

# Chapter 7. Orthogonality and least squares

C.O.S. Sorzano

Biomedical Engineering

December 3, 2013



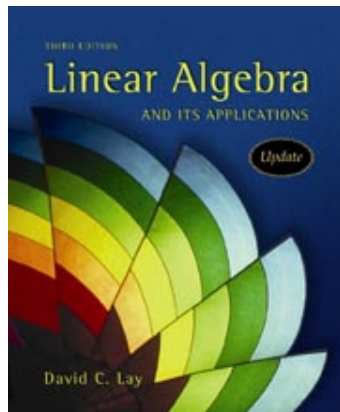
CEU

*Universidad  
San Pablo*

## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

# References



D. Lay. Linear algebra and its applications (3rd ed). Pearson (2006). Chapter 6.

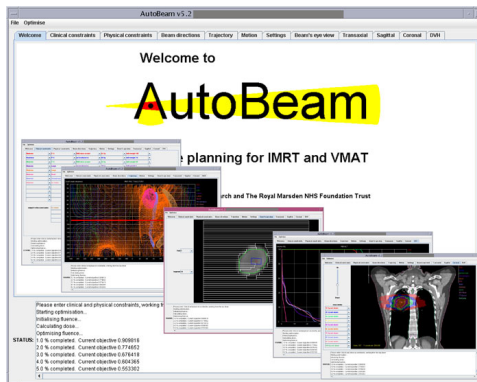
# A little bit of history

Least squares was first used to solve problems in geodesy ([Andrien-Marie Legendre](#), 1805) and astronomy ([Carl Friedrich Gauss](#), 1809). Gauss made the connection of this method to the distribution of measurement errors. Currently it is one of the best understood and most widely spread methods.



# Applications

In this example Least Squares are used to plan a radiation therapy.



Bedford, J. L. *Sinogram analysis of aperture optimization by iterative least-squares in volumetric modulated arc therapy*. Physics in Medicine and Biology, 2013, 58, 1235-1250

# Applications

Traditionally, control applications were formulated in a least-squares setup. Currently, they have found more sophisticated goal functions that can be regarded as evolved versions of least squares.



## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

# Inner product

## Definition 1.1 (Inner product or dot product)

Let  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$  be two vectors. The **inner product** or **dot product** between these two vectors is defined as

$$\mathbf{u} \cdot \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle \triangleq \sum_{i=1}^n u_i v_i$$

## Theorem 1.1

If we considered  $\mathbf{u}$  and  $\mathbf{v}$  to be column vectors ( $\in \mathcal{M}_{n \times 1}$ ), then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$$

## Example

Let  $\mathbf{u} = (2, -5, -1)$  and  $\mathbf{v} = (3, 2, -3)$ .

$$\mathbf{u} \cdot \mathbf{v} = 2 \cdot 3 + (-5) \cdot 2 + 1 \cdot (-3) = -1$$



# Inner product

## Theorem 1.2

*For any three vectors  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$  and any scalar  $r \in \mathbb{R}$  it is verified that*

- ①  $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- ②  $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
- ③  $(r\mathbf{u}) \cdot \mathbf{v} = r(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (r\mathbf{v})$
- ④  $\mathbf{u} \cdot \mathbf{u} \geq 0$
- ⑤  $\mathbf{u} \cdot \mathbf{u} = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$

## Corollary

$$(r_1\mathbf{u}_1 + r_2\mathbf{u}_2 + \dots + r_p\mathbf{u}_p) \cdot \mathbf{v} = r_1(\mathbf{u}_1 \cdot \mathbf{v}) + r_2(\mathbf{u}_2 \cdot \mathbf{v}) + \dots + r_p(\mathbf{u}_p \cdot \mathbf{v})$$

# Length

## Definition 1.2 (Length of a vector)

Given any vector  $\mathbf{v}$ , its length is defined as

$$\|\mathbf{v}\| \triangleq \sqrt{\mathbf{v} \cdot \mathbf{v}}$$

## Theorem 1.3

Given any vector  $\mathbf{v} \in \mathbb{R}^n$

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}$$

## Example

The length of  $\mathbf{v} = (1, -2, 2, 0)$  is

$$\|\mathbf{v}\| = \sqrt{1^2 + (-2)^2 + 2^2 + 0^2} = 3$$

# Length

## Theorem 1.4

For any vector  $\mathbf{v}$  and any scalar  $r$  it is verified that

$$\|r\mathbf{v}\| = |r|\|\mathbf{v}\|$$

Proof

It will be given only for  $\mathbf{v} \in \mathbb{R}^n$ :

$$\begin{aligned}\|r\mathbf{v}\| &= \sqrt{(rv_1)^2 + (rv_2)^2 + \dots + (rv_n)^2} = \sqrt{r^2(v_1^2 + v_2^2 + \dots + v_n^2)} \quad (q.e.d.) \\ &= \sqrt{r^2} \sqrt{v_1^2 + v_2^2 + \dots + v_n^2} = |r|\|\mathbf{v}\|\end{aligned}$$

## Example (continued)

Find a vector of unit length that has the same direction as  $\mathbf{v} = (1, -2, 2, 0)$ .

Solution

$$\mathbf{u}_{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \left(\frac{1}{3}, -\frac{2}{3}, \frac{2}{3}, 0\right) \Rightarrow \|\mathbf{u}_{\mathbf{v}}\| = \sqrt{\frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0} = 1$$

# Distance

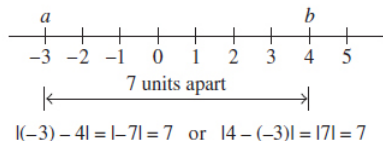
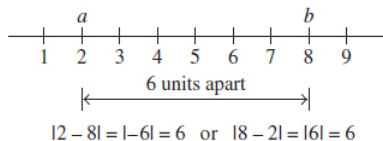
## Definition 1.3 (Distance in $\mathbb{R}$ )

The distance between any two numbers  $a, b \in \mathbb{R}$  can be defined as

$$d(a, b) = |a - b|$$

## Example

Calculate the distance between 2 and 8 as well as between -3 and 4.



**FIGURE 3** Distances in  $\mathbb{R}$ .

# Distance

## Definition 1.4 (Distance in $\mathbb{R}^n$ )

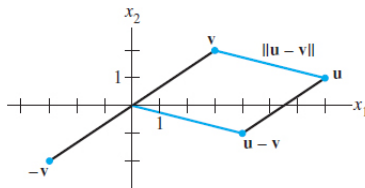
The distance between any two vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$  can be defined as

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

## Example

Calculate the distance between  $\mathbf{u} = (7, 1)$  and  $\mathbf{v} = (3, 2)$

$$d(\mathbf{u}, \mathbf{v}) = \|(7, 1) - (3, 2)\| = \|(4, -1)\| = \sqrt{4^2 + 1^2} = \sqrt{17}$$



**FIGURE 4** The distance between  $\mathbf{u}$  and  $\mathbf{v}$  is the length of  $\mathbf{u} - \mathbf{v}$ .

## Example

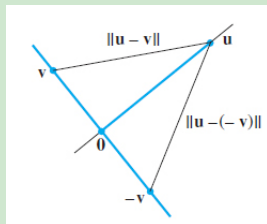
For any two vectors in  $\mathbb{R}^3$ ,  $\mathbf{u}$  and  $\mathbf{v}$ , the distance can be calculated through

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

# Orthogonality

## Example

Any two vectors in  $\mathbb{R}^2$ ,  $\mathbf{u}$  and  $\mathbf{v}$ , are orthogonal if  $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{u}, -\mathbf{v})$



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

$$d^2(\mathbf{u}, \mathbf{v}) = d^2(\mathbf{u}, -\mathbf{v}) \Rightarrow -2\mathbf{u} \cdot \mathbf{v} = 2\mathbf{u} \cdot \mathbf{v} \Rightarrow \mathbf{u} \cdot \mathbf{v} = 0$$

# Orthogonality

## Definition 1.5 (Orthogonality between two vectors)

Any two different vectors,  $\mathbf{u}$  and  $\mathbf{v}$ , in a vector space  $V$  are **orthogonal** iff

$$\mathbf{u} \cdot \mathbf{v} = 0$$

## Corollary

$\mathbf{0}$  is orthogonal to any other vector.

## Theorem 1.5 (Pythagorean theorem)

Any two vectors,  $\mathbf{u}$  and  $\mathbf{v}$ , in a vector space  $V$  are orthogonal iff

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



# Orthogonality

## Definition 1.6 (Orthogonality between vector and vector space)

Let  $\mathbf{u}$  be a vector in a vector space  $V$  and  $W$  a vector subspace of  $V$ .  $\mathbf{u}$  is **orthogonal** to  $W$  if  $\mathbf{u}$  is orthogonal to all vectors in  $W$ . The set of all vectors orthogonal to  $W$  is denoted as  $W^\perp$  (the **orthogonal complement** of  $W$ ).

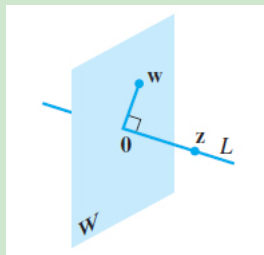
## Example

Let  $W$  be a plane in  $\mathbb{R}^3$  passing through the origin and  $L$  be a line, passing through the origin and perpendicular to  $W$ . For any vector  $\mathbf{w} \in W$  and any vector  $\mathbf{z} \in L$  we have

$$\mathbf{w} \cdot \mathbf{z} = 0$$

Therefore,

$$L = W^\perp \Leftrightarrow W = L^\perp$$



# Orthogonality

## Theorem 1.6

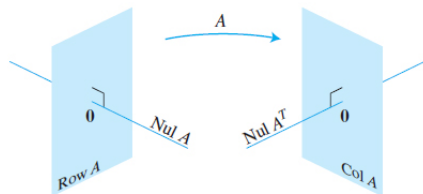
Let  $W$  be a vector subspace of a vector space  $V$ .

- 1  $\mathbf{x} \in W^\perp$  iff  $\mathbf{x}$  is orthogonal to every vector in a set that spans  $W$ .
- 2  $W^\perp$  is a vector subspace of  $V$ .

## Theorem 1.7

Let  $A \in \mathcal{M}_{m \times n}$ , then

- 1  $(\text{Row}\{A\})^\perp = \text{Nul}\{A\}$
- 2  $(\text{Col}\{A\})^\perp = \text{Nul}\{A^T\}$



**FIGURE 8** The fundamental subspaces determined by an  $m \times n$  matrix  $A$ .

# Orthogonality

*Proof*  $\text{Nul}\{A\} \subseteq (\text{Row}\{A\})^\perp$

Consider the rows of  $A$ ,  $\mathbf{a}_i$  ( $i = 1, 2, \dots, m$ ) as column vectors, then for any vector  $\mathbf{x} \in \text{Nul}\{A\}$  we know

$$A\mathbf{x} = \mathbf{0} \Rightarrow \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{a}_1^T \mathbf{x} \\ \mathbf{a}_2^T \mathbf{x} \\ \vdots \\ \mathbf{a}_m^T \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \mathbf{a}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_m \cdot \mathbf{x} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Consequently,  $\mathbf{x}$  is orthogonal to all the rows of  $A$ , which span  $\text{Row}\{A\}$  and by the previous theorem,  $\mathbf{x} \in (\text{Row}\{A\})^\perp$

*Proof*  $\text{Nul}\{A\} \supseteq (\text{Row}\{A\})^\perp$

Conversely, let  $\mathbf{x} \in (\text{Row}\{A\})^\perp$ , then by the previous theorem we know that

$$\mathbf{a}_i \cdot \mathbf{x} = 0 \quad \text{for } i = 1, 2, \dots, m \Rightarrow A\mathbf{x} = \mathbf{0}$$

So,  $\mathbf{x} \in \text{Nul}\{A\}$

# Orthogonality

*Proof*  $(\text{Col}\{A\})^\perp = \text{Nul}\{A^T\}$

Let's define  $B = A^T$ . By the first part of this theorem, we know

$$(\text{Row}\{B\})^\perp = \text{Nul}\{B\} \Rightarrow (\text{Row}\{A^T\})^\perp = \text{Nul}\{A^T\} \Rightarrow (\text{Col}\{A\})^\perp = \text{Nul}\{A^T\}$$

## Theorem 1.8

*For any two vectors  $\mathbf{u}$  and  $\mathbf{v}$  in a vector space  $V$ , the angle between the two can be measured through the dot product:*

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

## Exercises

From Lay (3rd ed.), Chapter 6, Section 1:

- 6.1.15
- 6.1.22
- 6.1.24
- 6.1.26
- 6.1.28
- 6.1.30
- 6.1.32 (computer)

## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

# Orthogonal sets

## Definition 2.1 (Orthogonal set)

Let  $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$  be a set of vectors.  $S$  is an orthogonal set iff

$$\mathbf{u}_i \cdot \mathbf{u}_j = 0 \quad \forall i, j \in \{1, 2, \dots, p\} \ i \neq j$$

## Example

Let  $\mathbf{u}_1 = (3, 1, 1)$ ,  $\mathbf{u}_2 = (-1, 2, 1)$ ,  $\mathbf{u}_3 = (-\frac{1}{2}, -2, \frac{7}{2})$ . Check whether the set  $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  is orthogonal.

