Chapter 6. Orthogonality and Least Squares

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

서울과학기술대학교
SEOULTECH SEOUL NATIONAL UNIVERSITY OF SCIENCE & TECHNOLOGY

1. *6.1. Inner Product, Length, And Orthogonality*
2. *6.2. Orthogonal Sets*
3. *6.3. Orthogonal Projections*
4. *6.4 Gram-Schmidt process*
5. *6.5 Least-Squares Problems*

6.1. Inner Product, Length, And Orthogonality

Inner Product

- If $\mathbf{u}$ and $\mathbf{v}$ are vector in $\mathbb{R}^n$, then we regard $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices. (*i.e. column vector*)
- The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets.
- The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** of $\mathbf{u}$ and $\mathbf{v}$, and it is written as $\mathbf{u} \cdot \mathbf{v}$
- This inner product is also referred to as a **dot product**.
- If $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ then, the inner product of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 + v_2 + \cdots + u_n v_n$$

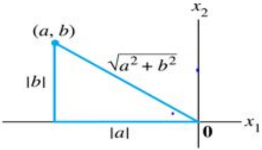
- **Theorem 1**: Let $\mathbf{u}$, $\mathbf{v}$, and W be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
  a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
  b) $(\mathbf{u} + \mathbf{v}) \cdot W = \mathbf{u} \cdot W + \mathbf{v} \cdot W$
  c) $(c \cdot \mathbf{u}) \cdot \mathbf{v} = c(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (c\mathbf{v})$
  d) $\mathbf{u} \cdot \mathbf{u} \geq 0$, and $\mathbf{u} \cdot \mathbf{u} = 0$ if and only if $\mathbf{u} = 0$

- **Remarks**
  - Properties (b) and (c) can be combined several times to produce the following useful property. $(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p) \cdot W = c_1(\mathbf{u}_1 \cdot W) \cdots + c_p(\mathbf{u}_p \cdot W)$
  - If $\mathbf{v}$ is in $\mathbb{R}^n$, with entries $v_1, \cdots v_n$, then the square root of $\mathbf{v} \cdot \mathbf{v}$ is defined, because $\mathbf{v} \cdot \mathbf{v}$ is nonnegative.

The length of Vector

- **Definition**: The **length** or **norm** of $\mathbf{v}$ is the nonnegative scalar
  - $\|\mathbf{v}\| := \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ and $\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$

- **Remarks**
  - Suppose $\mathbf{v}$ is in $\mathbb{R}^2$, say $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. If we identify $\mathbf{v}$ with a geometric point in the plane, as usual, $\|\mathbf{v}\|$ coincides with the standard notion of the length of the line segment from the origin to $\mathbf{v}$.
  - This follows from the Pythagorean Theorem applied to a triangle such as the one shown in the following figure.



Interpretation of $\|\mathbf{v}\|$ as length.

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

- **Example 2:** Let $\mathbf{v} = (1, -2, 2, 0)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{u}$.

- **Solution:**
  - First, compute the length of $\mathbf{v}$: $\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$. $\|\mathbf{v}\| = \sqrt{9} = 3$.
  - Then, multiply $\mathbf{v}$ by $1/\|\mathbf{v}\|$ to obtain

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

  - To check that $\|\mathbf{u}\| = 1$, it suffices to show that $\|\mathbf{u}\|^2 = 1$

$$\begin{aligned} \|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u} &= (\frac{1}{3})^2 + (-\frac{2}{3})^2 + (\frac{2}{3})^2 + (0)^2 \\ &= \frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0 = 1 \end{aligned}$$

Distance in $\mathbb{R}^n$

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

$$dist(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

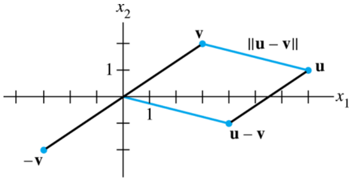
- **Example 4:** Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$

- **Solution:**
  - Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

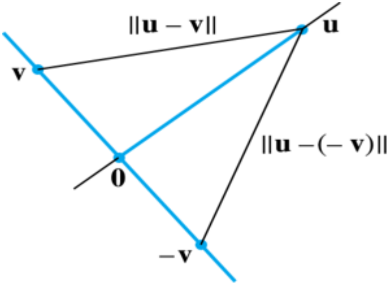
  - The vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$ are shown in the figure on the next slide.
  - When the vector $\mathbf{u} - \mathbf{v}$ is added to $\mathbf{v}$, the result is $\mathbf{u}$.
  - Notice that the parallelogram in the figure below shows that the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u}$ to $\mathbf{u}$ is the same as the distance $\mathbf{u} - \mathbf{v}$ to 0.



