 + ds_2 + e = t_2 \\ \qquad\qquad \vdots \\ cr_6^2 + ds_6 + e = t_6 \end{cases} \Rightarrow \begin{pmatrix} r_i^2 & s_i & 1 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = t_i$$

Examples

- Suppose r, s and t are parameters satisfying
  - $t = cr^2 + ds + e.$

ith experiment	1	2	3	4	5	6
r_i	0	0	1	1	2	2
s_i	0	1	2	0	1	2
t_i	0.5	1.6	2.8	0.8	5.1	5.9

  - $$\begin{cases} cr_1^2 + ds_1 + e = t_1 \\ cr_2^2 + ds_2 + e = t_2 \\ \qquad\qquad \vdots \\ cr_6^2 + ds_6 + e = t_6 \end{cases} \Rightarrow \begin{pmatrix} r_1^2 & s_1 & 1 \\ r_2^2 & s_2 & 1 \\ \vdots & \vdots & \vdots \\ r_6^2 & s_6 & 1 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_6 \end{pmatrix}$$

Examples

- Suppose r, s and t are parameters satisfying
  - $t = cr^2 + ds + e.$
  - Solve $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{pmatrix}$
    - The system $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent.
    - Solve $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ to get the least squares solutions.

Examples

- Suppose r, s and t are parameters satisfying
  - $t = cr^2 + ds + e.$
  - $$\begin{pmatrix} 0 & 0 & 1 & 1 & 4 & 4 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix}$$
    $$= \begin{pmatrix} 0 & 0 & 1 & 1 & 4 & 4 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{pmatrix}.$$

Examples

- Suppose r, s and t are parameters satisfying
  - $t = cr^2 + ds + e$.
  - $$\begin{pmatrix} 34 & 14 & 10 \\ 14 & 10 & 6 \\ 10 & 6 & 6 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 47.6 \\ 24.1 \\ 16.7 \end{pmatrix}$$
    $$\Rightarrow \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 0.9275 \\ 0.9225 \\ 0.3150 \end{pmatrix}$$
  - The data can be modeled by
    - $t = 0.9275r^2 + 0.9225s + 0.3150.$
  - Although no data satisfies the above equation, it is the best to fit the whole set of data.

Projection

- Consider the linear system $\boldsymbol{Ax} = \boldsymbol{b}$.
  - A least squares solution $\boldsymbol{u}$ gives the distance from $\boldsymbol{b}$ to V, the column space of $\boldsymbol{A}$:
    - $\|\boldsymbol{b} - \boldsymbol{Au}\| \leq \|\boldsymbol{b} - \boldsymbol{v}\|$ for all $\boldsymbol{v} \in V$.

    So $\boldsymbol{Au}$ is the projection of $\boldsymbol{b}$ onto V.

  We obtain another method to find the **projection** of a vector $\boldsymbol{b}$ onto a vector space V:

  1. Suppose $V = \mathrm{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$.
  2. Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \end{pmatrix}$, each $\boldsymbol{a}_j$ is a column vector.
  3. Find a least squares solution $\boldsymbol{u}$ to $\boldsymbol{Ax} = \boldsymbol{b}$;
     - i.e., a solution $\boldsymbol{u}$ to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.
  4. The projection of $\boldsymbol{b}$ onto V is $\boldsymbol{p} = \boldsymbol{Au}$.

Example

- Find the projection of $(1,1,1,1)$ onto
  - $V = \operatorname{span}\{(1,-1,1,-1),(1,2,0,1),(2,1,1,0)\}$.
- Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.
  - Find a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$; i.e., a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.
    - $$\begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Example

- Find the projection of $(1,1,1,1)$ onto
  - $V = \operatorname{span}\{(1,-1,1,-1),(1,2,0,1),(2,1,1,0)\}$.
- Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.
  - Find a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$; i.e., a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.
    - $$\begin{pmatrix} 4 & -2 & 2 \\ -2 & 6 & 4 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \\ 4 \end{pmatrix}$$
      $$\Rightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -t + 2/5 \\ -t + 4/5 \\ t \end{pmatrix}$$

Example

- Find the projection of $(1,1,1,1)$ onto
  - $V = \text{span}\{(1,-1,1,-1),(1,2,0,1),(2,1,1,0)\}$.
- Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.
  - The projection of $\boldsymbol{b}$ onto V is
    - $\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} -t+2/5 \\ -t+4/5 \\ t \end{pmatrix} = \begin{pmatrix} 6/5 \\ 6/5 \\ 2/5 \end{pmatrix}$
  - Since the projection is unique, we may choose any parameter t.

