Section 6.5 – Least Squares Solutions

Main Ideas in this section:

- How to "solve" $A\mathbf{x} = \mathbf{b}$ when $A\mathbf{x} = \mathbf{b}$ has no solution
- A Least Squares solution is an $\mathbf{x}$ which makes $A\mathbf{x}$ as close to $\mathbf{b}$ as possible... i.e., it's an $\mathbf{x}$ which minimizes $||\mathbf{b} - A\mathbf{x}||$.

• If A is $m \times n$ and $\mathbf{b} \in \mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

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

$\star$ Key ideas:
- $A\mathbf{x} \in col(A)$ for all possible $\mathbf{x}$
- Seeking $\mathbf{x}$ such that $A\mathbf{x}$ is closest to $\mathbf{b}$
- (If $\mathbf{b} \in col(A)$, the LS solution solves $A\mathbf{x} = \mathbf{b}$ exactly.)

Two main approaches to finding LS solutions:

1 Projection of b onto $col(A)$

Let $\hat{\mathbf{b}} = \text{proj}_{col(A)}\mathbf{b}$. Then $\exists\ \mathbf{x} \in \mathbb{R}^n$, call it $\hat{\mathbf{x}}$, such that $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$. This $\hat{\mathbf{x}}$ is the LS solution.

Problem: how to compute $\hat{\mathbf{b}}$*?*

1. **Contrived example** Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

   Find a LS solution to the inconsistent system $A\mathbf{x} = \mathbf{b}$.

   Solution A has orthogonal columns, so the columns of A provide an OG basis for $col(A)$. Thus,

$$\hat{\mathbf{b}} = \left(\frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\right)\mathbf{a}_1 + \left(\frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\right)\mathbf{a}_2$$

2

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

partial proof: Let $\hat{\mathbf{x}}$ be a LS solution.
Then $\mathbf{b} - A\hat{\mathbf{x}}$ is OG to $col(A)$ ...

$\star$ To find a LS solution using the normal equations (NE), compute $A^T A$ and $A^T\mathbf{b}$, then solve the new system $(A^T A)\mathbf{x} = (A^T\mathbf{b})$.

2. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Find a LS solution to the inconsistent system $A\mathbf{x} = \mathbf{b}$.

Solution: Columns of A are not OG, hence no obvious OG basis for $col(A)$, hence no easy computation of $\hat{\mathbf{b}} = \text{proj}_{col(A)}\mathbf{b}$. A couple options:

3. Let $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ -3 \\ 0 \\ 4 \end{bmatrix}$. Find a LS solution to the inconsistent system $A\mathbf{x} = \mathbf{b}$.

Solution: Columns of A are not OG. Use NEs.

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

$$A^T \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 0 \\ 4 \end{bmatrix}$$

$$[A^T A | A^T \mathbf{b}] = \left[\begin{array}{cccc|c} 4 & 2 & 1 & 1 & 3 \\ 2 & 2 & 0 & 0 & -1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 4 \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 4 \\ 0 & 2 & 0 & -2 & -9 \\ 0 & 0 & 1 & -1 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Question: Why did the last example have an infinite number of LS solutions, and the prior example have exactly one LS solution?

Theorem 14 The matrix A^TA is invertible if and only if the columns of A are linearly independent. In this case, the equation $A\mathbf{x} = \mathbf{b}$ has only one least-squares solution $\hat{\mathbf{x}}$, and it is given by

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

Note: this provides a fast solution method when $(A^TA)^{-1}$ is easy to find.

The **least-squares error** is the distance between $\mathbf{b}$ and $A\hat{\mathbf{x}}$.

4. Find a LS solution of $A\mathbf{x} = \mathbf{b}$ for

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

then find the least-squares error of the LS solution.

[Note: the columns of A are orthogonal.]

Summary:

- A LS 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 $col(A)$.
- (Equivalently,) A LS solution to $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that makes $A\hat{\mathbf{x}}$ as close to $\mathbf{b}$ as possible.
- (Equivalently,) A LS solution to $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ that minimizes the distance $||\mathbf{b} - A\mathbf{x}||$, known as the LS error.
- The Normal Equations $A^T A\mathbf{x} = A^T\mathbf{b}$ are often utilized to find LS solutions.
- If the columns of A are OG, the LS solution can be found directly by computing the orthogonal projection $\hat{\mathbf{b}} = \mathrm{proj}_{col(A)}\mathbf{b}$, then solving $A\mathbf{x} = \hat{\mathbf{b}}$.
- For any A and $\mathbf{b}$ of compatible dimensions, there can be many LS solutions or one, depending on whether the columns of A are LD or LI, respectively.