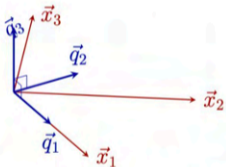




## Section 6.4 : The Gram-Schmidt Process

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



Vectors  $\vec{x}_1, \vec{x}_2, \vec{x}_3$  are given linearly independent vectors. We wish to construct an orthonormal basis  $\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}$  for the space that they span.

## Topics and Objectives

### Topics

1. Gram Schmidt Process
2. The  $QR$  decomposition of matrices and its properties

### Learning Objectives

1. Apply the iterative Gram Schmidt Process, and the QR decomposition, to construct an orthogonal basis.
2. Compute the  $QR$  factorization of a matrix.

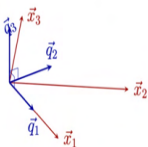
**Motivating Question** The vectors below span a subspace  $W$  of  $\mathbb{R}^4$ . Identify an orthogonal basis for  $W$ .

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

Week	Dates	Lecture	Studio	Web	Thu	Fri
1	5/8 - 5/12	5.1	WS1.1	1.2	WS1.2	1.3
2	5/15 - 5/19	Break	WS1.3	1.4	WS1.4	1.5
3	5/22 - 5/26	5.7	WS1.5,7	1.8	WS1.8	1.9
4	5/29 - 6/2	5.8,2.1	WS1.8,2.1	Exam 1 Review	Cancelled	2.2
5	6/5 - 6/9	2.3,2.4	WS2.2,4	2.5	WS2.5	2.8
6	6/12 - 6/16	2.9	WS2.8	2.9,3.1	WS2.9,3.1	3.2
7	6/19 - 6/23	3.3	WS2.2	4.9	WS3.8,4.9	5.1
8	6/26 - 6/30	5.2	WS5.5,2	Exam 2 Review	Cancelled	5.3
9	7/3 - 7/7	5.3	WS5.3	5.5	WS5.5	6.1
10	7/10 - 7/15	6.2,6.3	WS6.1	6.2	WS6.2	6.3
11	7/18 - 7/22	Break	Break	Break	Break	Break
12	7/25 - 7/29	6.4	WS6.3	6.4,6.5	WS6.4	6.5
13	7/31 - 8/4	6.6	WS6.5,6.6	Exam 3 Review	Cancelled	PageRank
14	8/6 - 8/12	7.1	WS7.1,6.8	7.2	WS7.2,7.2	7.3
15	8/13 - 8/19	7.3,7.4	WS7.3	7.4	WS7.4	7.4
16	8/22 - 8/26	Last Lecture	Last Studio	Reading Period		
17	8/29 - 9/2	Final Exams	MATH 155A Common Final Exam	Tuesday, April 30th at 6:00pm		

Section 6.4: The Gram-Schmidt Process

Chapter 6: Orthogonality and Least Squares  
Math 155A Linear Algebra



Vectors  $\vec{x}_1, \vec{x}_2, \vec{x}_3$  are given linearly independent vectors. We wish to construct an orthonormal basis  $\{\vec{q}_1, \vec{q}_2, \vec{q}_3\}$  for the space that they span.

Topics

1. Gram Schmidt Process
2. The QR decomposition of matrices and its properties

Learning Objectives

1. Apply the iterative Gram Schmidt Process, and the QR decomposition, to construct an orthogonal basis.
2. Compute the QR factorization of a matrix.

**Motivating Question** The vectors below span a subspace  $W$  of  $\mathbb{R}^3$ . Identify an orthogonal basis for  $W$ .

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \vec{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

THEOREM 8

The Orthogonal Decomposition Theorem

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

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (1)$$

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

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p \quad (2)$$

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

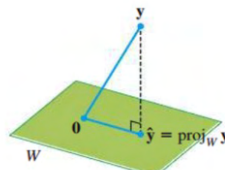


FIGURE 1

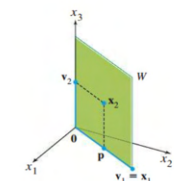


FIGURE 1 Construction of an orthogonal basis  $\{v_1, v_2\}$ .

$$\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

THEOREM 11

The Gram-Schmidt Process

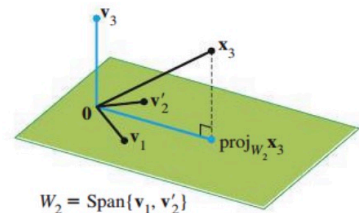
Given a basis  $\{x_1, \dots, x_p\}$  for a nonzero subspace  $W$  of  $\mathbb{R}^n$ , define

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

Then  $\{v_1, \dots, v_p\}$  is an orthogonal basis for  $W$ . In addition

$$\text{Span}\{v_1, \dots, v_k\} = \text{Span}\{x_1, \dots, x_k\} \quad \text{for } 1 \leq k \leq p \quad (1)$$

FIGURE 2 The construction of  $v_3$  from  $x_3$  and  $W_2$ .



Q: Given this basis  $B$  of a plane  $W = \text{span}\{\vec{x}_1, \vec{x}_2\}$

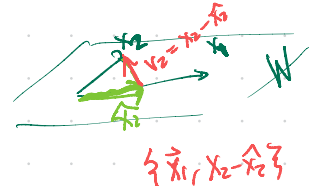
1.  $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$

$\uparrow$   
 $\vec{x}_1$   
 $\uparrow$   
 $\vec{x}_2$

Find  $\{v_1, v_2\}$  an orthogonal basis of  $W$

①  $\text{span}\{v_1, v_2\} = \text{span}\{\vec{x}_1, \vec{x}_2\} = W$

②  $v_1 \cdot v_2 = 0$  Check:  $\left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix} \right\}$



Set  $v_1 = \vec{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$   
 $\vec{v}_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} - \frac{30}{10} \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} - \begin{bmatrix} 9 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$   
 $v_2 = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$   
 dot prod =  $-3 + 0 + 30 = 27$

WANT  $\{v_1, v_2, v_3\}$  orthogonal basis  
 For  $\text{Span}\{x_1, x_2, x_3\} = W$ .

Example

The vectors below span a subspace  $W$  of  $\mathbb{R}^3$ . Construct an orthogonal basis for  $W$ .

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Step 1: set  $v_1 = \vec{x}_1$

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Step 2: Find  $v_2$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$$

The Gram-Schmidt Process

Given a basis  $\{\vec{x}_1, \dots, \vec{x}_p\}$  for a subspace  $W$  of  $\mathbb{R}^n$ , iteratively define

$$\begin{aligned} \vec{v}_1 &= \vec{x}_1 \\ \vec{v}_2 &= \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 = x_2 - \hat{x}_2 \\ \vec{v}_3 &= \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 = x_3 - \hat{x}_3 \\ &\vdots \\ \vec{v}_p &= \vec{x}_p - \frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \dots - \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}} \vec{v}_{p-1} = x_p - \hat{x}_p \end{aligned}$$

Then,  $\{\vec{v}_1, \dots, \vec{v}_p\}$  is an orthogonal basis for  $W$ .

Orthogonal basis of  $W$

$$\{v_1, v_2, v_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \end{bmatrix}, \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \end{bmatrix} \right\}$$

Step 3: Find  $v_3$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$

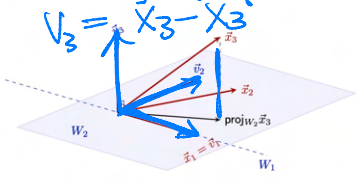
$$= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \end{bmatrix}}{\begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \end{bmatrix} \cdot \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \end{bmatrix}} \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

alt (Sally's way)

$$\{v_1, v_2, v_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \right\}$$

Geometric Interpretation

Suppose  $\vec{x}_1, \vec{x}_2, \vec{x}_3$  are linearly independent vectors in  $\mathbb{R}^3$ . We wish to construct an orthogonal basis for the space that they span.



We construct vectors  $\vec{v}_1, \vec{v}_2, \vec{v}_3$ , which form our orthogonal basis.  
 $W_1 = \text{Span}\{\vec{v}_1\}$ ,  $W_2 = \text{Span}\{\vec{v}_1, \vec{v}_2\}$ .

$$= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{12} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} - \begin{bmatrix} -1/2 \\ 1/6 \\ 1/6 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \end{bmatrix} = v_3$$



Q do G-S & normalize.

$R = Q^T A$

3x2 Q 2x2 R

**Definition**  
A set of vectors form an **orthonormal basis** if the vectors are mutually orthogonal and have unit length.

