Chapter 5

Orthogonality

Section 5.1

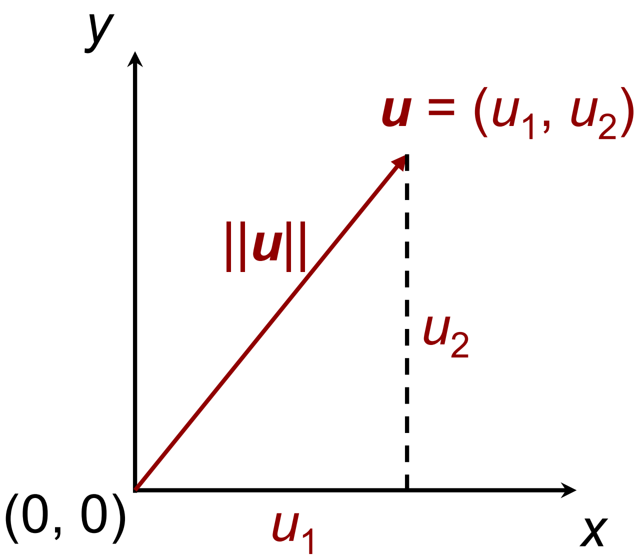
The Dot Product

Lengths of vectors in $\mathbb{R}^2$ (Discussion 5.1.1.2)

Let $\boldsymbol{u} = (u_1, u_2)$ be a vector in $\mathbb{R}^2$.

Then the length of $\boldsymbol{u}$ is given by

$$||\boldsymbol{u}|| = \sqrt{u_1^2 + u_2^2}\,.$$

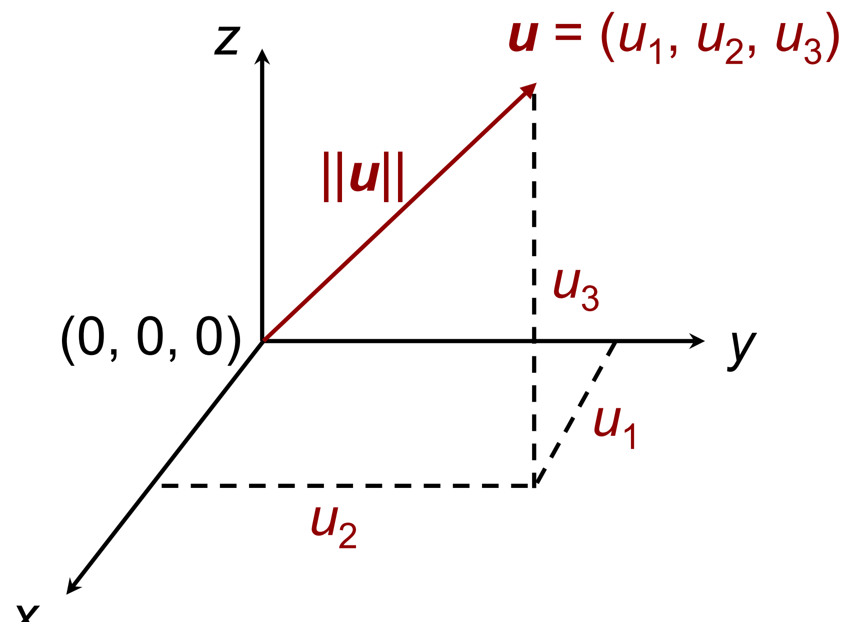


Lengths of vectors in $\mathbb{R}^3$ (Discussion 5.1.1.2)

Let $\boldsymbol{u} = (u_1, u_2, u_3)$ be a vector in $\mathbb{R}^3$.

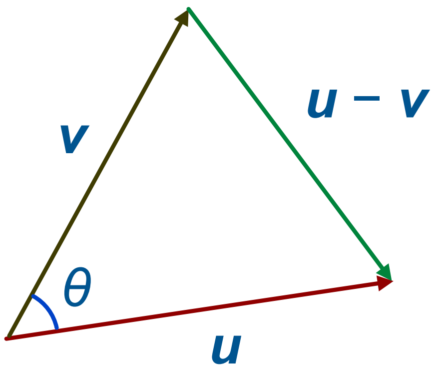
Then the length of $\boldsymbol{u}$ is given by

$$||\boldsymbol{u}|| = \sqrt{u_1^2 + u_2^2 + u_3^2}\,.$$



Distances and angles (Discussion 5.1.1.3)

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ and let θ be the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$.



The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}||.$$

The cosine rule of trigonometry states that

$$||\boldsymbol{u} - \boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - 2\,||\boldsymbol{u}||\,||\boldsymbol{v}||\cos(\theta)$$

and hence

$$\theta = \cos^{-1}\left(\frac{||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - ||\boldsymbol{u} - \boldsymbol{v}||^2}{2\,||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

Distances and angles (Discussion 5.1.1.3)

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

then $d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2}$

and $$\theta = \cos^{-1}\left(\frac{||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - ||\boldsymbol{u} - \boldsymbol{v}||^2}{2\,||\boldsymbol{u}||\,||\boldsymbol{v}||}\right)$$

$$= \cos^{-1}\left(\frac{u_1 v_1 + u_2 v_2}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

Distances and angles (Discussion 5.1.1.3)

If $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$ are vectors in $\mathbb{R}^3$,

then $d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}$

and $$\theta = \cos^{-1}\left(\frac{||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - ||\boldsymbol{u} - \boldsymbol{v}||^2}{2\,||\boldsymbol{u}||\,||\boldsymbol{v}||}\right)$$

$$= \cos^{-1}\left(\frac{u_1v_1 + u_2v_2 + u_3v_3}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

The dot product (Definition 5.1.2)

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

1. The dot product (or inner product) of $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be the value

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n .$$

2. The norm (or length) of $\boldsymbol{u}$ is defined to be

$$\|\boldsymbol{u}\| = \sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} .$$

Vectors of norm 1 are called unit vectors.

The dot product (Definition 5.1.2)

3. The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| = \sqrt{(\boldsymbol{u} - \boldsymbol{v}) \cdot (\boldsymbol{u} - \boldsymbol{v})}$$
$$= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}\,.$$

4. The angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\cos^{-1}\left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

The angle is well-defined because $-1 \leq \dfrac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||} \leq 1$.

(See Question 5.4(a).)

The dot product & matrix product (Remark 5.1.3)

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors in $\mathbb{R}^n$.

Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are written as row vectors, i.e.

$$\boldsymbol{u} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}.$$

Then

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}.$$

The dot product & matrix product (Remark 5.1.3)

Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are written as column vectors, i.e.

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Then

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{v}.$$

An example (Example 5.1.4)

Let $\boldsymbol{u} = (1, -2, 2, -1)$ and $\boldsymbol{v} = (1, 0, 2, 0)$.

$$\boldsymbol{u} \cdot \boldsymbol{v} = 1 \cdot 1 + (-2) \cdot 0 + 2 \cdot 2 + (-1) \cdot 0 = 5,$$

$$\|\boldsymbol{u}\| = \sqrt{1^2 + (-2)^2 + 2^2 + (-1)^2} = \sqrt{10},$$

$$\|\boldsymbol{v}\| = \sqrt{1^2 + 0^2 + 2^2 + 0^2} = \sqrt{5},$$

$$d(\boldsymbol{u}, \boldsymbol{v}) = \sqrt{(1-1)^2 + (-2-0)^2 + (2-2)^2 + (-1-0)^2} = \sqrt{5}$$

and the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\cos^{-1}\left(\frac{5}{\sqrt{10}\sqrt{5}}\right) = \cos^{-1}\left(\frac{1}{\sqrt{2}}\right) = \frac{\pi}{4}.$$

Some basic properties (Theorem 5.1.5)

Let $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ be vectors in $\mathbb{R}^n$ and c a scalar.

1. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$.
2. $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$ and $\boldsymbol{w} \cdot (\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{w} \cdot \boldsymbol{u} + \boldsymbol{w} \cdot \boldsymbol{v}$.
3. $(c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v}) = c(\boldsymbol{u} \cdot \boldsymbol{v})$.
4. $\|c\boldsymbol{u}\| = |c|\,\|\boldsymbol{u}\|$.
5. $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$; and $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if and only if $\boldsymbol{u} = \boldsymbol{0}$.

To Prove Part 5: Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$.

Then $\boldsymbol{u} \cdot \boldsymbol{u} = u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0$.

Furthermore, $\boldsymbol{u} \cdot \boldsymbol{u} = 0 \;\Leftrightarrow\; u_1^2 + u_2^2 + \cdots + u_n^2 = 0$

$\Leftrightarrow\; u_i = 0$ for $i = 1, 2, \ldots, n$

$\Leftrightarrow\; \boldsymbol{u} = \boldsymbol{0}$.

Section 5.2

Orthogonal and Orthonormal Bases

Orthogonality (Definition 5.2.1 & Remark 5.2.2)

1. Two vector $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ are called orthogonal if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.
2. A set S of vectors in $\mathbb{R}^n$ is called an orthogonal set if every pair of distinct vectors in S are orthogonal.
3. A set S of vectors in $\mathbb{R}^n$ is called an orthonormal set if S is an orthogonal set and every vector in S is a unit vector.

A unit vector is a vector of norm 1.

Given two nonzero vector $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$,

$$\boldsymbol{u} \cdot \boldsymbol{v} = 0 \quad \Rightarrow \quad \cos^{-1}\left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|}\right) = \cos^{-1}(0) = \frac{\pi}{2}.$$

The concept of orthogonal in $\mathbb{R}^n$ is the same as the concept of perpendicular in $\mathbb{R}^2$ and $\mathbb{R}^3$.

