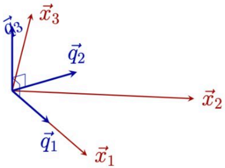
LINEAR ALGEBRA

Week 11

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.

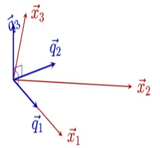
Motivating Question The vectors below span a subspace W of $\mathbb{R}^4$. Identify an orthogonal basis for W.

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

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.

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

Course Schedule

Cancellations due to inclement weather will likely result in cancelling review lectures and possibly moving through course material at

Week	Dates	Mon Lecture	Tue Studio	Wed Lecture	Thu Studio	Fri Lecture
1	8/21 – 8/25	1.1	WS1.1	1.2	WS1.2	1.3
2	8/28 – 9/1	1.4	WS1.3,1.4	1.5	WS1.5	1.7
3	9/4 – 9/8	Break	WS1.7	1.8	WS1.8	1.9
4	9/11 – 9/15	2.1	WS1.9,2.1	**Exam 1**, Review	Cancelled	2.2
5	9/18 – 9/22	2.3,2.4	WS2.2,2.3	2.5	WS2.4,2.5	2.8
6	9/25 – 9/29	2.9	WS2.8,2.9	3.1,3.2	WS3.1,3.2	3.3
7	10/2 – 10/6	4.9	WS3.3,4.9	5.1,5.2	WS5.1,5.2	5.2
8	10/9 – 10/13	Break	Break	**Exam 2**, Review	Cancelled	5.3
9	10/16 – 10/20	5.3	WS5.3	5.5	WS5.5	6.1
10	10/23 – 10/27	6.1,6.2	WS6.1	6.2	WS6.2	6.3
11	10/30 – 11/3	6.4	WS6.3,6.4	6.4,6.5	WS6.4,6.5	6.5
12	11/6 – 11/10	6.6	WS6.5,6.6	**Exam 3**, Review	Cancelled	PageRank
13	11/13 – 11/17	7.1	WSPageRank	7.2	WS7.1,7.2	7.3
14	11/20 – 11/24	7.3,7.4	WS7.2,7.3	Break	Break	Break
15	11/27 – 12/1	7.4	WS7.3,7.4	7.4	WS7.4	7.4
16	12/4 – 12/8	Last lecture	Last Studio	Reading Period		
17	12/11 – 12/15	Final Exams: **MATH 1554 Common Final Exam Tuesday, December 12th at 6pm**				

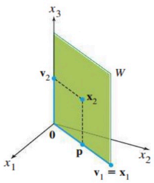


FIGURE 1
Construction of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$.

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

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

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

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

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

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

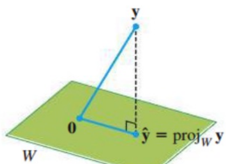


FIGURE 1

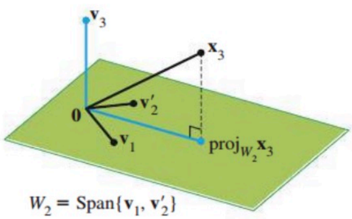


FIGURE 2 The construction of $\mathbf{v}_3$ from $\mathbf{x}_3$ and W_2.

THEOREM 11 **The Gram–Schmidt Process**

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

$$\mathbf{v}_1 = \mathbf{x}_1$$
$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1$$
$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2$$
$$\vdots$$
$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p\cdot\mathbf{v}_{p-1}}{\mathbf{v}_{p-1}\cdot\mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

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

$$\text{Span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

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

Example

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

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

The Gram-Schmidt Process

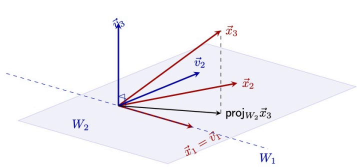
Given a basis $\{\vec{x}_1, \ldots, \vec{x}_p\}$ for a subspace W of $\mathbb{R}^n$, iteratively define

$$\begin{aligned}
\vec{v}_1 &= \vec{x}_1 \\
\vec{v}_2 &= \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 \\
\vec{v}_3 &= \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 \\
&\vdots \\
\vec{v}_p &= \vec{x}_p - \frac{\vec{x}_p \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{x}_p \cdot \vec{v}_{p-1}}{\vec{v}_{p-1} \cdot \vec{v}_{p-1}}\vec{v}_{p-1}
\end{aligned}$$

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

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.
$W_1 = \text{Span}\{\vec{v}_1\}$, $W_2 = \text{Span}\{\vec{v}_1, \vec{v}_2\}$.

