Introduction to Linear Algebra

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 305

1. Orthogonality
   - Inner Product, Length, Orthogonality
   - Orthogonal Sets
   - Orthogonal Projections

Subsection 1

Inner Product, Length, Orthogonality

The Inner Product

- If $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors in $\mathbb{R}^n$, then we regard $\boldsymbol{u}$ and $\boldsymbol{v}$ as $n \times 1$ matrices.
- The transpose $\boldsymbol{u}^T$ is a $1 \times n$ matrix, and the matrix product $\boldsymbol{u}^T\boldsymbol{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets.
- The number $\boldsymbol{u}^T\boldsymbol{v}$ is called the **inner product** of $\boldsymbol{u}$ and $\boldsymbol{v}$, and often it is written as $\boldsymbol{u} \cdot \boldsymbol{v}$.
- This inner product is also referred to as a **dot product**.
- The inner product of $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{u}^T\boldsymbol{v} = [u_1 \ u_2 \ \cdots \ u_n] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Example

- Compute $\boldsymbol{u} \cdot \boldsymbol{v}$ and $\boldsymbol{v} \cdot \boldsymbol{u}$ for $\boldsymbol{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$.
- We have

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= \boldsymbol{u}^T \boldsymbol{v} = [2 \ \ -5 \ \ -1] \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} \\ &= 2 \cdot 3 + (-5) \cdot 2 + (-1) \cdot (-3) = -1; \\ \boldsymbol{v} \cdot \boldsymbol{u} &= \boldsymbol{v}^T \boldsymbol{u} = [3 \ \ 2 \ \ -3] \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} \\ &= 3 \cdot 2 + 2 \cdot (-5) + (-3) \cdot (-1) = -1. \end{aligned}$$

Properties of Inner Product

Theorem

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

(a) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$;

(b) $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$;

(c) $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v}) = \boldsymbol{u} \cdot (c\boldsymbol{v})$;

(d) $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$, and $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if and only if $\boldsymbol{u} = \boldsymbol{0}$.

- Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1\boldsymbol{u}_1 + \cdots + c_p\boldsymbol{u}_p) \cdot \boldsymbol{w} = c_1(\boldsymbol{u}_1 \cdot \boldsymbol{w}) + \cdots + c_p(\boldsymbol{u}_p \cdot \boldsymbol{w}).$$

Length of a Vector

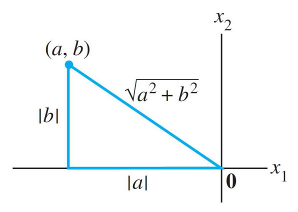
- If $\boldsymbol{v}$ is in $\mathbb{R}^n$, with entries $v_1, \ldots, v_n$, then the square root of $\boldsymbol{v} \cdot \boldsymbol{v}$ is defined because $\boldsymbol{v} \cdot \boldsymbol{v}$ is nonnegative.

Definition

The **length** (or **norm**) of $\boldsymbol{v}$ is the nonnegative scalar $\|\boldsymbol{v}\|$ defined by

$$\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\boldsymbol{v}\|^2 = \boldsymbol{v} \cdot \boldsymbol{v}.$$

- Suppose $\boldsymbol{v}$ is in $\mathbb{R}^2$, say, $\boldsymbol{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. If we identify v with a geometric point in the plane, as usual, then $\|\boldsymbol{v}\|$ coincides with the standard notion of the length of the line segment from the origin to $\boldsymbol{v}$.



Unit Vectors and Normalization

- For any scalar c, the length of $c\boldsymbol{v}$ is $|c|$ times the length of $\boldsymbol{v}$.
- That is,
$$\|c\boldsymbol{v}\| = |c|\|\boldsymbol{v}\|.$$
- A vector whose length is 1 is called a **unit vector**.
- If we divide a nonzero vector $\boldsymbol{v}$ by its length - that is, multiply by $\frac{1}{\|\boldsymbol{v}\|}$ - we obtain a unit vector $\boldsymbol{u}$ because the length of $\boldsymbol{u}$ is $\frac{1}{\|\boldsymbol{v}\|}\|\boldsymbol{v}\|$.
- The process of creating $\boldsymbol{u} = \frac{1}{\|\boldsymbol{v}\|}\boldsymbol{v}$ from $\boldsymbol{v}$ is sometimes called **normalizing** $\boldsymbol{v}$.
- We say that $\boldsymbol{u}$ is a **unit in the same direction as** $\boldsymbol{v}$.

Example

- Let $\boldsymbol{v} = (1, -2, 2, 0)$. Find a unit vector $\boldsymbol{u}$ in the same direction as $\boldsymbol{v}$.
- First, compute the length of $\boldsymbol{v}$:

$$\begin{aligned} \|\boldsymbol{v}\|^2 &= \boldsymbol{v} \cdot \boldsymbol{v} = 1^2 + (-2)^2 + 2^2 + 0^2 = 9; \\ \|\boldsymbol{v}\| &= \sqrt{9} = 3. \end{aligned}$$

Then, multiply $\boldsymbol{v}$ by $\frac{1}{\|\boldsymbol{v}\|}$ to obtain

$$\boldsymbol{u} = \frac{1}{\|\boldsymbol{v}\|}\boldsymbol{v} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ \frac{2}{3} \\ 0 \end{bmatrix}.$$

Example

- Let W be the subspace of $\mathbb{R}^2$ spanned by $\boldsymbol{x} = (\frac{2}{3}, 1)$. Find a unit vector $\boldsymbol{z}$ that is a basis for W.
- W consists of all multiples of $\boldsymbol{x}$.