Solution

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0$$

$$\mathbf{u}_1 \cdot \mathbf{u}_3 = 3 \cdot (-\frac{1}{2}) + 1 \cdot (-2) + 1 \cdot (\frac{7}{2}) = 0$$

$$\mathbf{u}_2 \cdot \mathbf{u}_3 = (-1) \cdot (-\frac{1}{2}) + 2 \cdot (-2) + 1 \cdot (\frac{7}{2}) = 0$$

# Orthogonal sets

## Theorem 2.1

*If  $S$  is an orthogonal set of non-null vectors, then  $S$  is linearly independent and, consequently, it is a basis of the subspace spanned by  $S$ .*

Proof  
*Let  $\mathbf{u}_i$  ( $i = 1, 2, \dots, p$ ) be the elements of  $S$ . Let us assume that  $S$  is linearly dependent. Then, there exists coefficients  $c_1, c_2, \dots, c_p$  not all of them null such that*

$$\mathbf{0} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_p\mathbf{u}_p$$

*Now, we compute the inner product with  $\mathbf{u}_1$*

$$\begin{aligned} \mathbf{0} \cdot \mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1 \\ 0 &= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1) + c_2(\mathbf{u}_2 \cdot \mathbf{u}_1) + \dots + c_p(\mathbf{u}_p \cdot \mathbf{u}_1) = c_1\|\mathbf{u}_1\|^2 \Rightarrow c_1 = 0 \end{aligned}$$

*Multiplying by  $\mathbf{u}_i$  ( $i = 2, 3, \dots, p$ ) we can show that all  $c_i$ 's are 0, and, therefore, the set  $S$  is linearly independent.*



# Orthogonal basis

## Definition 2.2 (Orthogonal basis)

*A set of vectors  $B$  is an orthogonal basis of a vector space  $V$  if it is an orthogonal set and it is a basis of  $V$ .*

## Theorem 2.2

*Let  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$  be an orthogonal basis for a vector space  $V$ , for each  $\mathbf{x} \in V$  we have*

$$\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \frac{\mathbf{x} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 + \dots + \frac{\mathbf{x} \cdot \mathbf{u}_p}{\|\mathbf{u}_p\|^2} \mathbf{u}_p$$

### Proof

*If  $\mathbf{x}$  is in  $V$ , then it can be expressed as a linear combination of the vectors in a basis of  $V$*

$$\mathbf{x} = c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \dots + c_p \mathbf{u}_p$$

*Now, we calculate the dot product with  $\mathbf{u}_1$*

$$\mathbf{x} \cdot \mathbf{u}_1 = (c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \dots + c_p \mathbf{u}_p) \cdot \mathbf{u}_1 = c_1 \|\mathbf{u}_1\|^2 \Rightarrow c_1 = \frac{\mathbf{x} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2}$$

# Orthogonal basis

## Example

Let  $\mathbf{u}_1 = (3, 1, 1)$ ,  $\mathbf{u}_2 = (-1, 2, 1)$ ,  $\mathbf{u}_3 = (-\frac{1}{2}, -2, \frac{7}{2})$ , and  $B = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  be an orthogonal basis of  $\mathbb{R}^3$ . Let  $\mathbf{x} = (6, 1, -8)$ . The coordinates of  $\mathbf{x}$  in  $B$  are given by

$$\begin{array}{lll} \mathbf{x} \cdot \mathbf{u}_1 = 11 & \mathbf{x} \cdot \mathbf{u}_2 = -12 & \mathbf{x} \cdot \mathbf{u}_3 = -33 \\ \|\mathbf{u}_1\|^2 = 11 & \|\mathbf{u}_2\|^2 = 6 & \|\mathbf{u}_3\|^2 = \frac{33}{2} \end{array}$$

$$\begin{aligned} \mathbf{x} &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{\frac{33}{2}}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$

The coordinates of  $\mathbf{x}$  in the basis  $B$  are

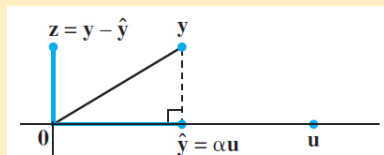
$$[\mathbf{x}]_B = (1, -2, -2)$$

# Orthogonal projections

## Orthogonal projection onto a vector

Consider a vector  $\mathbf{y}$  and another one  $\mathbf{u}$ . Let us assume we want to decompose  $\mathbf{y}$  as the sum of two orthogonal vectors  $\hat{\mathbf{y}}$  (along the direction of  $\mathbf{u}$ ) and another vector  $\mathbf{z}$  (orthogonal to  $\mathbf{u}$ ):

$$\begin{aligned}\mathbf{y} &= \hat{\mathbf{y}} + \mathbf{z} = \alpha \mathbf{u} + \mathbf{z} \Rightarrow \\ \mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}}\end{aligned}$$



**FIGURE 2**

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

We need to find  $\alpha$  that makes  $\mathbf{u}$  and  $\mathbf{z}$  orthogonal.

$$0 = \mathbf{z} \cdot \mathbf{u} = (\mathbf{y} - \alpha \mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha \|\mathbf{u}\|^2 \Rightarrow \alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\|\mathbf{u}\|^2}$$

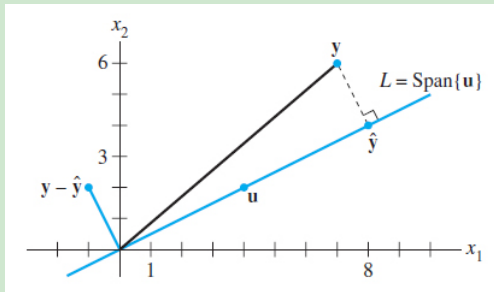
$\hat{\mathbf{y}}$  is the **orthogonal projection** of  $\mathbf{y}$  onto  $\mathbf{u}$ .

# Orthogonal projections

## Example

Let  $\mathbf{y} = (7, 6)$  and  $\mathbf{u} = (4, 2)$ . Then,

$$\left. \begin{array}{l} \mathbf{y} \cdot \mathbf{u} = 40 \\ \|\mathbf{u}\|^2 = 20 \end{array} \right\} \Rightarrow \begin{aligned} \hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}}{\|\mathbf{u}\|^2} \mathbf{u} = \frac{40}{20} \mathbf{u} = 2\mathbf{u} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} \\ \mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix} 7 \\ 6 \end{pmatrix} - \begin{pmatrix} 8 \\ 4 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \\ d(\mathbf{y}, \hat{\mathbf{y}}) &= \|\mathbf{y} - \hat{\mathbf{y}}\| = \|\mathbf{z}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5} \end{aligned}$$



# Orthonormal set

## Definition 2.3 (Orthonormal set)

$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$  is an **orthonormal set** if it is an orthogonal set and all  $\mathbf{u}_i$  vectors have unit length.

## Example

Show that the set  $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  is orthonormal, with

$$\mathbf{u}_1 = \frac{1}{\sqrt{11}} \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} \quad \mathbf{u}_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \quad \mathbf{u}_3 = \frac{1}{\sqrt{66}} \begin{pmatrix} -1 \\ -4 \\ 7 \end{pmatrix}$$

### Solution

Let's check that they are orthogonal:

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = \frac{1}{\sqrt{11}} \frac{1}{\sqrt{6}} (3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1) = 0$$

$$\mathbf{u}_1 \cdot \mathbf{u}_3 = \frac{1}{\sqrt{11}} \frac{1}{\sqrt{66}} (3 \cdot (-1) + 1 \cdot (-4) + 1 \cdot 7) = 0$$

$$\mathbf{u}_2 \cdot \mathbf{u}_3 = \frac{1}{\sqrt{6}} \frac{1}{\sqrt{66}} ((-1) \cdot (-1) + (2) \cdot (-4) + (1) \cdot 7) = 0$$

# Orthonormal set

## Example (continued)

Now, let's check that they have unit length:

$$\|\mathbf{u}_1\| = \sqrt{\left(\frac{1}{\sqrt{11}}\right)^2 (3^2 + 1^2 + 1^2)} = \sqrt{\frac{9+1+1}{11}} = 1$$

$$\|\mathbf{u}_2\| = \sqrt{\left(\frac{1}{\sqrt{6}}\right)^2 ((-1)^2 + 2^2 + 1^2)} = \sqrt{\frac{1+4+1}{6}} = 1$$

$$\|\mathbf{u}_3\| = \sqrt{\left(\frac{1}{\sqrt{66}}\right)^2 ((-1)^2 + (-4)^2 + 7^2)} = \sqrt{\frac{1+16+49}{66}} = 1$$

## Theorem 2.3

*If  $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  is an orthonormal set, then it is an orthonormal basis of  $\text{Span}\{S\}$ .*

## Example

$\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  is an orthonormal basis of  $\mathbb{R}^n$ .

# Orthonormal basis

## Theorem 2.4

Let  $S = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  is an orthogonal set of vectors, then the set  $S' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n\}$  where

$$\mathbf{u}'_i = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|}$$

is a orthonormal set (this operation is called **vector normalization**).

Proof

Let's check that the  $\mathbf{u}'_i$  vectors are orthogonal:

$$\mathbf{u}'_i \cdot \mathbf{u}'_j = \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|} \cdot \frac{\mathbf{u}_j}{\|\mathbf{u}_j\|} = \frac{1}{\|\mathbf{u}_i\|\|\mathbf{u}_j\|} \mathbf{u}_i \cdot \mathbf{u}_j$$

But this product is obviously 0 because the  $\mathbf{u}_i$  vectors are orthogonal. Let's check now that the  $\mathbf{u}'_i$  vectors have unit length:

$$\|\mathbf{u}'_i\| = \left\| \frac{\mathbf{u}_i}{\|\mathbf{u}_i\|} \right\| = \frac{\|\mathbf{u}_i\|}{\|\mathbf{u}_i\|} = 1$$

# Orthonormal matrix

## Theorem 2.5

Let  $U \in \mathcal{M}_{m \times n}$  be a square matrix. The columns of  $U$  form an orthonormal set iff

$$U^T U = I_n$$

It is said that  $U$  is an **orthonormal matrix**.

Proof

Let's consider the columns of  $U$

$$U = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n)$$

Let's calculate now  $U^T U$

$$U^T U = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \dots \\ \mathbf{u}_n^T \end{pmatrix} (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n) = \begin{pmatrix} \mathbf{u}_1^T \mathbf{u}_1 & \mathbf{u}_1^T \mathbf{u}_2 & \dots & \mathbf{u}_1^T \mathbf{u}_n \\ \mathbf{u}_2^T \mathbf{u}_1 & \mathbf{u}_2^T \mathbf{u}_2 & \dots & \mathbf{u}_2^T \mathbf{u}_n \\ \dots & \dots & \dots & \dots \\ \mathbf{u}_n^T \mathbf{u}_1 & \mathbf{u}_n^T \mathbf{u}_2 & \dots & \mathbf{u}_n^T \mathbf{u}_n \end{pmatrix}$$

The condition  $U^T U = I_n$  simply states  $\begin{cases} \mathbf{u}_i^T \mathbf{u}_j = 0 & i \neq j \\ \mathbf{u}_i^T \mathbf{u}_j = 1 & i = j \end{cases}$ , which is the definition of an orthonormal set.