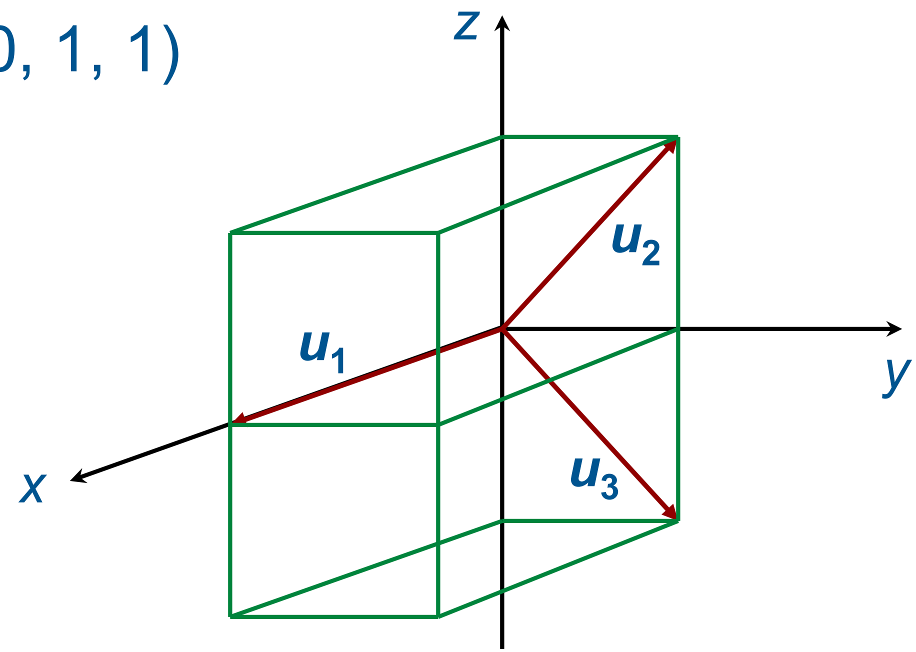
Examples (Example 5.2.3)

1.
$$\begin{aligned}&(1, 2, 2, -1) \cdot (1, 1, -1, 1)\\ &= 1 \cdot 1 + 2 \cdot 1 + 2 \cdot (-1) + (-1) \cdot 1\\ &= 0.\end{aligned}$$

So $(1, 2, 2, -1)$ and $(1, 1, -1, 1)$ are orthogonal.

2. Let $\boldsymbol{u}_1 = (2, 0, 0)$, $\boldsymbol{u}_2 = (0, 1, 1)$ and $\boldsymbol{u}_3 = (0, 1, -1)$.

Since $\boldsymbol{u}_1 \cdot \boldsymbol{u}_2 = 0$,
$\boldsymbol{u}_1 \cdot \boldsymbol{u}_3 = 0$
and $\boldsymbol{u}_2 \cdot \boldsymbol{u}_3 = 0$,
$\{\, \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \,\}$ is an orthogonal set.



Examples (Example 5.2.3)

Let $\boldsymbol{v}_1 = \dfrac{1}{||\boldsymbol{u}_1||}\boldsymbol{u}_1 = \dfrac{1}{2}(2, 0, 0) = (1, 0, 0),$

$\boldsymbol{v}_2 = \dfrac{1}{||\boldsymbol{u}_2||}\boldsymbol{u}_2 = \dfrac{1}{\sqrt{2}}(0, 1, 1) = \left(0, \dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}}\right),$

$\boldsymbol{v}_3 = \dfrac{1}{||\boldsymbol{u}_3||}\boldsymbol{u}_3 = \dfrac{1}{\sqrt{2}}(0, 1, -1) = \left(0, \dfrac{1}{\sqrt{2}}, -\dfrac{1}{\sqrt{2}}\right).$

Then $||\boldsymbol{v}_i|| = \left|\left|\dfrac{1}{||\boldsymbol{u}_i||}\boldsymbol{u}_i\right|\right| = \dfrac{1}{||\boldsymbol{u}_i||}||\boldsymbol{u}_i|| = 1$ for all i

and $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = \left(\dfrac{1}{||\boldsymbol{u}_i||}\boldsymbol{u}_i\right) \cdot \left(\dfrac{1}{||\boldsymbol{u}_j||}\boldsymbol{u}_j\right) = \dfrac{1}{||\boldsymbol{u}_i||\,||\boldsymbol{u}_j||}(\boldsymbol{u}_i \cdot \boldsymbol{u}_j) = 0$

for $i \neq j$.

Thus $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \}$ is an orthonormal set.

Examples (Example 5.2.3)

The process of multiplying a nonzero vector $\boldsymbol{u}$ by $\frac{1}{||\boldsymbol{u}||}$ (so that the resultant vector $\frac{1}{||\boldsymbol{u}||}\boldsymbol{u}$ is a unit vector) is called normalizing.

3. Consider the standard basis $E = \{ \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n \}$ for $\mathbb{R}^n$ where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)$, $\boldsymbol{e}_2 = (0, 1, 0, \ldots, 0)$, ..., $\boldsymbol{e}_n = (0, \ldots, 0, 1)$.

   It is easy to check that

   $||\boldsymbol{e}_i|| = 1$ for all i

   and $\boldsymbol{e}_i \cdot \boldsymbol{e}_j = 0$ for $i \neq j$.

   So E is an orthonormal set.

Orthogonal sets (Theorem 5.2.4)

Let S be an orthogonal set of nonzero vectors in a vector space. Then S is linearly independent.

Proof: Let $S = \{ \boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k} \}$.

Consider the vector equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \mathbf{0}. \qquad (*)$$

For any $i = 1, 2, \ldots, k$,

$$\begin{aligned}
&(c_1\boldsymbol{u_1} + \cdots + c_{i-1}\boldsymbol{u_{i-1}} + c_i\boldsymbol{u_i} + c_{i+1}\boldsymbol{u_{i+1}} + \cdots + c_k\boldsymbol{u_k}) \cdot \boldsymbol{u_i} \\
&= c_1(\boldsymbol{u_1} \cdot \boldsymbol{u_i}) + \cdots + c_{i-1}(\boldsymbol{u_{i-1}} \cdot \boldsymbol{u_i}) + c_i(\boldsymbol{u_i} \cdot \boldsymbol{u_i}) \\
&\qquad + c_{i+1}(\boldsymbol{u_{i+1}} \cdot \boldsymbol{u_i}) + \cdots + c_k(\boldsymbol{u_k} \cdot \boldsymbol{u_i}) \\
&= 0 + \cdots + 0 + c_i(\boldsymbol{u_i} \cdot \boldsymbol{u_i}) + 0 + \cdots + 0 \\
&= c_i(\boldsymbol{u_i} \cdot \boldsymbol{u_i}).
\end{aligned}$$

by Theorem 5.1.5

Since S is orthogonal, $\boldsymbol{u_j} \cdot \boldsymbol{u_i} = 0$ for $i \neq j$.

Orthogonal sets (Theorem 5.2.4)

Taking dot product on both sides of $(*)$ with $\boldsymbol{u_i}$, we have

$$c_i(\boldsymbol{u_i} \cdot \boldsymbol{u_i}) = (c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}) \cdot \boldsymbol{u_i} = \boldsymbol{0} \cdot \boldsymbol{u_i} = 0.$$

Given that $\boldsymbol{u_i} \neq \boldsymbol{0}$, (by Theorem 5.1.5) $\boldsymbol{u_i} \cdot \boldsymbol{u_i} \neq 0$.

So $c_i(\boldsymbol{u_i} \cdot \boldsymbol{u_i}) = 0$ implies $c_i = 0$.

Since $(*)$ has only the trivial solution, S is linearly independent.

Orthogonal & orthonormal bases (Definition 5.2.5 & Remark 5.2.6)

A basis S for a vector space is called an orthogonal basis if S is orthogonal.

A basis S for a vector space is called an orthonormal basis if S is orthonormal.

Suppose S is a set of nonzero vectors from a vector space V.

If we want to show that S is an orthogonal (respectively, orthonormal) basis for V, then we only need to check

(i) S is orthogonal (respectively, orthonormal);

(ii) $|S| = \dim(V)$ (if we know the dimension)
or $\mathrm{span}(S) = V$ (if we don't know the dimension).

Examples (Example 5.2.7)

1. The standard basis $E = \{ \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \}$ for $\mathbb{R}^n$ is an orthogonal basis as well as an orthonormal basis.

2. Let $\boldsymbol{u}_1 = (2, 0, 0)$, $\boldsymbol{u}_2 = (0, 1, 1)$, $\boldsymbol{u}_3 = (0, 1, -1)$;

   and $\boldsymbol{v}_1 = (1, 0, 0)$, $\boldsymbol{v}_2 = \left(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$,

   $$\boldsymbol{v}_3 = \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

   The set $\{ \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \}$ and $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \}$ are orthogonal bases for $\mathbb{R}^3$.

   The set $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \}$ is an orthonormal basis for $\mathbb{R}^3$.

Orthogonal bases (Theorem 5.2.8.1)

Let $S = \{ \boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k} \}$ be an orthogonal basis for a vector space V. Then for any $\boldsymbol{w} \in V$,

$$\boldsymbol{w} = \frac{\boldsymbol{w} \cdot \boldsymbol{u_1}}{\boldsymbol{u_1} \cdot \boldsymbol{u_1}} \boldsymbol{u_1} + \frac{\boldsymbol{w} \cdot \boldsymbol{u_2}}{\boldsymbol{u_2} \cdot \boldsymbol{u_2}} \boldsymbol{u_2} + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u_k}}{\boldsymbol{u_k} \cdot \boldsymbol{u_k}} \boldsymbol{u_k},$$

i.e. $(\boldsymbol{w})_S = \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u_1}}{\boldsymbol{u_1} \cdot \boldsymbol{u_1}}, \frac{\boldsymbol{w} \cdot \boldsymbol{u_2}}{\boldsymbol{u_2} \cdot \boldsymbol{u_2}}, \ldots, \frac{\boldsymbol{w} \cdot \boldsymbol{u_k}}{\boldsymbol{u_k} \cdot \boldsymbol{u_k}} \right)$.

Proof: Let $(\boldsymbol{w})_S = (c_1, c_2, \ldots, c_k)$,

i.e. $\boldsymbol{w} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$.