Orthonormal Bases

Definition

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

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

$$\vec{v}_1 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \quad \vec{v}_2 = \begin{bmatrix} -2 \\ 3 \\ 1 \end{bmatrix}.$$

QR Factorization

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 $\operatorname{Col} A$.
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.

In the interest of time:
- we will not consider the case where A has linearly dependent columns
- students are not expected to know the conditions for which A has a QR factorization

Examples (if time permits)

Construct the QR decomposition for A.

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

b) $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$

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 $R = Q^TA$.

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 Col A, show that $\mathbf{y} = Q\mathbf{x}$ for some $\mathbf{x}$. Also, given $\mathbf{y}$ in 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 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 $T : \mathbb{R}^n \to \mathbb{R}^n$ 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 $A = [\,\mathbf{x}_1 \;\; \cdots \;\; x_p\,]$. 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

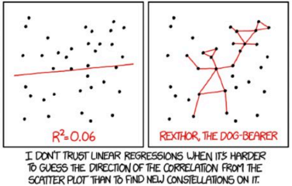
$$\mathbf{v}_{k+1} = \mathbf{x}_{k+1} - Q(Q^T\mathbf{x}_{k+1})$$

(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



https://xkcd.com/1725

Topics and Objectives

Topics

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

Learning Objectives

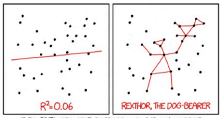
1. Compute general solutions, and least squares errors, to least squares problems using the normal equations and the QR decomposition.

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

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?

Course Schedule

Cancellations due to inclement weather will likely result in cancelling review lectures and possibly moving through course

Week	Dates	Mon Lecture	Tue Studio	Wed Lecture	Thu Studio	Fri Lecture
1	8/21 – 8/25	1.1	WS1.1	1.2	WS1.2	1.3
2	8/28 – 9/1	1.4	WS1.3,1.4	1.5	WS1.5	1.7
3	9/4 – 9/8	Break	WS1.7	1.8	WS1.8	1.9
4	9/11 – 9/15	2.1	WS1.9,2.1	**Exam 1**, Review	Cancelled	2.2
5	9/18 – 9/22	2.3,2.4	WS2.2,2.3	2.5	WS2.4,2.5	2.8
6	9/25 – 9/29	2.9	WS2.8,2.9	3.1,3.2	WS3.1,3.2	3.3
7	10/2 – 10/6	4.9	WS3.3,4.9	5.1,5.2	WS5.1,5.2	5.2
8	10/9 – 10/13	Break	Break	**Exam 2**, Review	Cancelled	5.3
9	10/16 – 10/20	5.3	WS5.3	5.5	WS5.5	6.1
10	10/23 – 10/27	6.1,6.2	WS6.1	6.2	WS6.2	6.3
11	10/30 – 11/3	6.4	WS6.3,6.4	6.4,6.5	WS6.4,6.5	6.5
12	11/6 – 11/10	6.6	WS6.5,6.6	**Exam 3**, Review	Cancelled	PageRank
13	11/13 – 11/17	7.1	WSPageRank	7.2	WS7.1,7.2	7.3
14	11/20 – 11/24	7.3,7.4	WS7.2,7.3	Break	Break	Break
15	11/27 – 12/1	7.4	WS7.3,7.4	7.4	WS7.4	7.4
16	12/4 – 12/8	Last lecture	Last Studio	Reading Period		
17	12/11 – 12/15	Final Exams: **MATH 1554 Common Final Exam Tuesday, December 12th at 6pm**				