**Theorem**  
Any  $m \times n$  matrix  $A$  with linearly independent columns has the QR factorization  $A = QR$   
where  
1.  $Q$  is  $m \times n$ , its columns are an orthonormal basis for Col  $A$ .  
2.  $R$  is  $n \times n$ , upper triangular, with positive entries on its diagonal, and the length of the  $j^{th}$  column of  $R$  is equal to the length of the  $j^{th}$  column of  $A$ .

**Example**  
The two vectors below form an orthogonal basis for a subspace  $W$ . Obtain an orthonormal basis for  $W$ .

$v_1 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}$

From G-S. Find  $v_1, v_2$  & normalize

$v_1 = x_1 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$

$v_2 = x_2 - \hat{x}_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$

$= \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} - \frac{0}{13} \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} = \hat{x}_2$

$R = Q^T A = \begin{bmatrix} 3/\sqrt{13} & 2/\sqrt{13} & 0 \\ -2/\sqrt{13} & 3/\sqrt{13} & 1/\sqrt{13} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 2 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{13} & 0 \\ 0 & \sqrt{14} \end{bmatrix}$

G-S  $\left\{ \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix} \right\}$  now normalize

$\left\{ \begin{bmatrix} 3/\sqrt{13} \\ 2/\sqrt{13} \\ 0 \end{bmatrix}, \begin{bmatrix} -2/\sqrt{14} \\ 3/\sqrt{14} \\ 1/\sqrt{14} \end{bmatrix} \right\}$   
orthonormal basis for  $W$

Examples (if time permits)

Construct the QR decomposition for  $A$ .

a)  $A = \begin{bmatrix} 3 & -2 \\ 2 & 3 \\ 0 & 1 \end{bmatrix}$

b)  $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \left\{ v_1, v_2, v_3 \right\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} \right\}$

reason 1 ( $R = Q^T A$ )

$A = QR$

$\Rightarrow Q^T A = Q^T QR$

$= I R$

$= R$

$\|v_1\| = \sqrt{4} = 2 \quad \|v_2\| = \sqrt{12} = 2\sqrt{3} \quad \|v_3\| = \sqrt{6}$

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

$R = \begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix} = Q^T A = \begin{bmatrix} 2 & 3/2 \\ 0 & \sqrt{3}/2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \sqrt{3} & 1 \\ 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} = R$

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

## 6.4 EXERCISES

In Exercises 1–6, the given set is a basis for a subspace  $W$ . Use the Gram–Schmidt process to produce an orthogonal basis for  $W$ .

1.  $\begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$

2.  $\begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \\ -7 \end{bmatrix}$

3.  $\begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix}$

4.  $\begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 14 \\ -7 \end{bmatrix}$

5.  $\begin{bmatrix} 1 \\ -4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 \\ -7 \\ -4 \\ 1 \end{bmatrix}$

6.  $\begin{bmatrix} 3 \\ -1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -5 \\ 9 \\ -9 \\ 3 \end{bmatrix}$

7. Find an orthonormal basis of the subspace spanned by the vectors in Exercise 3.  
 8. Find an orthonormal basis of the subspace spanned by the vectors in Exercise 4.

Find an orthogonal basis for the column space of each matrix in Exercises 9–12.

9.  $\begin{bmatrix} 3 & -5 & 1 \\ 1 & 1 & 1 \\ -1 & 5 & -2 \\ 3 & -7 & 8 \end{bmatrix}$

10.  $\begin{bmatrix} -1 & 6 & 6 \\ 3 & -8 & 3 \\ 1 & -2 & 6 \\ 1 & -4 & -3 \end{bmatrix}$

11.  $\begin{bmatrix} 1 & 2 & 5 \\ -1 & 1 & -4 \\ -1 & 4 & -3 \\ 1 & -4 & 7 \\ 1 & 2 & 1 \end{bmatrix}$

12.  $\begin{bmatrix} 1 & 3 & 5 \\ -1 & -3 & 1 \\ 0 & 2 & 3 \\ 1 & 5 & 2 \\ 1 & 5 & 8 \end{bmatrix}$

In Exercises 13 and 14, the columns of  $Q$  were obtained by applying the Gram–Schmidt process to the columns of  $A$ . Find an upper triangular matrix  $R$  such that  $A = QR$ . Check your work.

13.  $A = \begin{bmatrix} 5 & 9 \\ 1 & 7 \\ -3 & -5 \\ 1 & 5 \end{bmatrix}, Q = \begin{bmatrix} 5/6 & -1/6 \\ 1/6 & 5/6 \\ -3/6 & 1/6 \\ 1/6 & 3/6 \end{bmatrix}$

14.  $A = \begin{bmatrix} -2 & 3 \\ 5 & 7 \\ 2 & -2 \\ 4 & 6 \end{bmatrix}, Q = \begin{bmatrix} -2/7 & 5/7 \\ 5/7 & 2/7 \\ 2/7 & -4/7 \\ 4/7 & 2/7 \end{bmatrix}$

15. Find a QR factorization of the matrix in Exercise 11.  
 16. Find a QR factorization of the matrix in Exercise 12.

In Exercises 17 and 18, all vectors and subspaces are in  $\mathbb{R}^n$ . Mark each statement True or False. Justify each answer.

17. a. If  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  is an orthogonal basis for  $W$ , then multiplying  $\mathbf{v}_3$  by a scalar  $c$  gives a new orthogonal basis  $\{\mathbf{v}_1, \mathbf{v}_2, c\mathbf{v}_3\}$ .  
 b. The Gram–Schmidt process produces from a linearly independent set  $\{\mathbf{x}_1, \dots, \mathbf{x}_p\}$  an orthogonal set  $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  with the property that for each  $k$ , the vectors  $\mathbf{v}_1, \dots, \mathbf{v}_k$  span the same subspace as that spanned by  $\mathbf{x}_1, \dots, \mathbf{x}_k$ .  
 c. If  $A = QR$ , where  $Q$  has orthonormal columns, then  $R = Q^T A$ .  
 18. a. If  $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$  with  $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$  linearly independent, and if  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  is an orthogonal set in  $W$ , then  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  is a basis for  $W$ .

19. Suppose  $A = QR$ , where  $Q$  is  $m \times n$  and  $R$  is  $n \times n$ . Show that if the columns of  $A$  are linearly independent, then  $R$  must be invertible. [Hint: Study the equation  $R\mathbf{x} = \mathbf{0}$  and use the fact that  $A = QR$ .]

20. Suppose  $A = QR$ , where  $R$  is an invertible matrix. Show that  $A$  and  $Q$  have the same column space. [Hint: Given  $\mathbf{y}$  in  $\text{Col } A$ , show that  $\mathbf{y} = Q\mathbf{x}$  for some  $\mathbf{x}$ . Also, given  $\mathbf{y}$  in  $\text{Col } Q$ , show that  $\mathbf{y} = A\mathbf{x}$  for some  $\mathbf{x}$ .]  
 21. Given  $A = QR$  as in Theorem 12, describe how to find an orthogonal  $m \times m$  (square) matrix  $Q_1$  and an invertible  $n \times n$  upper triangular matrix  $R$  such that

$$A = Q_1 \begin{bmatrix} R \\ 0 \end{bmatrix}$$

The MATLAB `qr` command supplies this “full” QR factorization when  $\text{rank } A = n$ .

22. Let  $\mathbf{u}_1, \dots, \mathbf{u}_p$  be an orthogonal basis for a subspace  $W$  of  $\mathbb{R}^n$ , and let  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  be defined by  $T(\mathbf{x}) = \text{proj}_W \mathbf{x}$ . Show that  $T$  is a linear transformation.  
 23. Suppose  $A = QR$  is a QR factorization of an  $m \times n$  matrix  $A$  (with linearly independent columns). Partition  $A$  as  $[A_1 \ A_2]$ , where  $A_1$  has  $p$  columns. Show how to obtain a QR factorization of  $A_1$ , and explain why your factorization has the appropriate properties.  
 24. [M] Use the Gram–Schmidt process as in Example 2 to produce an orthogonal basis for the column space of

$$A = \begin{bmatrix} -10 & 13 & 7 & -11 \\ 2 & 1 & -5 & 3 \\ -6 & 3 & 13 & -3 \\ 16 & -16 & -2 & 5 \\ 2 & 1 & -5 & -7 \end{bmatrix}$$

25. [M] Use the method in this section to produce a QR factorization of the matrix in Exercise 24.  
 26. [M] For a matrix program, the Gram–Schmidt process works better with orthonormal vectors. Starting with  $\mathbf{x}_1, \dots, \mathbf{x}_p$  as in Theorem 11, let  $A = [\mathbf{x}_1 \ \dots \ \mathbf{x}_p]$ . Suppose  $Q$  is an  $n \times k$  matrix whose columns form an orthonormal basis for the subspace  $W_k$  spanned by the first  $k$  columns of  $A$ . Then for  $\mathbf{x}$  in  $\mathbb{R}^n$ ,  $QQ^T \mathbf{x}$  is the orthogonal projection of  $\mathbf{x}$  onto  $W_k$  (Theorem 10 in Section 6.3). If  $\mathbf{x}_{k+1}$  is the next column of  $A$ , then equation (2) in the proof of Theorem 11 becomes

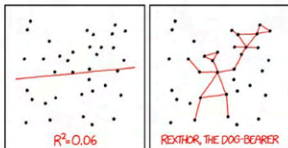
$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - Q(Q^T \mathbf{x}_{k+1})$$

(The parentheses above reduce the number of arithmetic operations.) Let  $\mathbf{u}_{k+1} = \mathbf{v}_{k+1}/\|\mathbf{v}_{k+1}\|$ . The new  $Q$  for the

## Section 6.5 : Least-Squares Problems

Chapter 6 : Orthogonality and Least Squares

Math 1554 Linear Algebra



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.