The distance between $\mathbf{u}$ and $\mathbf{v}$ is the length of $\mathbf{u} - \mathbf{v}$.

Orthogonal Vector

- Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors $\mathbf{u}$ and $\mathbf{v}$.
- See the figure below. The two lines shown in the figure are geometrically perpendicular if and only if the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u}$ to $-\mathbf{v}$



- This is the same as requiring the squares of the distances to be the same.

- Now,

$$\begin{aligned}[dist(\mathbf{u}, -\mathbf{v})]^2 &= \|\mathbf{u} - (-\mathbf{v})\|^2 = \|\mathbf{u} + \mathbf{v}\| \\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) + \mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v}\end{aligned}$$

- The same calculations with $\mathbf{v}$ and $-\mathbf{v}$ interchanged show that

$$\begin{aligned}[dist(\mathbf{u}, \mathbf{v})]^2 &= \|\mathbf{u}\|^2 + \| - \mathbf{v}\|^2 + 2\mathbf{u} \cdot (-\mathbf{v}) \\ &= \|\mathbf{u}\|^2 + \|\mathbf{u}\|^2 - 2\mathbf{u} \cdot \mathbf{v}\end{aligned}$$

- The two squared distances are equal if and only if $2\mathbf{u} \cdot \mathbf{v} = -2\mathbf{u} \cdot \mathbf{v}$, which happens if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.
- This calculation shows that when vectors $\mathbf{u}$ and $\mathbf{v}$ are identified with geometric points, the corresponding lines through the points and the origin are perpendicular if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

- **Definition:** Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.
- **Remark:** The zero vector is orthogonal to every vector $\mathbb{R}^n$ because $0^T\mathbf{v} = 0$ for all $\mathbf{v}$
- **Theorem 2 (The pythogorean Theorem):** Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if
$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Orthogonal Complements

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

- **Theorem 3:** 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:

$$(RowA)^{\perp} = NulA \text{ and } (ColA)^{\perp} = NulA^T$$

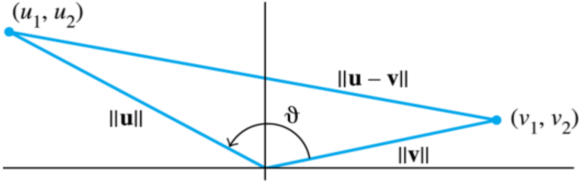
- **Proof :**
  - The row-column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in $Nul\ A$, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$).
  - Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to $Row\ A$.
  - Conversely, if $\mathbf{x}$ is orthogonal to $Row\ A$, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = 0$.
  - This proves the first statement of the theorem.
  - 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 $RowA^T = ColA$.

Angles in $\mathbb{R}^2$ *and* $\mathbb{R}^3$

- If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in either $\mathbb{R}^2$ or $\mathbb{R}^3$, then there is a nice connection between their inner product and the angle θ between the two line segments from the origin to the points identified with $\mathbf{u}$ and $\mathbf{v}$.
- The formula is

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| cos\theta \tag{2}$$

- To verify this formula for vectors in $\mathbb{R}^2$ consider the triangle shown in the next figure with sides of lengths, $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{u} - \mathbf{v}\|$.



The angle between two vectors.

- By the law of cosines,

$$\|\mathbf{u}-\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\|\mathbf{v}\|cos\theta$$

which can be rearranged to produce the next equations.

$$\begin{aligned}\|\mathbf{u}\|\|\mathbf{v}\|cos\,\theta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}||^2\right] \\ &= \frac{1}{2}\left[u_1^2 + u_2^2 + v_1^2 + v_2^2 - (u_1 - v_1)^2 - (u_2 = v_2)^2\right] \\ &= u_1 v_1 + u_2 v_2 \\ &= \mathbf{u}\cdot\mathbf{v}\end{aligned}$$

- The verification for $\mathbb{R}^3$ is similar. When $n > 3$, formula (2) may be used to *define* the angle between two vectors in $\mathbb{R}^n$
- In statistics, the value of $cos\theta$ defined by (2) for suitable vectors **u** and **v** is called a *correlation coefficient*.