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

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

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

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

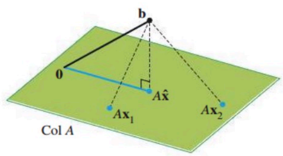


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

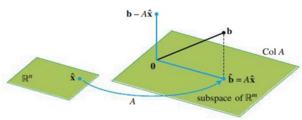


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

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:

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

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

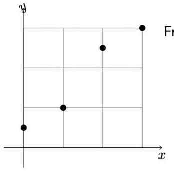
$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \qquad (4)$$

Inconsistent Systems

Suppose we want to construct a line of the form

$$y = mx + b$$

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?

The Least Squares Solution to a Linear System

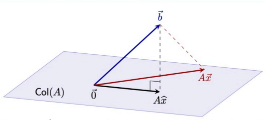
Definition: Least Squares Solution

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

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

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

A Geometric Interpretation

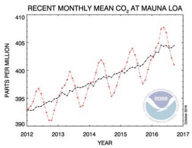


The vector $\vec{b}$ is closer to $A\hat{x}$ than to $A\vec{x}$ for all other $\vec{x} \in$ Col$A$.

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

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.)

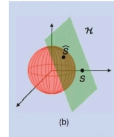
Previous data is the important time series of mean CO_2 in the atmosphere. The data is collected at the Mauna Loa observatory on the island of Hawaii (The Big Island). One of the most important observatories in the world, it is located at the top of the Mauna Kea volcano, 4,205 meters altitude.

Important Examples: Underdetermined Systems (Short/Fat Matrices)

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

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

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



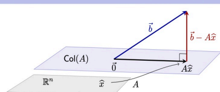
The Normal Equations

Theorem (Normal Equations for Least Squares)

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

$$\underbrace{A^T A\vec{x} = A^T\vec{b}}_{\text{Normal Equations}}$$

Derivation



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

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

Example

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

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

Solution:

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

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

The normal equations $A^T A\vec{x} = A^T\vec{b}$ become:

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

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

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

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \qquad \vec{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Hint: the columns of A are orthogonal.

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

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

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

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

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

Solution. The QR decomposition of A is

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

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 $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

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

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

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

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

6.5 EXERCISES

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

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

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

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

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

In Exercises 5 and 6, describe all least-squares solutions of the equation $A\mathbf{x} = \mathbf{b}$.

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

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

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 $\mathbf{b}$ onto Col A and (b) a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

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

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

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

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

13. Let $A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \\ 3 & 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 11 \\ -9 \\ 5 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 5 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Could $\mathbf{u}$ possibly be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (Answer this without computing a least-squares solution.)

14. Let $A = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 4 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$. Compute $A\mathbf{u}$ and $A\mathbf{v}$, and compare them with $\mathbf{b}$. Is it possible that at least one of $\mathbf{u}$ or $\mathbf{v}$ could be a least-squares solution of $A\mathbf{x} = \mathbf{b}$? (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 $A\mathbf{x} = \mathbf{b}$.

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

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

In Exercises 17 and 18, A is an $m \times n$ matrix and $\mathbf{b}$ 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 $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.

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

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

d. Any solution of $A^TA\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

e. If the columns of A are linearly independent, then the equation $A\mathbf{x} = \mathbf{b}$ has exactly one least-squares solution.

18. a. If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.

b. The least-squares solution of $A\mathbf{x} = \mathbf{b}$ is the point in the column space of A closest to $\mathbf{b}$.

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

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

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 $A\mathbf{x} = \mathbf{b}$ is to compute $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$.

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

a. Show that if $A\mathbf{x} = \mathbf{0}$, then $A^TA\mathbf{x} = \mathbf{0}$.

b. Suppose $A^TA\mathbf{x} = \mathbf{0}$. Explain why $\mathbf{x}^TA^TA\mathbf{x} = 0$, and use this to show that $A\mathbf{x} = \mathbf{0}$.