<https://xkcd.com/1725>

## Topics and Objectives

### Topics

1. Least Squares Problems
2. Different methods to solve Least Squares Problems

### Learning Objectives

1. Compute general solutions, and least squares errors, to least squares problems using the normal equations and the  $QR$  decomposition.

**Motivating Question** A series of measurements are corrupted by random errors. How can the dominant trend be extracted from the measurements with random error?

# Topics and Objectives

## Section 6.5 : Least-Squares Problems

### Topics

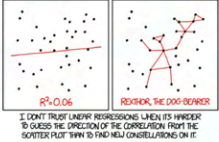
- Least Squares Problems
- Different methods to solve Least Squares Problems

### Learning Objectives

- Compute general solutions, and least squares errors, to least squares problems using the normal equations and the  $QR$  decomposition.

**Motivating Question** A series of measurements are corrupted by random errors. How can the dominant trend be extracted from the measurements with random error?

Chapter 6 : Orthogonality and Least Squares  
Math 1554 Linear Algebra



<https://kcd.com/1725>

Section 6.5 Slide 328

8	2/26 - 3/1	5.2	WS5.1.5.2	Exam 2, Review	Cancelled	5.3
9	3/4 - 3/8	5.3	WS5.3	5.5	WS5.5	6.1
10	3/11 - 3/15	6.1,6.2	WS6.1	6.2	WS6.2	6.3
11	3/18 - 3/22	Break	Break	Break	Break	Break
12	3/25 - 3/29	6.4	WS6.3	6.4,6.5	WS6.4	6.5
13	4/1 - 4/5	6.6	WS6.5,6.6	Exam 3, Review	Cancelled	PageRank

### 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  $\hat{\mathbf{x}}$  in  $\mathbb{R}^n$  such that

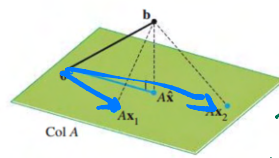
$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

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

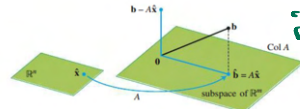
$$A = [v_1 \ v_2] \quad \hat{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \hat{\mathbf{b}} = \text{proj}_{\text{Col } A} \mathbf{b}$$

$$A\hat{\mathbf{x}} = x_1\vec{v}_1 + x_2\vec{v}_2 = \hat{\mathbf{b}}$$

$$\Rightarrow \mathbf{b} \in \text{Col } A \text{ give consistent } A\hat{\mathbf{x}} = \mathbf{b}$$



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



**FIGURE 2** The least-squares solution  $\hat{\mathbf{x}}$  is in  $\mathbb{R}^n$ .

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

then

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

is as small as possible.

**EXAMPLE 1** Find a least-squares solution of the inconsistent system  $A\mathbf{x} = \mathbf{b}$  for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

**THEOREM 14**

Let  $A$  be an  $m \times n$  matrix. The following statements are logically equivalent:

- The equation  $A\mathbf{x} = \mathbf{b}$  has a unique least-squares solution for each  $\mathbf{b}$  in  $\mathbb{R}^m$ .
- The columns of  $A$  are linearly independent.
- The matrix  $A^T A$  is invertible.

When these statements are true, the least-squares solution  $\hat{\mathbf{x}}$  is given by

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} \quad (4)$$

Step 1. Find  $A^T A$ ,  $A^T \mathbf{b}$

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

$$A^T \mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

Step 2: Solve  $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ .

$$[A^T A | A^T \mathbf{b}] = \left[ \begin{array}{cc|c} 17 & 1 & 19 \\ 1 & 5 & 11 \end{array} \right] \sim \left[ \begin{array}{cc|c} 1 & 5 & 11 \\ 0 & 84 & 168 \end{array} \right] \sim \left[ \begin{array}{cc|c} 1 & 5 & 11 \\ 0 & 1 & 2 \end{array} \right]$$

$$A\hat{\mathbf{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ least squares soln.}$$

$$= \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \hat{\mathbf{b}}$$

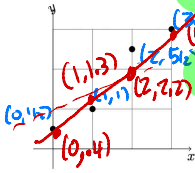
$$\rightarrow \left[ \begin{array}{cc|c} 1 & 0 & 11 \\ 0 & 1 & 2 \end{array} \right] \quad \hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Suppose we want to construct a line of the form

$$y = \alpha x + \beta$$

that best fits the data below:

$$(0, 1/2), (1, 1), (2, 5/2), (3, 3)$$



From the data, we can construct the system:

$$y = \alpha x + \beta$$

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 2.5 \\ 3 \end{bmatrix}$$

~~$\alpha \approx 1/2$~~   ~~$\beta \approx 1$~~   
 ~~$\alpha \approx 1$~~   ~~$\beta \approx 1/2$~~

$$A\vec{x} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 2.5 \end{bmatrix}$$

Definition: Least Squares Solution

Let  $A$  be a  $m \times n$  matrix. A least squares solution to  $A\vec{x} = \vec{b}$  is the solution  $\hat{\vec{x}}$  for which

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

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

Can we 'solve' this inconsistent system?

Section 6.5 Slide 327

Section 6.5 Slide 328

Step 0: Need  $Ax=b$ , Find  $A, b$ .

Plug in data into model

$$y = \alpha x + \beta$$

- @ (0, 1/2)  $1/2 = \alpha \cdot 0 + \beta$
- @ (1, 1)  $1 = \alpha \cdot 1 + \beta$
- @ (2, 5/2)  $5/2 = \alpha \cdot 2 + \beta$
- @ (3, 3)  $3 = \alpha \cdot 3 + \beta$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 2 \\ 2 & 3 \end{bmatrix} \quad \vec{x} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$b = \begin{bmatrix} 1/2 \\ 1 \\ 5/2 \\ 3 \end{bmatrix}$$

coeff of  $\alpha$   
coeff of  $\beta$   
y-values

$$A\vec{x} = \vec{b}$$

consistent or inconsistent?

Step 1: Set up  $A^T A \vec{x} = A^T b$  the normal eqns.

$$A^T A = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1 \\ 5/2 \\ 3 \end{bmatrix} = \begin{bmatrix} 15 \\ 7 \end{bmatrix}$$

Step 2 Solve

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

$$[A^T A | A^T b] = \left[ \begin{array}{cc|c} 14 & 6 & 15 \\ 6 & 4 & 7 \end{array} \right]$$

~ ... ?

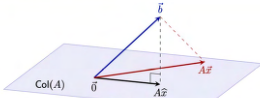
Option 2  $\vec{x} = (A^T A)^{-1} A^T b$

$$(A^T A)^{-1} = \begin{bmatrix} 14 & 6 \\ 6 & 4 \end{bmatrix}^{-1} = \frac{1}{20} \begin{bmatrix} 4 & -6 \\ -6 & 14 \end{bmatrix}$$

$$(A^T A)^{-1} A^T b = \frac{1}{20} \begin{bmatrix} 4 & -6 \\ -6 & 14 \end{bmatrix} \begin{bmatrix} 15 \\ 7 \end{bmatrix}$$

$$= \frac{1}{20} \begin{bmatrix} 18 \\ 8 \end{bmatrix} = \begin{bmatrix} 9/10 \\ 4/10 \end{bmatrix} = \begin{bmatrix} 0.9 \\ 0.4 \end{bmatrix}$$

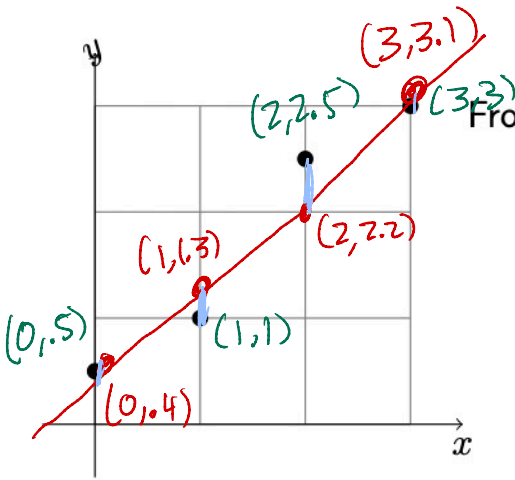
A Geometric Interpretation



The vector  $\vec{b}$  is closer to  $A\vec{x}$  than to  $A\vec{z}$  for all other  $\vec{z} \in \text{Col}(A)$ .