Orthogonal Matrices

Orthonormal Sets

- Recall the advantages of an **orthonormal** set:
  - Suppose $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an **orthonormal** subset of $\mathbb{R}^n$ $(k \leq n)$.
    - Let $V = \text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$.
    1. S is a basis for V.
    2. For any vector $\boldsymbol{v} \in V$,
       - $(\boldsymbol{v})_S = (\boldsymbol{v}\cdot\boldsymbol{v}_1, \boldsymbol{v}\cdot\boldsymbol{v}_2, \ldots, \boldsymbol{v}\cdot\boldsymbol{v}_k)$.
       - $\boldsymbol{v} = (\boldsymbol{v}\cdot\boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{v}\cdot\boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{v}\cdot\boldsymbol{v}_k)\boldsymbol{v}_k$
    3. The **projection** of $\boldsymbol{w} \in \mathbb{R}^n$ onto V:
       - $\boldsymbol{p} = (\boldsymbol{w}\cdot\boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{w}\cdot\boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{w}\cdot\boldsymbol{v}_k)\boldsymbol{v}_k$.

Orthonormal Sets

- Recall the advantages of an **orthonormal** set:
  - Suppose $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an **orthonormal** subset of $\mathbb{R}^n$ $(k \leq n)$.
    - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k \end{pmatrix}$.

    1. The columns of $\boldsymbol{A}$ are **linearly independent**.
       - $\operatorname{rank}(\boldsymbol{A}) = k$.
    2. $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = (\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{v}_j)_{k\times k} = (\boldsymbol{v}_i \cdot \boldsymbol{v}_j)_{k\times k} = \boldsymbol{I}_k$.
    3. Least squares solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:
       - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \Rightarrow \boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Definition

- **Definition.** Let $\boldsymbol{A}$ be a **square** matrix.
  - $\boldsymbol{A}$ is called an **orthogonal matrix** if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.
    - Equivalently, $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$, or $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- **Theorem**. Let $\boldsymbol{A}$ be a **square** matrix of order n.
  - $\boldsymbol{A}$ is an **orthogonal** matrix

    $\Leftrightarrow$ columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
    $\Leftrightarrow$ rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
- **Examples**.
  - The identity matrix $\boldsymbol{I}_n$ is an orthogonal matrix.
    - $(\boldsymbol{I}_n)^{\mathrm{T}}\boldsymbol{I}_n = \boldsymbol{I}_n\boldsymbol{I}_n = \boldsymbol{I}_n$.

Definition

- **Definition.** Let $\boldsymbol{A}$ be a **square** matrix.
  - $\boldsymbol{A}$ is called an **orthogonal matrix** if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.
    - Equivalently, $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$, or $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- **Theorem**. Let $\boldsymbol{A}$ be a **square** matrix of order n.
  - $\boldsymbol{A}$ is an **orthogonal** matrix

    $\Leftrightarrow$ columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
    $\Leftrightarrow$ rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
- **Examples**.
  - $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is an orthogonal matrix.
    - $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

80 / 105

Definition

- **Definition.** Let $\boldsymbol{A}$ be a **square** matrix.
  - $\boldsymbol{A}$ is called an **orthogonal matrix** if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.
    - Equivalently, $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$, or $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- **Theorem**. Let $\boldsymbol{A}$ be a **square** matrix of order n.
  - $\boldsymbol{A}$ is an **orthogonal** matrix

    $\Leftrightarrow$ columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
    $\Leftrightarrow$ rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.