20. Let A be an $m \times n$ matrix such that A^TA 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.]

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

a. Use Exercise 19 to show that A^TA is an invertible matrix.

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

c. Determine the rank of A.

22. Use Exercise 19 to show that rank A^TA = rank A. [*Hint:* How many columns does A^TA have? How is this connected with the rank of A^TA?]

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

24. Find a formula for the least-squares solution of $A\mathbf{x} = \mathbf{b}$ 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 $\{w_k\}$ into the zero signal, where $y_k = \cos(\pi k/4)$ and $w_k = \cos(3\pi k/4)$. The following calculations will design a filter with approximately those properties. The filter equation is

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

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

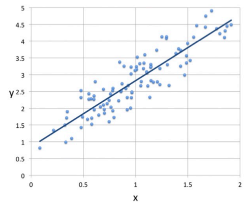
$$\begin{array}{c} \\ k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array} \begin{array}{c} \begin{array}{ccc} y_{k+2} & y_{k+1} & y_k \end{array} \\ \begin{bmatrix} 0 & .7 & 1 \\ -.7 & 0 & .7 \\ -1 & -.7 & 0 \\ -.7 & -1 & -.7 \\ 0 & -.7 & -1 \\ .7 & 0 & -.7 \\ 1 & .7 & 0 \\ .7 & 1 & .7 \end{bmatrix}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

$$\begin{array}{c} \\ k=0 \\ k=1 \\ \vdots \\ \\ \\ \\ \\ k=7 \end{array} \begin{array}{c} \begin{array}{ccc} w_{k+2} & w_{k+1} & w_k \end{array} \\ \begin{bmatrix} 0 & -.7 & 1 \\ .7 & 0 & -.7 \\ -1 & .7 & 0 \\ .7 & -1 & .7 \\ 0 & .7 & -1 \\ -.7 & 0 & .7 \\ 1 & -.7 & 0 \\ -.7 & 1 & -.7 \end{bmatrix} \end{array} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Write an equation $A\mathbf{x} = \mathbf{b}$, where A is a 16×3 matrix formed from the two coefficient matrices above and where $\mathbf{b}$ 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 $A\mathbf{x} = \mathbf{b}$. (The .7 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

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



Topics and Objectives

Topics

1. Least Squares Lines
2. Linear and more complicated models

Learning Objectives

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

1. Apply least-squares and multiple regression to construct a linear model from a set of data points.
2. 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

Course Schedule

Cancellations due to inclement weather will likely result in cancelling review lectures and possibly moving through course

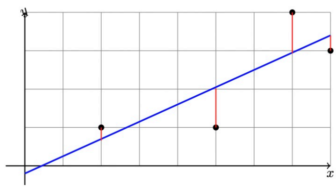
Week	Dates	Mon Lecture	Tue Studio	Wed Lecture	Thu Studio	Fri Lecture
1	8/21 – 8/25	1.1	WS1.1	1.2	WS1.2	1.3
2	8/28 – 9/1	1.4	WS1.3,1.4	1.5	WS1.5	1.7
3	9/4 – 9/8	Break	WS1.7	1.8	WS1.8	1.9
4	9/11 – 9/15	2.1	WS1.9,2.1	**Exam 1**, Review	Cancelled	2.2
5	9/18 – 9/22	2.3,2.4	WS2.2,2.3	2.5	WS2.4,2.5	2.8
6	9/25 – 9/29	2.9	WS2.8,2.9	3.1,3.2	WS3.1,3.2	3.3
7	10/2 – 10/6	4.9	WS3.3,4.9	5.1,5.2	WS5.1,5.2	5.2
8	10/9 – 10/13	Break	Break	**Exam 2**, Review	Cancelled	5.3
9	10/16 – 10/20	5.3	WS5.3	5.5	WS5.5	6.1
10	10/23 – 10/27	6.1,6.2	WS6.1	6.2	WS6.2	6.3
11	10/30 – 11/3	6.4	WS6.3,6.4	6.4,6.5	WS6.4,6.5	6.5
12	11/6 – 11/10	6.6	WS6.5,6.6	**Exam 3**, Review	Cancelled	PageRank
13	11/13 – 11/17	7.1	WSPageRank	7.2	WS7.1,7.2	7.3
14	11/20 – 11/24	7.3,7.4	WS7.2,7.3	Break	Break	Break
15	11/27 – 12/1	7.4	WS7.3,7.4	7.4	WS7.4	7.4
16	12/4 – 12/8	Last lecture	Last Studio	Reading Period		
17	12/11 – 12/15	Final Exams: **MATH 1554 Common Final Exam Tuesday, December 12th at 6pm**				