1. If  $\vec{b} \in \text{Col}(A)$ , then  $\hat{\vec{x}}$  is ...
2. Seek  $\hat{\vec{x}}$  so that  $A\hat{\vec{x}}$  is as close to  $\vec{b}$  as possible. That is,  $\hat{\vec{x}}$  should solve  $A\hat{\vec{x}} = \vec{b}$  where  $\vec{b}$  is ...

that best fits the data below.



From the data, we can construct the system:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1 \\ 2.5 \\ 3 \end{bmatrix}$$

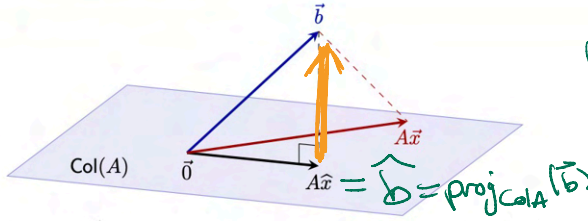
Can we 'solve' this inconsistent system?

Section 6.5 Slide 327

$$\begin{aligned} \|b - \hat{b}\|^2 &= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 \\ &\quad + (y_4 - \hat{y}_4)^2 \\ &= \text{least squares error, squared.} \end{aligned}$$

$Ax=b$  solve.

## A Geometric Interpretation



The vector  $\hat{b}$  is closer to  $A\hat{x}$  than to  $A\bar{x}$  for all other  $\bar{x} \in \text{Col}A$ .

1. If  $\hat{b} \in \text{Col}A$ , then  $\hat{x}$  is ...
2. Seek  $\hat{x}$  so that  $A\hat{x}$  is as close to  $\hat{b}$  as possible. That is,  $\hat{x}$  should solve  $A\hat{x} = \hat{b}$  where  $\hat{b}$  is ...

$$b - \hat{b} \in (\text{Col}A)^\perp = \text{Null}A^T$$

Section 6.5 Slide 329

$$\Rightarrow A^T(b - \hat{b}) = \vec{0} \quad \left( \begin{array}{l} \text{says } b - \hat{b} \\ \text{is in Null}A^T \end{array} \right)$$

$$\Rightarrow A^T b - A^T \hat{b} = \vec{0}$$

$$\Rightarrow A^T b - A^T A \hat{x} = \vec{0}$$

$$\Rightarrow \boxed{A^T A \hat{x} = A^T b}$$

Why  $A^T A$ ?

Why  $A^T b$ ?

Solving  $A^T A \hat{x} = A^T b$   
normal eqns.

Why when you solve  
the normal eqns  
to you get  $\hat{x}$   
least squares soln?

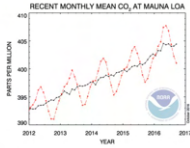
$$b - A\hat{x} \in (\text{Col}A)^\perp$$

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



# Important Examples: Overdetermined Systems (Tall/Thin Matrices)

A variety of factors impact the measured quantity.



In the above figure, the dashed red line with diamond symbols represents the monthly mean values, centered on the middle of each month. The black line with the square symbols represents the same, after correction for the average seasonal cycle. (NOAA graph.)

Section 6.5 Slide 330



Previous data is the important time series of mean CO<sub>2</sub> in the atmosphere. The data is collected at the Mauna Loa observatory on the island of Hawaii (The Big Island). One of the most important observatories in the world, it is located at the top of the Mauna Kea volcano, 4,205 meters altitude.

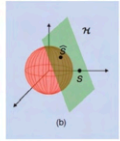
Section 6.5 Slide 331

# Important Examples: Underdetermined Systems (Short/Fat Matrices)

There are too few measurements, and many solutions to  $A\vec{x} = \vec{b}$ . Choose  $\vec{x}$  solving the system, with the smallest length.

- 1.  $A\vec{x} = \vec{b}$ .
- 2. For all  $\vec{x}$  with  $A\vec{x} = \vec{b}$ ,  $\|\vec{x}\| \leq \|\vec{x}'\|$ .

This is the least squares problem of 'Big Data.' (But not addressed in this course.)



Section 6.5 Slide 332

## The Normal Equations

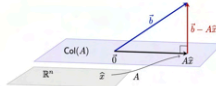
### Theorem (Normal Equations for Least Squares)

The least squares solutions to  $A\vec{x} = \vec{b}$  coincide with the solutions to

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

Normal Equations

## Derivation

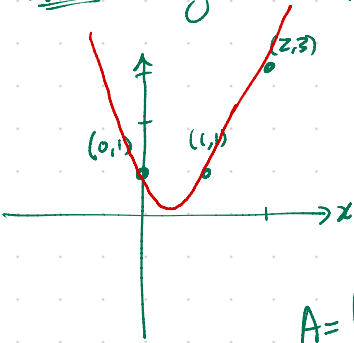


The least-squares solution  $\hat{x}$  is in  $\mathbb{R}^n$ .

1.  $\hat{x}$  is the least squares solution, is equivalent to  $\vec{b} - A\hat{x}$  is orthogonal to  $\text{Col}(A)$ .
2. A vector  $\vec{v}$  is in  $\text{Null}(A^T)$  if and only if  $\vec{v} \cdot \vec{w} = 0$ .
3. So we obtain the Normal Equations:

Section 6.5 Slide 334

Example. Model  $y = \alpha x^2 + \beta x + \gamma$  data  $(0,1), (1,1), (2,3)$ .



Step 0: plug in data into model.

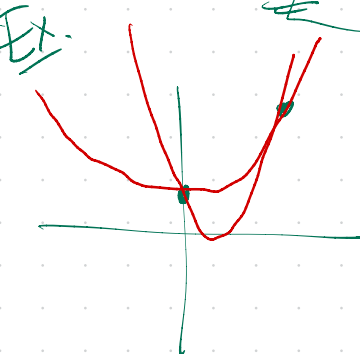
$$\begin{aligned} @ (0,1) & 1 = 0\alpha + 0\beta + \gamma \\ @ (1,1) & 1 = \alpha + \beta + \gamma \\ @ (2,3) & 3 = 4\alpha + 2\beta + \gamma \end{aligned}$$

A coeff matrix.

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

b constants

Solve  $Ax = b$  constants



Do many parabolas that work!

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

Section 6.5 Slide 333



Example: Given the data set  $(-6, -1)$ ,  $(-2, 2)$ ,  $(1, 1)$ ,  $(7, 6)$ , find the coefficient matrix  $A$  for the least squares problem  $Ax = b$  which can be used to find the line of best fit  $y = \alpha x + \beta$

**Step 0:** Plug in data into model!

$$\begin{aligned} \textcircled{c} (-6, -1) & \quad -6\alpha + \beta = -1 \\ \textcircled{c} (-2, 2) & \quad -2\alpha + \beta = 2 \\ \textcircled{c} (1, 1) & \quad \alpha + \beta = 1 \\ \textcircled{c} (7, 6) & \quad 7\alpha + \beta = 6 \end{aligned}$$

$$A = \begin{bmatrix} -6 & 1 \\ -2 & 1 \\ 1 & 1 \\ 7 & 1 \end{bmatrix}$$

$$b = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

"design matrix" of model fitting problem.

**Theorem**

**Theorem (Unique Solutions for Least Squares)**  
 Let  $A$  be any  $m \times n$  matrix. These statements are equivalent for each  $b \in \mathbb{R}^m$ .  
 1. The equation  $Ax = b$  has a unique least-squares solution.  
 2. The columns of  $A$  are linearly independent.  
 3. The matrix  $A^T A$  is invertible.  
 And, if these statements hold, the least squares solution is  $x = (A^T A)^{-1} A^T b$ .