- **Examples**.
  - $\begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ is an orthogonal matrix.
    - Verification is left as an exercise.

81 / 105

Definition

- **Definition.** Let $\boldsymbol{A}$ be a **square** matrix.
  - $\boldsymbol{A}$ is called an **orthogonal matrix** if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.
    - Equivalently, $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$, or $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- **Theorem**. Let $\boldsymbol{A}$ be a **square** matrix of order n.
  - $\boldsymbol{A}$ is an **orthogonal** matrix

    $\Leftrightarrow$ columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
    $\Leftrightarrow$ rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
- **Properties**.
  - If $\boldsymbol{A}$ is an orthogonal matrix, then
    - $\boldsymbol{I} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}$.

    So $\boldsymbol{A}^{\mathrm{T}}$ $(= \boldsymbol{A}^{-1})$ is also an orthogonal matrix.

Definition

- **Definition.** Let $\boldsymbol{A}$ be a **square** matrix.
  - $\boldsymbol{A}$ is called an **orthogonal matrix** if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$.
    - Equivalently, $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$, or $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.
- **Theorem**. Let $\boldsymbol{A}$ be a **square** matrix of order n.
  - $\boldsymbol{A}$ is an **orthogonal** matrix

    $\Leftrightarrow$ columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
    $\Leftrightarrow$ rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
- **Properties**.
  - If $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal matrices of the same size,
    - $(\boldsymbol{AB})^{\mathrm{T}}(\boldsymbol{AB}) = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{AB} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}$.

    So $\boldsymbol{AB}$ is also an orthogonal matrix.

Properties

- **Proof of Theorem**.
  - Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{pmatrix}$
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_n \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{a}_n \end{pmatrix}$

Properties

- **Proof of Theorem**.
  - Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{pmatrix}$
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{a}_n \\ \boldsymbol{a}_2 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_2 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2 \cdot \boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{a}_n \cdot \boldsymbol{a}_1 & \boldsymbol{a}_n \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \cdot \boldsymbol{a}_n \end{pmatrix}$

Properties

- **Proof of Theorem**.
  - Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{pmatrix}$
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = (\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{a}_j)_{n\times n} = (\boldsymbol{a}_i \cdot \boldsymbol{a}_j)_{n\times n}$.

$$\begin{aligned} \boldsymbol{A} \text{ is orthogonal} &\Leftrightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_n \\ &\Leftrightarrow \boldsymbol{a}_i \cdot \boldsymbol{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \\ &\Leftrightarrow \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \text{ are orthonormal.} \end{aligned}$$

86 / 105

Properties

- **Proof of Theorem**.
  - Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_n \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{b}_1^{\mathrm{T}} & \boldsymbol{b}_2^{\mathrm{T}} & \cdots & \boldsymbol{b}_n^{\mathrm{T}} \end{pmatrix}$
    - $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = (\boldsymbol{b}_i\boldsymbol{b}_j^{\mathrm{T}})_{n\times n} = (\boldsymbol{b}_i \cdot \boldsymbol{b}_j)_{n\times n}$.

$$\begin{aligned} \boldsymbol{A} \text{ is orthogonal} &\Leftrightarrow \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}_n \\ &\Leftrightarrow \boldsymbol{b}_i \cdot \boldsymbol{b}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \\ &\Leftrightarrow \boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n \text{ are orthonormal.} \end{aligned}$$

87 / 105

Properties

- More generally, for any $m \times n$ matrix $\boldsymbol{A}$:
  - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_n$

    $\Leftrightarrow$ the **columns** of $\boldsymbol{A}$ form an **orthonormal** set.
  - $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}_m$

    $\Leftrightarrow$ the **rows** of $\boldsymbol{A}$ form an **orthonormal** set.
- Let $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ be an **orthonormal** subset of $\mathbb{R}^n$.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_k$.
    - Let $\boldsymbol{P}$ be an $n \times n$ **orthogonal** matrix.