Then for $i = 1, 2, \ldots, k$,

$$\boldsymbol{w} \cdot \boldsymbol{u_i} = (c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}) \cdot \boldsymbol{u_i} = c_i(\boldsymbol{u_i} \cdot \boldsymbol{u_i})$$

and hence $c_i = \dfrac{\boldsymbol{w} \cdot \boldsymbol{u_i}}{\boldsymbol{u_i} \cdot \boldsymbol{u_i}}$.

S is orthogonal.

Orthonormal bases (Theorem 5.2.8.2)

Let $T = \{ \boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k} \}$ be an orthonormal basis for a vector space V. Then for any $\boldsymbol{w} \in V$,

$$\boldsymbol{w} = (\boldsymbol{w} \cdot \boldsymbol{v_1})\boldsymbol{v_1} + (\boldsymbol{w} \cdot \boldsymbol{v_2})\boldsymbol{v_2} + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v_k})\boldsymbol{v_k}$$

i.e. $(\boldsymbol{w})_S = (\boldsymbol{w} \cdot \boldsymbol{v_1},\ \boldsymbol{w} \cdot \boldsymbol{v_2},\ \ldots,\ \boldsymbol{w} \cdot \boldsymbol{v_k})$.

Proof: Let $(\boldsymbol{w})_S = (c_1, c_2, \ldots, c_k)$,

i.e. $\boldsymbol{w} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k}$.

Then for any $i = 1, 2, \ldots, k$,

$$\boldsymbol{w} \cdot \boldsymbol{v_i} = (c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_k\boldsymbol{v_k}) \cdot \boldsymbol{v_i} = c_i(\boldsymbol{v_i} \cdot \boldsymbol{v_i}) = c_i.$$

T is orthonormal.

Examples (Example 5.2.9.1)

Let $S = \{ \boldsymbol{v_1}, \boldsymbol{v_2} \}$ where $\boldsymbol{v_1} = \left(\frac{3}{5}, \frac{4}{5}\right)$ and $\boldsymbol{v_2} = \left(\frac{4}{5}, -\frac{3}{5}\right)$.

Note that S is an orthonormal basis for $\mathbb{R}^2$.

For any $\boldsymbol{w} = (x, y) \in \mathbb{R}^2$,

$$\boldsymbol{w} \cdot \boldsymbol{v_1} = \frac{3x + 4y}{5} \quad \text{and} \quad \boldsymbol{w} \cdot \boldsymbol{v_2} = \frac{4x - 3y}{5}.$$

So $\boldsymbol{w} = \dfrac{3x + 4y}{5}\boldsymbol{v_1} + \dfrac{4x - 3y}{5}\boldsymbol{v_2}$

and $(\boldsymbol{w})_S = \left(\dfrac{3x + 4y}{5}, \dfrac{4x - 3y}{5}\right)$.

Examples (Example 5.2.9.2)

Let $S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \}$ where $\boldsymbol{u}_1 = (1, 1, 1)$, $\boldsymbol{u}_2 = (1, 0, -1)$, $\boldsymbol{u}_3 = (1, -2, 1)$.

Note that S is an orthogonal basis for $\mathbb{R}^3$.

For $\boldsymbol{w} = (1, -1, 0) \in \mathbb{R}^3$,

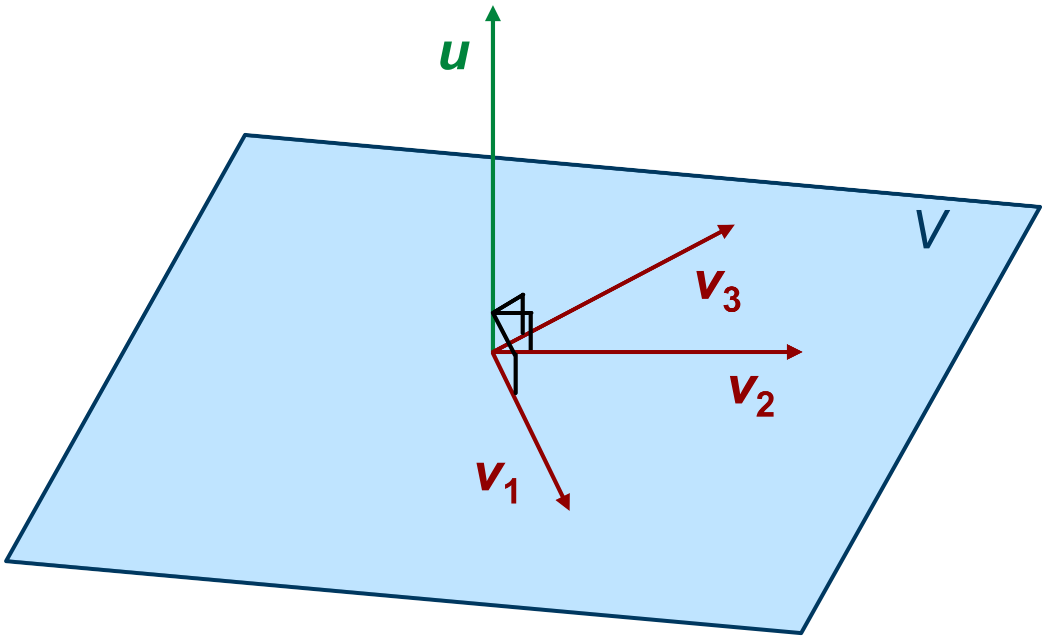
$$(\boldsymbol{w})_S = \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}, \frac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2}, \frac{\boldsymbol{w} \cdot \boldsymbol{u}_3}{\boldsymbol{u}_3 \cdot \boldsymbol{u}_3} \right)$$

$$= \left(0, \frac{1}{2}, \frac{1}{2} \right).$$

Orthogonality (Definition 5.2.10 & Example 5.2.11.1)

Let V be a subspace of $\mathbb{R}^n$.

A vector $\boldsymbol{u} \in \mathbb{R}^n$ is said to be orthogonal (or perpendicular) to V if $\boldsymbol{u}$ is orthogonal to all vectors in V.



Orthogonality (Example 5.2.11.1)

Let $V = \{ (x, y, z) \mid ax + by + cz = 0 \}$, where not all a, b, c are zero, which is a plane in $\mathbb{R}^3$ containing the origin.

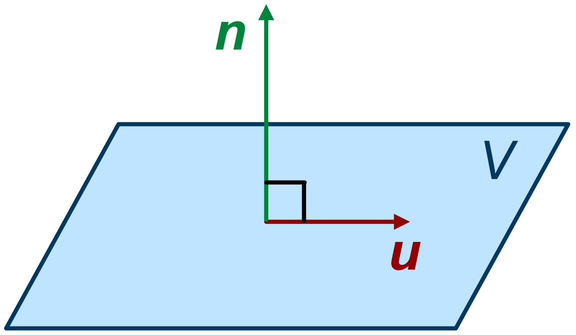
Let $\boldsymbol{n} = (a, b, c)$.

For any vector $\boldsymbol{u} = (u_1, u_2, u_3) \in V$,

$$\boldsymbol{n} \cdot \boldsymbol{u} = au_1 + bu_2 + cu_3 = 0.$$

Thus $\boldsymbol{n}$ is orthogonal to V.

(The vector $\boldsymbol{n}$ is called a normal vector of V.)



Orthogonality (Remark 5.2.12 & Example 5.2.11.2)

Let $V = \text{span}\{ \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k \}$ be a subspace of $\mathbb{R}^n$.

A vector $\boldsymbol{v} \in \mathbb{R}^n$ is orthogonal to V if only if $\boldsymbol{v} \cdot \boldsymbol{u}_i = 0$ for $i = 1, 2, \ldots, k$.

Let $V = \text{span}\{ \boldsymbol{u}_1, \boldsymbol{u}_2 \}$, where $\boldsymbol{u}_1 = (1, 1, 1, 0)$ and $\boldsymbol{u}_2 = (0, -1, -1, 1) \}$, be a subspace of $\mathbb{R}^4$.

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

$\boldsymbol{v}$ is orthogonal to V

$\Leftrightarrow \quad \boldsymbol{v} \cdot \boldsymbol{u}_1 = 0$ and $\boldsymbol{v} \cdot \boldsymbol{u}_2 = 0$

$$\Leftrightarrow \quad \begin{cases} w + x + y \phantom{{}+z} = 0 \\ - x - y + z = 0 \end{cases}$$

$\Leftrightarrow \quad (w, x, y, z) = (-t, \ -s + t, \ s, \ t)$ for some $s, t \in \mathbb{R}$.

Projections (Definition 5.2.13)

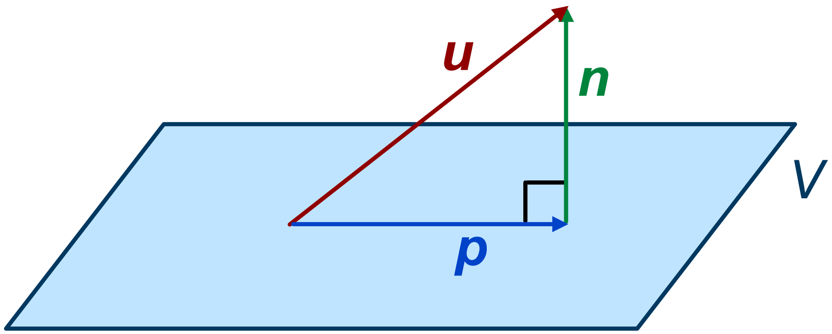
Let V be a subspace of $\mathbb{R}^n$.

Every $\boldsymbol{u} \in \mathbb{R}^n$ can be written uniquely as

$$\boldsymbol{u} = \boldsymbol{n} + \boldsymbol{p}$$

where $\boldsymbol{p}$ is a vector in V
and $\boldsymbol{n}$ is a vector orthogonal to V.