The Least Squares Line

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

1. Black circles are data.
2. Blue line is the **least squares** line.
3. Lengths of red lines are the ___________.

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



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 : $X\vec{\beta} = \vec{y}$.

The normal equations are

$$X^TX = \begin{bmatrix}1 & 1 & 1 & 1\\ & & & \end{bmatrix}\begin{bmatrix}1 & \\ 1 & \\ 1 & \\ 1 & \end{bmatrix} = \begin{bmatrix}4 & 22\\ 22 & 142\end{bmatrix}$$

$$X^T\vec{y} = \begin{bmatrix}1 & 1 & 1 & 1\\ & & & \end{bmatrix}\begin{bmatrix} \\ \\ \\ \end{bmatrix} = \begin{bmatrix}9\\ 59\end{bmatrix}$$

So the least-squares solution is given by

$$\begin{bmatrix}4 & 22\\ 22 & 142\end{bmatrix}\begin{bmatrix}\beta_0\\ \beta_1\end{bmatrix} = \begin{bmatrix}9\\ 59\end{bmatrix}$$

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

As we may have guessed, β_0 is negative, and β_1 is positive.

Least Squares Fitting for Other Curves

We can consider least squares fitting for the form

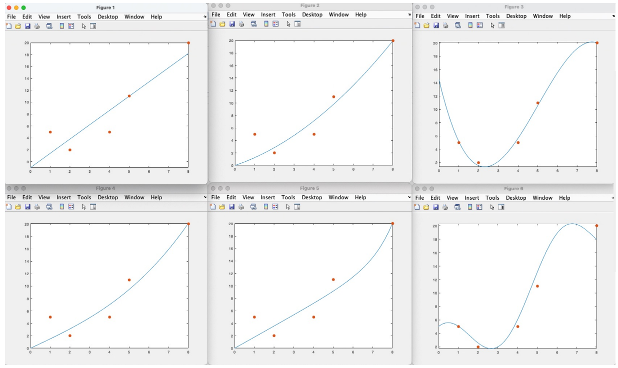
$$y = \beta_0 + \beta_1 f_1(x) + \beta_1 f_2(x) + \cdots + \beta_k f_k(x).$$

where the functions f_j are known. Should have only a few functions!
Keep in mind this is a **linear problem in the** β **variables.**