      $(\boldsymbol{PA})^{\mathrm{T}}(\boldsymbol{PA}) = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{PA} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_k$.
    - $\boldsymbol{PA} = \begin{pmatrix} \boldsymbol{Pu}_1 & \cdots & \boldsymbol{Pu}_k \end{pmatrix}$.
  - $\{\boldsymbol{Pu}_1, \dots, \boldsymbol{Pu}_k\}$ is also an orthonormal set.

Properties

- Let $S = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ and $T = \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ be **orthonormal** bases for a vector space V.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \end{pmatrix}$.
    - Then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}_k$.
  - For any $\boldsymbol{w} \in V$, $\boldsymbol{w} = \boldsymbol{A}[\boldsymbol{w}]_S = \boldsymbol{B}[\boldsymbol{w}]_T$.
    - $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}[\boldsymbol{w}]_S = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}[\boldsymbol{w}]_T = [\boldsymbol{w}]_T$.
  - Let $\boldsymbol{P}$ be the transition matrix from S to T.
    - Then $\boldsymbol{P}[\boldsymbol{w}]_S = [\boldsymbol{w}]_T$.

$\therefore$ $\boldsymbol{P} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}$ is the **transition matrix** from S to T.

Properties

- Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ and $T = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be **orthonormal** bases for a vector space V.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \end{pmatrix}$.
    - Then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}_k$.
  - For any $\boldsymbol{w} \in V$, $\boldsymbol{w} = \boldsymbol{A}[\boldsymbol{w}]_S = \boldsymbol{B}[\boldsymbol{w}]_T$.
    - $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}[\boldsymbol{w}]_T = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}[\boldsymbol{w}]_S = [\boldsymbol{w}]_S$.
  - Let $\boldsymbol{Q}$ be the transition matrix from T to S.
    - Then $\boldsymbol{Q}[\boldsymbol{w}]_T = [\boldsymbol{w}]_S$.

  $\therefore$ $\boldsymbol{Q} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}$ is the **transition matrix** from T to S.

Properties

- Let $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ and $T = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be **orthonormal** bases for a vector space V.
  - Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_k \end{pmatrix}$.
    - Then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{I}_k$.
  - $\boldsymbol{P} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}$ is the **transition matrix** from S to T;
    $\boldsymbol{Q} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}$ is the **transition matrix** from T to S.
    - $\boldsymbol{P}^{\mathrm{T}} = (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{Q}$.
    - It is also known that $\boldsymbol{P}^{-1} = \boldsymbol{Q}$; so $\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$.

  $\therefore$ $\boldsymbol{P}$ (and hence $\boldsymbol{Q}$) is an **orthogonal** matrix.
- **Theorem**. Let S and T be two **orthonormal** bases for a vector space V.
  - Let $\boldsymbol{P}$ be the **transition matrix** from S to T.
    - Then $\boldsymbol{P}$ is an **orthogonal** matrix.

Examples

- Let $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be the standard basis for $\mathbb{R}^3$.

  $\boldsymbol{e}_1 = (1,0,0)$, $\boldsymbol{e}_2 = (0,1,0)$, $\boldsymbol{e}_3 = (0,0,1)$.

  Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, where $\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)$,

  $\boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1)$, $\boldsymbol{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1)$.

  - Let $\boldsymbol{P}$ be the transition matrix from S to E:
    - $\begin{cases} \boldsymbol{u}_1 = \frac{1}{\sqrt{3}}\boldsymbol{e}_1 + \frac{1}{\sqrt{3}}\boldsymbol{e}_2 + \frac{1}{\sqrt{3}}\boldsymbol{e}_3 \\ \boldsymbol{u}_2 = \frac{1}{\sqrt{2}}\boldsymbol{e}_1 + \ \ 0\boldsymbol{e}_2 - \frac{1}{\sqrt{2}}\boldsymbol{e}_3 \\ \boldsymbol{u}_3 = \frac{1}{\sqrt{6}}\boldsymbol{e}_1 - \frac{2}{\sqrt{6}}\boldsymbol{e}_2 + \frac{1}{\sqrt{6}}\boldsymbol{e}_3 \end{cases}$
    - $\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$ is an orthogonal matrix.