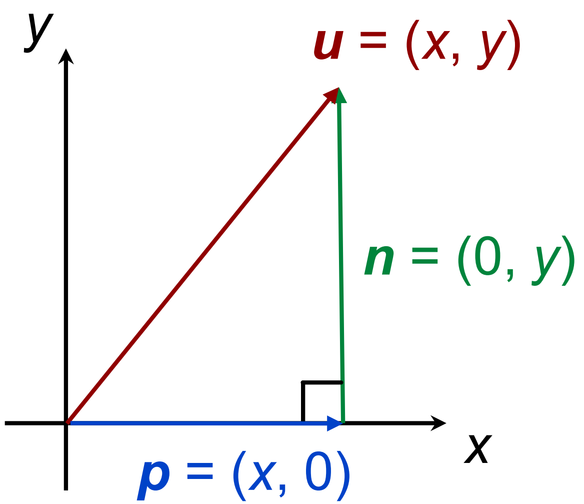
The vector $\boldsymbol{p}$ is called the (orthogonal) projection of $\boldsymbol{u}$ onto V.

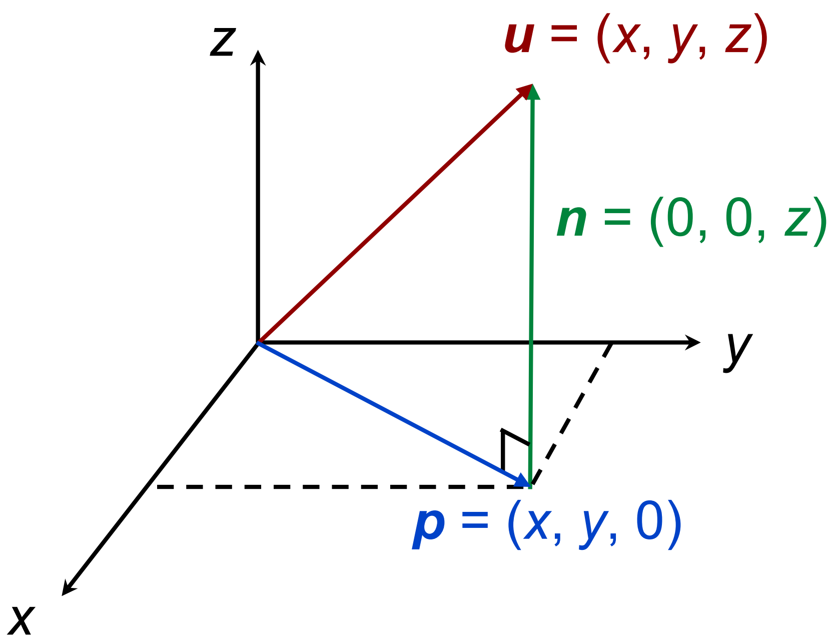


Examples (Example 5.2.14)

The projection of $\boldsymbol{u} = (x, y)$ onto the x-axis is $\boldsymbol{p} = (x, 0)$.

In here, $\boldsymbol{n} = (0, y)$.





The projection of $\boldsymbol{u} = (x, y, z)$ onto the xy-plane is $\boldsymbol{p} = (x, y, 0)$.

In here, $\boldsymbol{n} = (0, 0, z)$.

Orthogonal bases & projections (Theorem 5.2.15.1)

Let V be a subspace of $\mathbb{R}^n$ and $\{\, \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k \,\}$ an orthogonal basis for V.

Then for any $\boldsymbol{w} \in \mathbb{R}^n$,

$$\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2}\boldsymbol{u}_2 + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\boldsymbol{u}_k$$

is the projection of $\boldsymbol{w}$ onto V.

Proof: Define $\boldsymbol{p} = \frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2}\boldsymbol{u}_2 + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\boldsymbol{u}_k \in V$ and $\boldsymbol{n} = \boldsymbol{w} - \boldsymbol{p}$.

Since $\boldsymbol{w} = \boldsymbol{n} + \boldsymbol{p}$ where $\boldsymbol{p}$ is a vector in V, to show that $\boldsymbol{p}$ is a projection of $\boldsymbol{w}$ onto V, it suffices to show $\boldsymbol{n}$ is orthogonal to V.

Orthogonal bases & projections (Theorem 5.2.15.1)

To show $\boldsymbol{n}$ is orthogonal to V:

For $i = 1, 2, ..., k$,

$$\begin{aligned}
\boldsymbol{n} \cdot \boldsymbol{u}_i &= (\boldsymbol{w} - \boldsymbol{p}) \cdot \boldsymbol{u}_i \\
&= \boldsymbol{w} \cdot \boldsymbol{u}_i - \boldsymbol{p} \cdot \boldsymbol{u}_i \\
&= \boldsymbol{w} \cdot \boldsymbol{u}_i - \left(\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} \boldsymbol{u}_1 + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2} \boldsymbol{u}_2 + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k} \boldsymbol{u}_k \right) \cdot \boldsymbol{u}_i \\
&= \boldsymbol{w} \cdot \boldsymbol{u}_i - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} (\boldsymbol{u}_1 \cdot \boldsymbol{u}_i) - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2} (\boldsymbol{u}_2 \cdot \boldsymbol{u}_i) - \cdots \\
&\qquad - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k} (\boldsymbol{u}_k \cdot \boldsymbol{u}_i) \\
&= \boldsymbol{w} \cdot \boldsymbol{u}_i - \frac{\boldsymbol{w} \cdot \boldsymbol{u}_i}{\boldsymbol{u}_i \cdot \boldsymbol{u}_i} (\boldsymbol{u}_i \cdot \boldsymbol{u}_i) \\
&= 0.
\end{aligned}$$

So $\boldsymbol{n}$ is orthogonal to V.

Orthonormal bases & projections (Theorem 5.2.15.2 & Remark 5.2.17)

Let V be a subspace of $\mathbb{R}^n$ and $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \}$ an orthonormal basis for V.

Then for any $\boldsymbol{w} \in \mathbb{R}^n$,

$$(\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{w} \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$$

is the projection of $\boldsymbol{w}$ onto V.

(Theorem 5.2.8 can be regarded as a particular case of Theorem 5.2.15 when $\boldsymbol{w}$ is contained in V, i.e. $\boldsymbol{w} = \boldsymbol{p}$ and $\boldsymbol{n} = \boldsymbol{0}$.)

An example (Example 5.2.16)

Let $V = \text{span}\{\, \boldsymbol{u}_1, \boldsymbol{u}_2 \,\}$

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

Note that $\{\, \boldsymbol{u}_1, \boldsymbol{u}_2 \,\}$ is an orthogonal, basis for V.

For $\boldsymbol{w} = (1, 1, 0) \in \mathbb{R}^3$, the projection of $\boldsymbol{w}$ onto V is

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

Projections (Discussion 5.2.18.1)

Let $\{ \boldsymbol{u}_1, \boldsymbol{u}_2 \}$ be a basis for a vector space V (where V is either $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$ containing the origin).

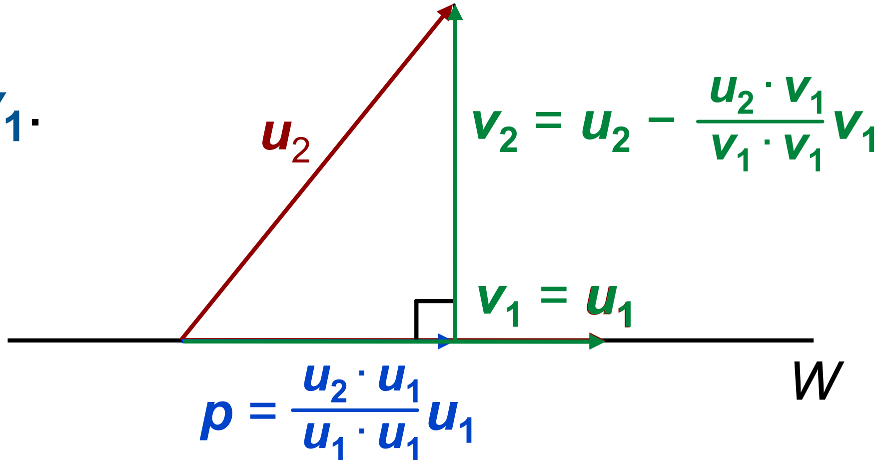
Let $W = \text{span}\{ \boldsymbol{u}_1 \}$ which is a subspace of V (W is a line through the origin).

The projection of $\boldsymbol{u}_2$ onto W is $\boldsymbol{p} = \dfrac{\boldsymbol{u}_2 \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1$.

Let $\boldsymbol{v}_1 = \boldsymbol{u}_1$

and $\boldsymbol{v}_2 = \boldsymbol{u}_2 - \boldsymbol{p} = \boldsymbol{u}_2 - \dfrac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1$.

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



Projections (Discussion 5.2.18.1)

Let $\{ \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \}$ be a basis for $\mathbb{R}^3$.

Let $V = \text{span}\{ \boldsymbol{u}_1, \boldsymbol{u}_2 \}$ which is a subspace of $\mathbb{R}^3$ (V is a plane containing the origin).

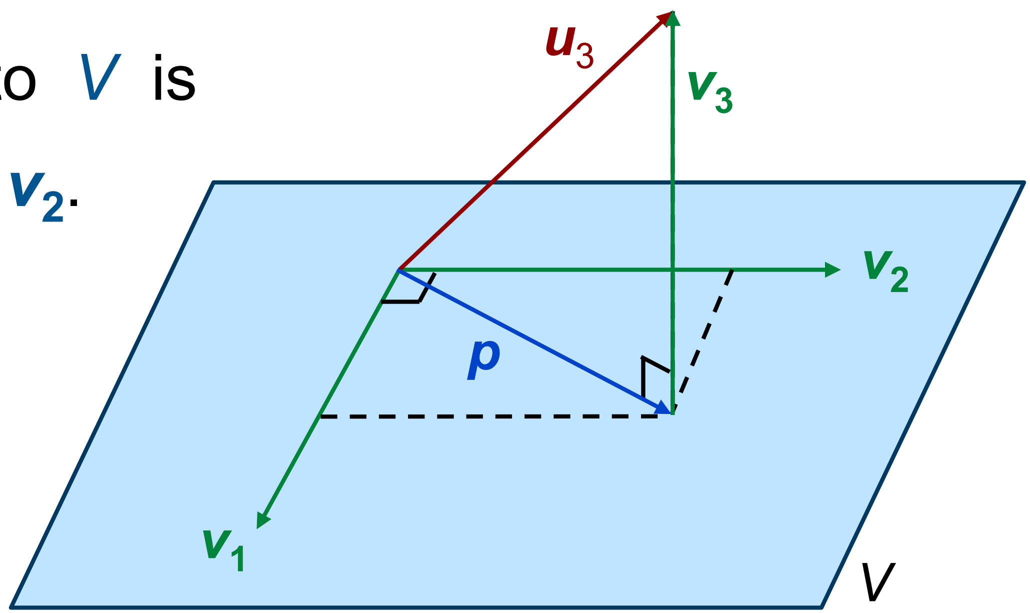
With $\boldsymbol{v}_1 = \boldsymbol{u}_1$ and $\boldsymbol{v}_2 = \boldsymbol{u}_2 - \dfrac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1$,
$\{ \boldsymbol{v}_1, \boldsymbol{v}_2 \}$ be an orthogonal basis for V.