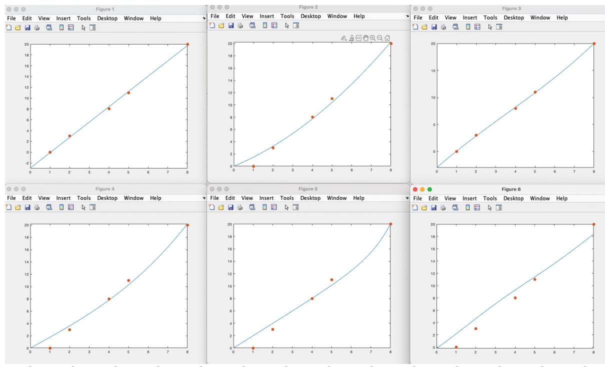
```
clc
format shortG

%% dataÅ
data=[1 5 ; 2 2 ; 4 5 ; 5 11 ; 8 20];
%data=[1 0 ; 2 3 ; 4 8 ; 5 11 ; 8 20];
%data=[1 1 ; 2 1 ; 4 8 ; 5 11 ; 8 20];
% x-values in column 1
x=data(:,1);
% y-values in column 2
y=data(:,2);
% plot points with 28pt dots
%plotdata=plot(x,y,'.','markersize',28);

%% first model: linear y=alpha*x+beta
A=[x(1) 1 ; x(2) 1; x(3) 1 ; x(4) 1 ; x(5) 1];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f1=@(x) alpha.*x+beta;
errorLine=sqrt((A*xhat-b)'*(A*xhat-b))

%% second model: quadratic y=alpha*x+beta*x^2
A=[x(1) x(1)^2 ; x(2) x(2)^2 ; x(3) x(3)^2 ; x(4) x(4)^2 ; x(5) x(5)^2];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f2=@(x) alpha.*x+beta.*x.^2;
errorQuad=sqrt((A*xhat-b)'*(A*xhat-b))

%% third model: general cubic y=a0+a1*x+a2*x^2+a3*x^3
A=[1 x(1) x(1)^2 x(1)^3 ; 1 x(2) x(2)^2 x(2)^3 ; 1 x(3) x(3)^2 x(3)^3 ; 1 x(4) x(4)^2 x(4)^3 ; 1 x(5)
x(5)^2 x(5)^3];
b=y;
xhat=inv(A'*A)*A'*b;
a0=xhat(1);
a1=xhat(2);
a2=xhat(3);
a3=xhat(4);
f3=@(x) a0+a1.*x+a2.*x.^2+a3.*x.^3;
errorGeneralCubic=sqrt((A*xhat-b)'*(A*xhat-b))

%% fourth model: truncated cubic y=a1*x+a3*x^3
A=[x(1) x(1)^3 ; x(2) x(2)^3 ; x(3) x(3)^3 ; x(4) x(4)^3 ; x(5) x(5)^3];
b=y;
xhat=inv(A'*A)*A'*b;
a1=xhat(1);
a3=xhat(2);
f4=@(x) a1.*x+a3.*x.^3;
errorTruncatedCubic=sqrt((A*xhat-b)'*(A*xhat-b))

%% fifth model: linear+exponential y=alpha*x+beta*exp(x)
A=[x(1) exp(x(1)) ; x(2) exp(x(2)) ; x(3) exp(x(3)) ; x(4) exp(x(4)) ; x(5) exp(x(5))];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f5=@(x) alpha.*x+beta.*exp(x);
errorExp=sqrt((A*xhat-b)'*(A*xhat-b))

%% sixth model: linear+cosine y=alpha*x+beta*cos(x)
A=[x(1) cos(x(1)) ; x(2) cos(x(2)) ; x(3) cos(x(3)) ; x(4) cos(x(4)) ; x(5) cos(x(5))];
b=y;
xhat=inv(A'*A)*A'*b;
alpha=xhat(1);
beta=xhat(2);
f6=@(x) alpha.*x+beta.*cos(x);
errorCos=sqrt((A*xhat-b)'*(A*xhat-b))

%% draw figure
figure(1)
fplot(f1,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(2)
fplot(f2,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(3)
fplot(f3,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(4)
fplot(f4,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(5)
fplot(f5,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off

figure(6)
fplot(f6,[0 8])
hold on
plot(x,y,'.','markersize',28)
hold off
```

https://piazza.com/class/lldv6iezpqh7jy/post/776

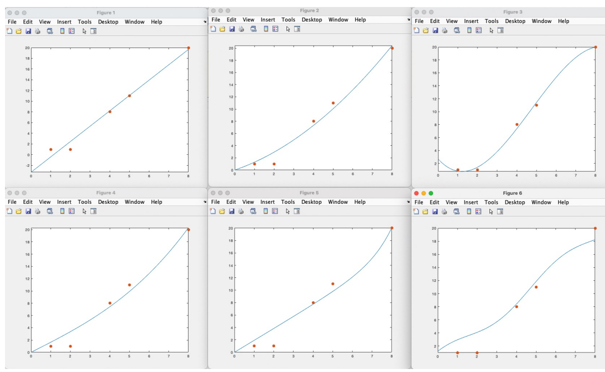
```
data =

     1     5
     2     2
     4     5
     5    11
     8    20

data =

     1     0
     2     3
     4     8
     5    11
     8    20

data =

     1     1
     2     1
     4     8
     5    11
     8    20
```



```
A =

     1     1
     2     1
     4     1
     5     1
     8     1

xhat =

      2.4667
     -1.0667

errorLine =

      3.5024
```