Examples

- Let $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ be the standard basis for $\mathbb{R}^3$.

  $\boldsymbol{e}_1 = (1,0,0)$, $\boldsymbol{e}_2 = (0,1,0)$, $\boldsymbol{e}_3 = (0,0,1)$.

  Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, where $\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)$,

  $\boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1)$, $\boldsymbol{u}_3 = \frac{1}{\sqrt{6}}(1,-2,1)$.

  - $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$ is the transition matrix from E to S:
    - $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$
    - $\begin{cases} \boldsymbol{e}_1 = \frac{1}{\sqrt{3}}\boldsymbol{u}_1 + \frac{1}{\sqrt{2}}\boldsymbol{u}_2 + \frac{1}{\sqrt{6}}\boldsymbol{u}_3 \\ \boldsymbol{e}_2 = \frac{1}{\sqrt{3}}\boldsymbol{u}_1 + \ \ 0\boldsymbol{u}_2 - \frac{2}{\sqrt{6}}\boldsymbol{u}_3 \\ \boldsymbol{e}_3 = \frac{1}{\sqrt{3}}\boldsymbol{u}_1 - \frac{1}{\sqrt{2}}\boldsymbol{u}_2 + \frac{1}{\sqrt{6}}\boldsymbol{u}_3 \end{cases}$

Examples

- Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$,
  - $\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)$, $\boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1)$, $\boldsymbol{u}_3 = \frac{1}{6}(1,-2,1)$.
  - $\boldsymbol{v}_1 = (0,0,1)$, $\boldsymbol{v}_2 = \frac{1}{\sqrt{2}}(1,-1,0)$, $\boldsymbol{v}_3 = \frac{1}{\sqrt{2}}(1,1,0)$.

  Both S and T are orthonormal bases for $\mathbb{R}^3$. (Verify!)
  - $\boldsymbol{u}_1 = (\boldsymbol{u}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_3)\boldsymbol{v}_3$.
    - $\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \frac{2}{\sqrt{6}}\boldsymbol{v}_3$.
  - $\boldsymbol{u}_2 = (\boldsymbol{u}_2 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + (\boldsymbol{u}_2 \cdot \boldsymbol{v}_3)\boldsymbol{v}_3$.
    - $\boldsymbol{u}_2 = -\frac{1}{\sqrt{2}}\boldsymbol{v}_1 + \frac{1}{2}\boldsymbol{v}_2 + \frac{1}{2}\boldsymbol{v}_3$.
  - $\boldsymbol{u}_3 = (\boldsymbol{u}_3 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_3 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + (\boldsymbol{u}_3 \cdot \boldsymbol{v}_3)\boldsymbol{v}_3$.
    - $\boldsymbol{u}_3 = \frac{1}{\sqrt{6}}\boldsymbol{v}_1 + \frac{3}{\sqrt{12}}\boldsymbol{v}_2 - \frac{1}{\sqrt{12}}\boldsymbol{v}_3$.

Examples

- Let $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and $T = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$,
  - $\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)$, $\boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1,0,-1)$, $\boldsymbol{u}_3 = \frac{1}{6}(1,-2,1)$.
  - $\boldsymbol{v}_1 = (0,0,1)$, $\boldsymbol{v}_2 = \frac{1}{\sqrt{2}}(1,-1,0)$, $\boldsymbol{v}_3 = \frac{1}{\sqrt{2}}(1,1,0)$.

