Math 311: Topics in Applied Math 1
5: Orthogonality
5.3: Least Squares Problems

Summary

- Recall an *overdetermined* system: $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is an $m \times n$ matrix with $m > n$ (more equations than unknowns), $\mathbf{x} \in \mathbb{R}^n$ is an unknown vector, and $\mathbf{b} \in \mathbb{R}^m$ is a given vector. Such systems are generally inconsistent. They have no (exact) solution.
- The **residual** is $r(\mathbf{x}) = \mathbf{b} - \mathbf{Ax}$. The distance between $\mathbf{b}$ and $\mathbf{Ax}$ is $\|\mathbf{b} - \mathbf{Ax}\| = \|r(\mathbf{x})\|$.
- We'd like to find $\mathbf{x} \in \mathbb{R}^n$ for which $\|r(\mathbf{x})\|$ is a minimum or (equivalently) $\|r(\mathbf{x})\|^2$ is a minimum. A vector $\hat{\mathbf{x}}$ that accomplishes this is a **least squares solution** of the system $\mathbf{Ax} = \mathbf{b}$.
- Moreover, $\mathbf{p} = \mathbf{A}\hat{\mathbf{x}}$ is a vector in $R(\mathbf{A})$, the column space of $\mathbf{A}$, that is closest to $\mathbf{b}$. The vector $\mathbf{p}$ is the **projection** of $\mathbf{b}$ onto $R(\mathbf{A})$.
- THEOREM: Let S be a subspace of $\mathbb{R}^m$. For each $\mathbf{b} \in \mathbb{R}^m$, there is a unique $\mathbf{p} \in S$ that is closest to $\mathbf{b}$; i.e., $\|\mathbf{b} - \mathbf{y}\| > \|\mathbf{b} - \mathbf{p}\|$ for any $\mathbf{y} \neq \mathbf{p}$ in S. Such a $\mathbf{p} \in S$ will be closest to $\mathbf{b}$ if and only if $\mathbf{b} - \mathbf{p} \in S^\perp$, the orthogonal complement of S.
- By this theorem, $\hat{\mathbf{x}}$ is a least squares solution if and only if $\mathbf{b} - \mathbf{p} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} = r(\hat{\mathbf{x}}) \in R(A)^\perp = N\left(\mathbf{A}^T\right)$. Equivalently, $\mathbf{0} = \mathbf{A}^T r(\hat{\mathbf{x}}) = \mathbf{A}^T(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}})$, whence we must solve $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$, an $n \times n$ system of linear equations called the **normal equations**.
- THEOREM: If $\mathbf{A}$ is an $m \times n$ matrix of rank n, then the normal equations $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ have a unique solution $\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$, which is the unique least squares solution of the system $\mathbf{Ax} = \mathbf{b}$.
- The projection vector $\mathbf{p} = \mathbf{A}\hat{\mathbf{x}} = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$ is the vector in $R(\mathbf{A})$ closest to $\mathbf{b}$ in the least squares sense. The matrix $\mathbf{P} = \mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T$ is called the **projection matrix**.

Examples

243/1c

Find the least squares solution of the system

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 4 \\ -x_1 & + & x_2 & + & x_3 & = & 0 \\ & - & x_2 & + & x_3 & = & 1 \\ x_1 & & & + & x_3 & = & 2 \end{array}$$

by identifying the coeffient matrix $\mathbf{A}$ and the right-hand side vector $\mathbf{b}$, then applying the second theorem.

Solution

- The coefficient matrix and right-hand side vector are

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 2 \end{bmatrix}.$$

- Via the second theorem, the least squares solution is

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b} = \begin{bmatrix} 8/5 \\ 3/5 \\ 6/5 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 0.6 \\ 1.2 \end{bmatrix},$$

by entering the matrix expression on a calculator once $\mathbf{A}$ and $\mathbf{b}$ have been assigned. (We can also obtain the solution by obtaining the reduced row echelon form of the augmented system matrix $\left[\mathbf{A}^T\mathbf{A} \middle| \mathbf{A}^T\mathbf{b}\right]$, which is more efficient and more general. See 243/3b below in this regard.)

243/2c

For the least squares solution in 243/1c,
(a) determine the projection $\mathbf{p} = \mathbf{A}\hat{\mathbf{x}}$.
(b) calculate the residual $r(\hat{\mathbf{x}}) = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$.
(c) verify that $r(\hat{\mathbf{x}}) \in N\left(\mathbf{A}^T\right)$.

Solution

Again, use your calculator!