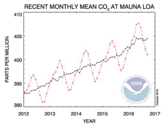


```
A =

     1     1
     2     4
     4    16
     5    25
     8    64

xhat =

      1.3882
     0.13705

errorQuad =

      2.4685
```



```
A =

     1     1
     2     8
     4    64
     5   125
     8   512

xhat =

      1.7102
    0.012391

errorCubic =

      2.3115
```

Least Squares Fitting for Other Curves



Black line is yearly CO_2 levels, and the monthly is the red line. To capture seasonality, would need a curve

$$\text{daily } CO_2 = \beta_0 + \beta_1 t + \beta_2 \sin(2\pi \tfrac{t}{12}) + \beta_3 \cos(2\pi \tfrac{t}{12})$$

Above, t is time, measured in months.

WolframAlpha and Mathematica Syntax

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`$[\{\{x_1, x_1, y_1\}, \{x_2, x_2, y_2\}, \dots, \{x_n, x_n, y_n\}\}]$

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

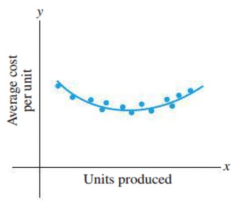


FIGURE 3
Average cost curve.

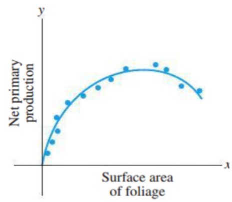


FIGURE 4
Production of nutrients.

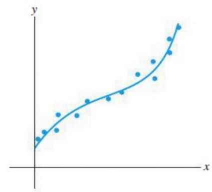


FIGURE 5
Data points along a cubic curve.

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

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

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.)

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

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

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

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.

1. $(0, 1), (1, 1), (2, 2), (3, 2)$
2. $(1, 0), (2, 1), (4, 2), (5, 3)$
3. $(-1, 0), (0, 1), (1, 2), (2, 4)$
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.

   a. Give the design matrix, the observation vector, and the unknown parameter vector.

   b. [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.

   a. 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 $(x_1, y_1), \dots, (x_n, y_n)$.

   b. [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

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

    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 $(15, 18.30)$.

    a. Describe a linear model that can be used to estimate M_A and M_B.

    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, ϑ) of a comet satisfies an equation of the form

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

    where β is a constant and e is the *eccentricity* of the orbit, with $0 \le e < 1$ 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 $\vartheta = 4.6$ (radians).[3]

ϑ	.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.

[3] 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 $\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ and $\overline{y} = \frac{1}{n}(y_1 + \cdots + y_n)$. Show that the least-squares line for the data $(x_1, y_1), \dots, (x_n, y_n)$ must pass through $(\overline{x}, \overline{y})$. That is, show that $\overline{x}$ and $\overline{y}$ satisfy the linear equation $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$. [*Hint:* Derive this equation from the vector equation $\mathbf{y} = X\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$. 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:

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

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

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \textstyle\sum x &= \textstyle\sum y \\ \hat{\beta}_0 \textstyle\sum x + \hat{\beta}_1 \textstyle\sum x^2 &= \textstyle\sum xy \end{aligned} \qquad (7)$$

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 $\hat{\boldsymbol{\beta}}$ of $\mathbf{y} = X\boldsymbol{\beta}$. Consider the following numbers.

(i) $\|X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares of the "regression term." Denote this number by SS(R).

(ii) $\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2$—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 $\|X\hat{\boldsymbol{\beta}}\|^2 = \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$. [*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):

$$\text{SS(E)} = \mathbf{y}^T\mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$$