  Any nonzero vector in W is a basis for W.

  To simplify the calculation, "scale" $\boldsymbol{x}$ to eliminate fractions, i.e., multiply $\boldsymbol{x}$ by 3 to get $\boldsymbol{y} = \left[\begin{array}{c} 2 \\ 3 \end{array}\right]$.

  Now compute $\|\boldsymbol{y}\|^2 = 2^2 + 3^2 = 13$. So $\|\boldsymbol{y}\| = \sqrt{13}$.

  We normalize $\boldsymbol{y}$ to get

$$\boldsymbol{z} = \frac{1}{\sqrt{13}} \left[\begin{array}{c} 2 \\ 3 \end{array}\right] = \left[\begin{array}{c} \frac{2}{\sqrt{13}} \\ \frac{3}{\sqrt{13}} \end{array}\right].$$

Distance Between Two Vectors

- Recall that if a and b are real numbers, the *distance* on the number line between a and b is the number $|a-b|$.

Definition

For $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$, the **distance between $\boldsymbol{u}$ and $\boldsymbol{v}$**, written as $\operatorname{dist}(\boldsymbol{u},\boldsymbol{v})$, is the length of the vector $\boldsymbol{u}-\boldsymbol{v}$. That is,

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

Example

- Compute the distance between the vectors $\boldsymbol{u} = (7,1)$ and $\boldsymbol{v} = (3,2)$.
- Calculate

$$\begin{aligned} \boldsymbol{u} - \boldsymbol{v} &= \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}; \\ \|\boldsymbol{u} - \boldsymbol{v}\| &= \sqrt{4^2 + (-1)^2} = \sqrt{17}. \end{aligned}$$

Example

- If $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$, then

$$\begin{aligned}
\operatorname{dist}(\boldsymbol{u}, \boldsymbol{v}) &= \|\boldsymbol{u} - \boldsymbol{v}\| \\
&= \sqrt{(\boldsymbol{u} - \boldsymbol{v}) \cdot (\boldsymbol{u} - \boldsymbol{v})} \\
&= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2}.
\end{aligned}$$

Orthogonality and Pythagorean Theorem

Definition

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

- The zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T \cdot \boldsymbol{v} = 0$ for all $\boldsymbol{v}$.

Theorem (The Pythagorean Theorem)

Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if and only if $\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

- We have

$$\begin{array}{rcl} \|\boldsymbol{u} + \boldsymbol{v}\|^2 & = & (\boldsymbol{u} + \boldsymbol{v}) \cdot (\boldsymbol{u} + \boldsymbol{v}) \\ & = & \boldsymbol{u} \cdot \boldsymbol{u} + \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{v} \\ & = & \|\boldsymbol{u}\|^2 + 2\boldsymbol{u} \cdot \boldsymbol{v} + \|\boldsymbol{v}\|^2. \end{array}$$

So we get that $\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$ if and only if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal.

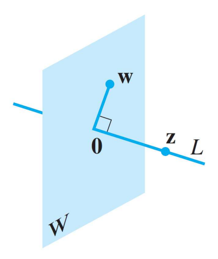
Orthogonal Complements

- If a vector $\boldsymbol{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\boldsymbol{z}$ is said to be **orthogonal** to W.
- The set of all vectors $\boldsymbol{z}$ that are orthogonal to W is called the **orthogonal complement** of W.
- The orthogonal complement of W is denoted by $W^\perp$ (and read as "W perpendicular" or simply "W perp").

Example

- Let W be a plane through the origin in $\mathbb{R}^3$.
- Let L be the line through the origin and perpendicular to W.
- If $\boldsymbol{z}$ and $\boldsymbol{w}$ are nonzero, $\boldsymbol{z}$ is on L, and $\boldsymbol{w}$ is in W, then

  the line segment from $\mathbf{0}$ to $\boldsymbol{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\boldsymbol{w}$, that is, $\boldsymbol{z} \cdot \boldsymbol{w} = 0$.



- So each vector on L is orthogonal to every $\boldsymbol{w}$ in W.
- In fact, L consists of all vectors that are orthogonal to the $\boldsymbol{w}$'s in W, and W consists of all vectors orthogonal to the $\boldsymbol{z}$'s in L.
- That is, $L = W^{\perp}$ and $W = L^{\perp}$.

Properties of Orthogonal Complements

Theorem

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

1. A vector $\boldsymbol{x}$ is in $W^\perp$ if and only if $\boldsymbol{x}$ is orthogonal to every vector in a set that spans W.
2. $W^\perp$ is a subspace of $\mathbb{R}^n$.

1. Let $W = \text{Span}\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$.
   We must show that $W^\perp = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}^\perp$.
   Suppose $\boldsymbol{x}$ is in $W^\perp$. Since $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p \in W$, we have $\boldsymbol{x} \cdot \boldsymbol{w}_1 = 0, \ldots, \boldsymbol{x} \cdot \boldsymbol{w}_p = 0$. This shows that $\boldsymbol{x}$ is in $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}^\perp$.
   Hence $W^\perp \subseteq \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}^\perp$.