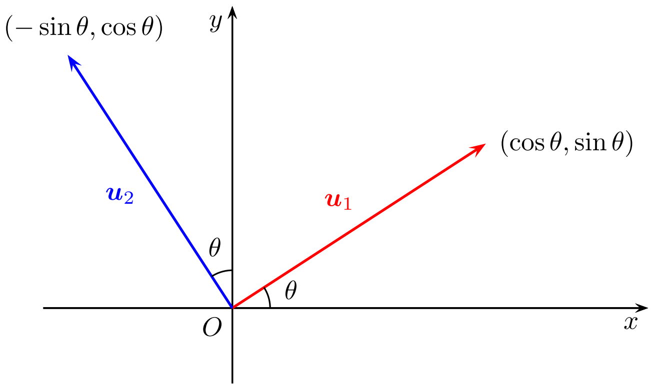
  Both S and T are orthonormal bases for $\mathbb{R}^3$. (Verify!)
  - The transition matrix from S to T:
    - $\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{2} & \frac{3}{\sqrt{12}} \\ \frac{2}{\sqrt{6}} & \frac{1}{2} & -\frac{1}{\sqrt{12}} \end{pmatrix}$.
  - The transition matrix from T to S:
    - $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}} = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{6}} & \frac{3}{\sqrt{12}} & -\frac{1}{\sqrt{12}} \end{pmatrix}$.

Classification

- What are the orthogonal matrices (numbers) of order 1?
  - (a) with $|a| = 1$: (1) and (-1).
- What are the orthogonal matrices of order 2?
  - (Exercise 2.57): $\det(\boldsymbol{A}) = \pm 1$.
    - If $\det(\boldsymbol{A}) = 1$, then $\boldsymbol{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.
    - If $\det(\boldsymbol{A}) = -1$, then $\boldsymbol{A} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$.
- **Exercise**.
  - Can you classify the orthogonal matrices of order 3?

Geometric Representation

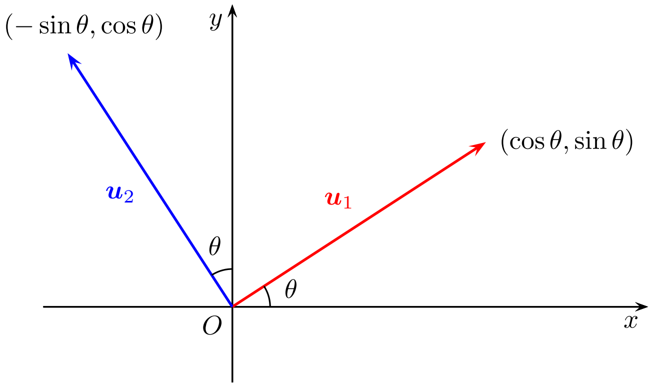
- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be an orthonormal basis for $\mathbb{R}^2$.



  - $\boldsymbol{u}_1 = \cos\theta\, \boldsymbol{e}_1 + \sin\theta\, \boldsymbol{e}_2$;
  - $\boldsymbol{u}_2 = -\sin\theta\, \boldsymbol{e}_1 + \cos\theta\, \boldsymbol{e}_2$.

Geometric Representation

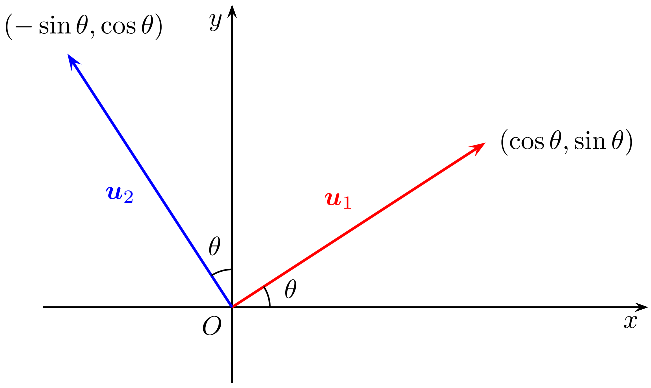
- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be an orthonormal basis for $\mathbb{R}^2$.



  - $\boldsymbol{P} = (\boldsymbol{u}_1 \quad \boldsymbol{u}_2) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is orthogonal.
    - The transition matrix from $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ to $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$.

