6.5 Least-Squares Problems

Several system of linear equations of the form $A\bar{x} = \bar{b}$ have no solutions (that is, they are inconsistent). When a solution is demanded and none exists, the best one can do is to find an $\bar{x}$ that makes $A\bar{x}$ as close as possible to $\bar{b}$.

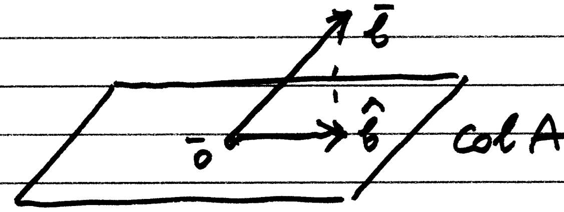
The general least-squares problem is to find an $\bar{x}$ that makes the distance $\|\bar{b} - A\bar{x}\|$ as small as possible.

Definition. If A is an $m \times n$ matrix and $\bar{b}$ is a given vector in R^n, a least-squares solution of the equation $A\bar{x} = \bar{b}$ is a vector $\hat{x}$ in R^n such that $\|\bar{b} - A\hat{x}\| \le \|\bar{b} - A\bar{x}\|$ for all $\bar{x} \in R^n$.

Solution of the General Least-Squares Problem

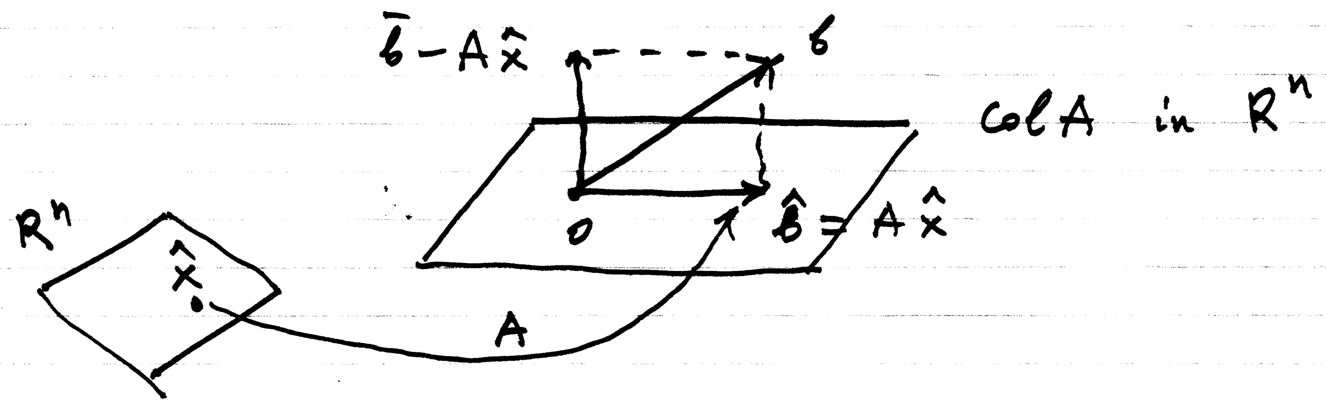
Given A and $\bar{b}$ above, apply the Best Approximation Theorem in Section 6.3 to the subspace Col A.

Let $\hat{b} = \text{proj}_{\text{Col}A}\, \bar{b}$.



Because $\hat{b}$ is in Col A, the equation $A\bar{x} = \hat{b}$ is consistent. Hence there is a vector $\hat{x}$ in R^n such that $A\hat{x} = \hat{b}$.

Since $\hat{b}$ is the closest to $\bar{b}$ vector in Col A, a vector $\hat{x}$ is a least-squares solution of $A\bar{x} = \bar{b}$ if and only if $A\hat{x} = \hat{b}$.



Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$. By the Orthogonal Decomposition Theorem in Section 6.3, the vector $\bar{b} - \hat{b}$ is orthogonal to Col A. So, $\bar{b} - \hat{b}$ is orthogonal to each column of A. If $\bar{a}_j$ denotes any column of A, then

$$\bar{a}_j \cdot (\bar{b} - A\hat{x}) = \bar{a}_j \cdot (\bar{b} - \hat{b}) = 0.$$

Equivalently, $\bar{a}_j^T(\bar{b} - A\hat{x}) = 0$.

Since each $\bar{a}_j^T$ is a row of A^T, we obtain

$$A^T(b - A\hat{x}) = \bar{0}.$$

Thus
$$A^T\bar{b} - A^T A\hat{x} = \bar{0}$$
$$A^T A\hat{x} = A^T\bar{b}. \qquad (*)$$

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

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

Example 1. Find a least-squares solution of the inconsistent system $A\bar{x} = \bar{b}$ for
$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Solution. The system $A\bar{x} = \bar{b}$ is inconsistent because
$$\begin{bmatrix} 4 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & .5 \\ 0 & 1 & 0 \\ 1 & 1 & 11 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & .5 \\ 0 & 1 & 0 \\ 0 & 0 & 10.5 \end{bmatrix}.$$

To use normal equations $A^T A\hat{x} = A^T\bar{b}$, compute:
$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, \quad A^T\bar{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

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

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

Since $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$,

$$\hat{x} = (A^T A)^{-1} A^T \bar{b} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We observe that the matrix $A^T A$ may be singular.

Example 2. Find a least-squares solution of $A\bar{x} = \bar{b}$:

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

Solution. It is easy to compute that

$$A^T A = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}, \quad A^T \bar{b} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}.$$

The augemented matrix for $A^T A \bar{x} = A^T \bar{b}$ is

$$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \sim \cdots \sim \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The general solution is

$x_1 = 3 - x_4, \quad x_2 = -5 + x_4, \quad x_3 = -2 + x_4, \quad x_4$ is free.

So, the general solution of $A^T A \bar{x} = A^T \bar{b}$ is

$$\hat{x} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Theorem 14. Let A be an $m \times n$ matrix. The following assertions are logically equivalent.

a. The equation $A\bar{x} = \bar{b}$ has a unique least-squares solution for each $\bar{b}$ in 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 $\hat{x}$ is given by

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

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

Example 3. For the system $A\bar{x} = \bar{b}$, where

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

determine the least-squares error.

Solution. From Example 1,

$$\bar{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}, \qquad A\hat{x} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}.$$

Hence $$\bar{b} - A\hat{x} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}.$$

$$\|\bar{b} - A\hat{x}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}.$$