Useful heuristic:  $A^T A$  plays the role of 'length-squared' of the matrix  $A$ . (See the sections on symmetric matrices and singular value decomposition.)

**Example**

Compute the least squares solution to  $Ax = b$ , where

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Hint: the columns of  $A$  are orthogonal.

$$\begin{aligned} A^T A &= \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} & A^T b &= \begin{bmatrix} -1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \\ & & & \sim \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 4 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1 \end{bmatrix} \\ & & & \Rightarrow x = \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} \end{aligned}$$

Solving new

$$Ax = \hat{b}$$

$\infty$ -many if  $A$  has lin dep cols.

$\hat{x} = \begin{pmatrix} * \\ * \end{pmatrix} \in \mathbb{R}^2$  least squares soln  
 unique!

$$ATAx = ATb$$

$\infty$ -many if  $ATA$  has free cols.

Claim # free vars of  $A =$  # free  $ATA$   
 # pivots of  $A =$  # pivots  $ATA$   
 # cols  $A =$  # cols of  $ATA$ .

Claim  $\text{Nul}(A) = \text{Nul}(A^T A)$ .

IF  $x \in \text{Nul } A$   
 $\Rightarrow Ax = \vec{0}$   
 $\Rightarrow A^T Ax = A^T \vec{0} = \vec{0}$

Then  $x \in \text{Nul } ATA$

IF  $x \in \text{Nul}(A^T A)$ , then  $\|Ax\| = 0 \Rightarrow Ax = \vec{0}$

$$A^T A x = \vec{0} \Rightarrow x^T A^T A x = x^T \vec{0} = 0 \Rightarrow (Ax) \cdot (Ax) = 0 \Rightarrow \|Ax\|^2 = 0$$

**Theorem (Least Squares and QR)**

Let  $m \times n$  matrix  $A$  have a QR decomposition. Then for each  $\vec{b} \in \mathbb{R}^m$  the equation  $A\vec{x} = \vec{b}$  has the unique least squares solution

$$R\hat{x} = Q^T \vec{b}.$$

(Remember,  $R$  is upper triangular, so the equation above is solved by back-substitution.)

**THEOREM 15**

Given an  $m \times n$  matrix  $A$  with linearly independent columns, let  $A = QR$  be a QR factorization of  $A$  as in Theorem 12. Then, for each  $\vec{b}$  in  $\mathbb{R}^m$ , the equation  $A\vec{x} = \vec{b}$  has a unique least-squares solution, given by

$$\hat{x} = R^{-1}Q^T \vec{b} \quad (6)$$

Example 3. Compute the least squares solution to  $A\vec{x} = \vec{b}$ , where

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

Solution. The QR decomposition of  $A$  is

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

$$Q^T \vec{b} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} -6 \\ 7 \\ -3 \\ 4 \end{bmatrix}$$

And then we solve by backwards substitution  $R\hat{x} = Q^T \vec{b}$

$$\begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -6 \\ -6 \\ 4 \end{bmatrix}$$

cols  $\downarrow$  are standard

Find L-S soln to  $Ax=b$ .

$$\left( \begin{matrix} \equiv \\ \equiv \\ \equiv \end{matrix} \right) \left( \begin{matrix} || \\ || \\ || \end{matrix} \right) = \left( \begin{matrix} \cdot \\ \cdot \\ \cdot \end{matrix} \right)$$

Section 6.5 Slide 302

Solve  ~~$ATA\hat{x} = AT\vec{b}$~~

$$R\hat{x} = Q^T \vec{b}$$

$$\begin{bmatrix} 2 & 4 & 5 & | & 12 \\ 0 & 2 & 3 & | & -6 \\ 0 & 0 & 2 & | & 4 \end{bmatrix} \sim \dots \sim \begin{bmatrix} 1 & 0 & 0 & | & * \\ 0 & 1 & 0 & | & * \\ 0 & 0 & 1 & | & * \end{bmatrix}$$

↑ Comes from  $ATA\hat{x} = AT\vec{b}$ , using  $A = QR$

$$\Rightarrow (QR)^T QR\hat{x} = (QR)^T \vec{b}$$

$$\Rightarrow R^T Q^T QR\hat{x} = R^T Q \vec{b}$$

$$\Rightarrow \cancel{R^T} R\hat{x} = \cancel{R^T} Q \vec{b}$$

$$\Rightarrow R\hat{x} = Q^T \vec{b}$$

## 6.5 EXERCISES

In Exercises 1–4, find a least-squares solution of  $Ax = b$  by (a) constructing the normal equations for  $\hat{x}$  and (b) solving for  $\hat{x}$ .

$$1. A = \begin{bmatrix} -1 & 2 \\ 2 & -3 \\ -1 & 3 \end{bmatrix}, b = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix}$$

$$2. A = \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix}, b = \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix}$$

$$3. A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix}, b = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 1 \end{bmatrix}$$

$$4. A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}$$

In Exercises 5 and 6, describe all least-squares solutions of the equation  $Ax = b$ .

$$5. A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}$$

$$6. A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 7 \\ 2 \\ 3 \\ 6 \\ 4 \end{bmatrix}$$

7. Compute the least-squares error associated with the least-squares solution found in Exercise 3.

8. Compute the least-squares error associated with the least-squares solution found in Exercise 4.

In Exercises 9–12, find (a) the orthogonal projection of  $b$  onto Col  $A$  and (b) a least-squares solution of  $Ax = b$ .

$$9. A = \begin{bmatrix} 1 & 5 \\ 3 & 1 \\ -2 & 4 \end{bmatrix}, b = \begin{bmatrix} 4 \\ -2 \\ -3 \end{bmatrix}$$

b. A least-squares solution of  $Ax = b$  is a vector  $\hat{x}$  that satisfies  $A\hat{x} = \bar{b}$ , where  $\bar{b}$  is the orthogonal projection of  $b$  onto Col  $A$ .

c. A least-squares solution of  $Ax = b$  is a vector  $\hat{x}$  such that  $\|b - A\hat{x}\| \leq \|b - Ax\|$  for all  $x$  in  $\mathbb{R}^n$ .

d. Any solution of  $A^T Ax = A^T b$  is a least-squares solution of  $Ax = b$ .

e. If the columns of  $A$  are linearly independent, then the equation  $Ax = b$  has exactly one least-squares solution.

10. a. If  $b$  is in the column space of  $A$ , then every solution of  $Ax = b$  is a least-squares solution.

b. The least-squares solution of  $Ax = b$  is the point in the column space of  $A$  closest to  $b$ .

c. A least-squares solution of  $Ax = b$  is a list of weights that, when applied to the columns of  $A$ , produces the orthogonal projection of  $b$  onto Col  $A$ .

d. If  $\hat{x}$  is a least-squares solution of  $Ax = b$ , then  $\hat{x} = (A^T A)^{-1} A^T b$ .

e. The normal equations always provide a reliable method for computing least-squares solutions.

f. If  $A$  has a QR factorization, say  $A = QR$ , then the best way to find the least-squares solution of  $Ax = b$  is to compute  $\hat{x} = R^{-1} Q^T b$ .

11. Let  $A$  be an  $m \times n$  matrix. Use the steps below to show that a vector  $x$  in  $\mathbb{R}^n$  satisfies  $Ax = 0$  if and only if  $A^T Ax = 0$ . This will show that  $\text{Nul } A = \text{Nul } A^T A$ .

a. Show that if  $Ax = 0$ , then  $A^T Ax = 0$ .

b. Suppose  $A^T Ax = 0$ . Explain why  $x^T A^T Ax = 0$ , and use this to show that  $Ax = 0$ .

12. Let  $A$  be an  $m \times n$  matrix such that  $A^T A$  is invertible. Show that the columns of  $A$  are linearly independent. [Careful: You may not assume that  $A$  is invertible; it may not even be square.]

13. Let  $A$  be an  $m \times n$  matrix whose columns are linearly independent. [Careful:  $A$  need not be square.]

a. Use Exercise 12 to show that  $A^T A$  is an invertible matrix.

b. Explain why  $A$  must have at least as many rows as columns.

c. Determine the rank of  $A$ .

14. Use Exercise 12 to show that  $\text{rank } A^T A = \text{rank } A$ . [Hint: How many columns does  $A^T A$  have? How is this connected with the rank of  $A^T A$ ?]