Suggested Excercises

- 6.1.11

6.2. Orthogonal Sets

Orthogonal Sets

- **Definition:** A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ 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 $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

- **Theorem 4:** If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ 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.

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

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

    because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$
  - Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1 \cdot \mathbf{u}_1$ not zero and so $c_1 = 0$. Likewise, $c_2, \ldots, c_p$ must be zero.
  - Thus S is linearly independent.

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

- **Theorem 5:** Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ are given by

$$c_j = \frac{\mathbf{y} \cdot \mathbf{u}_j}{\mathbf{u}_j \cdot \mathbf{u}_j} \ (j = 1, \ldots, p)$$

- **Proof:**
  - The orthogonality of $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ shows that

$$\mathbf{y} \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 = c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

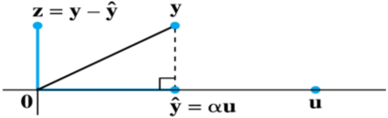
  - Since $\mathbf{u}_1 \cdot \mathbf{u}_1$ is not zero, the equation above can be solved for c_1.
  - To find c_j for $j = 2, \ldots, p$, compute $\mathbf{y} \cdot \mathbf{u}_j$ and solve for c_j

An Orthogonal Projection

- Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$.
- We wish to write

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See the following figure.



Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

- Given any scalar α, let $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, so that (1) is satisfied. Then, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathbf{u}$ if and only if

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

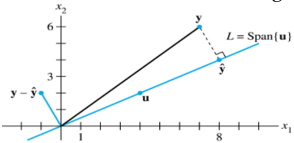
- That is, (1) is satisfied with $\mathbf{z}$ orthogonal to $\mathbf{u}$ if and only if $\alpha = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$ and $\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$
- The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$**, and the vector $\mathbf{z}$ is called the **component of $\mathbf{y}$ orthogonal to $\mathbf{u}$**.

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

$$\hat{\mathbf{y}} = proj_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} \tag{2}$$

- **Example 3:** Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in $Span\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

- **Solution:**
  - Compute $\mathbf{y} \cdot \mathbf{u} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$ and $\mathbf{u} \cdot \mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$
  - The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is $\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} = \frac{40}{20} \mathbf{u} = 2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$ and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is $\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$
  - The sum of these two vector is $\mathbf{y}$, i.e. $\mathbf{y} = \hat{\mathbf{y}} + (\mathbf{y} - \hat{\mathbf{y}})$. That is, $\begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$
  - The decomposition of $\mathbf{y}$ is illustrated in the following figure:



The orthogonal projection of $\mathbf{y}$ onto a line L through the origin.

- (*solution continued*)
  - If the calculation above are correct, then $\{\hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}}\}$ will be an orthogonal set.
  - As a check, compute

$$\hat{\mathbf{y}} \cdot (\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

- **Remark**: Since the line segment in the figure on the previous slide between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is perpendicular to L, by construction of $\hat{\mathbf{y}}$, the point identified with $\hat{\mathbf{y}}$ is the closest point of L to $\mathbf{y}$.

Orthonormal Sets

- **Definitions**
  - A set $\{\mathbf{u}_1, \ldots \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors.
  - If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \ldots \mathbf{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent, by Theorem 4.

- **Remark**
  - The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$.
  - Any nonempty subset of $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal, too.

- **Example 2:** Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$. where

$$v_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, v_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, v_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

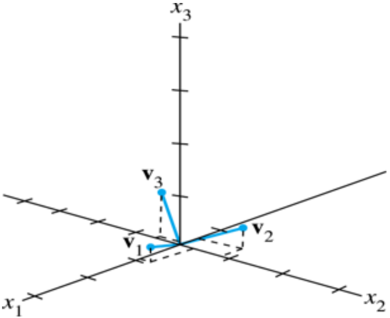
- **Solution:**
  - $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set because

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

  - $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors because

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

- Since the set is linearly independent, its three vectors form a basis for $\mathbb{R}^3$. See the following figure.



- When the vectors in an orthogonal set of nonzero vectors are *normalized* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set.

- **Theorem 6:** An $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.
- **Proof:**
  - To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$.
  - Let $U = \begin{bmatrix} \mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3 \end{bmatrix}$ and compute

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \\ \mathbf{u}_2^T \\ \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} \tag{4}$$

  - The entries in the matrix at the right are inner products, using transpose notation.
  - The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0,\ \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0,\ \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \tag{5}$$

  - The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1,\ \mathbf{u}_2^T\mathbf{u}_2 = 1,\ \mathbf{u}_3^T\mathbf{u}_3 = 1 \tag{6}$$

  - The theorem follows immediately from (4)–(6).