# Orthonormal matrix

## Theorem 2.6

Let  $U \in \mathcal{M}_{n \times n}$  be an orthonormal matrix and  $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ , then

- ①  $\|U\mathbf{x}\| = \|\mathbf{x}\|$
- ②  $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
- ③  $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0 \Leftrightarrow \mathbf{x} \cdot \mathbf{y} = 0$

## Example

Let  $U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{pmatrix}$  and  $\mathbf{x} = \begin{pmatrix} \sqrt{2} \\ 3 \end{pmatrix}$ .

$U$  is an orthonormal matrix because

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

# Orthonormal matrix

## Example (continued)

Let's calculate now  $U\mathbf{x}$

$$U\mathbf{x} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & -\frac{2}{3} \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \sqrt{2} \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$$

Let's check now that  $\|U\mathbf{x}\| = \|\mathbf{x}\|$

$$\|U\mathbf{x}\| = \|(3, -1, 1)\| = \sqrt{3^2 + (-1)^2 + 1^2} = \sqrt{11}$$

$$\|\mathbf{x}\| = \|(\sqrt{2}, 3)\| = \sqrt{(\sqrt{2})^2 + 3^2} = \sqrt{11}$$

## Theorem 2.7

Let  $U$  be an orthonormal and square matrix. Then,

- ①  $U^{-1} = U^T$
- ②  $U^T$  is also an orthonormal matrix (i.e., the rows of  $U$  also form an orthonormal set of vectors).

## Exercises

From Lay (3rd ed.), Chapter 6, Section 2:

- 6.2.1
- 6.2.10
- 6.2.15
- 6.2.25
- 6.2.26
- 6.2.29
- 6.2.35 (computer)

## 7 Orthogonality and least squares

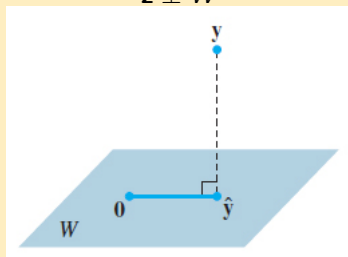
- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

# Orthogonal projections

## Definition 3.1 (Orthogonal projection)

*The orthogonal projection of a point  $\mathbf{y}$  onto a vector subspace  $W$  is a point  $\hat{\mathbf{y}} \in W$  such that*

$$\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$$
$$\mathbf{z} \perp W$$



# Orthogonal projections

## Example

Let  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_5\}$  be an orthogonal basis of  $\mathbb{R}^5$ . Consider the subspace  $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ . Given any vector  $\mathbf{y} \in \mathbb{R}^5$ , we can decompose it as the sum of a vector in  $W$  and a vector perpendicular to  $W$

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

## Solution

If  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_5\}$  is a basis of  $\mathbb{R}^5$ , then any vector  $\mathbf{y} \in \mathbb{R}^5$  can be written as

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \dots + c_5\mathbf{u}_5$$

We may decompose this sum as

$$\begin{aligned}\hat{\mathbf{y}} &= c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \\ \mathbf{z} &= c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5\end{aligned}$$

# Orthogonal projections

## Example (continued)

It is obvious that  $\hat{\mathbf{y}} \in W$ . Now we need to show that  $\mathbf{z} \in W^\perp$ . For doing so, we will show that

$$\mathbf{z} \cdot \mathbf{u}_1 = 0$$

$$\mathbf{z} \cdot \mathbf{u}_2 = 0$$

To show the first equation we note that

$$\begin{aligned}\mathbf{z} \cdot \mathbf{u}_1 &= (c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3(\mathbf{u}_3 \cdot \mathbf{u}_1) + c_4(\mathbf{u}_4 \cdot \mathbf{u}_1) + c_5(\mathbf{u}_5 \cdot \mathbf{u}_1) \\ &= c_3 \cdot 0 + c_4 \cdot 0 + c_5 \cdot 0 \\ &= 0\end{aligned}$$

We would proceed analogously for  $\mathbf{z} \cdot \mathbf{u}_2 = 0$ .

# Orthogonal projections

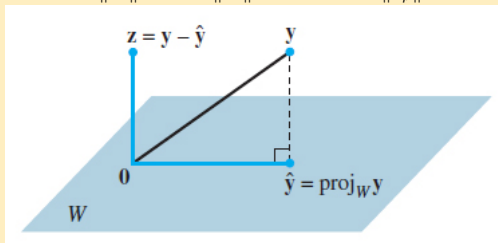
## Theorem 3.1 (Orthogonal Decomposition Theorem)

Let  $W$  be a vector subspace of a vector space  $V$ . Then, any vector  $\mathbf{y} \in V$  can be written uniquely as

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

with  $\hat{\mathbf{y}} \in W$  and  $\mathbf{z} \in W^\perp$ . In fact, if  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$  is an orthogonal basis of  $W$ , then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\|\mathbf{u}_p\|^2} \mathbf{u}_p$$





# Orthogonal projections

## Proof

$\hat{\mathbf{y}}$  is obviously in  $W$  since it has been written as a linear combination of vectors in a basis of  $W$ .  $\mathbf{z}$  is perpendicular to  $W$  because

$$\begin{aligned}\mathbf{z} \cdot \mathbf{u}_1 &= \left( \mathbf{y} - \left( \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\|\mathbf{u}_p\|^2} \mathbf{u}_p \right) \right) \cdot \mathbf{u}_1 \\&= \mathbf{y} \cdot \mathbf{u}_1 - \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} (\mathbf{u}_1 \cdot \mathbf{u}_1) - \frac{\mathbf{y} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} (\mathbf{u}_2 \cdot \mathbf{u}_1) - \dots - \frac{\mathbf{y} \cdot \mathbf{u}_p}{\|\mathbf{u}_p\|^2} (\mathbf{u}_p \cdot \mathbf{u}_1) \\&\quad [\{\mathbf{u}_i\} \text{ is an orthogonal set}] \\&= \mathbf{y} \cdot \mathbf{u}_1 - \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} (\mathbf{u}_1 \cdot \mathbf{u}_1) \\&= \mathbf{y} \cdot \mathbf{u}_1 - \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \|\mathbf{u}_1\|^2 \\&= \mathbf{y} \cdot \mathbf{u}_1 - \mathbf{y} \cdot \mathbf{u}_1 \\&= 0\end{aligned}$$

We could proceed analogously for all elements in the basis of  $W$ .

# Orthogonal projections

We need to show now that the decomposition is unique. Let us assume that it is not unique. Consequently, there exist different vectors such that

$$\begin{aligned}\mathbf{y} &= \hat{\mathbf{y}} + \mathbf{z} \\ \mathbf{y} &= \hat{\mathbf{y}}' + \mathbf{z}'\end{aligned}$$

We subtract both equations

$$\mathbf{0} = (\hat{\mathbf{y}} - \hat{\mathbf{y}}') + (\mathbf{z} - \mathbf{z}') \Rightarrow \hat{\mathbf{y}} - \hat{\mathbf{y}}' = \mathbf{z}' - \mathbf{z}$$

Let  $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}'$ . It is obvious that  $\mathbf{v} \in W$  because it is written as a linear combination of vectors in  $W$ . On the other side,  $\mathbf{v} = \mathbf{z}' - \mathbf{z}$ , i.e., it is a linear combination of vectors in  $W^\perp$ , so  $\mathbf{v} \in W^\perp$ . The only vector that belongs to  $W$  and  $W^\perp$  at the same time is

$$\mathbf{v} = \mathbf{0} \Rightarrow \begin{cases} \hat{\mathbf{y}} = \hat{\mathbf{y}}' \\ \mathbf{z} = \mathbf{z}' \end{cases}.$$

and consequently, the orthogonal decomposition is unique. Additionally, the uniqueness of the decomposition depends only on  $W$  and not on the particular basis chosen for  $W$ .

# Orthogonal projections

## Example

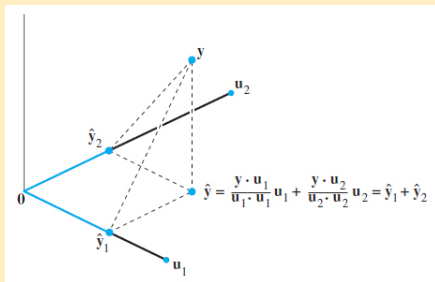
Let  $\mathbf{u}_1 = (2, 5, -1)$  and  $\mathbf{u}_2 = (-2, 1, 1)$ . Let  $W$  be the subspace spanned by  $\mathbf{u}_1$  and  $\mathbf{u}_2$ . Let  $\mathbf{y} = (1, 2, 3) \in \mathbb{R}^3$ . The orthogonal projection of  $\mathbf{y}$  onto  $W$  is

$$\begin{aligned}\hat{\mathbf{y}} &= \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 \\&= \frac{1 \cdot 2 + 2 \cdot 5 + 3 \cdot (-1)}{2^2 + 5^2 + (-1)^2} \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{1 \cdot (-2) + 2 \cdot 1 + 3 \cdot 1}{(-2)^2 + 1^2 + 1^2} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} \\&= \frac{9}{30} \begin{pmatrix} 2 \\ 5 \\ -1 \end{pmatrix} + \frac{15}{30} \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{pmatrix} \\ \mathbf{z} &= \mathbf{y} - \hat{\mathbf{y}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{pmatrix} = \begin{pmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{pmatrix}\end{aligned}$$

# Orthogonal projections

## Geometrical interpretation

$\hat{\mathbf{y}}$  can be understood as the sum of the orthogonal projection of  $\mathbf{y}$  onto each one of the elements of the basis of  $W$ .



## Theorem 3.2

*If  $\mathbf{y}$  belongs to  $W$ , then the orthogonal projection of  $\mathbf{y}$  onto  $W$  is itself:*

$$\hat{\mathbf{y}} = \mathbf{y}$$

# Properties of orthogonal projections

## Theorem 3.3 (Best approximation theorem)

*The orthogonal projection of  $\mathbf{y}$  onto  $W$  is the point in  $W$  with minimum distance to  $\mathbf{y}$ , i.e.,*

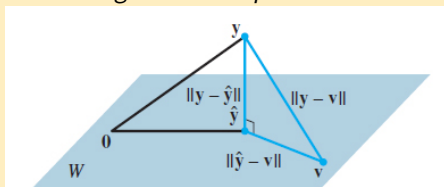
$$\|\mathbf{y} - \hat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{v}\|$$

*for all  $\mathbf{v} \in W, \mathbf{v} \neq \hat{\mathbf{y}}$ .*

Proof

*We know that  $\mathbf{y} - \hat{\mathbf{y}}$  is orthogonal to  $W$ . For any vector  $\mathbf{v} \in W, \mathbf{v} \neq \hat{\mathbf{y}}$ , we have that  $\hat{\mathbf{y}} - \mathbf{v}$  is in  $W$ . Now consider the orthogonal decomposition of the vector  $\mathbf{y} - \mathbf{v}$*

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



**FIGURE 4** The orthogonal projection of  $\mathbf{y}$  onto  $W$  is the closest point in  $W$  to  $\mathbf{y}$ .

# Properties of orthogonal projections

Due to the orthogonal decomposition theorem (Theorem 3.1), this decomposition is unique and due to the Pythagorean theorem (Theorem 1.5) we have

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

Since  $\mathbf{v} \neq \hat{\mathbf{y}}$  we have  $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$  and consequently

$$\|\mathbf{y} - \mathbf{v}\|^2 > \|\mathbf{y} - \hat{\mathbf{y}}\|^2$$

# Properties of orthogonal projections

## Theorem 3.4

*If  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$  is an orthonormal basis of  $W$ , then the orthogonal projection of  $\mathbf{y}$  onto  $W$  is*

$$\hat{\mathbf{y}} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \dots + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p$$

*If we construct the orthonormal matrix  $U = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_p)$ , then*

$$\hat{\mathbf{y}} = UU^T \mathbf{y}$$

### Proof

*By Theorem 3.1 we know that for all orthogonal bases it is verified*

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\|\mathbf{u}_1\|^2} \mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\|\mathbf{u}_2\|^2} \mathbf{u}_2 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\|\mathbf{u}_p\|^2} \mathbf{u}_p$$

*Since the basis is in this case orthonormal, then  $\|\mathbf{u}\| = 1$  and consequently*

$$\hat{\mathbf{y}} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \dots + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p$$

# Properties of orthogonal projections

On the other side we have

$$U^T \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_p^T \end{pmatrix} \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^T \mathbf{y} \\ \mathbf{u}_2^T \mathbf{y} \\ \vdots \\ \mathbf{u}_p^T \mathbf{y} \end{pmatrix} = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{y} \rangle \\ \langle \mathbf{u}_2, \mathbf{y} \rangle \\ \vdots \\ \langle \mathbf{u}_p, \mathbf{y} \rangle \end{pmatrix}$$

Then,

$$UU^T \mathbf{y} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_p \end{pmatrix} \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{y} \rangle \\ \langle \mathbf{u}_2, \mathbf{y} \rangle \\ \vdots \\ \langle \mathbf{u}_p, \mathbf{y} \rangle \end{pmatrix} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \dots + \langle \mathbf{y}, \mathbf{u}_p \rangle \mathbf{u}_p$$

(q.e.d.)