Geometric Representation

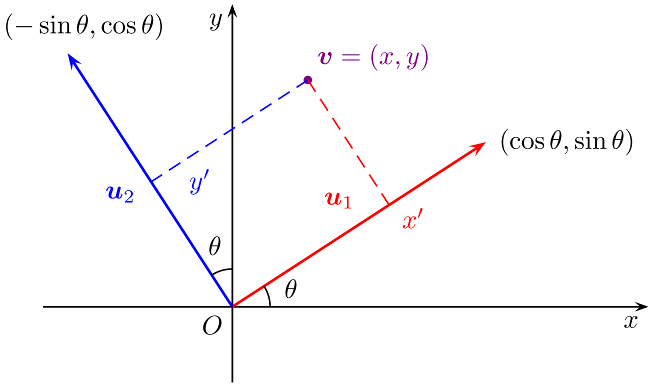
- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be an orthonormal basis for $\mathbb{R}^2$.



  - $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$ is also orthogonal.
    - The transition matrix from $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ to $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$.

Geometric Representation

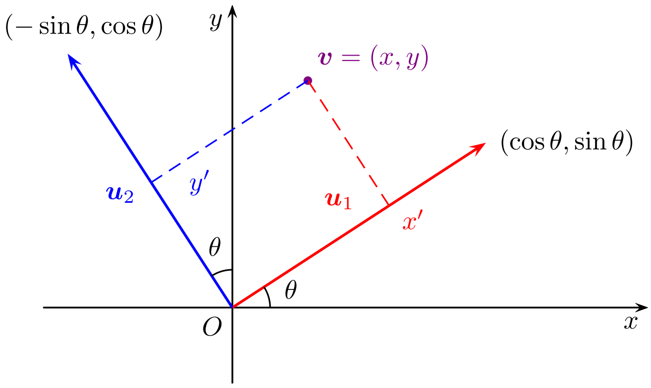
- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be an orthonormal basis for $\mathbb{R}^2$.



  - Let $\boldsymbol{v} = (x, y)$ and $(\boldsymbol{v})_S = (x', y')$, $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2\}$.
    - $\boldsymbol{v} = \boldsymbol{P}[\boldsymbol{v}]_S \Rightarrow [\boldsymbol{v}]_S = \boldsymbol{P}^{-1}\boldsymbol{v} = \boldsymbol{P}^{\mathrm{T}}\boldsymbol{v}$.

Geometric Representation

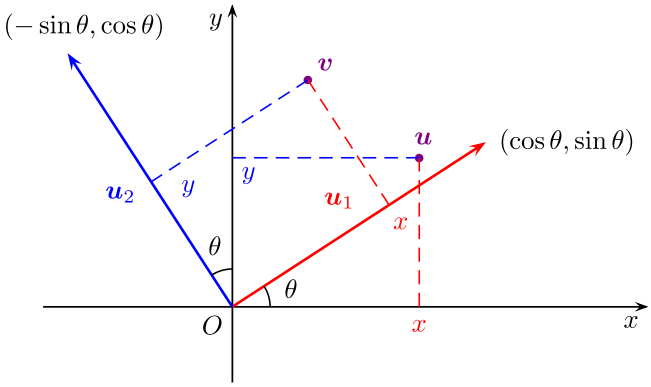
- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be an orthonormal basis for $\mathbb{R}^2$.



  - $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$.
    - The coordinates of $\boldsymbol{v}$ using $x'y'$-coordinate system.

Geometric Representation

- Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ be an orthonormal basis for $\mathbb{R}^2$.



  - Let $\boldsymbol{u} \in \mathbb{R}^2$, and $\boldsymbol{v} = \boldsymbol{P}\boldsymbol{u} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \boldsymbol{u}$.
    - $\boldsymbol{v}$ is the **rotation** of $\boldsymbol{u}$ about O by θ anticlockwise.

Geometric Representation

- Let $\boldsymbol{P}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Then for any $\boldsymbol{u} \in \mathbb{R}^2$,
  - $\boldsymbol{P}_\theta \boldsymbol{u}$ = **rotation** of $\boldsymbol{u}$ about O by θ anticlockwise.