15. Suppose  $A$  is  $m \times n$  with linearly independent columns and  $b$  is in  $\mathbb{R}^m$ . Use the normal equations to produce a formula for  $\hat{b}$ , the projection of  $b$  onto Col  $A$ . [Hint: Find  $\hat{x}$  first. The formula does not require an orthogonal basis for Col  $A$ .]

$$10. A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix}, b = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$$

$$11. A = \begin{bmatrix} 4 & 0 & 1 \\ 1 & -5 & 1 \\ 6 & 1 & 0 \\ 1 & -1 & -5 \end{bmatrix}, b = \begin{bmatrix} 9 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$12. A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 5 \\ 6 \\ 6 \end{bmatrix}$$

$$13. \text{ Let } A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \\ 3 & 4 \end{bmatrix}, b = \begin{bmatrix} 11 \\ -9 \\ 5 \end{bmatrix}, u = \begin{bmatrix} 5 \\ -1 \end{bmatrix}, \text{ and } v =$$

$$\begin{bmatrix} 5 \\ -2 \end{bmatrix}. \text{ Compute } Au \text{ and } Av, \text{ and compare them with } b. \text{ Could } u \text{ possibly be a least-squares solution of } Ax = b? \text{ (Answer this without computing a least-squares solution.)}$$

$$14. \text{ Let } A = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix}, b = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}, u = \begin{bmatrix} 4 \\ -5 \end{bmatrix}, \text{ and } v =$$

$$\begin{bmatrix} 6 \\ -5 \end{bmatrix}. \text{ Compute } Au \text{ and } Av, \text{ and compare them with } b. \text{ Is it possible that at least one of } u \text{ or } v \text{ could be a least-squares solution of } Ax = b? \text{ (Answer this without computing a least-squares solution.)}$$

In Exercises 15 and 16, use the factorization  $A = QR$  to find the least-squares solution of  $Ax = b$ .

$$15. A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2/3 & -1/3 \\ 2/3 & 2/3 \\ 1/3 & -2/3 \end{bmatrix} \begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}, b = \begin{bmatrix} 7 \\ 3 \\ 1 \end{bmatrix}$$

$$16. A = \begin{bmatrix} 1 & -1 \\ 1 & 4 \\ 1 & -1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 5 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix}$$

In Exercises 17 and 18,  $A$  is an  $m \times n$  matrix and  $b$  is in  $\mathbb{R}^m$ . Mark each statement True or False. Justify each answer.

17. a. The general least-squares problem is to find an  $x$  that makes  $Ax$  as close as possible to  $b$ .

24. Find a formula for the least-squares solution of  $Ax = b$  when the columns of  $A$  are orthonormal.

25. Describe all least-squares solutions of the system

$$x + y = 2 \\ x + y = 4$$

26. [M] Example 3 in Section 4.8 displayed a low-pass linear filter that changed a signal  $\{y_k\}$  into  $\{y_{k+1}\}$  and changed a higher-frequency signal  $\{u_k\}$  into the zero signal, where  $y_k = \cos(\pi k/4)$  and  $u_k = \cos(3\pi k/4)$ . The following calculations will design a filter with approximately those properties. The filter equation is

$$a_0 y_{k+2} + a_1 y_{k+1} + a_2 y_k = z_k \quad \text{for all } k \quad (8)$$

Because the signals are periodic, with period 8, it suffices to study equation (8) for  $k = 0, \dots, 7$ . The action on the two signals described above translates into two sets of eight equations, shown below:

$$\begin{matrix} k=0 \\ k=1 \\ \vdots \\ k=7 \end{matrix} \begin{matrix} y_{k+2} & y_{k+1} & y_k \\ \begin{bmatrix} 0 & .7 & 1 \\ -7 & 0 & .7 \\ -1 & -7 & 0 \\ -7 & -1 & -7 \\ 0 & -7 & -1 \\ .7 & 0 & -7 \\ 1 & .7 & 0 \\ .7 & 1 & .7 \end{bmatrix} \end{matrix} \begin{matrix} a_0 \\ a_1 \\ a_2 \end{matrix} = \begin{matrix} y_{k+1} \\ \begin{bmatrix} .7 \\ 0 \\ -7 \\ -1 \\ -7 \\ 0 \\ .7 \\ 1 \end{bmatrix} \end{matrix}$$

$$\begin{matrix} k=0 \\ k=1 \\ \vdots \\ k=7 \end{matrix} \begin{matrix} u_{k+2} & u_{k+1} & u_k \\ \begin{bmatrix} 0 & -.7 & 1 \\ .7 & 0 & -.7 \\ -1 & .7 & 0 \\ .7 & -1 & .7 \\ 0 & .7 & -1 \\ -.7 & 0 & .7 \\ 1 & -.7 & 0 \\ -.7 & 1 & -.7 \end{bmatrix} \end{matrix} \begin{matrix} a_0 \\ a_1 \\ a_2 \end{matrix} = \begin{matrix} u_{k+1} \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{matrix}$$

Write an equation  $Ax = b$ , where  $A$  is a  $16 \times 3$  matrix formed from the two coefficient matrices above and where  $b$  in  $\mathbb{R}^{16}$  is formed from the two right sides of the equations. Find  $a_0, a_1,$  and  $a_2$  given by the least-squares solution of  $Ax = b$ . (The  $z_k$  in the data above was used as an approximation for  $\sqrt{2}/2$ , to illustrate how a typical computation in an applied problem might proceed. If .707 were used instead, the resulting filter coefficients would agree to at least seven decimal places with  $\sqrt{2}/4, 1/2,$  and  $\sqrt{2}/4$ , the values produced by exact arithmetic calculations.)

WEB

## Topics and Objectives

8	2/26 - 3/1	5.2	WS5.1.5.2	Exam 2, Review	Cancelled	5.3
9	3/4 - 3/8	5.3	WS5.3	5.5	WS5.5	6.1
10	3/11 - 3/15	6.1, 6.2	WS6.1	6.2	WS6.2	6.3
11	3/18 - 3/22	Break	Break	Break	Break	Break
12	3/25 - 3/29	6.4	WS6.3	6.4, 6.5	WS6.4	6.5
13	4/1 - 4/5	6.6	WS5.6.6	Exam 3, Review	Cancelled	Page 8

### Topics

- Least Squares Lines
- Linear and more complicated models

### Learning Objectives

For the topics covered in this section, students are expected to be able to do the following.

- Apply least-squares and multiple regression to construct a linear model from a set of data points.
- Apply least-squares to fit polynomials and other curves to data.

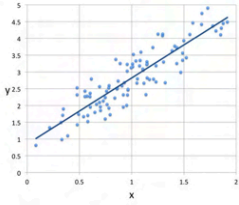
### Motivating Question

Compute the equation of the line  $y = \beta_0 + \beta_1 x$  that best fits the data

$x$	2	5	7	8
$y$	1	1	4	3

Section 6.6 Slide 348

## Chapter 6 : Orthogonality and Least Squares 6.6 : Applications to Linear Models



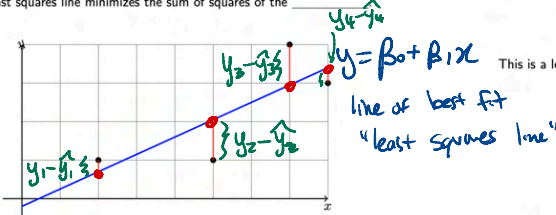
Slide 343

## The Least Squares Line

Graph below gives an approximate linear relationship between  $x$  and  $y$ .

- Black circles are data.
- Blue line is the **least squares line**.
- Lengths of red lines are the **residuals**.

The least squares line minimizes the sum of squares of the residuals.



Section 6.6 Slide 345

Section 6.6 Slide 348

**Example 1** Compute the least squares line  $y = \beta_0 + \beta_1 x$  that best fits the data

$x$	2	5	7	8
$y$	1	1	4	3

We want to solve

$$\begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix}$$

This is a least-squares problem:  $X^T \beta = \vec{y}$ .

$$\text{Error} = \sqrt{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2 + (y_4 - \hat{y}_4)^2}$$

Minimize the SUM of squares of residual errors.

residual is vertical distance between data point & predicted point on model.

The normal equations are

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 29 \end{bmatrix}$$

So the least-squares solution is given by

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 29 \end{bmatrix}$$

$$y = \beta_0 + \beta_1 x = \frac{-5}{21} + \frac{19}{42}x$$

As we may have guessed,  $\beta_0$  is negative, and  $\beta_1$  is positive.

## Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

$$y = \beta_0 + \beta_1 f_1(x) + \beta_2 f_2(x) + \dots + \beta_n f_n(x),$$

where the functions  $f_j$  are known. Should have only a few functional. Keep in mind this is a **linear problem in the  $\beta$  variables**.