- **Theorem 7:** Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in $\mathbb{R}^3$ Then,
  a. $\|Ux\| = \|x\|$
  b. $(Ux) \cdot (Uy) = x \cdot y$
  c. $(Ux) \cdot (Uy) = 0\$$ if and only if $\$x \cdot y = 0$

- **Remark**: Properties a and c say that the linear mapping $x \to Ux$ preserves lengths and orthogonality.

Suggested Excercises

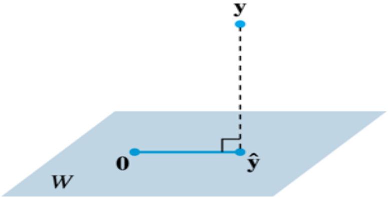
- 6.2.2

6.3. Orthogonal Projections

Orthogonal Projections

- The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$

- Given a vector $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector in $\hat{\mathbf{y}}$ in W such that
  (1) $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y}$ - $\hat{\mathbf{y}}$ is orthogonal to W, and
  (2) $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$.
- See the following figure.
- These two properties of $\hat{\mathbf{y}}$ provide the key to finding the least-squares solutions of linear systems.



The Orthogonal Decomposition Theorem

- **Theorem 8:** Let W be a subspace of $\mathbb{R}^n$.
  - Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

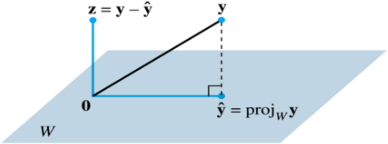
$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}, \tag{1}$$

    where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$.
  - In fact, if $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ is any orthogonal basis of W then,

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}} \mathbf{u_1} + \cdots + \frac{\mathbf{y} \cdot \mathbf{u_p}}{\mathbf{u_p} \cdot \mathbf{u_p}} \mathbf{u_p} \tag{2}$$

    and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$

- The vector $\hat{\mathbf{y}}$ in (1) is called the **orthogonal projection of y onto** W and often is written as $proj_W\,\mathbf{y}$. See the following figure:



The orthogonal projection of **y** onto W.

- **Proof for $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^{\perp}$:**
  - Let $\{\mathbf{u_1}, \dots, \mathbf{u_p}\}$ be any orthogonal basis for W, and define $\hat{\mathbf{y}}$ by (2). Then $\hat{\mathbf{y}}$ is in W because $\hat{\mathbf{y}}$ is a linear combination of the basis $\mathbf{u_1}, \dots, \mathbf{u_p}$
  - Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Since $\mathbf{u_1}$ is orthogonal to $\mathbf{u_1}, \dots, \mathbf{u_p}$, it follows from (2) that

$$\mathbf{z} \cdot \mathbf{u_1} = (\mathbf{y} - \hat{\mathbf{y}}) \cdot \mathbf{u_1} = \mathbf{y} \cdot \mathbf{u_1} - (\frac{\mathbf{y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}})\mathbf{u_1} \cdot u_1 = \mathbf{y} \cdot \mathbf{u_1} - \mathbf{y} \cdot \mathbf{u_1} = 0$$

  - Thus, $\mathbf{z}$ is orthogonal to $\mathbf{u_1}$. Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u_j}$ in the basis for W.
  - Hence $\mathbf{z}$ is orthogonal to every vector in W. That is, $\mathbf{z}$ is in $W^{\perp}$

- **Proof for unique representation:**
  - To show that the decomposition in (1) is unique, suppose $\mathbf{y}$ can also be written as $\mathbf{y}$ be also written as $\mathbf{y} = \mathbf{y_1} + \mathbf{z_1}$, with $\mathbf{y_1}$ in W and $\mathbf{z_1}$ in $W^{\perp}$
  - Then, $\mathbf{y} + \mathbf{z} = \mathbf{y_1} + \mathbf{z_1}$ (since both sides equal $\mathbf{y}$), and so $\mathbf{y} - \mathbf{y_1} = \mathbf{z_1} - \mathbf{z}$
  - This equality shows that the vector $\mathbf{v} = \mathbf{y} - \mathbf{y_1}$ is in W and in $W^{\perp}$ (because $\mathbf{z_1}$ and $\mathbf{z}$ are both in $W^{\perp}$, and $W^{\perp}$ is a subspace).
  - Hence $\mathbf{v} \cdot \mathbf{v} = 0$ which shows that $\mathbf{v} = 0$.
  - This proves that $\mathbf{y} = \mathbf{y_1}$ and also $\mathbf{z_1} = \mathbf{z}$
  - The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y_1}}$ depends only on W and not on the particular basis used in (2).