The projection of $\boldsymbol{u}_3$ onto V is

$$\boldsymbol{p} = \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 + \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2.$$

Define $\boldsymbol{v}_3 = \boldsymbol{u}_3 - \boldsymbol{p}$.

Then $\{ \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \}$ is an orthogonal basis for $\mathbb{R}^3$.



Gram-Schmidt Process (Theorem 5.2.19)

Let $\{ \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k \}$ be a basis for a vector space V.

Let $\boldsymbol{v}_1 = \boldsymbol{u}_1$,

$$\boldsymbol{v}_2 = \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1,$$

$$\boldsymbol{v}_3 = \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2} \boldsymbol{v}_2,$$

$$\vdots$$

$$\boldsymbol{v}_k = \boldsymbol{u}_k - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2} \boldsymbol{v}_2 - \cdots - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_{k-1}}{\boldsymbol{v}_{k-1} \cdot \boldsymbol{v}_{k-1}} \boldsymbol{v}_{k-1}.$$

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

Furthermore, $\left\{ \frac{1}{\|\boldsymbol{v}_1\|} \boldsymbol{v}_1, \frac{1}{\|\boldsymbol{v}_2\|} \boldsymbol{v}_2, \ldots, \frac{1}{\|\boldsymbol{v}_k\|} \boldsymbol{v}_k \right\}$ is an orthonormal basis for V.

An example (Example 5.2.20)

Let $\boldsymbol{u}_1 = (1, -1, 2)$, $\boldsymbol{u}_2 = (2, 1, 0)$ and $\boldsymbol{u}_3 = (0, 0, 1)$.

$\{ \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \}$ is a basis for $\mathbb{R}^3$.

Let $\boldsymbol{v}_1 = \boldsymbol{u}_1 = (1, -1, 2)$,

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

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

An example (Example 5.2.20)

Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$.

Furthermore, the following is an orthonormal basis for $\mathbb{R}^3$:

$$\left\{ \frac{1}{||\boldsymbol{v}_1||}\boldsymbol{v}_1,\ \frac{1}{||\boldsymbol{v}_2||}\boldsymbol{v}_2,\ \frac{1}{||\boldsymbol{v}_3||}\boldsymbol{v}_3 \right\}$$

$$= \left\{ \frac{1}{\sqrt{6}}(1, -1, 2),\ \frac{1}{\sqrt{29/6}}\left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right),\ \frac{1}{\sqrt{9/29}}\left(-\frac{6}{29}, \frac{12}{29}, \frac{9}{29}\right) \right\}$$

$$= \left\{ \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right), \left(\frac{11}{\sqrt{174}}, \frac{7}{\sqrt{174}}, -\frac{2}{\sqrt{174}}\right), \left(-\frac{2}{\sqrt{29}}, \frac{4}{\sqrt{29}}, \frac{3}{\sqrt{29}}\right) \right\}.$$

Section 5.3

Best Approximations

Best Approximations (Theorem 5.3.2)

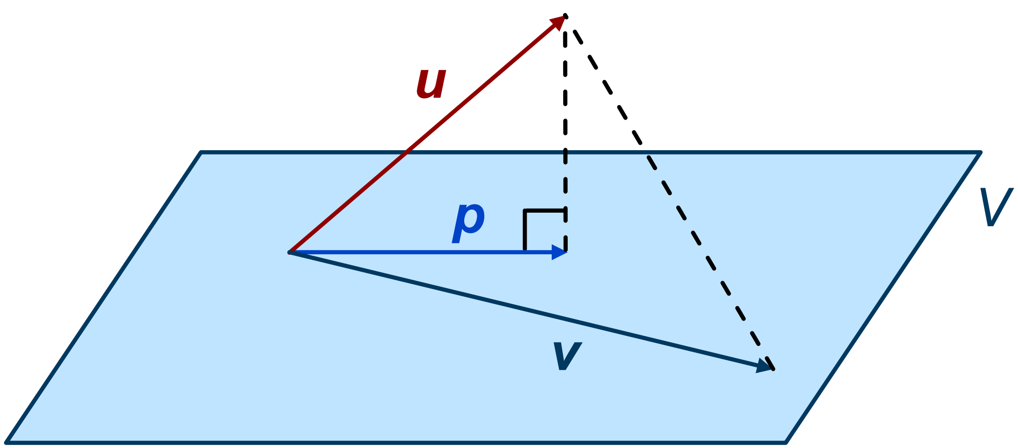
Let V be a subspace of $\mathbb{R}^n$.

Take any $\boldsymbol{u} \in \mathbb{R}^n$ and let $\boldsymbol{p}$ be the projection of $\boldsymbol{u}$ onto V.

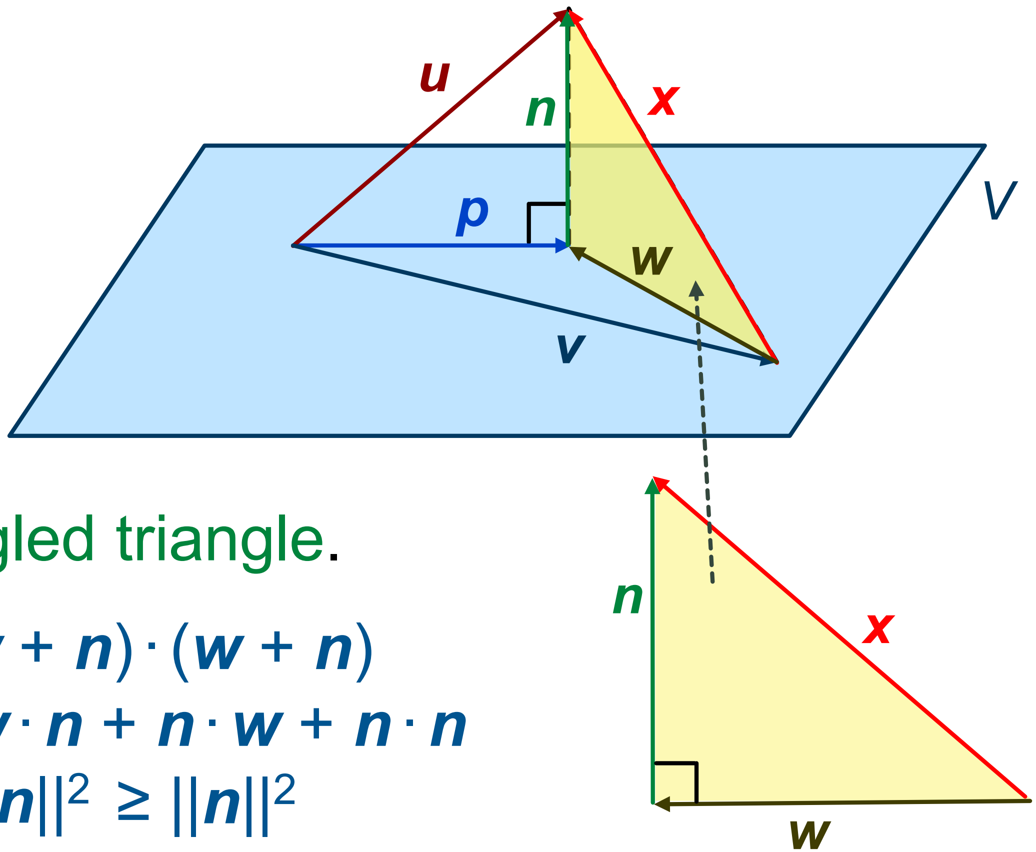
Then

$$d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in V,$$

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



Proof of Best Approximation (Theorem 5.3.2)



Proof: Define

$\boldsymbol{n} = \boldsymbol{u} - \boldsymbol{p}$,

$\boldsymbol{w} = \boldsymbol{p} - \boldsymbol{v}$,

and $\boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v}$.

Observe that $\boldsymbol{n}$, $\boldsymbol{w}$ and $\boldsymbol{x}$ form a right-angled triangle.

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

$\Rightarrow \quad \|\boldsymbol{x}\| \geq \|\boldsymbol{n}\|$.

Thus $d(\boldsymbol{u}, \boldsymbol{p}) = \|\boldsymbol{u} - \boldsymbol{p}\| = \|\boldsymbol{n}\| \leq \|\boldsymbol{x}\| = \|\boldsymbol{u} - \boldsymbol{v}\| = d(\boldsymbol{u}, \boldsymbol{v})$.

An example (Example 5.3.3)

Let $V = \text{span}\{ (1, 0, 1), (1, 1, 1) \}$ which is a plane in $\mathbb{R}^3$ containing the origin.

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

Solution: The shortest distance from $\boldsymbol{u}$ to V is $d(\boldsymbol{u}, \boldsymbol{p})$ where $\boldsymbol{p}$ is the projection of $\boldsymbol{u}$ onto V (by Theorem 5.3.2).

First, applying the Gram-Schmidt Process (Theorem 5.2.19), the vectors

$$(1, 0, 1) \quad \text{and} \quad (1, 1, 1) - \frac{(1, 1, 1) \cdot (1, 0, 1)}{(1, 0, 1) \cdot (1, 0, 1)}(1, 0, 1) = (0, 1, 0)$$

form an orthogonal basis for V.

