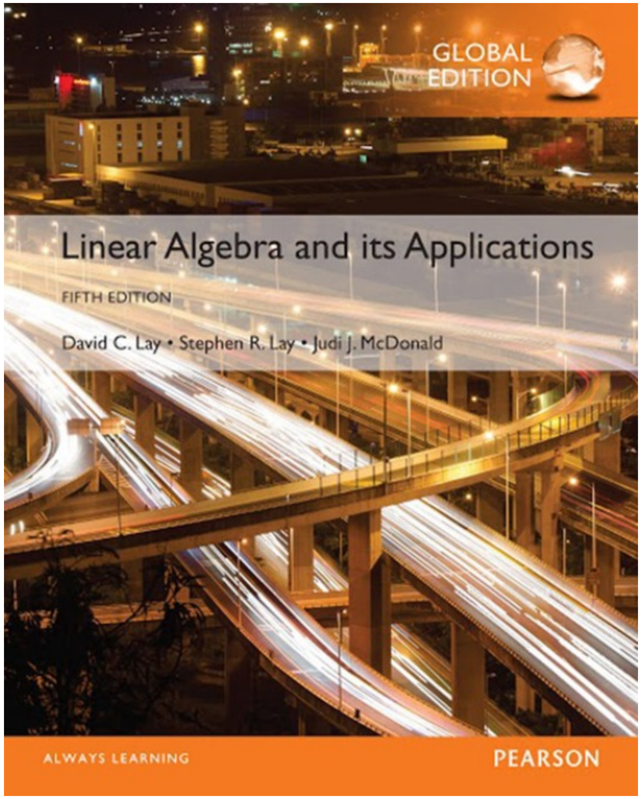
6 Orthogonality and Least Squares

6.5

LEAST-SQUARES PROBLEMS



LEASE-SQUARES PROBLEMS

- **Definition:** If A is $m \times n$ and **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

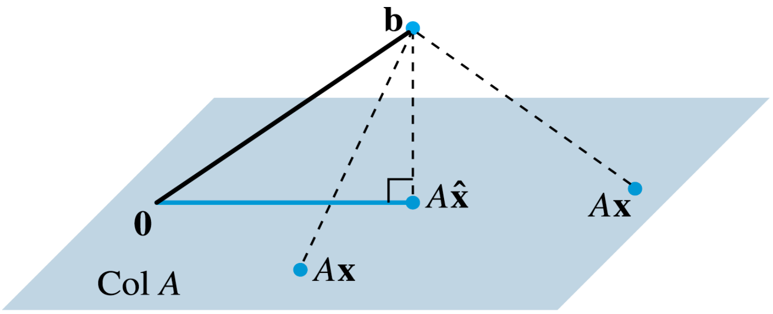
$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$

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

- The most important aspect of the least-squares problem is that no matter what **x** we select, the vector A**x** will necessarily be in the column space, Col A.

- So we seek an **x** that makes A**x** the closest point in Col A to **b**. See the figure on the next slide.

LEAST-SQUARES PROBLEMS



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

- **Solution of the General Least-Squares Problem**

- Given A and **b**, apply the Best Approximation Theorem to the subspace Col A.

- Let

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

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

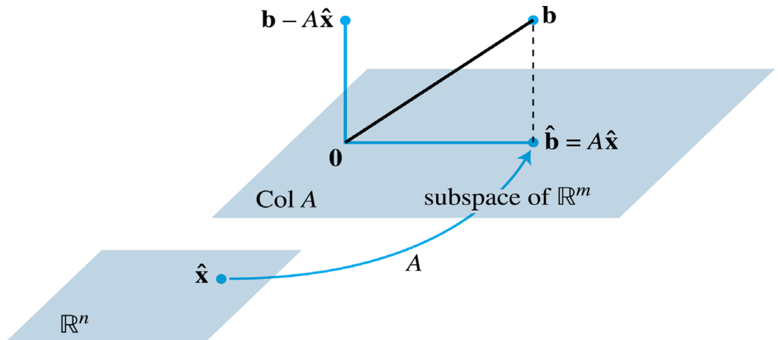
- Because $\hat{\mathbf{b}}$ is in the column space A, the equation $A\mathbf{x} = \hat{\mathbf{b}}$ *is* consistent, and there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$A\hat{\mathbf{x}} = \hat{\mathbf{b}} \tag{1}$$

- Since $\hat{\mathbf{b}}$ is the closest point in Col A to **b**, a vector $\hat{\mathbf{x}}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\hat{\mathbf{x}}$ satisfies (1).

- Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of A. See the figure on the next slide.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM



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

- Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.
- By the Orthogonal Decomposition Theorem, the projection $\hat{b}$ has the property that $b - \hat{b}$ is orthogonal to Col A, so $b - A\hat{x}$ is orthogonal to each column of A.
- If $\mathbf{a}_j$ is any column of A, then $a_j \cdot (b - A\hat{x}) = 0$,and $a_j^T(b - A\hat{x}) = 0$.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \qquad (2)$$

- Thus

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

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

- These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^T A\mathbf{x} = A^T\mathbf{b} \qquad (3)$$

- The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.
- A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Theorem 13:** The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.
- **Proof:** The set of least-squares solutions is nonempty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations.
- Conversely, suppose $\hat{\mathbf{x}}$ satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.
- Then $\hat{\mathbf{x}}$ satisfies (2), which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A.
- Since the columns of A span Col A, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of Col A.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Hence the equation

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

is a decomposition of **b** into the sum of a vector in Col A and a vector orthogonal to Col A.

- By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of **b** onto Col A.

- That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ and $\hat{\mathbf{x}}$ is a least-squares solution.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

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

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

- **Solution:** To use normal equations (3), compute:

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

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

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

- Then the equation $A^T A\mathrm{x} = A^T\mathrm{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Row operations can be used to solve the system on the previous slide, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

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

$$= \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Theorem 14:** Let A be an $m \times n$ matrix. The following statements are logically equivalent:
  a. The equation $A\mathrm{x} = \mathrm{b}$ has a unique least-squares solution for each **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 $\hat{\mathrm{x}}$ is given by

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

- When a least-squares solution $\hat{\mathrm{x}}$ is used to produce $A\hat{\mathrm{x}}$ as an approximation to **b**, the distance from **b** to $A\hat{\mathrm{x}}$ is called the **least-squares error** of this approximation.

PROOF

- The logical equivalence of a) and c) is obvious.
- In the following, we prove that b) and c) are logically equivalent.
- For this, we prove that columns of A are linearly independent if and only if columns of $A^T A$ are linearly independent.
- For this, we prove that equations $A\mathbf{x} = 0$ and $A^T A\mathbf{x} = 0$ have the same set of solutions.
- And for this, we show that $A\mathbf{x} = 0$ yields $A^T A\mathbf{x} = 0$, and $A^T A\mathbf{x} = 0$ yields $A\mathbf{x} = 0$.
  - We have: $A\mathbf{x} = 0$. Multiplying both sides by A^T yields: $A^T A\mathbf{x} = 0A^T = 0$.
  - We have: $A^T A\mathbf{x} = 0$. This yields: $\mathbf{x}^T A^T A\mathbf{x} = 0\mathbf{x}^T = 0 \implies (A\mathbf{x})^T A\mathbf{x} = 0 \implies \|A\mathbf{x}\|^2 = 0 \implies A\mathbf{x} = 0$.

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

- **Theorem 15:** Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR-factorization of A. Then, for each **b** in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

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

Proof: When columns of A are linearly independent, by the previous theorem, the least-square solution $\hat{\mathrm{x}}$ is unique and

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

Replacing A with QR and A^T with R^TQ^T proves the theorem!

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

- **Example 4:** Find a least-squares solution of $A\mathrm{x} = \mathrm{b}$ for

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

- **Solution:** Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{b} = \frac{b \cdot a_1}{a_1 \cdot a_1} a_1 + \frac{b \cdot a_2}{a_2 \cdot a_2} a_2 = \frac{8}{4} a_1 + \frac{45}{90} a_2 \qquad (5)$$

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

- Now that $\hat{\mathrm{b}}$ is known, we can solve $A\hat{\mathrm{x}} = \hat{\mathrm{b}}$.
- But this is trivial, since we already know weights to place on the columns of A to produce $\hat{\mathrm{b}}$.
- It is clear from (5) that

$$\hat{x} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$