- **Example 1:**
  - Let $\mathbf{u_1} = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u_2} = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$
  - Observe that $\{\mathbf{u_1}, \mathbf{u_2}\}$ is an orthogonal basis for $W = Span\{\mathbf{u_1}, \mathbf{u_2}\}$.
  - Write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

- **Solution:**
  - The orthogonal projection of $\mathbf{y}$ onto W is

$$\begin{aligned} \hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}} \mathbf{u_1} + \frac{\mathbf{y} \cdot \mathbf{u_2}}{\mathbf{u_2} \cdot \mathbf{u_2}} \mathbf{u_2} \\ &= \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} \end{aligned}$$

  - Also, $\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$
  - Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^{\perp}$

- (*solution continued:*)
  - To check the calculations, verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u_1}$ and $\mathbf{u_2}$ and hence to all of W
  - The desired decomposition of $\mathbf{y}$ is $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$

Properties of Orthogonal Projections

- If $\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$ be any orthogonal basis for W and if $\mathbf{y}$ happens to in W, then the formula for $proj_L\,\mathbf{y}$ is exactly the same as the representation of $\mathbf{y}$ given in Theorem 5 in Section 6.2. In this case, $proj_W\,\mathbf{y}$ =$\mathbf{y}$
- If $\mathbf{y}$ in $W = Span\{\mathbf{u_1}, \ldots, \mathbf{u_p}\}$, then $proj_W\,\mathbf{y} = \mathbf{y}$

The Best Approximation Theorem

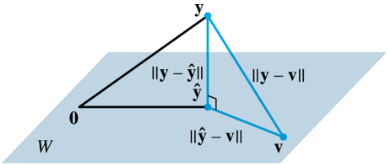
- **Theorem 9:** Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

  for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$

- Remarks
  - The vector $\hat{\mathbf{y}}$ in Theorem 9 is called the best approximation to $\mathbf{y}$ by elements of W.
  - The distance from $\mathbf{y}$ to $\mathbf{v}$, given by $\|\mathbf{y} - \mathbf{v}\|$, can be regarded as the "error' ' of using $\mathbf{v}$ in place of $\mathbf{y}$.
  - Theorem 9 says that this error is minimized when $\mathbf{v} = \hat{\mathbf{y}}$
  - Inequality (3) leads to a new proof that $\hat{\mathbf{y}}$ does not depend on the particular orthogonal basis used to compute it.
  - If a different orthogonal basis for W were used to construct an orthogonal projection of $\mathbf{y}$, then this projection would also be the closest point in W to $\mathbf{y}$, namely, $\hat{\mathbf{y}}$.

- **Proof for the theorem:**
  - Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$. See the following figure:



The orthogonal projection of **y** onto W is the closest point in W to **y**.

-
  - Then, $\hat{\mathbf{y}} - \mathbf{v}$ is in W. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W).
  - Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \mathbf{y}) + (\mathbf{y} - \mathbf{v}),$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \mathbf{y}\|^2 + \|\mathbf{y} - \mathbf{v}\|^2$$

  - (See the colored right triangle in the figure. The length of each side is labeled.)
  - Now $\|\mathbf{y} - \mathbf{v}\|^2 > 0$ because $\mathbf{y} - \mathbf{v} \neq 0$, and so inequality (3) follows immediately.

