6.5 & 6.6 – Least-Squares Problems

Math 220
Warnock - Class Notes

We will now look at the case where $A\mathbf{x}=\mathbf{b}$ has no solution. What would be "closest" possible solution $\mathbf{x}$? This is called the Least-Squares problem, and it mirrors our Best-Approximation Theorem from 6.3.

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 $\mathbb{R}^n$.

$A\vec{x} = \vec{b}$ has no solution

$\hat{b} = \text{proj}_{\text{Col}A} b$ $\quad A\vec{x} = \hat{b}$ is consistent since $\hat{b} \in \text{Col } A$

Let $\hat{x}$ be that solution

$b - \hat{b}$ is orthogonal to Col A (Orthogonal Decomp Theorem in 6.3)

$b - A\hat{x}$ " " " "

For any $\vec{a}_j$ (a column of A)

$$\vec{a}_j \cdot (\vec{b} - A\hat{x}) = 0 \Rightarrow \vec{a}_j^T (b - A\hat{x}) = \vec{0}$$

$$A^T(b - A\hat{x}) = \vec{0} \Rightarrow A^T\vec{b} - A^TA\hat{x} = \vec{0}$$

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

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

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

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

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

$$A^TA=\begin{bmatrix}-1 & 2 & -1\\ 2 & -3 & 3\end{bmatrix}\begin{bmatrix}-1 & 2\\ 2 & -3\\ -1 & 3\end{bmatrix}=\begin{bmatrix}6 & -11\\ -11 & 22\end{bmatrix}$$

$$A^T\vec{b}=\begin{bmatrix}-1 & 2 & -1\\ 2 & -3 & 3\end{bmatrix}\begin{bmatrix}4\\1\\2\end{bmatrix}=\begin{bmatrix}-4\\11\end{bmatrix}$$

$$\left[A^TA\right]^{-1}=\frac{1}{11}\begin{bmatrix}22 & 11\\ 11 & 6\end{bmatrix}=\begin{bmatrix}2 & 1\\ 1 & 6/11\end{bmatrix}$$

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

$$\begin{bmatrix}6 & -11\\ -11 & 22\end{bmatrix}\vec{x}=\begin{bmatrix}-4\\11\end{bmatrix}$$

$$\vec{x}=\underbrace{\begin{bmatrix}2 & 1\\ 1 & 6/11\end{bmatrix}}_{[A^TA]^{-1}}\underbrace{\begin{bmatrix}-4\\11\end{bmatrix}}_{[A^T\vec{b}]}=\boxed{\begin{bmatrix}3\\2\end{bmatrix}}$$

$$A\vec{x}=\begin{bmatrix}1\\0\\3\end{bmatrix}=\hat{b}$$

Ex 2: Find a least-squares solution of the inconsistent system $A\mathbf{x}=\mathbf{b}$ for

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

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

$$A^T\vec{b}=\begin{bmatrix}14\\4\\10\end{bmatrix}$$

$$\begin{bmatrix}4 & 2 & 2 & 14\\ 2 & 2 & 0 & 4\\ 2 & 0 & 2 & 10\end{bmatrix}\to\begin{bmatrix}1 & 0 & 1 & 5\\ 0 & 1 & -1 & -3\\ 0 & 0 & 0 & 0\end{bmatrix}$$

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

$$\hat{X}=\begin{bmatrix}5\\-3\\0\end{bmatrix}+x_3\begin{bmatrix}-1\\1\\1\end{bmatrix}$$

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 **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

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

The distance from **b** to *A***x** is called the least-squares error

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

Ex 3: Find the least-squares error of Ex 1.

$$\|\vec{b} - A\hat{x}\| = \left\| \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \right\| = \sqrt{3^2 + 1^2 + (-1)^2}$$
$$= \sqrt{11}$$

If the columns of A are orthogonal, the least-squares solution is even easier to find.

Ex 4: Verify the columns of A are orthogonal and find a least-squares solution of $A\mathbf{x} = \mathbf{b}$. ✓✓✓

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

$a_1 \cdot a_2 = 0 \quad (1-1)$
$a_1 \cdot a_3 = 0 \quad (-1+1)$
$a_2 \cdot a_3 = 0 \quad (1-1)$

$$\hat{b} = \frac{b \cdot a_1}{a_1 \cdot a_1} a_1 + \frac{b \cdot a_2}{a_2 \cdot a_2} a_2 + \frac{b \cdot a_3}{a_3 \cdot a_3} a_3$$

$$= \frac{2+5-6}{3} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} + \frac{2+6+6}{3} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \frac{-5+6-6}{3} \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/3 + 14/3 + 0 \\ 1/3 \quad 0 \quad + 5/3 \\ 0 + 14/3 - 5/3 \\ -1/3 + 14/3 + 5/3 \end{bmatrix} = \boxed{\begin{bmatrix} 5 \\ 2 \\ 3 \\ 6 \end{bmatrix} = \hat{b}}$$

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

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

$$\hat{x} = \begin{bmatrix} \frac{b \cdot a_1}{a_1 \cdot a_1} \\ \frac{b \cdot a_2}{a_2 \cdot a_2} \\ \frac{b \cdot a_3}{a_3 \cdot a_3} \end{bmatrix}$$

Practice Problems

1. Let $A = \begin{bmatrix} 1 & -3 & -3 \\ 1 & 5 & 1 \\ 1 & 7 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$. Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$, and compute the associated least-squares error.