# Properties of orthogonal projections

## Corollary

Let  $U = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_p)$  be a  $n \times p$  matrix with orthonormal columns and  $W = \text{Col}\{U\}$  its column space. Then,

$$\forall \mathbf{x} \in \mathbb{R}^p \quad U^T U \mathbf{x} = \mathbf{x} \quad \text{No effect}$$

$$\forall \mathbf{y} \in \mathbb{R}^n \quad U U^T \mathbf{y} = \hat{\mathbf{y}} \quad \text{Orthogonal projection of } \mathbf{y} \text{ onto } W$$

If  $U$  is a  $n \times n$ , then  $W = \mathbb{R}^n$  and the projection has no effect

$$\forall \mathbf{y} \in \mathbb{R}^n \quad U U^T \mathbf{y} = \hat{\mathbf{y}} = \mathbf{y} \quad \text{No effect}$$

## Exercises

From Lay (3rd ed.), Chapter 6, Section 3:

- 6.3.1
- 6.3.7
- 6.3.15
- 6.3.23
- 6.3.24
- 6.3.25 (computer)

## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

# Gram-Schmidt orthogonalization

Gram-Schmidt orthogonalization is a procedure aimed at producing an orthogonal basis of any subspace  $W$ .

## Example

Let  $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$  with  $\mathbf{x}_1 = (3, 6, 0)$  and  $\mathbf{x}_2 = (1, 2, 2)$ . Let's look for an orthogonal basis of  $W$ .

### Solution

We may keep the first vector for the basis

$$\mathbf{v}_1 = \mathbf{x}_1 = (3, 6, 0)$$

For the second vector in the basis, we need to keep the component of  $\mathbf{x}_2$  that is orthogonal to  $\mathbf{x}_1$ . For doing so we calculate the projection of  $\mathbf{x}_2$  onto  $\mathbf{x}_1$  ( $\mathbf{p}$ ), and we decompose  $\mathbf{x}_2$  as

$$\mathbf{x}_2 = \mathbf{p} + (\mathbf{x}_2 - \mathbf{p}) = (1, 2, 0) + (0, 0, 2)$$

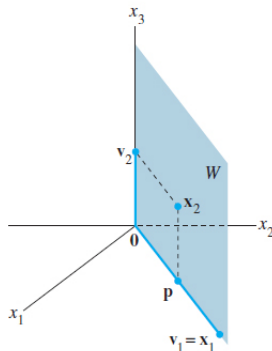
We, then, keep the orthogonal part of  $\mathbf{x}_2$

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = (0, 0, 2)$$

# Gram-Schmidt orthogonalization

## Example (continued)

The set  $\{\mathbf{v}_1, \mathbf{v}_2\}$  is an orthogonal basis of  $W$ .



**FIGURE 1**

Construction of an orthogonal basis  $\{\mathbf{v}_1, \mathbf{v}_2\}$ .

# Gram-Schmidt orthogonalization

## Example

Let  $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$  with  $\mathbf{x}_1 = (1, 1, 1, 1)$ ,  $\mathbf{x}_2 = (0, 1, 1, 1)$  and  $\mathbf{x}_3 = (0, 0, 1, 1)$ . Let's look for an orthogonal basis of  $W$ .

### Solution

We may keep the first vector for the basis. Then we construct a subspace ( $W_1$ ) with a single element in its basis

$$\mathbf{v}_1 = \mathbf{x}_1 = (1, 1, 1, 1) \quad W_1 = \text{Span}\{\mathbf{v}_1\}$$

For the second vector in the basis, we need to keep the component of  $\mathbf{x}_2$  that is orthogonal to  $W_1$ . With the already computed basis vectors, we construct a new subspace ( $W_2$ ) with two elements in its basis

$$\mathbf{v}_2 = \mathbf{x}_2 - \text{Proj}_{W_1}(\mathbf{x}_2) = \left(-\frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \quad W_2 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$$

For the third vector in the basis, we repeat the same procedure

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{Proj}_{W_2}(\mathbf{x}_3) = \left(0, -\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right) \quad W_3 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

# Gram-Schmidt orthogonalization

## Theorem 4.1 (Gram-Schmidt orthogonalization)

Given a basis  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$  for a vector subspace  $W$ . Define

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{x}_1 & W_1 &= \text{Span}\{\mathbf{v}_1\} \\ \mathbf{v}_2 &= \mathbf{x}_2 - \text{Proj}_{W_1}(\mathbf{x}_2) & W_2 &= \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\} \\ &\dots & & \\ \mathbf{v}_p &= \mathbf{x}_p - \text{Proj}_{W_{p-1}}(\mathbf{x}_p) & W_p &= \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\} = W\end{aligned}$$

Then  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  is an orthogonal basis of  $W$ .

# Gram-Schmidt orthogonalization

## Proof

Consider  $W_k = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  and let us assume that  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  is a basis of  $W_k$ . Now we construct

$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - \text{Proj}_{W_k}(\mathbf{x}_{k+1}) \quad W_{k+1} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}\}$$

By the orthogonal decomposition theorem (Theorem 3.1), we know that  $\mathbf{v}_{k+1}$  is orthogonal to  $W_k$ . Because  $\mathbf{x}_{k+1}$  is an element of a basis, we know that  $\mathbf{x}_{k+1} \notin W_k$ . Therefore,  $\mathbf{v}_{k+1}$  is not null and  $\mathbf{x}_{k+1} \in W_{k+1}$ . Finally, the set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{k+1}\}$  is a set of orthogonal, non-null vectors in the  $(k+1)$ -dimensional space  $W_{k+1}$ . Consequently, by Theorem 9.4 in Chapter 5, it must be a basis of  $W_{k+1}$ . This process can be iterated till  $k = p$ .



# Gram-Schmidt orthogonalization

## Orthonormal basis

Once we have an orthogonal basis, we simply have to normalize each vector to have an orthonormal basis.

## Example

Let  $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$  with  $\mathbf{x}_1 = (3, 6, 0)$  and  $\mathbf{x}_2 = (1, 2, 2)$ . Let's look for an orthonormal basis of  $W$ .

### Solution

In Slide 52 we learned that an orthogonal basis was given by

$$\mathbf{v}_1 = (3, 6, 0)$$

$$\mathbf{v}_2 = (0, 0, 2)$$

Now, we normalize these two vectors to obtain an orthonormal basis

$$\mathbf{v}'_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{45}}(3, 6, 0) = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}, 0\right)$$

$$\mathbf{v}'_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{2}(0, 0, 2) = (0, 0, 1)$$

# QR factorization of matrices

If we apply the Gram-Schmidt factorization to the columns of a matrix, we have the following factorization scheme. This factorization is used in practice to find eigenvalues and eigenvectors as well as to solve linear equation systems.

## Theorem 4.2 (QR factorization)

*Let  $A \in \mathcal{M}_{m \times n}$  with linearly independent columns. Then,  $A$  can be factored as*

$$A = QR$$

*where  $Q \in \mathcal{M}_{m \times n}$  is a matrix whose columns form an orthonormal basis of  $\text{Col}\{A\}$  and  $R \in \mathcal{M}_{n \times n}$  is an upper triangular invertible matrix with positive entries on its diagonal.*

### Proof

*Let's orthogonalize the columns of  $A$  following the Gram-Schmidt procedure and construct the orthonormal basis of  $\text{Col}\{A\}$ . Let  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  be such a basis. Let us construct the matrix*

$$Q = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_n)$$

# QR factorization of matrices

Let us call  $\mathbf{a}_i$  ( $i = 1, 2, \dots, n$ ) to the columns of  $A$ . By the Gram-Schmidt orthogonalization, we know that for any  $k$  between 1 and  $n$  we have

$$\text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\} = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$$

Consequently, we can express each column of  $A$  in the orthonormal basis:

$$\mathbf{a}_k = r_{1k}\mathbf{u}_1 + r_{2k}\mathbf{u}_2 + \dots + r_{kk}\mathbf{u}_k + 0 \cdot \mathbf{u}_{k+1} + \dots + 0 \cdot \mathbf{u}_n$$

If  $r_{kk}$  is negative, we can multiply both  $r_{kk}$  and  $\mathbf{u}_k$  by -1. We now collect all these coefficients in a vector  $\mathbf{r}_k = (r_{1k}, r_{2k}, \dots, r_{kk}, 0, 0, \dots, 0)$  to have

$$\mathbf{a}_k = Q\mathbf{r}_k$$

By gathering all these vectors in a matrix, we have the triangular matrix  $R$

$$R = (\mathbf{r}_1 \quad \mathbf{r}_2 \quad \dots \quad \mathbf{r}_n)$$

$R$  is invertible because the columns of  $A$  are linearly independent.

# QR factorization of matrices

## Example

Let's calculate the QR factorization of  $A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ . From Slide 54 we know that the vectors

$$\begin{aligned}\mathbf{v}_1 &= (1, 1, 1, 1) \\ \mathbf{v}_2 &= \left(-\frac{3}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \\ \mathbf{v}_3 &= \left(0, -\frac{2}{3}, \frac{1}{3}, \frac{1}{3}\right)\end{aligned}$$

Is an orthogonal basis of the column space of  $A$ . We now normalize these vectors to obtain the orthonormal basis in  $Q$

$$Q = \begin{pmatrix} \frac{1}{2} & -\frac{3}{\sqrt{12}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$

# QR factorization of matrices

## Example (continued)

To find  $R$  we multiply on both sides of the factorization by  $Q^T$

$$\begin{aligned} A = QR &\Rightarrow Q^T A = Q^T QR = R \\ R = Q^T A &= \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{3}{\sqrt{12}} & \frac{2}{\sqrt{12}} \\ 0 & 0 & \frac{1}{\sqrt{6}} \end{pmatrix} \end{aligned}$$

## Exercises

From Lay (3rd ed.), Chapter 6, Section 4:

- 6.4.7
- 6.4.13
- 6.4.19
- 6.4.22
- 6.4.24

## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- **Least squares (c)**
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

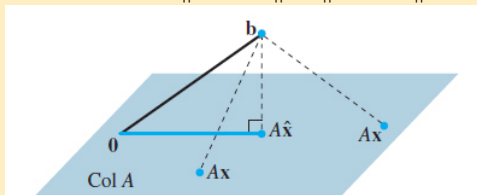
# Least squares

Let's assume we want to solve the equation system  $A\mathbf{x} = \mathbf{b}$ , but, due to noise, there is no solution. We may still look for a solution such that  $A\mathbf{x} \approx \mathbf{b}$ . In fact the goal will be to minimize  $d(A\mathbf{x}, \mathbf{b})$ .