Properties of Orthogonal Complements (Part 1 Cont'd)

- Suppose conversely, that $\boldsymbol{x}$ is in $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}^\perp$. Let $\boldsymbol{w}$ be in W. Since $W = \text{Span}\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}$, there exist $c_1, \ldots, c_p$ in $\mathbb{R}$, such that

$$\boldsymbol{w} = c_1\boldsymbol{w}_1 + \cdots + c_p\boldsymbol{w}_p.$$

Then we have

$$\begin{aligned} \boldsymbol{x} \cdot \boldsymbol{w} &= \boldsymbol{x} \cdot (c_1\boldsymbol{w}_1 + \cdots + c_p\boldsymbol{w}_p) \\ &= c_1(\boldsymbol{x} \cdot \boldsymbol{w}_1) + \cdots + c_p(\boldsymbol{x} \cdot \boldsymbol{w}_p) \\ &= c_1 \cdot 0 + \cdots + c_p \cdot 0 = 0. \end{aligned}$$

Therefore $\boldsymbol{x}$ is in $W^\perp$. We conclude that $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}^\perp \subseteq W^\perp$. Combining both inclusions we get

$$W^\perp = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_p\}^\perp.$$

Properties of Orthogonal Complements (Part 2)

2. Recall that to see that $W^\perp$ is a subspace of $\mathbb{R}^n$, we must show that it contains $\mathbf{0}$ and that it is closed under addition and scalar multiplication.
   That $\mathbf{0}$ is in $W^\perp$ is obvious, since, for all $\boldsymbol{w}$ in W, $\mathbf{0} \cdot \boldsymbol{w} = 0$.
   Suppose, next that $\boldsymbol{x}_1, \boldsymbol{x}_2$ are in $W^\perp$. Then we have, for all $\boldsymbol{w}$ in W,
   $$(\boldsymbol{x}_1 + \boldsymbol{x}_2) \cdot \boldsymbol{w} = (\boldsymbol{x}_1 \cdot \boldsymbol{w}) + (\boldsymbol{x}_2 \cdot \boldsymbol{w}) = 0 + 0 = 0.$$
   This proves that $\boldsymbol{x}_1 + \boldsymbol{x}_2$ is in $W^\perp$. So $W^\perp$ is closed under addition.
   Next let $\boldsymbol{x}$ be in $W^\perp$ and c in $\mathbb{R}$.
   Then we get, for all $\boldsymbol{w}$ in W,
   $$(c\boldsymbol{x}) \cdot \boldsymbol{w} = c(\boldsymbol{x} \cdot \boldsymbol{w}) = c \cdot 0 = 0.$$
   This shows that $c\boldsymbol{x}$ is in $W^\perp$. Therefore, $W^\perp$ is also closed under scalar multiplication.
   We now conclude that $W^\perp$ is a subspace of $\mathbb{R}^n$.

Row Space, Null Space and Complements

Theorem

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row}A)^\perp = \text{Nul}A \quad \text{and} \quad (\text{Col}A)^\perp = \text{Nul}A^T.$$

- The row-column rule for computing $A\boldsymbol{x}$ shows that if $\boldsymbol{x}$ is in Nul$A$, then $\boldsymbol{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$). Since the rows of A span the row space, $\boldsymbol{x}$ is orthogonal to Row$A$.
  Conversely, if $\boldsymbol{x}$ is orthogonal to Row$A$, then $\boldsymbol{x}$ is certainly orthogonal to each row of A. Hence $A\boldsymbol{x} = \mathbf{0}$.
  Since this statement is true for any matrix, it is true for A^T. That is, the orthogonal complement of the row space of A^T is the null space of A^T. This proves the second statement, because Row$A^T =$ Col$A$.

Subsection 2

Orthogonal Sets

Orthogonal Set of Vectors

- A set of vectors $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0$ whenever $i \neq j$.

  Example: Show that $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthogonal set, where

$$\boldsymbol{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}.$$

  We check the pairwise inner products:

$$\begin{aligned} \boldsymbol{u}_1 \cdot \boldsymbol{u}_2 &= 3(-1) + 1 \cdot 2 + 1 \cdot 1 = 0; \\ \boldsymbol{u}_1 \cdot \boldsymbol{u}_3 &= 3(-\tfrac{1}{2}) + 1(-2) + 1 \cdot \tfrac{7}{2} = 0; \\ \boldsymbol{u}_2 \cdot \boldsymbol{u}_3 &= -1(-\tfrac{1}{2}) + 2(-2) + 1 \cdot (\tfrac{7}{2}) = 0. \end{aligned}$$

  Each pair of distinct vectors is orthogonal, and so $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthogonal set.

Orthogonality and Linear Independence

Theorem

If $S = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

- If $\mathbf{0} = c_1\boldsymbol{u}_1 + \cdots + c_p\boldsymbol{u}_p$ for some scalars $c_1, \ldots, c_p$, then

$$\begin{array}{rcl} 0 & = & \mathbf{0} \cdot \boldsymbol{u}_1 \\ & = & (c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_p\boldsymbol{u}_p) \cdot \boldsymbol{u}_1 \\ & = & (c_1\boldsymbol{u}_1) \cdot \boldsymbol{u}_1 + (c_2\boldsymbol{u}_2) \cdot \boldsymbol{u}_1 + \cdots + (c_p\boldsymbol{u}_p) \cdot \boldsymbol{u}_1 \\ & = & c_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_1) + c_2(\boldsymbol{u}_2 \cdot \boldsymbol{u}_1) + \cdots + c_p(\boldsymbol{u}_p \cdot \boldsymbol{u}_1) \\ & = & c_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_1). \end{array}$$

  Since $\boldsymbol{u}_1$ is nonzero, $\boldsymbol{u}_1 \cdot \boldsymbol{u}_1$ is not zero. So $c_1 = 0$.