An example (Example 5.3.3)

Thus (by Theorem 5.2.15)

$$\boldsymbol{p} = \frac{(1, 2, 3) \cdot (1, 0, 1)}{(1, 0, 1) \cdot (1, 0, 1)}(1, 0, 1) + \frac{(1, 2, 3) \cdot (0, 1, 0)}{(0, 1, 0) \cdot (0, 1, 0)}(0, 1, 0)$$

$$= (2, 2, 2)$$

and the distance from $\boldsymbol{u}$ to V is

$$d(\boldsymbol{u}, \boldsymbol{p}) = \|\boldsymbol{u} - \boldsymbol{p}\| = \|(1, 2, 3) - (2, 2, 2)\|$$

$$= \|(-1, 0, 1)\| = \sqrt{2}.$$

Fitting experimental data (Remark 5.3.4 & Example 5.3.5)

In analyzing experimental results, scientists always face a problem of fitting experimental data to an equation.

For example, suppose r, s and t are physical quantities that satisfy the rule

$$t = cr^2 + ds + e$$

for some constants c, d and e.

An experiment was conducted in order to find the constants c, d and e.

Six measurements of t were taken with various setting of r and s.

i	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

Fitting experimental data (Example 5.3.5)

If there are no experimental errors, we have

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

Let $\boldsymbol{A} = \begin{bmatrix} r_1^2 & s_1 & 1 \\ r_2^2 & s_2 & 1 \\ \vdots & \vdots & \vdots \\ r_6^2 & s_6 & 1 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} c \\ d \\ e \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_6 \end{bmatrix}$.

By solving the linear system $\boldsymbol{Ax} = \boldsymbol{b}$, we can obtain the values c, d and e.

Fitting experimental data (Example 5.3.5)

However, due to experimental errors, we do not expect to get the exact values of t_i's.

The system $\boldsymbol{Ax} = \boldsymbol{b}$ is usually inconsistent.

We cannot obtain the values c, d, e directly.

The usual scheme is to get the approximate values of c, d, e that minimize the sum of squares of errors (SSE):

$$[t_1 - (cr_1^2 + ds_1 + e)]^2 + [t_2 - (cr_2^2 + ds_2 + e)]^2 + \cdots + [t_6 - (cr_6^2 + ds_6 + e)]^2$$

$$= \|\boldsymbol{b} - \boldsymbol{Ax}\|^2.$$

$$\boldsymbol{b} - \boldsymbol{Ax} = \begin{bmatrix} t_1 - (cr_1^2 + ds_1 + e) \\ t_2 - (cr_2^2 + ds_2 + e) \\ \vdots \\ t_6 - (cr_6^2 + ds_6 + e) \end{bmatrix}$$

To minimize the SSE is equivalent to find $\boldsymbol{x}$ that minimize $\|\boldsymbol{b} - \boldsymbol{Ax}\|$.

Least square solutions (Definition 5.3.6 & Discussion 5.3.7)

Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system where $\boldsymbol{A}$ is an $m \times n$ matrix.

A vector $\boldsymbol{u} \in \mathbb{R}^n$ is called a least square solution to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ if

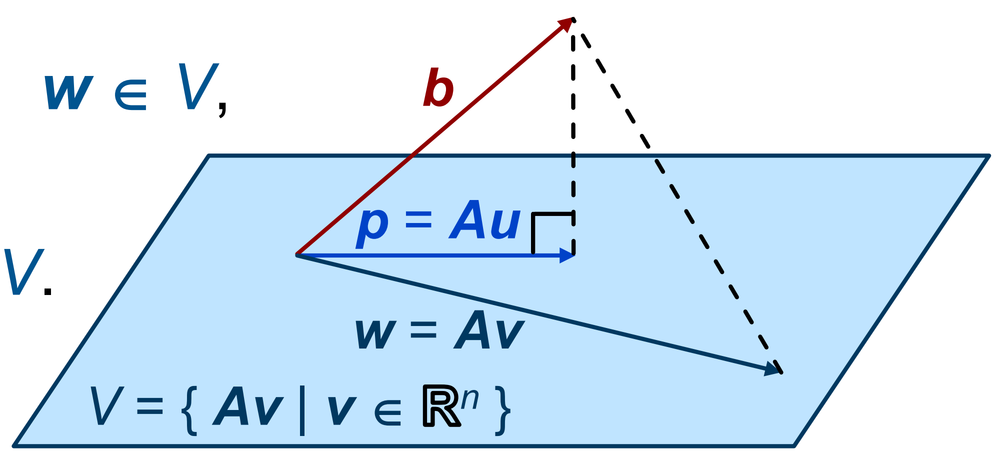
$$||\boldsymbol{b} - \boldsymbol{Au}|| \leq ||\boldsymbol{b} - \boldsymbol{Av}|| \text{ for all } \boldsymbol{v} \in \mathbb{R}^n. \qquad (\#)$$

Let $V = \{ \boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n \}$ and $\boldsymbol{p} = \boldsymbol{Au}$.

Then (#) is rewritten as

$$d(\boldsymbol{b}, \boldsymbol{p}) \leq d(\boldsymbol{b}, \boldsymbol{w}) \text{ for all } \boldsymbol{w} \in V,$$

i.e. $\boldsymbol{p} = \boldsymbol{Au}$ is the best approximation of $\boldsymbol{b}$ onto V.



Least square solutions (Discussion 5.3.7 & Theorem 5.3.8)

Recall that (by Theorem 4.1.16)

$$V = \{ \boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n \} = \text{the column space of } \boldsymbol{A}.$$

Then $\boldsymbol{u} \in \mathbb{R}^n$ is a least square solution to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$

if and only if $\boldsymbol{p} = \boldsymbol{Au}$ is the best approximation of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$

if and only if $\boldsymbol{p} = \boldsymbol{Au}$ is the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$ (by Theorem 5.3.2).

An example (Example 5.3.9)

Let $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and

$$V = \text{the column space of } \boldsymbol{A} = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

We know that (by Example 5.3.3) the projection of $\boldsymbol{b}$ onto V is $\boldsymbol{p} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$.

An example (Example 5.3.9)

Thus (by Theorem 5.3.8) $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ is a least square solution

to $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

which implies $\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$.

Least square solutions (Theorem 5.3.10)

Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system.

Then $\boldsymbol{u}$ is a least square solution to the system $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if $\boldsymbol{u}$ is a solution to

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

Proof: Write $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$

where $\boldsymbol{a}_j$ is the j^{th} column of $\boldsymbol{A}$.

Let V be the column space of $\boldsymbol{A}$,

i.e. $V = \mathrm{span}\{\, \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \,\} = \{\, \boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n \,\}$.

Least square solutions (Theorem 5.3.10)

$\boldsymbol{u}$ is a least square solution to $\boldsymbol{Ax} = \boldsymbol{b}$

$\Leftrightarrow$ $\boldsymbol{Au}$ is the projection of $\boldsymbol{b}$ onto V (by Theorem 5.3.8)

$\Leftrightarrow$ $\boldsymbol{b} - \boldsymbol{Au}$ is orthogonal to V (by Definition 5.2.13)

$\Leftrightarrow$ $\boldsymbol{b} - \boldsymbol{Au}$ is orthogonal to $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ (by Remark 5.2.12)

$\Leftrightarrow$ $\boldsymbol{a}_1 \cdot (\boldsymbol{b} - \boldsymbol{Au}) = 0, \ \boldsymbol{a}_2 \cdot (\boldsymbol{b} - \boldsymbol{Au}) = 0, \ \ldots, \ \boldsymbol{a}_n \cdot (\boldsymbol{b} - \boldsymbol{Au}) = 0$

$\Leftrightarrow$ $\boldsymbol{a}_1^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{Au}) = 0, \ \boldsymbol{a}_2^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{Au}) = 0, \ \ldots, \ \boldsymbol{a}_n^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{Au}) = 0$ (by Remark 5.1.3)

$$\Leftrightarrow \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{bmatrix} (\boldsymbol{b} - \boldsymbol{Au}) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \Leftrightarrow \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{Au}) = \boldsymbol{0}$$

$$\Leftrightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Au} = \boldsymbol{0}$$

$$\Leftrightarrow \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Au} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

Examples (Example 5.3.11.1)

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

Then a least square solution to $\boldsymbol{Ax} = \boldsymbol{b}$ is a solution to

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\Leftrightarrow \quad \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \end{bmatrix} \quad \Leftrightarrow \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

Examples (Example 5.3.11.2)

For the example of fitting experimental data (Example 5.3.5), the linear system is

$$\left\{\begin{aligned} e &= 0.5 \\ d + e &= 1.6 \\ c + 2d + e &= 2.8 \\ c \qquad + e &= 0.8 \\ 4c + d + e &= 5.1 \\ 4c + 2d + e &= 5.9 \end{aligned}\right. \Leftrightarrow \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{bmatrix}.$$

Examples (Example 5.3.11.2)

Then a least square solution to the linear system is a solution to

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 4 & 4 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 & 4 & 4 \\ 0 & 1 & 2 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 34 & 14 & 10 \\ 14 & 10 & 6 \\ 10 & 6 & 6 \end{bmatrix} \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 47.6 \\ 24.1 \\ 16.7 \end{bmatrix} \quad \Leftrightarrow \begin{bmatrix} c \\ d \\ e \end{bmatrix} = \begin{bmatrix} 0.9275 \\ 0.9225 \\ 0.3150 \end{bmatrix}.$$

Examples (Example 5.3.11.3)

We demonstrate how to find the projection using a least square solution.

Let $V = \text{span}\{\,(1, -1, 1, -1),\ (1, 2, 0, 1),\ (2, 1, 1, 0)\,\}$.

Find the projection of $(1, 1, 1, 1)$ onto V.

Solution: Let $\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$.

We first find a least square solution $\boldsymbol{u}$ to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$, then (by Theorem 5.3.8) $\boldsymbol{Au}$ is the projection of $\boldsymbol{b}$ onto V.

Examples (Example 5.3.11.3)

The equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ is

$$\begin{bmatrix} 4 & -2 & 2 \\ -2 & 6 & 4 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$$

which gives us a general solution

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -t + \frac{2}{5} \\ -t + \frac{4}{5} \\ t \end{bmatrix}$$ where t is an arbitrary parameter.