Section 6.6 Slide 347

Section 6.6 Slide 348

```

% show data
data =
    1  0
    2  3
    4  8
    5 11
    6 20

% fit a linear model
fit = fit(data, 'linear');
% plot the data and the fit
plot(data(:,1), data(:,2), 'o');
hold on;
plot(fit, 'r');

```

$$y = kx + B$$

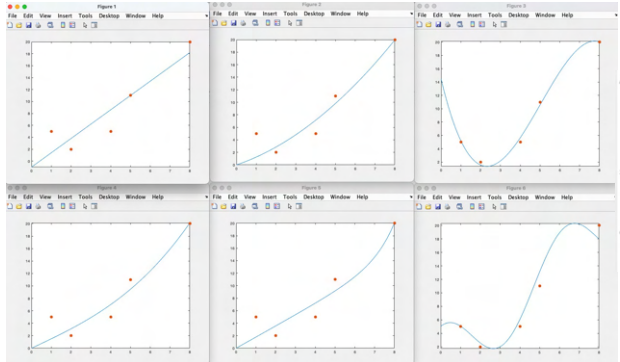
$$y = kx + Bx^2$$

$$y = a_0 + a_1x + a_2x^2 + a_3x^3$$

$$y = a_1x + a_2x^2$$

$$y = kx + B \cos x$$

$$y = kx + B \cos x$$



A =  
1 1  
2 1  
4 1  
5 1  
8 1

bhat =  
2.4667  
-1.8667

errorLine =  
3.5824

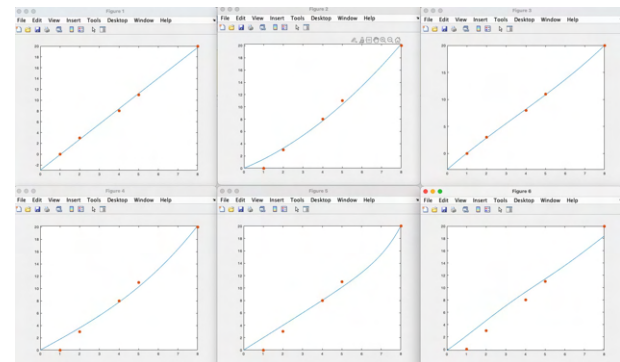
```

% fit a quadratic model
fit = fit(data, 'quadratic');
% plot the data and the fit
plot(data(:,1), data(:,2), 'o');
hold on;
plot(fit, 'r');

% fit a cubic model
fit = fit(data, 'cubic');
% plot the data and the fit
plot(data(:,1), data(:,2), 'o');
hold on;
plot(fit, 'r');

% fit a linear model with a cosine term
fit = fit(data, 'linear+cosine');
% plot the data and the fit
plot(data(:,1), data(:,2), 'o');
hold on;
plot(fit, 'r');

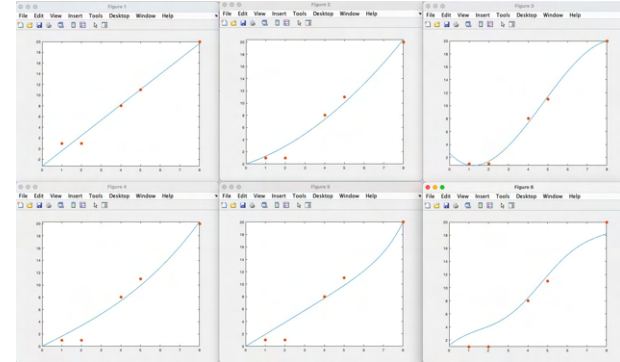
```



A =  
1 1  
2 4  
4 16  
5 25  
8 64

bhat =  
1.3882  
6.13785

errorQuad =  
2.4685



A =  
1 1  
2 4  
4 16  
5 25  
8 64

bhat =  
1.3782  
6.82294

errorCubic =  
3.3015

Model

$$y = \alpha x + \beta e^x$$

least squares problem

- e (1,5)
- e (2,2)
- e (4,5)
- e (5,11)
- e (8,20)

$$\begin{cases} x + e\beta = 5 \\ 2x + e^2\beta = 2 \\ 4x + e^4\beta = 5 \\ 5x + e^5\beta = 11 \\ 8x + e^8\beta = 20 \end{cases}$$

$$A = \begin{pmatrix} 1 & e \\ 2 & e^2 \\ 4 & e^4 \\ 5 & e^5 \\ 8 & e^8 \end{pmatrix} \quad b = \begin{pmatrix} 5 \\ 2 \\ 5 \\ 11 \\ 20 \end{pmatrix}$$

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

plug in data into model

$$\hat{x} = (A^T A)^{-1} A^T b = \begin{pmatrix} 1.7758 \\ 0.0019603 \end{pmatrix}$$

$$y = (1.77)x + (0.0019)e^x$$

New model

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$$

- e (1,5)
- e (2,2)
- e (4,5)
- e (5,11)
- e (8,20)

$$\begin{cases} a_0 + a_1 + a_2 + a_3 = 5 \\ a_0 + 2a_1 + 4a_2 + 8a_3 = 2 \\ a_0 + 4a_1 + 16a_2 + 64a_3 = 5 \\ a_0 + 5a_1 + 25a_2 + 125a_3 = 11 \\ a_0 + 8a_1 + 64a_2 + 512a_3 = 20 \end{cases}$$

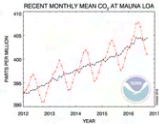
$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 8 \\ 1 & 4 & 16 & 64 \\ 1 & 5 & 25 & 125 \\ 1 & 8 & 64 & 512 \end{pmatrix} \quad b = \begin{pmatrix} 5 \\ 2 \\ 5 \\ 11 \\ 20 \end{pmatrix}$$

$A^T A \hat{x} = A^T b$  solve.

$$\hat{x} = \begin{pmatrix} * \\ * \\ * \\ * \end{pmatrix} = \begin{pmatrix} 14.445 \\ -12.543 \\ 3.5772 \\ -0.23276 \end{pmatrix}$$

model of best fit

$$y = 14.4 - 12.5x + 3.5x^2 - 0.2x^3$$



Black line is yearly CO<sub>2</sub> levels, and the monthly is the red line. To capture seasonality, would need a curve

$$\text{daily CO}_2 = \beta_0 + \beta_1 t + \beta_2 \sin(2\pi \frac{t}{12}) + \beta_3 \cos(2\pi \frac{t}{12})$$

Above,  $t$  is time, measured in months.

Least squares problems can be computed with WolframAlpha, Mathematica, and many other software.

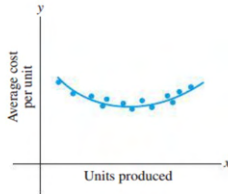
**WolframAlpha**

linear fit  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

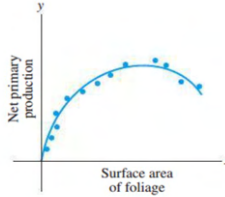
**Mathematica**

LeastSquares $\{\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}\}$

Almost any spreadsheet program does this as a function as well.



**FIGURE 3**  
Average cost curve.



**FIGURE 4**  
Production of nutrients.



**FIGURE 5**  
Data points along a cubic curve.

Theorem

**Theorem (Unique Solutions for Least Squares)**

Let  $A$  be any  $m \times n$  matrix. These statements are equivalent.

1. The equation  $A\vec{x} = \vec{b}$  has a unique least-squares solution for each  $\vec{b} \in \mathbb{R}^m$ .
2. The columns of  $A$  are linearly independent.
3. The matrix  $A^T A$  is invertible.

And, if these statements hold, the least square solution is

$$\hat{x} = (A^T A)^{-1} A^T \vec{b}.$$

**Theorem (Least Squares and QR)**

Let  $m \times n$  matrix  $A$  have a  $QR$  decomposition. Then for each  $\vec{b} \in \mathbb{R}^m$  the equation  $A\vec{x} = \vec{b}$  has the unique least squares solution

$$R\vec{x} = Q^T \vec{b}.$$

(Remember,  $R$  is upper triangular, so the equation above is solved by back-substitution.)

Useful heuristic:  $A^T A$  plays the role of 'length-squared' of the matrix  $A$ .  
(See the sections on symmetric matrices and singular value decomposition.)



## 6.6 EXERCISES

In Exercises 1–4, find the equation  $y = \beta_0 + \beta_1 x$  of the least-squares line that best fits the given data points.

