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

Summary

- Recall an overdetemined system: $Ax = b$ where $A$ is an $m \times n$ matrix with $m > n$ (more equations than unknowns), $x \in \mathbb{R}^n$ is an unknown vector, and $b \in \mathbb{R}^m$ is a given vector. Such systems are generally inconsistent. They have no (exact) solution.

- The residual is $r(x) = b - Ax$. The distance between $b$ and $Ax$ is $\|b - Ax\| = \|r(x)\|$.

- We’d like to find $x \in \mathbb{R}^n$ for which $\|r(x)\|$ is a minimum or (equivalently) $\|r(x)\|^2$ is a minimum. A vector $\hat{x}$ that accomplishes this is a least squares solution of the system $Ax = b$.

- Moreover, $p = A\hat{x}$ is a vector in $R(A)$, the column space of $A$, that is closest to $b$. The vector $p$ is the projection of $b$ onto $R(A)$.

- THEOREM: Let $S$ be a subspace of $\mathbb{R}^m$. For each $b \in \mathbb{R}^m$, there is a unique $p \in S$ that is closest to $b$; i.e., $\|b - y\| > \|b - p\|$ for any $y \neq p$ in $S$. Such a $p \in S$ will be closest to $b$ if and only if $b - p \in S^\perp$, the orthogonal complement of $S$.

- By this theorem, $\hat{x}$ is a least squares solution if and only if $b - \hat{x} = b - A\hat{x} = r(\hat{x}) \in R(A)^\perp = N(A^T)$.
- Equivalently, $0 = A^T r(\hat{x}) = A^T (b - A\hat{x})$, whence we must solve $A^T Ax = A^T b$, an $n \times n$ system of linear equations called the normal equations.

- THEOREM: If $A$ is an $m \times n$ matrix of rank $n$, then the normal equations $A^T Ax = A^T b$ have a unique solution $\hat{x} = (A^T A)^{-1} A^T b$, which is the unique least squares solution of the system $Ax = b$.

- The projection vector $p = A\hat{x} = A(A^T A)^{-1} A^T b$ is the vector in $R(A)$ closest to $b$ in the least squares sense. The matrix $P = A(A^T A)^{-1} A^T$ is called the projection matrix.

Examples

243/1c

Find the least squares solution of the system

\[
\begin{align*}
&x_1 + x_2 + x_3 = 4 \\
&-x_1 + x_2 + x_3 = 0 \\
&-x_2 + x_3 = 1 \\
&x_1 + x_3 = 2
\end{align*}
\]

by identifying the coefficient matrix $A$ and the right-hand side vector $b$, then applying the second theorem.

Solution

- The coefficient matrix and right-hand side vector are

\[
A = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 2 \end{bmatrix}.
\]

- Via the second theorem, the least squares solution is

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

by entering the matrix expression on a calculator once $A$ and $b$ have been assigned. (We can also obtain the solution by obtaining the reduced row echelon form of the augmented system matrix $[A^T A A^T b]$, 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 $p = A\hat{x}$.
(b) calculate the residual $r(\hat{x}) = b - A\hat{x}$.
(c) verify that $r(\hat{x}) \in N(A^T)$.

Solution

Again, use your calculator!

- (a) The projection is

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

- (b) The residual is

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

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

\[
A^T r(\hat{x}) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.
\]
For the system $Ax = b$, find all least squares solutions.

\[
A = \begin{bmatrix}
1 & 1 & 3 \\
-1 & 3 & 1 \\
1 & 2 & 4
\end{bmatrix}, \quad b = \begin{bmatrix}
-2 \\
0 \\
8
\end{bmatrix}
\]

**Solution**

- Let’s form the augmented system matrix for the normal equations, \([AT] \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} b \end{bmatrix} \), then obtain its reduced row echelon form $U$.

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

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

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

For the least squares solution in $\frac{243}{3}b$, (a) determine the projection $p = A\hat{x}$ of $b$ onto $R(A)$.

(b) calculate the residual $r(\hat{x}) = b - A\hat{x}$.

(c) verify that $r(\hat{x}) \in N(A^T)$; i.e., that $b - p$ is orthogonal to each of the column vectors of $A$.

**Solution**

- (a) The projection is

\[
p = A\hat{x} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}.
\]

- (b) The residual is

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

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

\[
A^Tr(\hat{x}) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\]

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

\[
c_0 = \bar{y}, \quad c_1 = \frac{\mathbf{x}^T \mathbf{y}}{\mathbf{x}^T \mathbf{x}}.
\]

**Solution**

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

\[
Ac = y
\]

\[
\begin{bmatrix}
1 & \cdots & 1 \\
1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
\vdots \\
x_n
\end{bmatrix}
\begin{bmatrix}
1 & \cdots & 1 \\
1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix} = \begin{bmatrix}
y_0 \\
y_1 \\
\vdots \\
y_n
\end{bmatrix},
\]

The normal equations are

\[
A^TAc = A^Ty
\]

\[
\begin{bmatrix}
1 & \cdots & 1 \\
1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
\vdots \\
x_n
\end{bmatrix}
\begin{bmatrix}
1 & \cdots & 1 \\
1 & \cdots & 1
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix} = \begin{bmatrix}
y_0 \\
y_1 \\
\vdots \\
y_n
\end{bmatrix},
\]

whence

\[
\begin{bmatrix}
\frac{1}{n} \sum x_k \\
\frac{1}{n} \sum x_k^2
\end{bmatrix}
\begin{bmatrix}
x_0 \\
x_1 \\
\vdots \\
x_n
\end{bmatrix} = \begin{bmatrix}
\frac{1}{n} \sum y_k \\
\frac{1}{n} \sum y_k x_k
\end{bmatrix}.
\]

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

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

2