Any one of the solutions is a least square solution to the system $\boldsymbol{Ax} = \boldsymbol{b}$.

Examples (Example 5.3.11.3)

Take $\boldsymbol{u} = \begin{bmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{bmatrix}$. Then $\boldsymbol{Au} = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{6}{5} \\ \frac{6}{5} \\ \frac{2}{5} \\ \frac{2}{5} \end{bmatrix}$

So $\left(\frac{6}{5}, \frac{6}{5}, \frac{2}{5}, \frac{2}{5}\right)$ is the projection of $(1, 1, 1, 1)$ onto V.

(Although in this example, there are infinitely many least square solutions, all of them will give us the same projection vector.)

Section 5.4

Orthogonal Matrices

Transition matrices (Discussion 5.4.1)

Let $S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k \}$ and $T = \{ \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \}$ be two bases for a vector space V.

Recall that the matrix

$$\boldsymbol{P} = \begin{bmatrix} [\boldsymbol{u}_1]_T & [\boldsymbol{u}_2]_T & \cdots & [\boldsymbol{u}_k]_T \end{bmatrix}$$

is the transition matrix from S to T.

For any $\boldsymbol{w} \in V$, $[\boldsymbol{w}]_T = \boldsymbol{P}[\boldsymbol{w}]_S$.

If both S and T are orthonormal bases, the transition matrix $\boldsymbol{P}$ has some interesting properties.

An example (Example 5.4.2)

Let $E = \{ \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3 \}$ be the standard bases for $\mathbb{R}^3$,
i.e. $\boldsymbol{e}_1 = (1, 0, 0)$, $\boldsymbol{e}_2 = (0, 1, 0)$, $\boldsymbol{e}_3 = (0, 0, 1)$,
and let $S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \}$ where

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1, 1, 1), \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1, 0, -1), \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{6}}(1, -2, 1).$$

Both E and S are orthonormal bases for $\mathbb{R}^3$.

$$\begin{aligned}
\boldsymbol{u}_1 &= \frac{1}{\sqrt{3}}\mathbf{e}_1 + \frac{1}{\sqrt{3}}\mathbf{e}_2 + \frac{1}{\sqrt{3}}\mathbf{e}_3, \\
\boldsymbol{u}_2 &= \frac{1}{\sqrt{2}}\mathbf{e}_1 \qquad\qquad - \frac{1}{\sqrt{2}}\mathbf{e}_3, \\
\boldsymbol{u}_3 &= \frac{1}{\sqrt{6}}\mathbf{e}_1 - \frac{2}{\sqrt{6}}\mathbf{e}_2 + \frac{1}{\sqrt{6}}\mathbf{e}_3.
\end{aligned}$$

The transition matrix from S to E is

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

An example (Example 5.4.2)

As S is an orthonormal basis for $\mathbb{R}^3$, (by Theorem 5.2.8)

$$\boldsymbol{e}_1 = (\boldsymbol{e}_1 \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{e}_1 \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + (\boldsymbol{e}_1 \cdot \boldsymbol{u}_3)\boldsymbol{u}_3 = \frac{1}{\sqrt{3}}\boldsymbol{u}_1 + \frac{1}{\sqrt{2}}\boldsymbol{u}_2 + \frac{1}{\sqrt{6}}\boldsymbol{u}_3,$$

$$\boldsymbol{e}_2 = (\boldsymbol{e}_2 \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{e}_2 \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + (\boldsymbol{e}_2 \cdot \boldsymbol{u}_3)\boldsymbol{u}_3 = \frac{1}{\sqrt{3}}\boldsymbol{u}_1 - \frac{2}{\sqrt{6}}\boldsymbol{u}_3,$$

$$\boldsymbol{e}_3 = (\boldsymbol{e}_3 \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{e}_3 \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + (\boldsymbol{e}_3 \cdot \boldsymbol{u}_3)\boldsymbol{u}_3 = \frac{1}{\sqrt{3}}\boldsymbol{u}_1 - \frac{1}{\sqrt{2}}\boldsymbol{u}_2 + \frac{1}{\sqrt{6}}\boldsymbol{u}_3.$$

The transition matrix from E to S is

$$\boldsymbol{Q} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

Note that $\boldsymbol{Q} = \boldsymbol{P}^{\mathrm{T}}$.

On the other hand, (by Theorem 3.7.5) $\boldsymbol{Q} = \boldsymbol{P}^{-1}$.

Thus $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$.

Orthogonal matrices (Definition 5.4.3 & Remark 5.4.4 & Example 5.3.5)

A square matrix $\boldsymbol{A}$ is called orthogonal if $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$.

To show that a square matrix $\boldsymbol{A}$ is an orthogonal matrix, (by Theorem 2.4.14) we only need to check that $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$ (or $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$).

The following are some examples of orthogonal matrices:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix} \qquad \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Orthogonal matrices (Theorem 5.4.6)

Let $\boldsymbol{A}$ be a square matrix of order n.

The following statements are equivalent:

1. $\boldsymbol{A}$ is orthogonal.
2. The rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
3. The columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.

Proof: We only prove $1 \Leftrightarrow 2$ in the following.

Write $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix}$ where $\boldsymbol{a}_i$ is the i^{th} row of $\boldsymbol{A}$.

Orthogonal matrices (Theorem 5.4.6)

Observe that

$$\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1^{\mathsf{T}} & \boldsymbol{a}_2^{\mathsf{T}} & \cdots & \boldsymbol{a}_n^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1\boldsymbol{a}_1^{\mathsf{T}} & \boldsymbol{a}_1\boldsymbol{a}_2^{\mathsf{T}} & \cdots & \boldsymbol{a}_1\boldsymbol{a}_n^{\mathsf{T}} \\ \boldsymbol{a}_2\boldsymbol{a}_1^{\mathsf{T}} & \boldsymbol{a}_2\boldsymbol{a}_2^{\mathsf{T}} & \cdots & \boldsymbol{a}_2\boldsymbol{a}_n^{\mathsf{T}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_n\boldsymbol{a}_1^{\mathsf{T}} & \boldsymbol{a}_n\boldsymbol{a}_2^{\mathsf{T}} & \cdots & \boldsymbol{a}_n\boldsymbol{a}_n^{\mathsf{T}} \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{a}_n \\ \boldsymbol{a}_2 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_2 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2 \cdot \boldsymbol{a}_n \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_n \cdot \boldsymbol{a}_1 & \boldsymbol{a}_n \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \cdot \boldsymbol{a}_n \end{bmatrix}.$$

Orthogonal matrices (Theorem 5.4.6)

$\boldsymbol{A}$ is orthogonal

$\Leftrightarrow$ $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$ (by Remark 5.4.4)

$$\Leftrightarrow \begin{bmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{a}_n \\ \boldsymbol{a}_2 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_2 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2 \cdot \boldsymbol{a}_n \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_n \cdot \boldsymbol{a}_1 & \boldsymbol{a}_n \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \cdot \boldsymbol{a}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$\Leftrightarrow$ for all i, j, $\boldsymbol{a}_i \cdot \boldsymbol{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

$\Leftrightarrow$ $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ are orthonormal

$\Leftrightarrow$ $\{ \boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n \}$ is an orthonormal basis for $\mathbb{R}^n$

(by Remark 5.2.6).

Transition matrices (Theorem 5.4.7)

Let S and T be two orthonormal bases for a vector space and let $\boldsymbol{P}$ be the transition matrix from S to T.

1. $\boldsymbol{P}$ is orthogonal.
2. $\boldsymbol{P}^{\mathrm{T}}$ is the transition matrix from T to S.

Proof: Let $S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k \}$ and $T = \{ \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k \}$.

Since T is an orthonormal basis for $\mathbb{R}^3$, (by Theorem 5.2.8)

$$\begin{aligned}
\boldsymbol{u}_1 &= (\boldsymbol{u}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_k)\boldsymbol{v}_k, \\
\boldsymbol{u}_2 &= (\boldsymbol{u}_2 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{u}_2 \cdot \boldsymbol{v}_k)\boldsymbol{v}_k, \\
&\vdots \\
\boldsymbol{u}_k &= (\boldsymbol{u}_k \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_k \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{u}_k \cdot \boldsymbol{v}_k)\boldsymbol{v}_k.
\end{aligned}$$

Transition matrices (Theorem 5.4.7)

Thus the transition matrix from S to T is

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{v}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_1 & \cdots & \boldsymbol{u}_k \cdot \boldsymbol{v}_1 \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_2 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_2 & \cdots & \boldsymbol{u}_k \cdot \boldsymbol{v}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_k & \boldsymbol{u}_2 \cdot \boldsymbol{v}_k & \cdots & \boldsymbol{u}_k \cdot \boldsymbol{v}_k \end{bmatrix}.$$

Similarly, the transition matrix from T to S is

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{v}_1 \cdot \boldsymbol{u}_1 & \boldsymbol{v}_2 \cdot \boldsymbol{u}_1 & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{u}_1 \\ \boldsymbol{v}_1 \cdot \boldsymbol{u}_2 & \boldsymbol{v}_2 \cdot \boldsymbol{u}_2 & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{u}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_1 \cdot \boldsymbol{u}_k & \boldsymbol{v}_2 \cdot \boldsymbol{u}_k & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{u}_k \end{bmatrix}.$$

Transition matrices (Theorem 5.4.7)

The transition matrix from S to T:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{v}_1 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_1 & \cdots & \boldsymbol{u}_k \cdot \boldsymbol{v}_1 \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_2 & \boldsymbol{u}_2 \cdot \boldsymbol{v}_2 & \cdots & \boldsymbol{u}_k \cdot \boldsymbol{v}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{u}_1 \cdot \boldsymbol{v}_k & \boldsymbol{u}_2 \cdot \boldsymbol{v}_k & \cdots & \boldsymbol{u}_k \cdot \boldsymbol{v}_k \end{pmatrix}$$

The transition matrix from T to S:

$$\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{v}_1 \cdot \boldsymbol{u}_1 & \boldsymbol{v}_2 \cdot \boldsymbol{u}_1 & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{u}_1 \\ \boldsymbol{v}_1 \cdot \boldsymbol{u}_2 & \boldsymbol{v}_2 \cdot \boldsymbol{u}_2 & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{u}_2 \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v}_1 \cdot \boldsymbol{u}_k & \boldsymbol{v}_2 \cdot \boldsymbol{u}_k & \cdots & \boldsymbol{v}_k \cdot \boldsymbol{u}_k \end{pmatrix}$$

For all i, j,

by Theorem 5.1.5.1

the (i, j)-entry of $\boldsymbol{P} = \boldsymbol{u}_j \cdot \boldsymbol{v}_i = \boldsymbol{v}_i \cdot \boldsymbol{u}_j =$ the (j, i)-entry of $\boldsymbol{Q}$.