  Similarly, $c_2, \ldots, c_p$ must be zero.

  Thus S is linearly independent.

Orthogonal Basis

Definition

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

- An orthogonal basis is much nicer than other bases because the weights in a linear combination can be computed easily:

Theorem

Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\boldsymbol{y}$ in W, the weights in the linear combination $\boldsymbol{y} = c_1\boldsymbol{u}_1 + \cdots + c_p\boldsymbol{u}_p$ are given by

$$c_j = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_j}{\boldsymbol{u}_j \cdot \boldsymbol{u}_j}, \quad j = 1, \ldots, p.$$

- The orthogonality of $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ shows that $\boldsymbol{y} \cdot \boldsymbol{u}_1 = (c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_p\boldsymbol{u}_p) \cdot \boldsymbol{u}_1 = c_1(\boldsymbol{u}_1 \cdot \boldsymbol{u}_1)$. Since $\boldsymbol{u}_1 \cdot \boldsymbol{u}_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \ldots, p$, compute $\boldsymbol{y} \cdot \boldsymbol{u}_j$ and solve for c_j.

Example

- $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$, with $\boldsymbol{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$, $\boldsymbol{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}$, is an orthogonal basis for $\mathbb{R}^3$.

  Express the vector $\boldsymbol{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.
- We compute:

$$\begin{array}{lll} \boldsymbol{y} \cdot \boldsymbol{u}_1 = 11, & \boldsymbol{y} \cdot \boldsymbol{u}_2 = -12, & \boldsymbol{y} \cdot \boldsymbol{u}_3 = -33, \\ \boldsymbol{u}_1 \cdot \boldsymbol{u}_1 = 11, & \boldsymbol{u}_2 \cdot \boldsymbol{u}_2 = 6, & \boldsymbol{u}_3 \cdot \boldsymbol{u}_3 = \frac{33}{2}. \end{array}$$

  Now we have

$$\begin{array}{lll} \boldsymbol{y} & = & \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} \boldsymbol{u}_1 + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2} \boldsymbol{u}_2 + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_3}{\boldsymbol{u}_3 \cdot \boldsymbol{u}_3} \boldsymbol{u}_3 \\ & = & \frac{11}{11} \boldsymbol{u}_1 + \frac{-12}{6} \boldsymbol{u}_2 + \frac{-33}{\frac{33}{2}} \boldsymbol{u}_3 \\ & = & \boldsymbol{u}_1 - 2\boldsymbol{u}_2 - 2\boldsymbol{u}_3. \end{array}$$

An Orthogonal Projection

- Given a nonzero vector $\boldsymbol{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\boldsymbol{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\boldsymbol{u}$ and the other orthogonal to $\boldsymbol{u}$.
- We wish to write $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z}$, where $\widehat{\boldsymbol{y}} = \alpha \boldsymbol{u}$ for some scalar α and $\boldsymbol{z}$ is some vector orthogonal to $\boldsymbol{u}$.
- Given any scalar α, let $\boldsymbol{z} = \boldsymbol{y} - \alpha \boldsymbol{u}$.
- Then $\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$ is orthogonal to $\boldsymbol{u}$ if and only if

$$0 = (\boldsymbol{y} - \alpha \boldsymbol{u}) \cdot \boldsymbol{u} = \boldsymbol{y} \cdot \boldsymbol{u} - (\alpha \boldsymbol{u}) \cdot \boldsymbol{u} = \boldsymbol{y} \cdot \boldsymbol{u} - \alpha (\boldsymbol{u} \cdot \boldsymbol{u}).$$

- So $\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z}$, with $\boldsymbol{z}$ orthogonal to $\boldsymbol{u}$ if and only if

$$\alpha = \frac{\boldsymbol{y} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}} \quad \text{and} \quad \widehat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}} \boldsymbol{u}.$$

- The vector $\widehat{\boldsymbol{y}}$ is called the **orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$**;
- The vector $\boldsymbol{z}$ is called the **component of $\boldsymbol{y}$ orthogonal to $\boldsymbol{u}$**.

Alternative Notation

- If c is any nonzero scalar and if $\boldsymbol{u}$ is replaced by $c\boldsymbol{u}$ in the definition of $\widehat{\boldsymbol{y}}$, then the orthogonal projection of $\boldsymbol{y}$ onto $c\boldsymbol{u}$ is exactly the same as the orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$.
- Hence this projection is determined by the subspace L spanned by $\boldsymbol{u}$ (the line through $\boldsymbol{u}$ and $\mathbf{0}$).
- Sometimes $\widehat{\boldsymbol{y}}$ is denoted by $\text{proj}_L\boldsymbol{y}$ and is called the **orthogonal projection of $\boldsymbol{y}$ onto L**.
- That is,

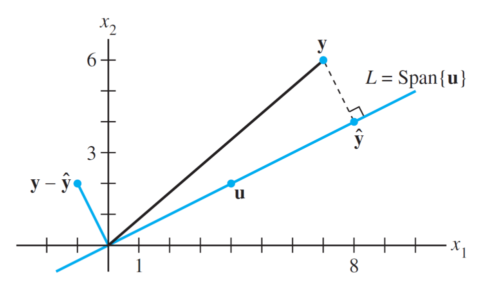
$$\widehat{\boldsymbol{y}} = \text{proj}_L\boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}.$$