## Definition 5.1 (Least squares solution)

Let  $A$  be a  $m \times n$  matrix and  $\mathbf{b} \in \mathbb{R}^m$ .  $\mathbf{x} \in \mathbb{R}^n$  is a **least squares solution** of the equation system  $A\mathbf{x} = \mathbf{b}$  iff

$$\forall \mathbf{x} \in \mathbb{R}^n \quad \|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$



**FIGURE 1** The vector  $\mathbf{b}$  is closer to  $A\hat{\mathbf{x}}$  than to  $A\mathbf{x}$  for other  $\mathbf{x}$ .



# Least squares

## Solution of the general least squares problem

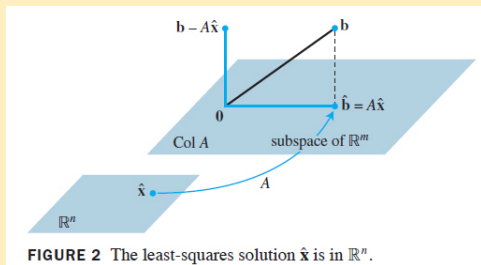
Applying the Best Approximation Theorem (Theorem 3.3), we may project  $\mathbf{b}$  onto the column space of  $A$

$$\hat{\mathbf{b}} = \text{Proj}_{\text{Col}\{A\}}\{\mathbf{b}\}$$

Then, we solve the equation system

$$A\mathbf{x} = \hat{\mathbf{b}}$$

that has at least one solution.



# Least squares

## Theorem 5.1

*The set of least-squares solutions of  $A\mathbf{x} = \mathbf{b}$  is the same as the set of solutions of the normal equations*

$$A^T A \mathbf{x} = A^T \mathbf{b}$$

*Proof: least-squares solutions  $\subset$  normal equations solutions*

*Let us assume that  $\hat{\mathbf{x}}$  is a least-squares solution. Then,  $\mathbf{b} - A\hat{\mathbf{x}}$  is orthogonal to  $\text{Col}\{A\}$ , and in particular, to each one of the columns of  $A$  ( $\mathbf{a}_i$ ,  $i = 1, 2, \dots, n$ ):*

$$\mathbf{a}_i \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \forall i \in \{1, 2, \dots, n\} \Rightarrow$$

$$\mathbf{a}_i^T (\mathbf{b} - A\hat{\mathbf{x}}) = 0 \quad \forall i \in \{1, 2, \dots, n\} \Rightarrow$$

$$A^T (\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \Rightarrow$$

$$A^T \mathbf{b} = A^T A \hat{\mathbf{x}}$$

*That is, every least-squares solution is also a solution of the normal equations.*

# Least squares

Proof: least-squares solutions  $\supset$  normal equations solutions

Let us assume that  $\hat{\mathbf{x}}$  is solution of the normal equations. Then,

$$\begin{aligned} A^T \mathbf{b} &= A^T A \hat{\mathbf{x}} \Rightarrow \\ A^T (\mathbf{b} - A \hat{\mathbf{x}}) &= \mathbf{0} \Rightarrow \\ \mathbf{a}_i^T (\mathbf{b} - A \hat{\mathbf{x}}) &= 0 \quad \forall i \in \{1, 2, \dots, n\} \end{aligned}$$

That is,  $\mathbf{b} - A \hat{\mathbf{x}}$  is orthogonal to the columns of  $A$  and, consequently, to  $\text{Col}\{A\}$ . Hence, the equation

$$\mathbf{b} = A \hat{\mathbf{x}} + (\mathbf{b} - A \hat{\mathbf{x}})$$

is the orthogonal decomposition of  $\mathbf{b}$  as a vector in  $\text{Col}\{A\}$  and a vector orthogonal to  $\text{Col}\{A\}$ . By the uniqueness of the orthogonal decomposition,  $A \hat{\mathbf{x}}$  must be the orthogonal projection of  $\mathbf{b}$  onto  $\text{Col}\{A\}$  so that

$$A \hat{\mathbf{x}} = \hat{\mathbf{b}}$$

and, therefore,  $\hat{\mathbf{x}}$  is a least-squares solution.

# Least squares

## Example

Find a least-squares solution to  $A\mathbf{x} = \mathbf{b}$  with  $A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}$  and  $\mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}$ .

### Solution

Let's solve the normal equations  $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$

$$A^T A = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix} \quad A^T \mathbf{b} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$
$$\begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix} \hat{\mathbf{x}} = \begin{pmatrix} 19 \\ 11 \end{pmatrix} \Rightarrow \hat{\mathbf{x}} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 19 \\ 11 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Let's check that  $\hat{\mathbf{x}}$  is not a solution of the original equation system but a least-squares solution

$$A\hat{\mathbf{x}} = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 3 \end{pmatrix} = \hat{\mathbf{b}} \neq \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}$$

# Least squares

## Definition 5.2 (Least-squares error)

*The least-squares error is defined as*

$$\sigma_{\epsilon}^2 \triangleq \|A\hat{\mathbf{x}} - \mathbf{b}\|^2 = \|\hat{\mathbf{b}} - \mathbf{b}\|^2$$

## Example (continued)

In this case:

$$\sigma_{\epsilon}^2 = \|(4, 4, 3) - (2, 0, 11)\| = \|(2, 4, -8)\| \approx 9.165$$

# Least squares

## Example

Unfortunately, the least-squares solution may not be unique as shown in the next example (arising in ANOVA). Find a least-squares solution to  $A\mathbf{x} = \mathbf{b}$  with

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix}.$$

Solution

$$A^T A = \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix} \quad A^T \mathbf{b} = \begin{pmatrix} 4 \\ -4 \\ 2 \\ 6 \end{pmatrix}$$

# Least squares

## Example (continued)

The augmented matrix is

$$\left( \begin{array}{cccc|c} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{array} \right) \sim \left( \begin{array}{cccc|c} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Any point of the form

$$\hat{\mathbf{x}} = \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \forall x_4 \in \mathbb{R}$$

is a least-squares solution of the problem.

## Theorem 5.2

*The matrix  $A^T A$  is invertible iff the columns of  $A$  are linearly independent. In this case, the equation system  $A\mathbf{x} = \mathbf{b}$  has a unique least-squares solution given by*

$$\hat{\mathbf{x}} = A^+ \mathbf{b}$$

*where  $A^+$  is the **Moore-Penrose pseudoinverse***

$$A^+ = (A^T A)^{-1} A^T$$



# Least squares and QR decomposition

Sometimes  $A^T A$  is ill-conditioned, this means that small perturbations in  $\mathbf{b}$  translate into large perturbations in  $\hat{\mathbf{x}}$ . The QR decomposition offers a numerically more stable way of finding the least-squares solution.

## Theorem 5.3

Let there be  $A \in \mathcal{M}_{m \times n}$  with linearly independent columns. Consider its QR decomposition ( $A = QR$ ). Then, for each  $\mathbf{b} \in \mathbb{R}^m$  there is a unique least-squares solution of  $A\mathbf{x} = \mathbf{b}$  given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$$

### Proof

If we substitute  $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$  into  $A\mathbf{x}$  we have

$$A\hat{\mathbf{x}} = AR^{-1}Q^T\mathbf{b} = QRR^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}.$$

But  $Q$  is an orthonormal basis of  $\text{Col}\{A\}$  (Theorem 4.2 and Corollary in Slide 49) and consequently  $QQ^T\mathbf{b}$  is the orthogonal projection of  $\mathbf{b}$  onto  $\text{Col}\{A\}$ , that is,  $\hat{\mathbf{b}}$ . So,  $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$  is a least-squares solution of  $A\mathbf{x} = \mathbf{b}$ . Additionally, since the columns of  $A$  are linearly independent, by Theorem 5.2, this solution is unique.

# Least squares and QR decomposition

Remind that numerically it is easier to solve  $R\hat{\mathbf{x}} = Q^T \mathbf{b}$  than  $\hat{\mathbf{x}} = R^{-1}Q^T \mathbf{b}$

L

et  $A = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix}$  and  $\mathbf{b} = \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix}$ . Its QR decomposition is

$$A = QR = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

$$Q^T \mathbf{b} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix} \Rightarrow \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix} \hat{\mathbf{x}} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix} \Rightarrow \hat{\mathbf{x}} = \begin{pmatrix} 10 \\ -6 \\ 2 \end{pmatrix}$$

## Exercises

From Lay (3rd ed.), Chapter 6, Section 5:

- 6.5.1
- 6.5.19
- 6.5.20
- 6.5.21
- 6.5.24

## 7 Orthogonality and least squares

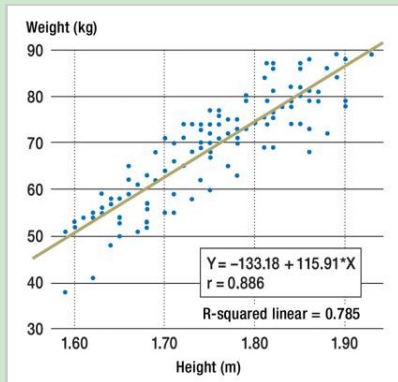
- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- **Least-squares linear regression (c)**
- Inner product spaces (d)
- Applications of inner product spaces (d)

# Least-squares linear regression

## Example

In many scientific and engineering problems, it is needed to explain some observations  $\mathbf{y}$  as a linear function of an independent variable  $\mathbf{x}$ . For instance, we may try to explain the weight of a person as a linear function of its height

$$\text{Weight} = \beta_0 + \beta_1 \text{Height}$$



# Least-squares linear regression

## Example (continued)

For each observation we have an equation

Height (m.)	Weight (kg.)
1.70	57
1.53	43
1.90	94
...	...

$$57 = \beta_0 + 1.70\beta_1$$

$$43 = \beta_0 + 1.53\beta_1$$

$$94 = \beta_0 + 1.90\beta_1$$

...

$$\begin{pmatrix} 1 & 1.70 \\ 1 & 1.53 \\ 1 & 1.90 \\ \dots & \dots \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 57 \\ 43 \\ 94 \\ \dots \end{pmatrix}$$

which is of the form

$$X\beta = \mathbf{y}$$

# Least-squares linear regression

## Least-squares regression

Each one of the observed **data points**  $(x_j, y_j)$  gives an equation. All together provide an equation system

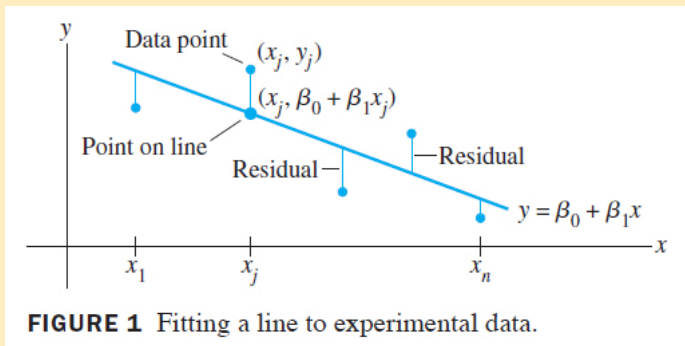
$$X\beta = \mathbf{y}$$

that is an overdetermined, linear equation system of the form  $A\mathbf{x} = \mathbf{b}$ . The matrix  $X$  is called the **system matrix** and it is related to the **independent (predictor)** variables (the height in this case). The vector  $\mathbf{y}$  is called the **observation vector** and collects the values of the **dependent (predicted)** variable (the weight in this case). The model

$$y = \beta_0 + \beta_1 x + \epsilon$$

is called the **linear regression of  $y$  on  $x$** .  $\beta_0$  and  $\beta_1$  are called the **regression coefficients**. The difference between the predicted value and the observed value for a particular observation ( $\epsilon$ ) is called the **residual** of that observation.