- (0, 1), (1, 1), (2, 2), (3, 2)
- (1, 0), (2, 1), (4, 2), (5, 3)
- (-1, 0), (0, 1), (1, 2), (2, 4)
- (2, 3), (3, 2), (5, 1), (6, 0)

5. Let  $X$  be the design matrix used to find the least-squares line to fit data  $(x_1, y_1), \dots, (x_n, y_n)$ . Use a theorem in Section 6.5 to show that the normal equations have a unique solution if and only if the data include at least two data points with different  $x$ -coordinates.

6. Let  $X$  be the design matrix in Example 2 corresponding to a least-squares fit of a parabola to data  $(x_1, y_1), \dots, (x_n, y_n)$ . Suppose  $x_1, x_2,$  and  $x_3$  are distinct. Explain why there is only one parabola that fits the data best, in a least-squares sense. (See Exercise 5.)

7. A certain experiment produces the data (1, 1.8), (2, 2.7), (3, 3.4), (4, 3.8), (5, 3.9). Describe the model that produces a least-squares fit of these points by a function of the form  $y = \beta_1 x + \beta_2 x^2$

Such a function might arise, for example, as the revenue from the sale of  $x$  units of a product, when the amount offered for sale affects the price to be set for the product.

- Give the design matrix, the observation vector, and the unknown parameter vector.
  - [M] Find the associated least-squares curve for the data.
8. A simple curve that often makes a good model for the variable costs of a company, as a function of the sales level  $x$ , has the form  $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ . There is no constant term because fixed costs are not included.
- Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above, with data  $(x_1, y_1), \dots, (x_n, y_n)$ .
  - [M] Find the least-squares curve of the form above to fit the data (4, 1.58), (6, 2.08), (8, 2.5), (10, 2.8), (12, 3.1), (14, 3.4), (16, 3.8), and (18, 4.32), with values in thousands. If possible, produce a graph that shows the data points and the graph of the cubic approximation.

9. A certain experiment produces the data (1, 7.9), (2, 5.4), and (3, -9). Describe the model that produces a least-squares fit of these points by a function of the form

$$y = A \cos x + B \sin x$$

10. Suppose radioactive substances A and B have decay constants of .02 and .07, respectively. If a mixture of these two substances at time  $t = 0$  contains  $M_A$  grams of A and  $M_B$  grams of B, then a model for the total amount  $y$  of the mixture present at time  $t$  is

$$y = M_A e^{-.02t} + M_B e^{-.07t} \quad (6)$$

Suppose the initial amounts  $M_A$  and  $M_B$  are unknown, but a scientist is able to measure the total amounts present at several times and records the following points  $(t_i, y_i)$ : (10, 21.34), (11, 20.68), (12, 20.05), (14, 18.87), and (15, 18.30).

- Describe a linear model that can be used to estimate  $M_A$  and  $M_B$ .
- [M] Find the least-squares curve based on (6).



Halley's Comet last appeared in 1986 and will reappear in 2061.

11. [M] According to Kepler's first law, a comet should have an elliptic, parabolic, or hyperbolic orbit (with gravitational attractions from the planets ignored). In suitable polar coordinates, the position  $(r, \theta)$  of a comet satisfies an equation of the form

$$r = \beta + e(r \cdot \cos \theta)$$

where  $\beta$  is a constant and  $e$  is the *eccentricity* of the orbit, with  $0 \leq e < 1$  for an ellipse,  $e = 1$  for a parabola, and  $e > 1$  for a hyperbola. Suppose observations of a newly discovered comet provide the data below. Determine the type of orbit, and predict where the comet will be when  $\theta = 4.6$  (radians).<sup>3</sup>

$\theta$	.88	1.10	1.42	1.77	2.14
$r$	3.00	2.30	1.65	1.25	1.01

12. [M] A healthy child's systolic blood pressure  $p$  (in millimeters of mercury) and weight  $w$  (in pounds) are approximately related by the equation

$$\beta_0 + \beta_1 \ln w = p$$

Use the following experimental data to estimate the systolic blood pressure of a healthy child weighing 100 pounds.

<sup>3</sup> The basic idea of least-squares fitting of data is due to K. F. Gauss (and, independently, to A. Legendre), whose initial rise to fame occurred in 1801 when he used the method to determine the path of the asteroid *Ceres*. Forty days after the asteroid was discovered, it disappeared behind the sun. Gauss predicted it would appear ten months later and gave its location. The accuracy of the prediction astonished the European scientific community.



$w$	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
$p$	91	98	103	110	112

13. [M] To measure the takeoff performance of an airplane, the horizontal position of the plane was measured every second, from  $t = 0$  to  $t = 12$ . The positions (in feet) were: 0, 8.8, 29.9, 62.0, 104.7, 159.1, 222.0, 294.5, 380.4, 471.1, 571.7, 686.8, and 809.2.

- a. Find the least-squares cubic curve  $y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$  for these data.  
b. Use the result of part (a) to estimate the velocity of the plane when  $t = 4.5$  seconds.

14. Let  $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$  and  $\bar{y} = \frac{1}{n}(y_1 + \cdots + y_n)$ . Show that the least-squares line for the data  $(x_1, y_1), \dots, (x_n, y_n)$  must pass through  $(\bar{x}, \bar{y})$ . That is, show that  $\bar{x}$  and  $\bar{y}$  satisfy the linear equation  $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ . [Hint: Derive this equation from the vector equation  $\mathbf{y} = X\hat{\beta} + \boldsymbol{\epsilon}$ . Denote the first column of  $X$  by  $\mathbf{1}$ . Use the fact that the residual vector  $\boldsymbol{\epsilon}$  is orthogonal to the column space of  $X$  and hence is orthogonal to  $\mathbf{1}$ .]

Given data for a least-squares problem,  $(x_1, y_1), \dots, (x_n, y_n)$ , the following abbreviations are helpful:

$$\sum x = \sum_{i=1}^n x_i, \quad \sum x^2 = \sum_{i=1}^n x_i^2, \\ \sum y = \sum_{i=1}^n y_i, \quad \sum xy = \sum_{i=1}^n x_i y_i$$

The normal equations for a least-squares line  $y = \hat{\beta}_0 + \hat{\beta}_1 x$  may be written in the form

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum x = \sum y \\ \hat{\beta}_0 \sum x + \hat{\beta}_1 \sum x^2 = \sum xy \quad (7)$$

15. Derive the normal equations (7) from the matrix form given in this section.  
16. Use a matrix inverse to solve the system of equations in (7) and thereby obtain formulas for  $\hat{\beta}_0$  and  $\hat{\beta}_1$  that appear in many statistics texts.

17. a. Rewrite the data in Example 1 with new  $x$ -coordinates in mean deviation form. Let  $X$  be the associated design matrix. Why are the columns of  $X$  orthogonal?  
b. Write the normal equations for the data in part (a), and solve them to find the least-squares line,  $y = \beta_0 + \beta_1 x^*$ , where  $x^* = x - 5.5$ .  
18. Suppose the  $x$ -coordinates of the data  $(x_1, y_1), \dots, (x_n, y_n)$  are in mean deviation form, so that  $\sum x_i = 0$ . Show that if  $X$  is the design matrix for the least-squares line in this case, then  $X^T X$  is a diagonal matrix.

Exercises 19 and 20 involve a design matrix  $X$  with two or more columns and a least-squares solution  $\hat{\beta}$  of  $\mathbf{y} = X\hat{\beta}$ . Consider the following numbers.

- (i)  $\|X\hat{\beta}\|^2$ —the sum of the squares of the “regression term.” Denote this number by  $SS(R)$ .  
(ii)  $\|\mathbf{y} - X\hat{\beta}\|^2$ —the sum of the squares for error term. Denote this number by  $SS(E)$ .  
(iii)  $\|\mathbf{y}\|^2$ —the “total” sum of the squares of the  $y$ -values. Denote this number by  $SS(T)$ .

Every statistics text that discusses regression and the linear model  $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$  introduces these numbers, though terminology and notation vary somewhat. To simplify matters, assume that the mean of the  $y$ -values is zero. In this case,  $SS(T)$  is proportional to what is called the *variance* of the set of  $y$ -values.

19. Justify the equation  $SS(T) = SS(R) + SS(E)$ . [Hint: Use a theorem, and explain why the hypotheses of the theorem are satisfied.] This equation is extremely important in statistics, both in regression theory and in the analysis of variance.  
20. Show that  $\|X\hat{\beta}\|^2 = \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$ . [Hint: Rewrite the left side and use the fact that  $\hat{\boldsymbol{\beta}}$  satisfies the normal equations.] This formula for  $SS(R)$  is used in statistics. From this and from Exercise 19, obtain the standard formula for  $SS(E)$ :  
 $SS(E) = \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$