Example

- Let $\boldsymbol{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$.
  Find the orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$.
  Then write $\boldsymbol{y}$ as the sum of two orthogonal vectors, one in $\mathrm{Span}\{\boldsymbol{u}\}$ and one orthogonal to $\boldsymbol{u}$.
- We compute: $\boldsymbol{y} \cdot \boldsymbol{u} = 40$, $\boldsymbol{u} \cdot \boldsymbol{u} = 20$.
  The orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}$ is
  $$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}}{\boldsymbol{u} \cdot \boldsymbol{u}} \boldsymbol{u} = \frac{40}{20} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}.$$
  The component of $\boldsymbol{y}$ orthogonal to $\boldsymbol{u}$ is
  $$\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$
  So $\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \widehat{\boldsymbol{y}} + \boldsymbol{z} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

Distance from $\boldsymbol{y}$ to Span$\{\boldsymbol{u}\}$

- The line segment between $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$ is perpendicular to $L = \text{Span}\{\boldsymbol{u}\}$, by construction of $\widehat{\boldsymbol{y}}$:



- So the point identified with $\widehat{\boldsymbol{y}}$ is the closest point of L to $\boldsymbol{y}$.
- It follows that the distance from $\boldsymbol{y}$ to L is the length of the perpendicular line segment from $\boldsymbol{y}$ to the orthogonal projection $\widehat{\boldsymbol{y}}$, i.e., $\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|$.

Orthonormal Sets

- A set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors.
- If W is the subspace spanned by such a set, then $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an orthonormal basis for W, since the set is automatically linearly independent.
- The simplest example of an orthonormal set is the standard basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ for $\mathbb{R}^n$.
- Any nonempty subset of $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is orthonormal, too.

Example

- Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} \frac{3}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \\ \frac{1}{\sqrt{11}} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -\frac{1}{\sqrt{66}} \\ -\frac{4}{\sqrt{66}} \\ \frac{7}{\sqrt{66}} \end{bmatrix}.$$

- We first check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is orthogonal:

$$\begin{aligned} \mathbf{v}_1 \cdot \mathbf{v}_2 &= -\frac{3}{\sqrt{66}} + \frac{2}{\sqrt{66}} + \frac{1}{\sqrt{66}} = 0; \\ \mathbf{v}_1 \cdot \mathbf{v}_3 &= -\frac{3}{\sqrt{726}} - \frac{4}{\sqrt{726}} + \frac{7}{\sqrt{726}} = 0; \\ \mathbf{v}_2 \cdot \mathbf{v}_3 &= \frac{1}{\sqrt{396}} - \frac{8}{\sqrt{396}} + \frac{7}{\sqrt{396}} = 0. \end{aligned}$$

Now we check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ consists of unit vectors:

$$\begin{aligned} \mathbf{v}_1 \cdot \mathbf{v}_1 &= \frac{9}{11} + \frac{1}{11} + \frac{1}{11} = 1; \\ \mathbf{v}_2 \cdot \mathbf{v}_2 &= \frac{1}{6} + \frac{4}{6} + \frac{1}{6} = 1; \\ \mathbf{v}_3 \cdot \mathbf{v}_3 &= \frac{1}{66} + \frac{16}{66} + \frac{49}{66} = 1. \end{aligned}$$

It follows that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ form an orthonormal basis for $\mathbb{R}^3$.

Matrix With Orthonormal Columns

Theorem

An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

- To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof of the general case is similar.

  Let $U = [\boldsymbol{u}_1\ \boldsymbol{u}_2\ \boldsymbol{u}_3]$. Compute

$$U^T U = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \boldsymbol{u}_3^T \end{bmatrix} [\boldsymbol{u}_1\ \boldsymbol{u}_2\ \boldsymbol{u}_3] = \begin{bmatrix} \boldsymbol{u}_1^T \boldsymbol{u}_1 & \boldsymbol{u}_1^T \boldsymbol{u}_2 & \boldsymbol{u}_1^T \boldsymbol{u}_3 \\ \boldsymbol{u}_2^T \boldsymbol{u}_1 & \boldsymbol{u}_2^T \boldsymbol{u}_2 & \boldsymbol{u}_2^T \boldsymbol{u}_3 \\ \boldsymbol{u}_3^T \boldsymbol{u}_1 & \boldsymbol{u}_3^T \boldsymbol{u}_2 & \boldsymbol{u}_3^T \boldsymbol{u}_3 \end{bmatrix}.$$

  The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if $\boldsymbol{u}_1^T \boldsymbol{u}_2 = \boldsymbol{u}_2^T \boldsymbol{u}_1 = 0$, $\boldsymbol{u}_1^T \boldsymbol{u}_3 = \boldsymbol{u}_3^T \boldsymbol{u}_1 = 0$ and $\boldsymbol{u}_2^T \boldsymbol{u}_3 = \boldsymbol{u}_3^T \boldsymbol{u}_2 = 0$. The columns of U all have unit length if and only if $\boldsymbol{u}_1^T \boldsymbol{u}_1 = 1$, $\boldsymbol{u}_2^T \boldsymbol{u}_2 = 1$ and $\boldsymbol{u}_3^T \boldsymbol{u}_3 = 1$. The theorem now follows by looking at the matrix.