# Least-squares linear regression



**FIGURE 1** Fitting a line to experimental data.

The residual of the  $j$ -th observation is defined as

$$\epsilon_j = y_j - (\beta_0 + \beta_1 x_j)$$



# Least-squares linear regression

The goal of least-squares regression is to minimize

$$\sum_{j=1}^n \epsilon_j^2 = \|\mathbf{y} - \mathbf{X}\beta\|^2$$

Let's analyze this term

$$\mathbf{X}\beta = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_2 x_2 \\ \dots \\ \beta_0 + \beta_n x_n \end{pmatrix} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \dots \\ \hat{y}_n \end{pmatrix}$$

Then

$$\|\mathbf{y} - \mathbf{X}\beta\|^2 = \left\| \begin{pmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \dots \\ y_n - \hat{y}_n \end{pmatrix} \right\|^2 = \sum_{j=1}^n (y_j - \hat{y}_j)^2 = \sum_{j=1}^n \epsilon_j^2$$

# Least-squares linear regression

## Example

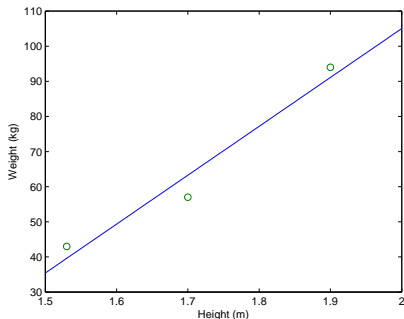
Suppose we have observed the following values of height and weight (1.70,57), (1.53,43), (1.90,94). We construct the system matrix  $X = \begin{pmatrix} 1 & 1.70 \\ 1 & 1.53 \\ 1 & 1.90 \end{pmatrix}$  and the

observation vector  $\mathbf{y} = \begin{pmatrix} 57 \\ 43 \\ 94 \end{pmatrix}$ . Now we look the normal equations

$$\begin{aligned} X\beta &= \mathbf{y} \Rightarrow X^T X \beta = X^T \mathbf{y} \\ X^T X &= \begin{pmatrix} 3.00 & 5.13 \\ 5.13 & 8.84 \end{pmatrix} & X^T \mathbf{y} &= \begin{pmatrix} 194.00 \\ 341.29 \end{pmatrix} & \hat{\beta} &= (X^T X)^{-1} X^T \mathbf{y} = \begin{pmatrix} -173.14 \\ 137.90 \end{pmatrix} \\ &&&&& \text{Weight} &= -173.39 + 139.21 \text{Height} \end{aligned}$$

# Least-squares linear regression

## Example



MATLAB:

```
X=[1 1.70; 1 1.53; 1 1.90];  
y=[57; 43; 94];  
beta=inv(X'*X)*X'*y  
x=1.5:0.01:2.00;  
yp=beta(1)+beta(2)*x;  
plot(x,yp,X(:,1),y,'o')  
xlabel('Height (m)')  
ylabel('Weight (kg)')
```

# Least-squares linear regression

## The general linear model

The linear model is not restricted to straight lines. We can use it to fit any kind of curves:

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \beta_2 f_2(x) + \dots$$

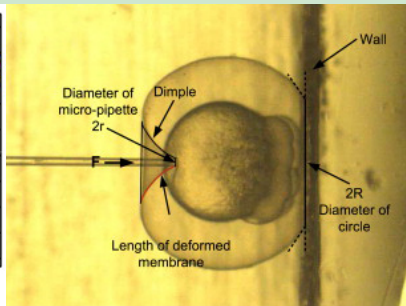
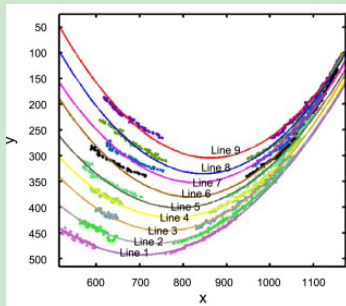
## Fitting a parabola

$$\begin{aligned} f_0(x) &= 1 \\ f_1(x) &= x \\ f_2(x) &= x^2 \end{aligned} \Rightarrow \begin{aligned} y_1 &= f_0(x_1) + \beta_1 f_1(x_1) + \beta_2 f_2(x_1) \\ y_2 &= f_0(x_2) + \beta_1 f_1(x_2) + \beta_2 f_2(x_2) \\ &\dots \\ y_n &= f_0(x_n) + \beta_1 f_1(x_n) + \beta_2 f_2(x_n) \end{aligned}$$
$$\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \dots & \dots & \dots \\ 1 & x_n & x_n^2 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_n \end{pmatrix} \Rightarrow \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

# Least-squares linear regression

## Fitting a parabola

In this example they model the deformation of the wall of the zebra fish embryo as a function of strain.



Z. Lua, P. C.Y. Chen, H. Luo, J. Nam, R. Ge, W. Lin. *Models of maximum stress and strain of zebrafish embryos under indentation*. J. Biomechanics 42 (5): 620–625 (2009)

# Least-squares linear regression

## Multivariate linear regression

The linear model is not restricted to one variable. By fitting several variables we may fit surfaces and hypersurfaces

$$y = \beta_0 f_0(x_1, x_2) + \beta_1 f_1(x_1, x_2) + \beta_2 f_2(x_1, x_2) + \dots$$

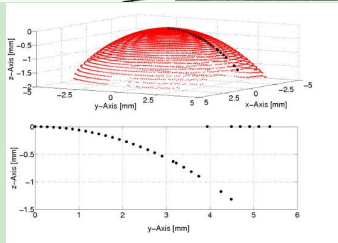
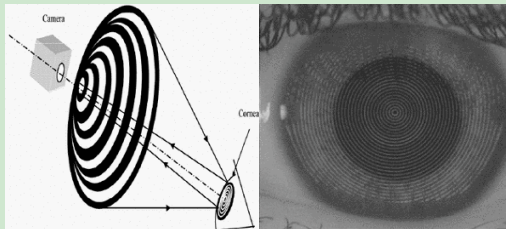
## Fitting a parabolic surface

$$\begin{aligned} f_0(x_1, x_2) &= 1 \\ f_1(x_1, x_2) &= x_1 \\ f_2(x_1, x_2) &= x_2 \\ f_3(x_1, x_2) &= x_1^2 \\ f_4(x_1, x_2) &= x_2^2 \\ f_5(x_1, x_2) &= x_1 x_2 \end{aligned} \Rightarrow X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{11}^2 & x_{12}^2 & x_{11}x_{12} \\ 1 & x_{21} & x_{22} & x_{21}^2 & x_{22}^2 & x_{21}x_{22} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & x_{n1}^2 & x_{n2}^2 & x_{n1}x_{n2} \end{pmatrix}$$

# Least-squares linear regression

## Fitting a parabolic surface

In this example they model the shape of cornea using videokeratoscopic images.



## Exercises

From Lay (3rd ed.), Chapter 6, Section 6:

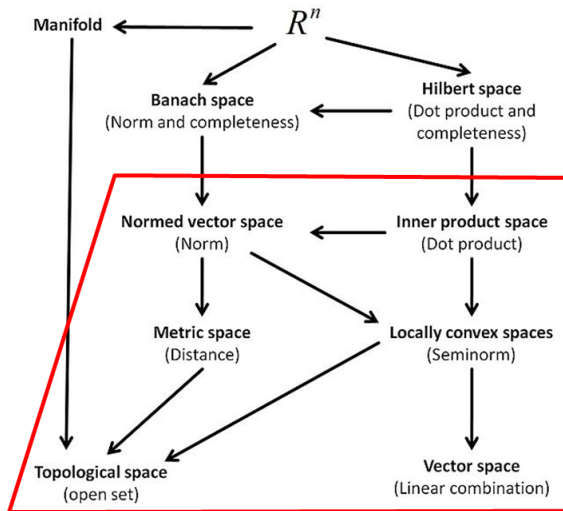
- 6.6.1
- 6.6.5
- 6.6.9
- 6.6.12 (computer)



## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- **Inner product spaces (d)**
- Applications of inner product spaces (d)

# Inner product spaces



# Inner product spaces

## Definition 7.1 (Inner product)

*An inner product in a vector space  $V$  is a function that assigns a real number to every pair of vectors  $\mathbf{u}$  and  $\mathbf{v}$ ,  $\langle \mathbf{u}, \mathbf{v} \rangle$  and that satisfies the following axioms for all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$  and all scalars  $c$ :*

- ❶  $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
- ❷  $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
- ❸  $\langle c\mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$
- ❹  $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$  and  $\langle \mathbf{u}, \mathbf{u} \rangle = 0$  iff  $\mathbf{u} = \mathbf{0}$ .

## Example

For instance in **Weighted Least Squares** (WLS) we may use an inner product in  $\mathbb{R}^2$  defined as:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 4u_1v_1 + 5u_2v_2$$

In this way we give less weight to distances in the first component with respect to distances in the second component.

# Inner product spaces

Now we have to prove that this function is effectively an inner product:

$$\textcircled{1} \quad \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$$

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= 4u_1v_1 + 5u_2v_2 && \text{[by definition]} \\ &= 4v_1u_1 + 5v_2u_2 && \text{[commutativity of scalar multiplication]} \\ &= \langle \mathbf{v}, \mathbf{u} \rangle && \text{[by definition]} \end{aligned}$$

$$\textcircled{2} \quad \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$$

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= 4(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 && \text{[by definition]} \\ &= 4u_1w_1 + 4v_1w_1 + 5u_2w_2 + 5v_2w_2 && \text{[distributivity of scalar]} \\ & && \text{[multiplication/addition]} \\ &= 4u_1w_1 + 5u_2w_2 + 4v_1w_1 + 5v_2w_2 && \text{[commutativity]} \\ & && \text{[of scalar addition]} \\ &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle && \text{[by definition]} \end{aligned}$$

# Inner product spaces

$$\textcircled{3} \quad \langle c\mathbf{u}, \mathbf{v} \rangle = c \langle \mathbf{u}, \mathbf{v} \rangle$$

$$\begin{aligned} \langle c\mathbf{u}, \mathbf{v} \rangle &= 4cu_1v_1 + 5cu_2v_2 && \text{[by definition]} \\ &= c4v_1u_1 + c5v_2u_2 && \text{[commutativity of scalar multiplication]} \\ &= c(4v_1u_1 + 5v_2u_2) && \text{[distributivity of scalar multiplication]} \\ &= c \langle \mathbf{u}, \mathbf{v} \rangle && \text{[by definition]} \end{aligned}$$

$$\textcircled{4} \quad \langle \mathbf{u}, \mathbf{u} \rangle \geq 0 \text{ and } \langle \mathbf{u}, \mathbf{u} \rangle = 0 \text{ iff } \mathbf{u} = \mathbf{0}.$$

$$\textcircled{1} \quad \langle \mathbf{u}, \mathbf{u} \rangle \geq 0$$

$$\langle \mathbf{u}, \mathbf{u} \rangle = 4u_1^2 + 5u_2^2 \quad \text{[by definition]}$$

which is obviously larger than 0.

$$\textcircled{2} \quad \langle \mathbf{u}, \mathbf{u} \rangle = 0 \text{ iff } \mathbf{u} = \mathbf{0}.$$

$$\langle \mathbf{u}, \mathbf{u} \rangle = 0 \Leftrightarrow 4u_1^2 + 5u_2^2 = 0 \Leftrightarrow u_1 = u_2 = 0$$

# Inner product spaces

## Example

Consider two vectors  $p$  and  $q$  the vector space of polynomials of degree  $n$  ( $\mathbb{P}_n$ ). Let  $t_0, t_1, \dots, t_n$  be  $n$  distinct real numbers and  $K$  any scalar. The inner product between  $p$  and  $q$  is defined as

$$\langle p, q \rangle = K (p(t_0)q(t_0) + p(t_1)q(t_1) + \dots + p(t_n)q(t_n))$$

Axioms 1-3 are easy to check. Let's prove Axiom 4

$$\textcircled{4} \quad \langle p, p \rangle \geq 0 \text{ and } \langle p, p \rangle = 0 \text{ iff } p = 0.$$

$$\textcircled{1} \quad \langle p, p \rangle \geq 0$$

$$\langle p, p \rangle = K (p^2(t_0) + p^2(t_1) + \dots + p^2(t_n)) \quad [\text{by definition}]$$

which is obviously larger than 0.

$$\textcircled{2} \quad \langle p, p \rangle = 0 \text{ iff } p = 0.$$

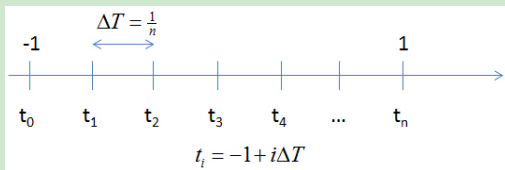
$$\begin{aligned} \langle p, p \rangle = 0 &\Leftrightarrow K (p^2(t_0) + p^2(t_1) + \dots + p^2(t_n)) \Leftrightarrow \\ &p(t_0) = p(t_1) = \dots = p(t_n) = 0 \end{aligned}$$

But  $p$  is a polynomial of degree  $n$  so, at most, it can have  $n$  zeros. However, the previous condition requires the polynomial to vanish at  $n+1$  points. This is impossible unless  $p = 0$ .