$\hat{x} = ?$

$A^TA = \begin{bmatrix} 3 & 9 & 0 \\ 9 & 83 & 28 \\ 0 & 28 & 14 \end{bmatrix}$ $\xrightarrow{\text{rref}}$ $\begin{bmatrix} 1 & 0 & -3/2 & 2 \\ 0 & 1 & 1/2 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

let $x_3 = 0$

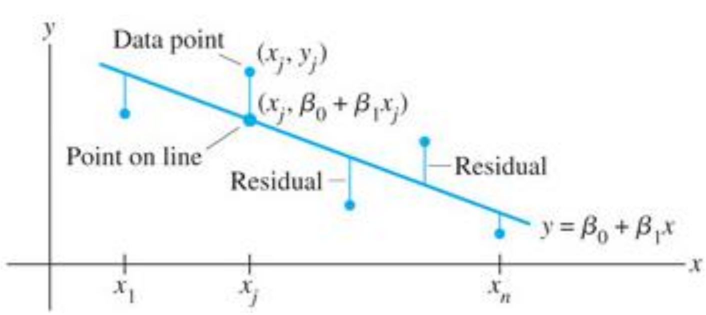
$A^T\vec{b} = \begin{bmatrix} -3 \\ -65 \\ -28 \end{bmatrix}$

$\hat{x} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = \vec{x}$

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

$A\hat{x} = \begin{bmatrix} 5 \\ -3 \\ -5 \end{bmatrix}$ $\quad \|A\hat{x} - b\| = 0 \quad b \in \text{Col} A$

Now we're going to look at finding a best-fit line for a set of data points, also known as linear-regression.



Predicted *y*-value		Observed *y*-value
$\beta_0 + \beta_1 x_1$	$=$	y_1
$\beta_0 + \beta_1 x_2$	$=$	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	$=$	y_n

$$X\beta = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Ex 5: Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points. $(1,1),(4,2),(8,4),(11,5)$

$\beta_0 + \beta_1 x = y$

$$\begin{array}{c} \\ \\ A \end{array}\begin{array}{c} \quad x\text{'s} \\ \begin{bmatrix} 1 & 1 \\ 1 & 4 \\ 1 & 8 \\ 1 & 11 \end{bmatrix} \\ \end{array} \begin{array}{c} \\ \vec{\beta} \\ \vec{x} \end{array} \begin{array}{c} \\ = \\ = \end{array} \begin{array}{c} \\ \begin{bmatrix} 1 \\ 2 \\ 4 \\ 5 \end{bmatrix} \\ \vec{b} \end{array}$$

$$A^T A = \begin{bmatrix} 4 & 24 \\ 24 & 202 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} 12 \\ 96 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 15/29 \\ 12/29 \end{bmatrix}$$

$$\boxed{y = \frac{12}{29} x + \frac{15}{29}}$$

$y = .413793x + .51724$

Ex 6: Find the quadratic regression equation $y = \beta_0 + \beta_1 x + \beta_2 x^2$ of the least-squares line that best fits the data points. $(-2,12),(-1,5),(0,3),(1,2),(2,4)$.

$\beta_0 + \beta_1 x + \beta_2 x^2$

$$\begin{array}{ccc} \beta_0 & \beta_1 & \beta_2 \end{array}$$
$$\begin{bmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \vec{\beta} = \begin{bmatrix} 12 \\ 5 \\ 3 \\ 2 \\ 4 \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 5 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{bmatrix}$$

$$A^T \vec{b} = \begin{bmatrix} 26 \\ -19 \\ 71 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 87/35 \\ -19/10 \\ +19/14 \end{bmatrix}$$

$$\underline{y = \frac{19}{14} x^2 - \frac{19}{10} x + \frac{87}{35}}$$

The General Linear Model

In some applications, it is necessary to fit data points with something other than a straight line. In the examples that follow, the matrix equation is still $X\beta = \mathbf{y}$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** $\in$, defined by $\in = \mathbf{y} - X\beta$, and write

$$\mathbf{y} = X\beta + \in$$

Any equation of this form is referred to as a **linear model**. Once X and $\mathbf{y}$ are determined, the goal is to minimize the length of $\in$, which amounts to finding a least-squares solution of $X\beta = \mathbf{y}$. In each case, the least-squares solution $\widehat{\beta}$ is a solution of the normal equations

$$X^T X\beta = X^T\mathbf{y}$$

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

$$y = A\cos x + B\sin x$$

$$\vec{\beta} = \begin{bmatrix} A \\ B \end{bmatrix}$$

$$\begin{array}{cc} A & B \end{array}$$

$$\begin{bmatrix} \cos 1 & \sin 1 \\ \cos 2 & \sin 2 \\ \cos 3 & \sin 3 \end{bmatrix} \quad \vec{\beta} = \begin{bmatrix} 7.9 \\ 5.4 \\ -.9 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} 2.3421 \\ 7.4475 \end{bmatrix}$$

$$\boxed{y = 2.3421\cos x + 7.4475\sin x}$$

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