Properties of Matrices With Orthonormal Columns

Theorem

Let U be an $m \times n$ matrix with orthonormal columns, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be in $\mathbb{R}^n$. Then:

(a) $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$;

(b) $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$;

(c) $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = 0$ if and only if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$.

(b) We have $(U\boldsymbol{x}) \cdot (U\boldsymbol{y}) = (U\boldsymbol{x})^T(U\boldsymbol{y}) = \boldsymbol{x}^T U^T U\boldsymbol{y} = \boldsymbol{x}^T\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$.

- Note that
  - Property (a) says that the linear mapping $\boldsymbol{x} \mapsto U\boldsymbol{x}$ preserves length;
  - Property (b) says that the linear mapping $\boldsymbol{x} \mapsto U\boldsymbol{x}$ preserves inner products;
  - Property (c) says that the linear mapping $\boldsymbol{x} \mapsto U\boldsymbol{x}$ preserves orthogonality.

Example

- Let $U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$.
  - (a) Check that U has orthonormal columns by computing $U^T U$.
  - (b) Verify that $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$.

(a)

$$U^T U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

(b)

$$\begin{aligned} U\boldsymbol{x} &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}; \\ \|U\boldsymbol{x}\| &= \sqrt{9+1+1} = \sqrt{11} = \|\boldsymbol{x}\|. \end{aligned}$$

Orthogonal Matrices

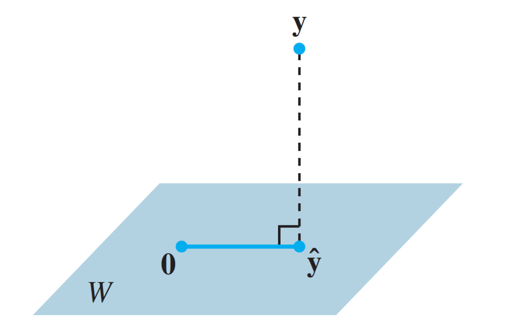
- An **orthogonal matrix** is a square invertible matrix U such that
$$U^{-1} = U^T.$$
- By a previous theorem, such a matrix has orthonormal columns.
- It is easy to see that any square matrix with orthonormal columns is an orthogonal matrix.
- Surprisingly, such a matrix must have orthonormal rows, too.

Subsection 3

Orthogonal Projections

Orthogonal Projection

- Consider a vector $\boldsymbol{y}$ and a subspace W in $\mathbb{R}^n$.
- There is a vector $\widehat{\boldsymbol{y}}$ in W such that:
  (1) $\widehat{\boldsymbol{y}}$ is the unique vector in W for which $\boldsymbol{y} - \widehat{\boldsymbol{y}}$ is orthogonal to W;
  (2) $\widehat{\boldsymbol{y}}$ is the unique vector in W closest to $\boldsymbol{y}$.



Decompositions

- Whenever a vector $\boldsymbol{y}$ is written as a linear combination of vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n$ in $\mathbb{R}^n$,

$$\boldsymbol{y} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_n\boldsymbol{u}_n,$$

  the terms in the sum for $\boldsymbol{y}$ can be grouped into two parts.
- So $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{z}_1 + \boldsymbol{z}_2,$$

  where $\boldsymbol{z}_1$ is a linear combination of some of the $\boldsymbol{u}_i$ and $\boldsymbol{z}_2$ is a linear combination of the rest of the $\boldsymbol{u}_i$.
- This idea is particularly useful when $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ is an orthogonal basis.
- Recall that $W^\perp$ denotes the set of all vectors orthogonal to a subspace W.

Example

- Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let $\boldsymbol{y} = c_1\boldsymbol{u}_1 + \cdots + c_5\boldsymbol{u}_5$.
  Consider the subspace $W = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$, and write $\boldsymbol{y}$ as the sum of a vector $\boldsymbol{z}_1$ in W and a vector $\boldsymbol{z}_2$ in $W^\perp$.
- Write
  $$\boldsymbol{y} = \underbrace{c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2}_{\boldsymbol{z}_1} + \underbrace{c_3\boldsymbol{u}_3 + c_4\boldsymbol{u}_4 + c_5\boldsymbol{u}_5}_{\boldsymbol{z}_2},$$
  where $\boldsymbol{z}_1 = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2$ is in $\text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ and $\boldsymbol{z}_2 = c_3\boldsymbol{u}_3 + c_4\boldsymbol{u}_4 + c_5\boldsymbol{u}_5$ is in $\text{Span}\{\boldsymbol{u}_3, \boldsymbol{u}_4, \boldsymbol{u}_5\}$.
  To show that $\boldsymbol{z}_2$ is in $W^\perp$, it suffices to show that $\boldsymbol{z}_2$ is orthogonal to the vectors in the basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ for W.
  Using properties of the inner product, we compute:
  $$\begin{aligned} \boldsymbol{z}_2 \cdot \boldsymbol{u}_1 &= (c_3\boldsymbol{u}_3 + c_4\boldsymbol{u}_4 + c_5\boldsymbol{u}_5) \cdot \boldsymbol{u}_1 \\ &= c_3\boldsymbol{u}_3 \cdot \boldsymbol{u}_1 + c_4\boldsymbol{u}_4 \cdot \boldsymbol{u}_1 + c_5\boldsymbol{u}_5 \cdot \boldsymbol{u}_1 = 0. \end{aligned}$$
  A similar calculation shows that $\boldsymbol{z}_2 \cdot \boldsymbol{u}_2 = 0$.