# Inner product spaces

## Example

Consider two vectors  $p$  and  $q$  the vector space of polynomials of degree  $n$  ( $\mathbb{P}_n$ ). Assume that we regularly space the  $n + 1$  points in the interval  $[-1, 1]$



and set  $K = \Delta T$ , then the inner product between the two polynomials becomes

$$\langle p, q \rangle = (p(t_0)q(t_0) + p(t_1)q(t_1) + \dots + p(t_n)q(t_n)) \Delta T = \sum_{i=0}^n p(t_i)q(t_i) \Delta T$$

Making  $\Delta T$  tend to 0, the inner product becomes

$$\langle p, q \rangle = \int_{-1}^1 p(t)q(t)dt$$

# Inner product spaces

Legendre polynomials are orthogonal polynomials in the interval  $[-1, 1]$

$$P_0(x) = 1$$

$$P_1(x) = x$$

$$P_2(x) = \frac{1}{2}(3x^2 - 1)$$

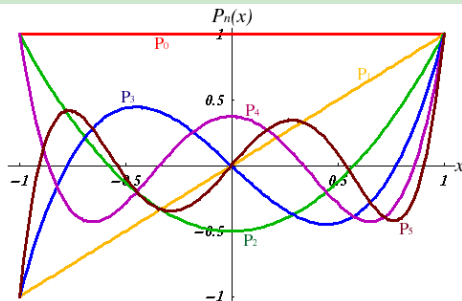
$$P_3(x) = \frac{1}{2}(5x^3 - 3x)$$

$$P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3)$$

$$P_5(x) = \frac{1}{8}(63x^5 - 70x^3 + 15x)$$

$$P_6(x) = \frac{1}{16}(231x^6 - 315x^4 + 105x^2 - 5)$$

$$P_7(x) = \frac{1}{16}(429x^7 - 693x^5 + 315x^3 - 35x)$$



Legendre polynomials are very useful for regression of high-order polynomials as shown in next slide.



# Inner product spaces

The screenshot shows a web browser window with the URL [www.ncbi.nlm.nih.gov/pubmed/?term=legendre+polynomials](http://www.ncbi.nlm.nih.gov/pubmed/?term=legendre+polynomials). The page is the PubMed search results interface. At the top, there's a search bar with 'legendre polynomials' entered. Below the search bar, the results are displayed in a list format. The first result is titled 'Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes.' and is by Borquis RR, Neto FR, Baldi F, Hurtado-Lugo N, de Camargo GM, Muñoz-Berrocal M, Tonhati H. The second result is 'Analysis of egg production in layer chickens using a random regression model with genomic relationships' by Wolc A, Arango J, Settler P, Fulton JE, O'Sullivan NP, Preisinger R, Fernando R, Garrick DJ, Dekkers JC. The third result is 'The mathematical modeling of the lactation curve for dairy traits of the donkey (Equus asinus)' by Bordonaro S, Dimauro C, Criscione A, Marietta D, Macciotta NP. The fourth result is 'Comparison of random regression models to estimate genetic parameters for milk production in Guzerat (Bos indicus) cows.' The right sidebar contains sections for 'Titles with your search terms' and '14 free full-text articles in PubMed Central'. The bottom of the page shows a 'Find related data' section with a database dropdown menu.

www.ncbi.nlm.nih.gov/pubmed/?term=legendre+polynomials

NCBI Resources How To Sign in to NCBI

PubMed US National Library of Medicine National Institutes of Health

Search legendre polynomials

RSS Save search Advanced Help

Show additional filters

Article types  
More ...

Text availability  
Abstract available  
Free full text available  
Full text available

Publication dates  
5 years  
10 years  
Custom range...

Species  
Humans  
Other Animals

Clear all

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to: Filters: Manage Filters

Results: 1 to 20 of 183

<< First < Prev Page 1 of 10 Next > Last >>

☐ Multiple-trait random regression models for the estimation of genetic parameters for milk, fat, and protein yield in buffaloes.  
1. Borquis RR, Neto FR, Baldi F, Hurtado-Lugo N, de Camargo GM, Muñoz-Berrocal M, Tonhati H. J Dairy Sci. 2013 Jul 4. doi:10.3168/jds.2012-6023. [Epub ahead of print]  
PMID: 23831097 [PubMed - as supplied by publisher]  
[Related citations](#)

☐ Analysis of egg production in layer chickens using a random regression model with genomic relationships.  
2. Wolc A, Arango J, Settler P, Fulton JE, O'Sullivan NP, Preisinger R, Fernando R, Garrick DJ, Dekkers JC. Poult Sci. 2013 Jun;92(6):1486-91. doi: 10.3382/ps.2012-02882.  
PMID: 23687143 [PubMed - indexed for MEDLINE]  
[Related citations](#)

☐ The mathematical modeling of the lactation curve for dairy traits of the donkey (Equus asinus).  
3. Bordonaro S, Dimauro C, Criscione A, Marietta D, Macciotta NP. J Dairy Sci. 2013 Jun;96(6):4005-14. doi: 10.3168/jds.2012-6180. Epub 2013 Apr 12.  
PMID: 23587386 [PubMed - in process]  
[Related citations](#)

☐ Comparison of random regression models to estimate genetic parameters for milk production in Guzerat (Bos indicus) cows.  
4.

Titles with your search terms

Random regression models on Legendre polynomials to estimate [J Anim Breed Genet. 2010]  
Comparison of random regression models with Legendre polynomials and I [J Dairy Sci. 2008]  
Genetic parameters of Legendre polynomials for first parity lactation curves. [J Dairy Sci. 2000]  
[See more...](#)

14 free full-text articles in PubMed Central

Box-Cox Transformation and Random Regression Models for Fecal egg Count [Front Genet. 2012]  
Genome-wide association studies for bivariate sparse longitudinal data. [Hum Hered. 2011]  
A Spectral Finite Element Approach to Modeling Soft Soil [Comput Methods Appl Mech Eng. 2011]  
[See all \(14\)...](#)

Find related data

Database: Select

# Length, distance and orthogonality

## Length, distance and orthogonality

The **length** of a vector  $\mathbf{u}$  in an inner product space is defined in the standard way

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

Similarly, the **distance between two vectors**  $\mathbf{u}$  and  $\mathbf{v}$  is defined as

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

Finally, two vectors  $\mathbf{u}$  and  $\mathbf{v}$  are said to be **orthogonal** iff

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0$$

# Length, distance and orthogonality

## Example

In the vector space of polynomials in the interval  $[0, 1]$ ,  $\mathbb{P}[0, 1]$ , let's define the inner product

$$\langle p, q \rangle = \int_0^1 p(t)q(t)dt$$

What is the length of the vector  $p(t) = 3t^2$ ?

Solution

$$\begin{aligned}\|p\| &= \sqrt{\langle p, p \rangle} = \sqrt{\int_0^1 p^2(t)dt} = \sqrt{\int_0^1 (3t^2)^2 dt} = \sqrt{\int_0^1 9t^4 dt} \\ &= \sqrt{9 \left. \frac{t^5}{5} \right|_0^1} = \sqrt{9 \left( \frac{1}{5} - 0 \right)} = \frac{3}{\sqrt{5}}\end{aligned}$$

# Gram-Schmidt orthogonalization

## Example

Gram-Schmidt is applied in the standard way. For instance, find an orthogonal basis of  $\mathbb{P}_2[-1, 1]$ . A basis that spans this space is

$$\{1, t, t^2\}$$

Let's orthogonalize it

$$p_0(t) = 1$$

$$p_1(t) = t - \frac{\langle t, p_0(t) \rangle}{\|p_0\|^2} p_0(t) = t - \frac{\int_{-1}^1 t dt}{\int_{-1}^1 dt} 1 = t - \frac{0}{2} 1 = t$$

$$\begin{aligned} p_2(t) &= t^2 - \frac{\langle t^2, p_0(t) \rangle}{\|p_0\|^2} p_0(t) - \frac{\langle t^2, p_1(t) \rangle}{\|p_1\|^2} p_1(t) \\ &= t^2 - \frac{\int_{-1}^1 t^2 dt}{\int_{-1}^1 dt} - \frac{\int_{-1}^1 t^2 t dt}{\int_{-1}^1 t^2 dt} t = t^2 - \frac{\frac{2}{3}}{2} = t^2 - \frac{1}{3} \end{aligned}$$

In Slide 97 we proposed the Legendre polynomial of degree 2 to be  $P_2(t) = \frac{1}{2}(3t^2 - 1)$ , we can easily show that  $P_2(t) = \frac{3}{2}p_2(t)$ . Consequently, if  $p_2(t)$  is orthogonal to  $p_0(t)$  and  $p_1(t)$  so is  $P_2(t)$ .

# Best approximation

## Example

What is the best approximation in  $\mathbb{P}_2[-1, 1]$  of  $p(t) = t^3$ ?

### Solution

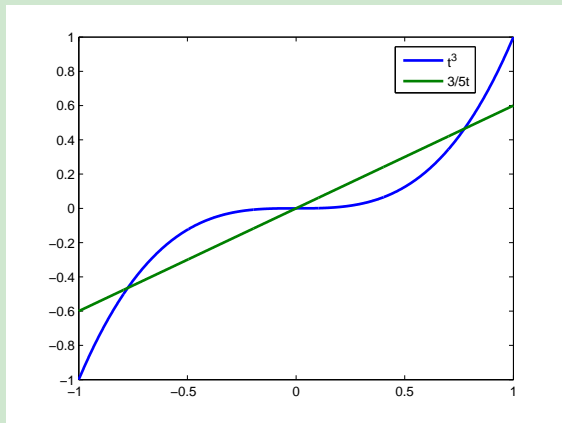
We know the answer is the orthogonal projection of  $p(t)$  onto  $\mathbb{P}_2[-1, 1]$ . An orthogonal basis of  $\mathbb{P}_2[-1, 1]$  is  $\{1, t, t^2 - \frac{1}{3}\}$ . Therefore, this projection can be calculated as

$$\hat{p}(t) = \text{Proj}_{\mathbb{P}_2[-1,1]} \{p(t)\} = \frac{\langle p, p_0 \rangle}{\|p_0\|^2} p_0(t) + \frac{\langle p, p_1 \rangle}{\|p_1\|^2} p_1(t) + \frac{\langle p, p_2 \rangle}{\|p_2\|^2} p_2(t)$$

Let's perform these calculations:

$$\begin{array}{l|l} \langle p, p_0(t) \rangle = \int_{-1}^1 t^3 dt = 0 & \|p_0\|^2 = \int_{-1}^1 dt = 2 \\ \langle p, p_1(t) \rangle = \int_{-1}^1 t^3 t dt = \frac{2}{5} & \|p_1\|^2 = \int_{-1}^1 t^2 dt = \frac{2}{3} \\ \langle p, p_2(t) \rangle = \int_{-1}^1 t^3 (t^2 - \frac{1}{3}) dt = 0 & \|p_2\|^2 = \int_{-1}^1 (t^2 - \frac{1}{3})^2 dt = \frac{8}{45} \end{array}$$
$$\hat{p}(t) = \frac{0}{2} + \frac{\frac{2}{5}}{\frac{2}{3}} t + \frac{0}{\frac{8}{45}} (t^2 - \frac{1}{3}) = \frac{3}{5} t$$

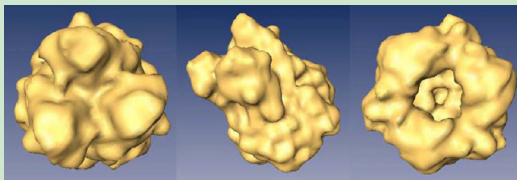
# Best approximation



# Best approximation

## Example

In this example we exploited the best approximation property of orthogonal wavelets to speed-up and make more robust angular alignments of projections in 3D Electron Microscopy.