- **Example 4:** The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $Span\{\mathbf{u_1}, \mathbf{u_2}\}$, where

$$y = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \mathbf{u_1} = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \mathbf{u_2} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

- **Solution :**
  - By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = proj_W\, \mathbf{y}$.
  - Since $\{\mathbf{u_1}, \mathbf{u_2}\}$ is an orthogonal basis for W,

$$\begin{aligned}
\hat{\mathbf{y}} &= \frac{15}{30}\mathbf{u_1} + \frac{-21}{6}\mathbf{u_2} = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} + -\frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} \\
\mathbf{y} - \hat{\mathbf{y}} &= \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix} \\
\|\mathbf{y} - \hat{\mathbf{y}}\|^2 &= 3^2 + 6^2 = 45
\end{aligned}$$

  - The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$

- **Theorem 10:**
  - If $\{\mathbf{u_1} \ldots, \mathbf{u_p}\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$proj_W \, \mathbf{y} = (\mathbf{y} \cdot \mathbf{u_1})\mathbf{u_1} + (\mathbf{y} \cdot \mathbf{u_2})\mathbf{u_2} + \cdots + (\mathbf{y} \cdot \mathbf{u_p})\mathbf{u_p} \tag{4}$$

  - If $U = [\mathbf{u_1} \, \mathbf{u_2} \ldots \mathbf{u_p}]$, then for all $\mathbf{y}$ in $\mathbb{R}^n$,

$$proj_W \, \mathbf{y} = \mathbf{U}\mathbf{U}^T \mathbf{y} \tag{5}$$

- **Proof:**
  - Formula (4) follows immediately from (2) in Theorem 8.
  - Also, (4) shows that $proj_W \; \mathbf{y}$ is a linear combination of the columns of U using the weights $\mathbf{y} \cdot \mathbf{u_1}, \mathbf{y} \cdot \mathbf{u_2}, \ldots, \mathbf{y} \cdot \mathbf{u_p}$.
  - The weights can be written as $\mathbf{u_1}^T \mathbf{y}, \mathbf{u_2}^T \mathbf{y}, \ldots, \mathbf{u_p}^T \mathbf{y}$, showing that they are the entries in $U^T \mathbf{y}$ and justifying (5).

6.4 Gram-Schmidt process

Gram-Schmidt process

- **Theorem 11: The Gram-Schmidt Process**
  - Given a basis $\{\mathbf{x_1}, \dots, \mathbf{x_p}\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$
\begin{aligned}
\mathbf{v_1} &= \mathbf{x_1} \\
\mathbf{v_2} &= \mathbf{x_2} - \frac{\mathbf{x_2} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} \\
\mathbf{v_3} &= \mathbf{x_3} - \frac{\mathbf{x_3} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} - \frac{\mathbf{x_3} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} \\
\mathbf{v_p} &= \mathbf{x_p} - \frac{\mathbf{x_p} \cdot \mathbf{v_1}}{\mathbf{v_1} \cdot \mathbf{v_1}}\mathbf{v_1} - \frac{\mathbf{x_p} \cdot \mathbf{v_2}}{\mathbf{v_2} \cdot \mathbf{v_2}}\mathbf{v_2} - \dots - \frac{\mathbf{x_p} \cdot \mathbf{v_{p-1}}}{\mathbf{v_{p-1}} \cdot \mathbf{v_{p-1}}}\mathbf{v_{p-1}}
\end{aligned}
$$

  - Then $\{\mathbf{v_1}, \dots, \mathbf{v_p}\}$ is an orthogonal basis for W.
  - In addition **Span** $\{\mathbf{v_1}, \dots, \mathbf{v_k}\}$ = **Span** $\{\mathbf{x_1}, \dots, \mathbf{x_k}\}$ for $1 \le k \le p$.

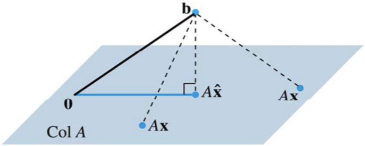
6.5 Least-Squares Problems

Least-Squares Problems

- **Definition:** If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\mathbf{x}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

for all $\mathbf{x}$ in $\mathbb{R}^n$.



The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

- **Example 1**

- **Theorem 14:** Let A be an $m \times n$ matrix. The following statements are logically equivalent:
  - *a.* The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each $\mathbf{b}$ in $\mathbb{R}^m$.
  - *b.* The columns of A are linearly independent.
  - *c.* The matrix $A^T A$ is invertible.
  - *d.* The normal equation has a unique solution.

- **Example 4**

Suggested Exercises

- 6.5.2
- 6.8.1

Acknowledgement

- This lecture note is based on the instructor's lecture notes (formatted as ppt files) provided by the publisher (Pearson Education) and the textbook authors (David Lay and others)
- The pdf conversion project for this chapter was possible thanks to the hard work by Bongseok Kim. Mr. Kim enrolls in the Department of Industrial Engineering since 2016, and expects to continue his studies in the Master's program of Data Science from 2022.