The Orthogonal Decomposition Theorem

Theorem (The Orthogonal Decomposition Theorem)

Let W be a subspace of $\mathbb{R}^n$. Then each $\boldsymbol{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z},$$

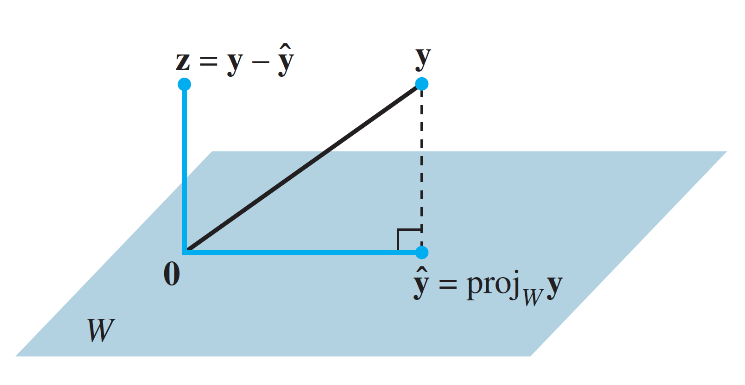
where $\widehat{\boldsymbol{y}}$ is in W and $\boldsymbol{z}$ is in $W^{\perp}$. In fact, if $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is any orthogonal basis of W, then

$$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p}\boldsymbol{u}_p$$

and $\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$.

- The vector $\widehat{\boldsymbol{y}}$ is called the **orthogonal projection of $\boldsymbol{y}$ onto** W and often is written as $\mathrm{proj}_W \boldsymbol{y}$.

The Orthogonal Decomposition Theorem (Illustration)



Proof of the Theorem (Existence)

- Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ be any orthogonal basis for W.

  Define $\widehat{\boldsymbol{y}}$ by
$$\widehat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p}\boldsymbol{u}_p.$$
  Then $\widehat{\boldsymbol{y}}$ is in W because $\widehat{\boldsymbol{y}}$ is a linear combination of the basis $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$. Let $\boldsymbol{z} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$. Since $\boldsymbol{u}_1$ is orthogonal to $\boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$, it follows that
$$\begin{array}{rcl} \boldsymbol{z} \cdot \boldsymbol{u}_1 & = & (\boldsymbol{y} - \widehat{\boldsymbol{y}}) \cdot \boldsymbol{u}_1 = \boldsymbol{y} \cdot \boldsymbol{u}_1 - (\frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1})\boldsymbol{u}_1 \cdot \boldsymbol{u}_1 - 0 \cdots - 0 \\ & = & \boldsymbol{y} \cdot \boldsymbol{u}_1 - \boldsymbol{y} \cdot \boldsymbol{u}_1 = 0. \end{array}$$
  Thus $\boldsymbol{z}$ is orthogonal to $\boldsymbol{u}_1$.

  Similarly, $\boldsymbol{z}$ is orthogonal to each $\boldsymbol{u}_j$ in the basis for W.

  Hence $\boldsymbol{z}$ is orthogonal to every vector in W. That is, $\boldsymbol{z}$ is in $W^{\perp}$.

Proof of the Theorem (Uniqueness)

- To show that the decomposition is unique, suppose $\boldsymbol{y}$ can also be written as
$$\boldsymbol{y} = \widehat{\boldsymbol{y}}_1 + \boldsymbol{z}_1,$$
with $\widehat{\boldsymbol{y}}_1$ in W and $\boldsymbol{z}_1$ in $W^\perp$. Then $\widehat{\boldsymbol{y}} + \boldsymbol{z} = \widehat{\boldsymbol{y}}_1 + \boldsymbol{z}_1$ (since both sides equal $\boldsymbol{y}$). So
$$\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}}_1 = \boldsymbol{z}_1 - \boldsymbol{z}.$$
This equality shows that the vector $\boldsymbol{v} = \widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}}_1$ is in W and in $W^\perp$ (because $\boldsymbol{z}_1$ and $\boldsymbol{z}$ are both in $W^\perp$, and $W^\perp$ is a subspace). Hence $\boldsymbol{v} \cdot \boldsymbol{v} = 0$, which shows that $\boldsymbol{v} = \boldsymbol{0}$. This proves that $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{y}}_1$ and also $\boldsymbol{z}_1 = \boldsymbol{z}$.

Example

- Let $\boldsymbol{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$. Write $\boldsymbol{y}$ as the sum of a vector in W and a vector orthogonal to W.
- The orthogonal projection of $\boldsymbol{y}$ onto W is

$$\begin{aligned}
\widehat{\boldsymbol{y}} &= \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_2}{\boldsymbol{u}_2\cdot\boldsymbol{u}_2}\boldsymbol{u}_2 \\
&= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}; \\
\boldsymbol{y} - \widehat{\boldsymbol{y}} &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{bmatrix}.
\end{aligned}$$

The desired decomposition of $\boldsymbol{y}$ is $\boldsymbol{y} = \widehat{\boldsymbol{y}} + (\boldsymbol{y} - \widehat{\boldsymbol{y}})$.