Thus the transition matrix from T to S is $\boldsymbol{Q} = \boldsymbol{P}^{\mathrm{T}}$.

On the other hand, (by Theorem 3.7.5) $\boldsymbol{P}^{-1}$ is the transition matrix from T to S.

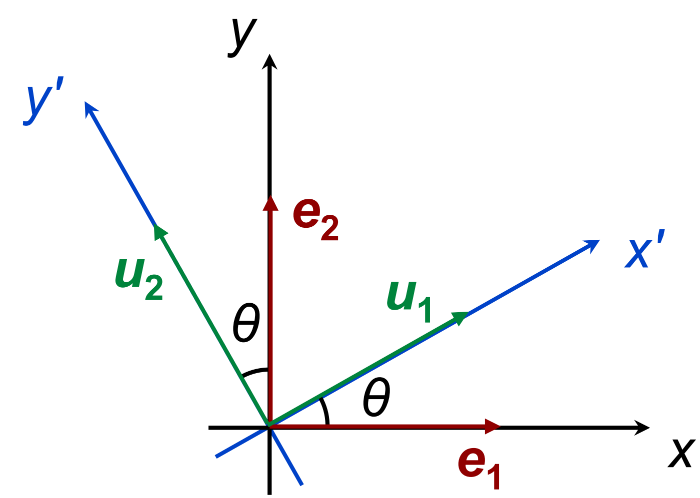
So $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$ and hence $\boldsymbol{P}$ is orthogonal.

Rotation of *xy*-coordinates (Example 5.4.8.1)

Let $E = \{ \mathbf{e}_1, \mathbf{e}_2 \}$ be the standard bases for $\mathbb{R}^2$ where

$\mathbf{e}_1 = (1, 0)$ is in the same direction as the x-axis,
$\mathbf{e}_2 = (0, 1)$ is in the same direction as the y-axis.

Consider a new $x'y'$-coordinate system obtained by rotating the original xy-coordinates anti-clockwise about the origin through an angle θ.



Let $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ be the unit vectors such that

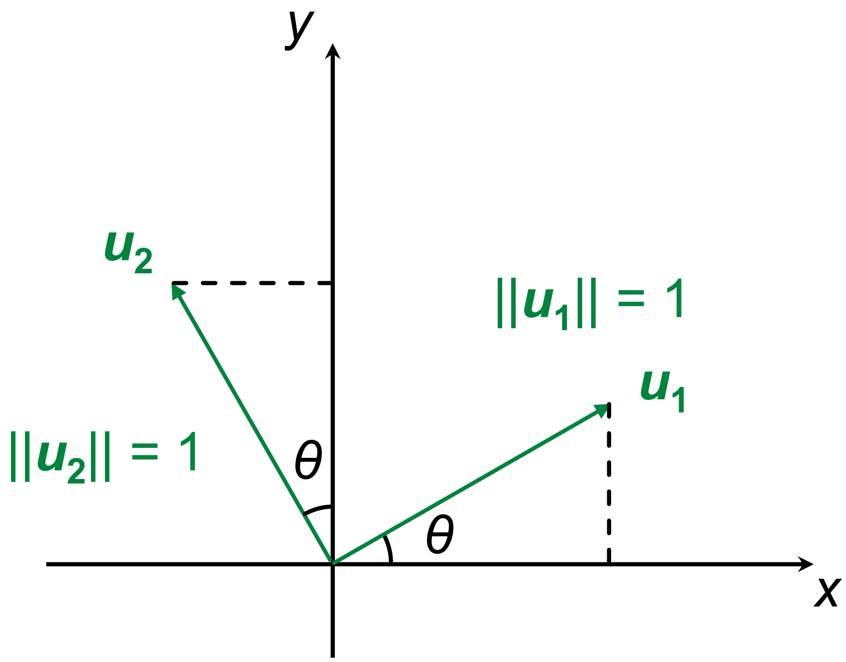
$\boldsymbol{u}_1$ is in the direction of the x'-axis,
$\boldsymbol{u}_2$ is in the direction of the y'-axis.

$S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2 \}$ is an orthonormal basis for $\mathbb{R}^2$.

Rotation of *xy*-coordinates (Example 5.4.8.1)

$$\boldsymbol{u}_1 = (\cos(\theta),\ \sin(\theta)) = \cos(\theta)\,\mathbf{e_1} + \sin(\theta)\,\mathbf{e_2},$$

$$\boldsymbol{u}_2 = (-\sin(\theta),\ \cos(\theta)) = -\sin(\theta)\,\mathbf{e_1} + \cos(\theta)\,\mathbf{e_2}.$$



The transition matrix from S to E is

$$\boldsymbol{P} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Thus (by Theorem 5.4.7) the transition matrix from E to S is

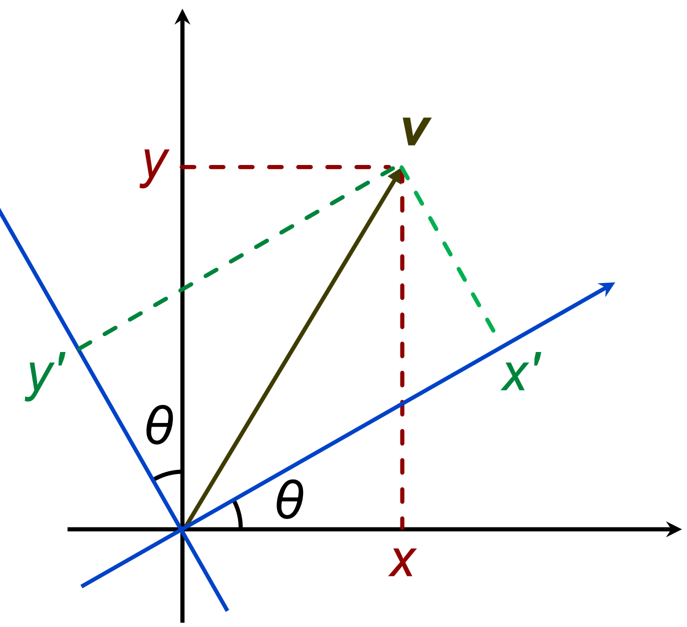
$$\boldsymbol{P}^{\mathrm{T}} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Rotation of *xy*-coordinates (Example 5.4.8.1)

Let $\boldsymbol{v} = (x, y) \in \mathbb{R}^2$

and let $(\boldsymbol{v})_S = (x', y')$.

In here, (x', y') is the coordinates of $\boldsymbol{v}$ using the new $x'y'$-coordinate system.



Since the transition matrix from E to S is $\boldsymbol{P}^{\mathrm{T}}$,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = [\boldsymbol{v}]_S = \boldsymbol{P}^{\mathrm{T}}[\boldsymbol{v}]_E = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

So $\quad x' = x\cos(\theta) + y\sin(\theta),$

$\qquad y' = -x\sin(\theta) + y\cos(\theta).$

An example (Example 5.4.8.2)

Let $S = \{ \boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3 \}$, where

$$\boldsymbol{u}_1 = \frac{1}{\sqrt{3}}(1, 1, 1), \quad \boldsymbol{u}_2 = \frac{1}{\sqrt{2}}(1, 0, -1), \quad \boldsymbol{u}_3 = \frac{1}{\sqrt{6}}(1, -2, 1),$$

and $T = \{ \boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3 \}$, where

$$\boldsymbol{v}_1 = (0, 0, 1), \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}(1, -1, 0), \quad \boldsymbol{v}_3 = \frac{1}{\sqrt{2}}(1, 1, 0).$$

Both S and T are orthonormal based for $\mathbb{R}^3$.

$$\boldsymbol{u}_1 = (\boldsymbol{u}_1 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + (\boldsymbol{u}_1 \cdot \boldsymbol{v}_3)\boldsymbol{v}_3 = \frac{1}{\sqrt{3}}\boldsymbol{v}_1 + \frac{2}{\sqrt{6}}\boldsymbol{v}_3,$$

$$\boldsymbol{u}_2 = (\boldsymbol{u}_2 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + (\boldsymbol{u}_2 \cdot \boldsymbol{v}_3)\boldsymbol{v}_3 = \frac{-1}{\sqrt{2}}\boldsymbol{v}_1 + \frac{1}{2}\boldsymbol{v}_2 + \frac{1}{2}\boldsymbol{v}_3,$$

$$\boldsymbol{u}_3 = (\boldsymbol{u}_3 \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}_3 \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + (\boldsymbol{u}_3 \cdot \boldsymbol{v}_3)\boldsymbol{v}_3 = \frac{1}{\sqrt{6}}\boldsymbol{v}_1 + \frac{3}{\sqrt{12}}\boldsymbol{v}_2 + \frac{-1}{\sqrt{12}}\boldsymbol{v}_3.$$

An example (Example 5.4.8.2)

The transition matrix form S to T is

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{2} & \frac{3}{\sqrt{12}} \\ \frac{2}{\sqrt{6}} & \frac{1}{2} & \frac{-1}{\sqrt{12}} \end{bmatrix}.$$

The transition matrix form T to S is

$$\boldsymbol{P}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ \frac{-1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{6}} & \frac{3}{\sqrt{12}} & \frac{-1}{\sqrt{12}} \end{bmatrix}.$$