Practices before the class (April 14)

- (T/F) If $\mathbf{x}$ is not in a subspace W, then $\mathbf{x}-\operatorname{proj}_W \mathbf{x}$ is not zero.
- (T/F) The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$.
- (T/F) If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution.
- (T/F) 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 $\operatorname{Col} A$.
- (T/F) Any solution of $A^T A\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

Practices before the class (April 14)

- (T/F) If $\mathbf{x}$ is not in a subspace W, then $\mathbf{x}-\operatorname{proj}_W \mathbf{x}$ is not zero. True. If $\mathbf{x}$ is not in a subspace W, then $\mathbf{x}$ cannot equal $\operatorname{proj}_W \mathbf{x}$, because $\operatorname{proj}_W \mathbf{x}$ is in W.
- (T/F) The general least-squares problem is to find an $\mathbf{x}$ that makes $A\mathbf{x}$ as close as possible to $\mathbf{b}$. True.
- (T/F) If $\mathbf{b}$ is in the column space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least-squares solution. True. If $\mathbf{b}$ is in the column space of A, then $\|\mathbf{b} - A\mathbf{x}\|=0$ for $\mathbf{x}$ satisfying the equation $A\mathbf{x} = \mathbf{b}$.
- (T/F) 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 $\operatorname{Col} A$. True. See the notes for § 6.5.
- (T/F) Any solution of $A^T A\mathbf{x} = A^T\mathbf{b}$ is a least-squares solution of $A\mathbf{x} = \mathbf{b}$. True. Check Theorem 13.

6.5 Least-Squares Problems

- Think of $A\mathbf{x}$ as an approximation to $\mathbf{b}$. The smaller the distance between $\mathbf{b}$ and $A\mathbf{x}$, given by $\|\mathbf{b} - A\mathbf{x}\|$, the better the approximation.
- The **general least-squares problem** is to find an $\mathbf{x}$ that makes $\|\mathbf{b} - A\mathbf{x}\|$ as small as possible.
- The adjective "least-squares" arises from the fact that $\|\mathbf{b} - A\mathbf{x}\|$ is the square root of a sum of squares.

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 $\mathbb{R}^n$.

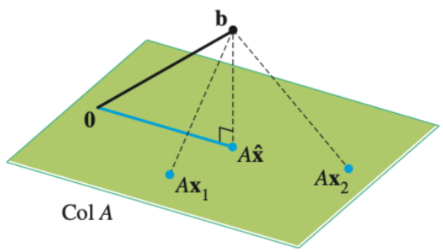

FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Solution of the General Least-Squares Problem

The following steps help us to understand Theorem 13.

- Given A and $\mathbf{b}$ as above, apply the Best Approximation Theorem in Section 6.3 to the subspace $\operatorname{Col} A$. Let

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

- Because $\hat{\mathbf{b}}$ is in the column space of 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 $\operatorname{Col} A$ to $\mathbf{b}$, 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.

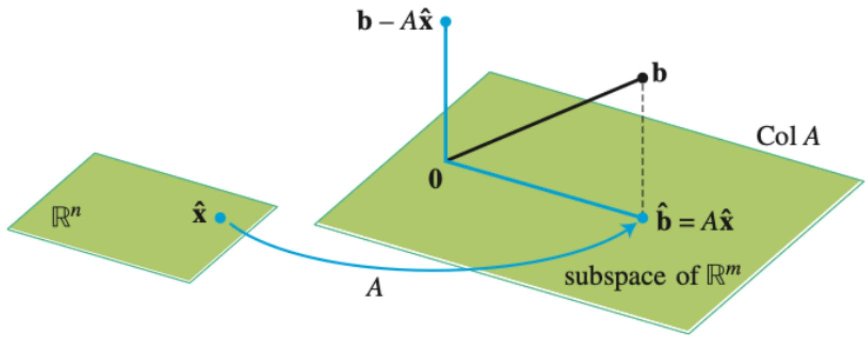

FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

- Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. By the **Orthogonal Decomposition Theorem** in Section 6.3, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to $\operatorname{Col} A$, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of A.
- If $\mathbf{a}_j$ is any column of A, then $\mathbf{a}_j \cdot (\mathbf{b} - A\hat{\mathbf{x}}) = 0$, and $\mathbf{a}_j^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$. Since each $\mathbf{a}_j^T$ is a row of A^T,

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

- Thus

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

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

$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{3}$$

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 $\hat{\mathbf{x}}$.

Theorem 13 The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

Example 1 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 $\hat{\mathbf{x}}$.