The Best Approximation Theorem

Theorem (The Best Approximation Theorem)

Let W be a subspace of $\mathbb{R}^n$, let $\boldsymbol{y}$ be any vector in $\mathbb{R}^n$, and let $\widehat{\boldsymbol{y}}$ be the orthogonal projection of $\boldsymbol{y}$ onto W. Then $\widehat{\boldsymbol{y}}$ is the closest point in W to $\boldsymbol{y}$, in the sense that

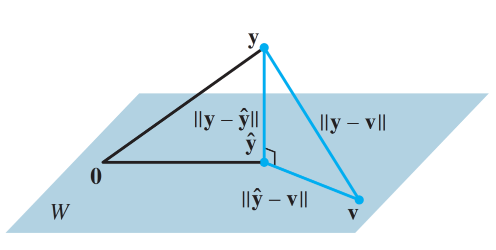
$$\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| < \|\boldsymbol{y} - \boldsymbol{v}\|$$

for all $\boldsymbol{v}$ in W distinct from $\widehat{\boldsymbol{y}}$.

- The vector $\widehat{\boldsymbol{y}}$ is called **the best approximation to $\boldsymbol{y}$ by elements of** W.
- The distance from $\boldsymbol{y}$ to $\boldsymbol{v}$, given by $\|\boldsymbol{y} - \boldsymbol{v}\|$, can be regarded as the "error" of using $\boldsymbol{v}$ in place of $\boldsymbol{y}$.
- Then the theorem says that this error is minimized when $\boldsymbol{v} = \widehat{\boldsymbol{y}}$.

Proof of Best Approximation Theorem

- Take $\boldsymbol{v}$ in W distinct from $\widehat{y}$. Then $\widehat{\boldsymbol{y}} - \boldsymbol{v}$ is in W.



By the Orthogonal Decomposition Theorem, $\boldsymbol{y} - \widehat{\boldsymbol{y}}$ is orthogonal to W. In particular, $\boldsymbol{y} - \widehat{\boldsymbol{y}}$ is orthogonal to $\widehat{\boldsymbol{y}} - \boldsymbol{v}$ (which is in W).
Since $\boldsymbol{y} - \boldsymbol{v} = (\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\widehat{\boldsymbol{y}} - \boldsymbol{v})$ the Pythagorean Theorem gives

$$\|\boldsymbol{y} - \boldsymbol{v}\|^2 = \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|^2 + \|\widehat{\boldsymbol{y}} - \boldsymbol{v}\|^2.$$

Now $\|\widehat{\boldsymbol{y}} - \boldsymbol{v}\|^2 > 0$ because $\widehat{\boldsymbol{y}} - \boldsymbol{v} \neq \mathbf{0}$. So the inequality follows.

Example

- Let $\boldsymbol{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $W = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$.

  Find the closest point in W to $\boldsymbol{y}$.
- We have, by the theorem,

$$\begin{aligned} \widehat{\boldsymbol{y}} &= \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_2}{\boldsymbol{u}_2\cdot\boldsymbol{u}_2}\boldsymbol{u}_2 \\ &= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{3}{5} \\ \frac{3}{2} \\ -\frac{3}{10} \end{bmatrix} + \begin{bmatrix} -1 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}. \end{aligned}$$

Example

- The distance from a point $\boldsymbol{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\boldsymbol{y}$ to the nearest point in W.
  Find the distance from $\boldsymbol{y}$ to $W = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$, where
  $$\boldsymbol{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \boldsymbol{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$
- By the theorem, the distance from $\boldsymbol{y}$ to W is $\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|$ where $\widehat{\boldsymbol{y}} = \text{proj}_W \boldsymbol{y}$. Since $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is an orthogonal basis for W,
  $$\begin{aligned} \widehat{\boldsymbol{y}} &= \frac{15}{30}\boldsymbol{u}_1 + \frac{-21}{6}\boldsymbol{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}, \\ \boldsymbol{y} - \widehat{\boldsymbol{y}} &= \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}, \\ \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| &= \sqrt{3^2 + 6^2} = \sqrt{45}. \end{aligned}$$

The Case of Orthonormal Bases

- We finally see how the formula for $\text{proj}_W \boldsymbol{y}$ is simplified when the basis for W is an orthonormal set:

Theorem

If $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W \boldsymbol{y} = (\boldsymbol{y} \cdot \boldsymbol{u}_1)\boldsymbol{u}_1 + (\boldsymbol{y} \cdot \boldsymbol{u}_2)\boldsymbol{u}_2 + \cdots + (\boldsymbol{y} \cdot \boldsymbol{u}_p)\boldsymbol{u}_p.$$

If $U = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \ \cdots \ \boldsymbol{u}_p]$, then

$$\text{proj}_W \boldsymbol{y} = UU^T \boldsymbol{y}, \quad \text{for all } y \text{ in } \mathbb{R}^n.$$

- The first formula follows immediately from the Orthogonal Decomposition Theorem. Also, it shows that $\text{proj}_W \boldsymbol{y}$ is a linear combination of the columns of U using the weights $\boldsymbol{y} \cdot \boldsymbol{u}_1$, $\boldsymbol{y} \cdot \boldsymbol{u}_2$, ..., $\boldsymbol{y} \cdot \boldsymbol{u}_p$. The weights can be written as $\boldsymbol{u}_1^T \boldsymbol{y}$, ...; $\boldsymbol{u}_2^T \boldsymbol{y}$, ..., $\boldsymbol{u}_p^T \boldsymbol{y}$, showing that they are the entries in $U^T \boldsymbol{y}$.