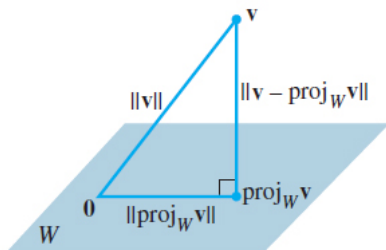
C.O.S.Sorzano, S. Jonic, C. El-Bez, J.M. Carazo, S. De Carlo, P. Thévenaz, M. Unser. *A multiresolution approach to orientation assignment in 3-D electron microscopy of single particles*. Journal of Structural Biology 146(3): 381-392 (2004, cover article)

# Pythagorean theorem

## Theorem 7.1 (Pythagorean theorem)

Given any vector  $\mathbf{v}$  in an inner product space  $V$  and a subspace of it  $W \subseteq V$  we have

$$\|\mathbf{v}\|^2 = \|\text{Proj}_W\{\mathbf{v}\}\|^2 + \|\mathbf{v} - \text{Proj}_W\{\mathbf{v}\}\|^2$$



**FIGURE 2**

The hypotenuse is the longest side.



# The Cauchy-Schwarz inequality

## Theorem 7.2 (The Cauchy-Schwarz inequality)

For all  $\mathbf{u}, \mathbf{v} \in V$  it is verified

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

Proof

If  $\mathbf{u} = \mathbf{0}$ , then

$$|\langle \mathbf{0}, \mathbf{v} \rangle| = 0 \quad \text{and} \quad \|\mathbf{0}\| \|\mathbf{v}\| = 0 \|\mathbf{v}\| = 0$$

So the inequality becomes an equality.

If  $\mathbf{u} \neq \mathbf{0}$ , then consider  $W = \text{Span}\{\mathbf{u}\}$  and

$$\|\text{Proj}_W\{\mathbf{v}\}\| = \left\| \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} \mathbf{u} \right\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|^2} \|\mathbf{u}\| = \frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|}$$

But by the Pythagorean Theorem (Theorem 7.1) we have  $\|\text{Proj}_W\{\mathbf{v}\}\| \leq \|\mathbf{v}\|$ .  
Consequently,

$$\frac{|\langle \mathbf{v}, \mathbf{u} \rangle|}{\|\mathbf{u}\|} \leq \|\mathbf{v}\| \Rightarrow |\langle \mathbf{v}, \mathbf{u} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\| \quad (q.e.d.)$$

# The Triangle inequality

## Theorem 7.3 (The Triangle inequality)

For all  $\mathbf{u}, \mathbf{v} \in V$  it is verified

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Proof

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle && \text{[By definition]} \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle + 2 \langle \mathbf{u}, \mathbf{v} \rangle && \text{[Properties of inner product]} \\ &\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| && \langle \mathbf{u}, \mathbf{v} \rangle \leq |\langle \mathbf{u}, \mathbf{v} \rangle| \\ &\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| && \text{Cauchy-Schwarz} \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \\ &\Rightarrow \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| && \text{[Taking square root]} \end{aligned}$$

(q.e.d.)

## Exercises

From Lay (3rd ed.), Chapter 6, Section 7:

- 6.7.1
- 6.7.13
- 6.7.16
- 6.7.18

## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)

# Weighted Least Squares

## Weighted Least Squares

Let us assume we have a table of collected data and we want to fit a least squares model. However, we want to give more importance to some observations because we are more confident about them or they are more important. We encode the importance as a weight value (the larger the weight, the more importance the observation has)

<b>X</b>	<b>Y</b>	<b>W</b>
$x_1$	$y_1$	$w_1$
$x_2$	$y_2$	$w_2$
$x_3$	$y_3$	$w_3$
...	...	...

Let us call  $\hat{y}_j$  the prediction of the model for the  $j$ -th observation and  $\epsilon_j$  the committed error

$$y_j = \hat{y}_j + \epsilon_j$$

# Weighted Least Squares

The goal is now to minimize the weighted sum of square errors

$$\sum_{j=1}^n (w_j \epsilon_j)^2 = \sum_{j=1}^n (w_j (y_j - \hat{y}_j))^2 = \sum_{j=1}^n (w_j y_j - w_j \hat{y}_j)^2$$

Let us collect all observed values into a vector  $\mathbf{y}$  and do analogously with the predictions  $\hat{\mathbf{y}}$ . Let us define the diagonal matrix

$$W = \begin{pmatrix} w_1 & 0 & 0 & \dots & 0 \\ 0 & w_2 & 0 & \dots & 0 \\ 0 & 0 & w_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & w_n \end{pmatrix}$$

Then, the previous objective function becomes

$$\sum_{j=1}^n (w_j y_j - w_j \hat{y}_j)^2 = \|W\mathbf{y} - W\hat{\mathbf{y}}\|^2$$

# Weighted Least Squares

Now, suppose that  $\hat{\mathbf{y}}$  is calculated from the columns of a matrix  $A$ , that is,  $\hat{\mathbf{y}} = A\mathbf{x}$ . The objective function becomes

$$\sum_{j=1}^n (w_j y_j - w_j \hat{y}_j)^2 = \|W\mathbf{y} - WA\mathbf{x}\|^2$$

The minimum of this objective function is attained for  $\hat{\mathbf{x}}$  that is the least-squares solution of the equation system

$$WA\mathbf{x} = W\mathbf{y}$$

The normal equations of the problem are

$$(WA)^T WA\mathbf{x} = (WA)^T W\mathbf{y}$$

# Weighted Least Squares

## Example

In this work they used Weighted Least Squares to calibrate a digital system to measure maximum respiratory pressures.



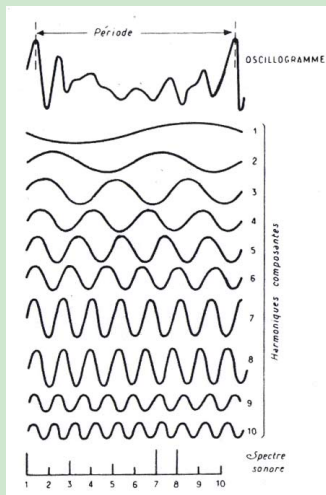
J.L. Ferreira, F.H. Vasconcelos, C.J. Tierra-Criollo. *A Case Study of Applying Weighted Least Squares to Calibrate a Digital Maximum Respiratory Pressures Measuring System*. Applied Biomedical Engineering, Chapter 18 (2011)



# Fourier Series

## Example

Fourier tools are, maybe, the most widespread tool to analyze signals and its frequency components. Fourier decomposition states that any signal can be obtained by summing sine waves of different amplitude, phase and frequency.



## Theorem 8.1

Consider the vector space of continuous functions in the interval  $[0, 2\pi]$ ,  $C[0, 2\pi]$ . The set

$$S = \{1, \cos(t), \sin(t), \cos(2t), \sin(2t), \dots, \cos(Nt), \sin(Nt)\}$$

is orthogonal with respect to the inner product defined as

$$\langle f(t), g(t) \rangle = \int_0^{2\pi} f(t)g(t)dt$$

Proof

$$\begin{aligned}\langle \cos(nt), \cos(mt) \rangle &= \int_0^{2\pi} \cos(nt) \cos(mt) dt \\ &= \int_0^{2\pi} \frac{1}{2} (\cos((n+m)t) + \cos((n-m)t)) dt \\ &= \frac{1}{2} \left( \frac{\sin((n+m)t)}{n+m} + \frac{\sin((n-m)t)}{n-m} \right) \Big|_0^{2\pi} \\ &= 0\end{aligned}$$

where we have used  $\cos(A)\cos(B) = \frac{1}{2}(\cos(A+B) + \cos(A-B))$ .

Analogously we could prove that

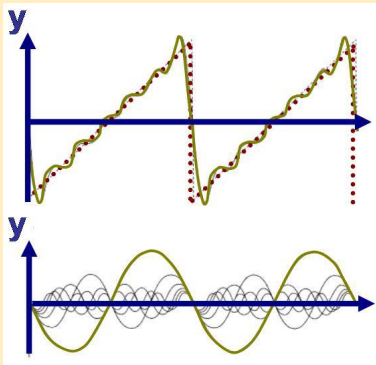
$$\begin{aligned}\langle \cos(nt), \sin(mt) \rangle &= 0 \\ \langle \cos(nt), 1 \rangle &= 0 \\ \langle \sin(nt), 1 \rangle &= 0 \\ \|\cos(nt)\|^2 &= \pi \\ \|\sin(nt)\|^2 &= \pi \\ \|1\|^2 &= 2\pi\end{aligned}$$

# Fourier Series

## Theorem 8.2 (Fourier series)

Given any function  $f(t) \in C[0, 2\pi]$ ,  $f(t)$  can be approximated as closely as desired by a sum of the form simply by orthogonally projecting it onto  $W = \text{Span}\{S\}$

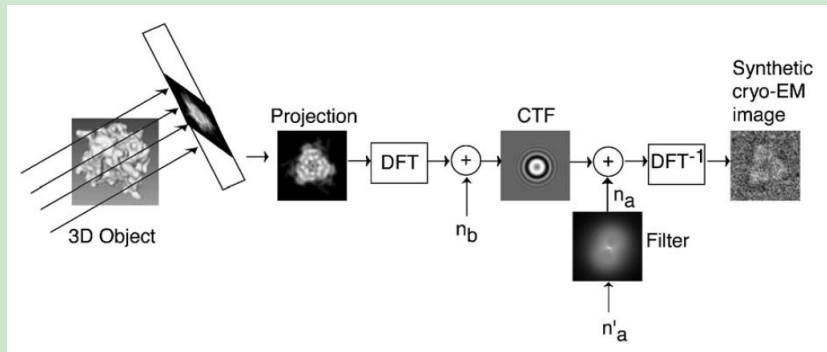
$$f(t) \approx \text{Proj}_W\{f(t)\} = \frac{\langle f(t), 1 \rangle}{\|1\|^2} + \sum_{n=1}^N \left( \frac{\langle f(t), \cos(nt) \rangle}{\|\cos(nt)\|^2} \cos(nt) + \frac{\langle f(t), \sin(nt) \rangle}{\|\sin(nt)\|^2} \sin(nt) \right)$$



# Fourier Series

## Example

In this work we used Fourier space to simulate and to align electron microscopy images



S. Jonic, C.O.S.Sorzano, P. Thévenaz, C. El-Bez, S. De Carlo, M. Unser. *Spline-Based image-to-volume registration for three-dimensional electron microscopy*. Ultramicroscopy, 103: 303-317 (2005)

## Exercises

From Lay (3rd ed.), Chapter 6, Section 8:

- 6.8.1
- 6.8.6
- 6.8.8
- 6.8.11

## 7 Orthogonality and least squares

- Inner product, length and orthogonality (a)
- Orthogonal sets, bases and matrices (a)
- Orthogonal projections (b)
- Gram-Schmidt orthogonalization (b)
- Least squares (c)
- Least-squares linear regression (c)
- Inner product spaces (d)
- Applications of inner product spaces (d)