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

ANS: (a) The normal equation for $\hat{\vec{x}}$ is

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

We compute

$$A^T A = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -2 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 12 & 8 \\ 8 & 10 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -5 \\ 8 \\ 1 \end{bmatrix} = \begin{bmatrix} -24 \\ -2 \end{bmatrix}$$

Thus the normal equations are

$$\begin{bmatrix} 12 & 8 \\ 8 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -24 \\ -2 \end{bmatrix}$$

The augmented matrix

$$\begin{bmatrix} 12 & 8 & -24 \\ 8 & 10 & -2 \end{bmatrix} \sim \begin{bmatrix} 3 & 2 & -6 \\ 4 & 5 & -1 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 5 \\ 3 & 2 & -6 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 5 \\ 0 & \cancel{-7}\ 1 & \cancel{-21}\ 3 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 3 \end{bmatrix}$$

Thus $\hat{\vec{x}} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ is the least-square solution

Example 2 Describe all least-squares solutions of the equation $A\mathbf{x} = \mathbf{b}$.

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

ANS: The normal equations are $A^TA\vec{x} = A^T\vec{b}$. where

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

$$A^T\vec{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix} = \begin{bmatrix} 14 \\ 4 \\ 10 \end{bmatrix}$$

The augmented matrix

$$\left[\begin{array}{ccc|c} 4 & 2 & 2 & 14 \\ 2 & 2 & 0 & 4 \\ 2 & 0 & 2 & 10 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -1 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Thus

$$\begin{cases} x_1 = 5 - x_3 \\ x_2 = -3 + x_3 \\ x_3 = x_3 \end{cases}$$

$\hat{x} = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ are the least-squares solutions to $A\vec{x} = \vec{b}$

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

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

Alternative Calculations of Least-Squares Solutions

The next example shows how to find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ when the columns of A are orthogonal.

Example 3 Find (a) the orthogonal projection of $\mathbf{b}$ onto $\operatorname{Col} A$ and (b) a least-squares solution of $A\mathbf{x} = \mathbf{b}$.

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

ANS: Because the columns $\vec{a}_1, \vec{a}_2, \vec{a}_3$ for A are orthogonal, the orthogonal projection of $\vec{b}$ onto Col A is

$$\hat{\vec{b}} = \operatorname{proj}_{\operatorname{Col} A} \vec{b} = \frac{\langle \vec{b}, \vec{a}_1 \rangle}{\langle \vec{a}_1, \vec{a}_1 \rangle} \vec{a}_1 + \frac{\langle \vec{b}, \vec{a}_2 \rangle}{\langle \vec{a}_2, \vec{a}_2 \rangle} \vec{a}_2 + \frac{\langle \vec{b}, \vec{a}_3 \rangle}{\langle \vec{a}_3, \vec{a}_3 \rangle} \vec{a}_3$$

$$= \frac{1}{3}\vec{a}_1 + \frac{14}{3}\vec{a}_2 - \frac{5}{3}\vec{a}_3$$

$$= \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} + \frac{14}{3}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{5}{3}\begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 5 \\ 2 \\ 3 \\ 6 \end{bmatrix}$$

(b) We can solve

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

to find the least squares solution

From the above equation

$$\hat{\vec{b}} = \frac{1}{3}\vec{a}_1 + \frac{14}{3}\vec{a}_2 - \frac{5}{3}\vec{a}_3$$

We know the solution $\hat{\vec{x}}$ is obtained from the weights:

$$\hat{\vec{x}} = \begin{bmatrix} 1/3 \\ 14/3 \\ -5/3 \end{bmatrix}$$

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

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

Example 4 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.)

ANS: $A\vec{u} = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -5 \end{bmatrix} = \begin{bmatrix} 3 \\ 8 \\ 2 \end{bmatrix}$

$\vec{b} - A\vec{u} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 8 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}$ and $\|\vec{b} - A\vec{u}\| = \sqrt{24}$

$A\vec{v} = \begin{bmatrix} 2 & 1 \\ -3 & -4 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \\ 8 \end{bmatrix}$

$\vec{b} - A\vec{v} = \begin{bmatrix} 5 \\ 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 7 \\ 2 \\ 8 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \\ -4 \end{bmatrix}$ and $\|\vec{b} - A\vec{v}\| = \sqrt{24}$

Notice that the columns of A are linearly independent so $A\vec{x} = \vec{b}$ has a unique least-square solution by Thm 14.

Since $A\vec{u}$ and $A\vec{v}$ are equally close to $\vec{b}$, and the orthogonal projection is the unique closet point in $\text{Col}A$ to $\vec{b}$. Thus neither $\vec{u}$ nor $\vec{v}$ can be a least-squares solution to $A\vec{x} = \vec{b}$.