- (a) The projection is

$$\mathbf{p} = \mathbf{A}\hat{\mathbf{x}} = \begin{bmatrix} 17/5 \\ 1/5 \\ 3/5 \\ 14/5 \end{bmatrix} = \begin{bmatrix} 3.4 \\ 0.2 \\ 0.6 \\ 2.8 \end{bmatrix}.$$

- (b) The residual is

$$r(\hat{\mathbf{x}}) = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} 3/5 \\ -1/5 \\ 2/5 \\ -4/5 \end{bmatrix} = \begin{bmatrix} 0.60 \\ -0.20 \\ 0.40 \\ -0.80 \end{bmatrix}.$$

- (c) Finally, we verify that the residual is in the null space of $\mathbf{A}^T$.

$$\mathbf{A}^T r(\hat{\mathbf{x}}) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

243/3b

For the system $\mathbf{Ax} = \mathbf{b}$, find all least squares solutions.

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 3 \\ -1 & 3 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} -2 \\ 0 \\ 8 \end{bmatrix}$$

Solution

- Let's form the augmented system matrix for the normal equations, $\left[\mathbf{A}^T\mathbf{A} \middle| \mathbf{A}^T\mathbf{b}\right]$, then obtain its reduced row echelon form $\mathbf{U}$.

$$\begin{bmatrix} 3 & 0 & 6 & 6 \\ 0 & 14 & 14 & 14 \\ 6 & 14 & 26 & 26 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{U}$$

  Since $\mathbf{U}$ is row equivalent to the augmented system matrix, we conclude that least squares solutions are of the form

$$\hat{\mathbf{x}} = \begin{bmatrix} 2-2t \\ 1-t \\ t \end{bmatrix}, \quad t \in \mathbb{R}.$$

243/4b

For the least squares solution in 243/3b,
(a) determine the projection $\mathbf{p} = \mathbf{A}\hat{\mathbf{x}}$ of $\mathbf{b}$ onto $R(\mathbf{A})$.
(b) calculate the residual $r(\hat{\mathbf{x}}) = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}}$.
(c) verify that $r(\hat{\mathbf{x}}) \in N\left(\mathbf{A}^T\right)$; i.e., that $\mathbf{b} - \mathbf{p}$ is orthogonal to each of the column vectors of $\mathbf{A}$.

Solution

- (a) The projection is

$$\mathbf{p} = \mathbf{A}\hat{\mathbf{x}} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}.$$

- (b) The residual is

$$r(\hat{\mathbf{x}}) = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} -5 \\ -1 \\ 4 \end{bmatrix}.$$

- (c) We verify that the residual is in the null space of $\mathbf{A}^T$; i.e., that $\mathbf{b} - \mathbf{p}$ is orthogonal to each of the column vectors of $\mathbf{A}$.

$$\mathbf{A}^T r(\hat{\mathbf{x}}) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

243/7

Given a collection of points $\{(x_k, y_k)\}_{k=1}^n$, let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \bar{x} = \frac{1}{n}\sum_{k=1}^n x_k, \quad \bar{y} = \frac{1}{n}\sum_{k=1}^n y_k$$

and let $y = c_0 + c_1 x$ be the linear function that gives the best least squares fit to the points. If $\bar{x} = 0$, show that

$$c_0 = \bar{y}, \qquad c_1 = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x}}.$$

Solution

For brevity, let Σ signify $\sum_{k=1}^n$. The linear system is

$$\begin{aligned} \mathbf{Ac} &= \mathbf{y} \\ \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} &= \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \end{aligned}$$

The normal equations are

$$\begin{aligned} \mathbf{A}^T\mathbf{Ac} &= \mathbf{A}^T\mathbf{y} \\ \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} &= \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \\ \begin{bmatrix} n & \Sigma x_k \\ \Sigma x_k & \Sigma x_k^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} &= \begin{bmatrix} \Sigma y_k \\ \Sigma x_k y_k \end{bmatrix}, \text{ whence} \\ \begin{bmatrix} 1 & \frac{1}{n}\Sigma x_k \\ \frac{1}{n}\Sigma x_k & \frac{1}{n}\Sigma x_k^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{n}\Sigma y_k \\ \frac{1}{n}\Sigma x_k y_k \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{n}\Sigma x_k^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} &= \begin{bmatrix} \frac{1}{n}\Sigma y_k \\ \frac{1}{n}\Sigma x_k y_k \end{bmatrix}, \end{aligned}$$

since $\frac{1}{n}\Sigma x_k = \bar{x} = 0$.

Therefore, $c_0 = \frac{1}{n}\Sigma y_k = \bar{y}$ and $c_1 = \dfrac{\frac{1}{n}\Sigma x_k y_k}{\frac{1}{n}\Sigma x_k^2} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x}}$.