  Fix angles α and β. Then for any $\boldsymbol{u} \in \mathbb{R}^2$,

$$\begin{aligned} \boldsymbol{P}_\alpha \boldsymbol{u} &= \text{rotation of } \boldsymbol{u} \text{ about } O \text{ by } \alpha \text{ anticlockwise} \\ \boldsymbol{P}_\beta(\boldsymbol{P}_\alpha \boldsymbol{u}) &= \text{rotation of } \boldsymbol{P}_\alpha \boldsymbol{u} \text{ about } O \text{ by } \beta \text{ anticlockwise} \\ &= \text{rotation of } \boldsymbol{u} \text{ about } O \text{ by } \alpha + \beta \\ &= \boldsymbol{P}_{\alpha+\beta} \boldsymbol{u}. \end{aligned}$$

  - Therefore, $\boldsymbol{P}_\beta \boldsymbol{P}_\alpha = \boldsymbol{P}_{\alpha+\beta}$.

$$\begin{aligned} &\begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \\ = &\begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}. \end{aligned}$$

Geometric Representation

- Let $\boldsymbol{P}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Then for any $\boldsymbol{u} \in \mathbb{R}^2$.
  - $\boldsymbol{P}_\theta \boldsymbol{u} =$ rotation of $\boldsymbol{u}$ about O by θ anticlockwise.

  Fix angles α and β. Then for any $\boldsymbol{u} \in \mathbb{R}^2$,

$$\begin{aligned} \boldsymbol{P}_\alpha \boldsymbol{u} &= \text{rotation of } \boldsymbol{u} \text{ about } O \text{ by } \alpha \text{ anticlockwise} \\ \boldsymbol{P}_\beta(\boldsymbol{P}_\alpha \boldsymbol{u}) &= \text{rotation of } \boldsymbol{P}_\alpha \boldsymbol{u} \text{ about } O \text{ by } \beta \text{ anticlockwise} \\ &= \text{rotation of } \boldsymbol{u} \text{ about } O \text{ by } \alpha + \beta \\ &= \boldsymbol{P}_{\alpha+\beta} \boldsymbol{u}. \end{aligned}$$

  - Therefore, $\boldsymbol{P}_\beta \boldsymbol{P}_\alpha = \boldsymbol{P}_{\alpha+\beta}$.

$$\begin{aligned} &\begin{pmatrix} \cos\beta\cos\alpha - \sin\beta\sin\alpha & -\cos\beta\sin\alpha - \sin\beta\cos\alpha \\ \sin\beta\cos\alpha + \cos\beta\sin\alpha & -\sin\beta\sin\alpha + \cos\beta\cos\alpha \end{pmatrix} \\ = &\begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix}. \end{aligned}$$

104 / 105

Geometric Representation

- Let $\boldsymbol{P}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Then for any $\boldsymbol{u} \in \mathbb{R}^2$.
  - $\boldsymbol{P}_\theta \boldsymbol{u} =$ rotation of $\boldsymbol{u}$ about O by θ anticlockwise.

  Fix angles α and β. Then for any $\boldsymbol{u} \in \mathbb{R}^2$,

$$\begin{aligned} \boldsymbol{P}_\alpha \boldsymbol{u} &= \text{rotation of } \boldsymbol{u} \text{ about } O \text{ by } \alpha \text{ anticlockwise} \\ \boldsymbol{P}_\beta(\boldsymbol{P}_\alpha \boldsymbol{u}) &= \text{rotation of } \boldsymbol{P}_\alpha \boldsymbol{u} \text{ about } O \text{ by } \beta \text{ anticlockwise} \\ &= \text{rotation of } \boldsymbol{u} \text{ about } O \text{ by } \alpha + \beta \\ &= \boldsymbol{P}_{\alpha+\beta} \boldsymbol{u}. \end{aligned}$$

  - Therefore, $\boldsymbol{P}_\beta \boldsymbol{P}_\alpha = \boldsymbol{P}_{\alpha+\beta}$.
  - **Sum Laws for Sine and Cosine**:
    - $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$;
    - $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$.

105 / 105