More on orthogonality

Example 54. (review) Find the least squares solution to $A\boldsymbol{x}=\boldsymbol{b}$, where

$$A=\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{b}=\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Solution. First, $A^TA=\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}=\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$ and $A^T\boldsymbol{b}=\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}=\begin{bmatrix} 19 \\ 11 \end{bmatrix}$.

Hence, the normal equations $A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$ take the form $\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}\hat{\boldsymbol{x}}=\begin{bmatrix} 19 \\ 11 \end{bmatrix}$. Solving, we find $\hat{\boldsymbol{x}}=\begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

Check. The error $A\hat{\boldsymbol{x}}-\boldsymbol{b}=\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ is indeed orthogonal to $\operatorname{col}(A)$. Because $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}=0$ and $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}\cdot\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}=0$.

Orthogonal projections

The **(orthogonal) projection** $\hat{\boldsymbol{b}}$ of a vector $\boldsymbol{b}$ onto a subspace W is the vector in W closest to $\boldsymbol{b}$.

We can compute $\hat{\boldsymbol{b}}$ as follows:

- Write $W=\operatorname{col}(A)$ for some matrix A.
- Then $\hat{\boldsymbol{b}}=A\hat{\boldsymbol{x}}$ where $\hat{\boldsymbol{x}}$ is a least squares solution to $A\boldsymbol{x}=\boldsymbol{b}$. (i.e. $\hat{\boldsymbol{x}}$ solves $A^TA\hat{\boldsymbol{x}}=A^T\boldsymbol{b}$)

Why? Why is $A\hat{\boldsymbol{x}}$ the projection of $\boldsymbol{b}$ onto $\operatorname{col}(A)$?

Because, if $\hat{\boldsymbol{x}}$ is a least squares solution then $A\hat{\boldsymbol{x}}-\boldsymbol{b}$ is as small as possible (and any element in $\operatorname{col}(A)$ is of the form $A\boldsymbol{x}$ for some $\boldsymbol{x}$).

Note. This is a recipe for computing any orthogonal projection! That's because every subspace W can be written as $\operatorname{col}(A)$ for some choice of the matrix A (take, for instance, A so that its columns are a basis for W).

Assuming A^TA is invertible (which, as discussed in the lemma below, is automatically the case if the columns of A are independent), we have $\hat{\boldsymbol{x}}=(A^TA)^{-1}A^T\boldsymbol{b}$ and hence:

(projection matrix) The projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\operatorname{col}(A)$ is (assuming cols of A are independent)

$$\hat{\boldsymbol{b}}=\underbrace{A(A^TA)^{-1}A^T}_{P}\boldsymbol{b}.$$

The matrix $P=A(A^TA)^{-1}A^T$ is the **projection matrix** for projecting onto $\operatorname{col}(A)$.

Example 55.

(a) What is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $W=\operatorname{span}\left\{\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix},\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right\}$?

(b) What is the matrix P for projecting onto $W=\operatorname{span}\left\{\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix},\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}\right\}$?

(c) **(once more)** Using P, what is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto W?

(d) Using P, what is the orthogonal projection of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto W?

Armin Straub
straub@southalabama.edu

Solution.

(a) In other words, what is the orthogonal projection of $\boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $\operatorname{col}(A)$ with $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$.

In Example 54, we found that the system $A\boldsymbol{x} = \boldsymbol{b}$ has the least squares solution $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

The projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto $\operatorname{col}(A)$ thus is $A\hat{\boldsymbol{x}} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Check. The error $\hat{\boldsymbol{b}} - \boldsymbol{b} = \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ needs to be orthogonal to $\operatorname{col}(A)$. Indeed: $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} = 0$ and $\begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = 0$.

(b) Note that $W = \operatorname{col}(A)$ for $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and that $A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$. Thus $(A^T A)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$.

$P = A(A^T A)^{-1} A^T = \frac{1}{84}\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 & -2 & 4 \\ -2 & 17 & 8 \\ 4 & 8 & 5 \end{bmatrix}$

(c) The orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto W is $P\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 & -2 & 4 \\ -2 & 17 & 8 \\ 4 & 8 & 5 \end{bmatrix}\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 84 \\ 84 \\ 63 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$.

Note. Of course, that agrees with what our computations in the first part. Note that computing P is more work than what we did in in the first part. However, after having computed P once, we can easily project many vectors onto W.

(d) The orthogonal projection of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto W is $P\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 & -2 & 4 \\ -2 & 17 & 8 \\ 4 & 8 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 20 \\ -2 \\ 4 \end{bmatrix}$.

Check. The error $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{21}\begin{bmatrix} 20 \\ -2 \\ 4 \end{bmatrix} = \frac{1}{21}\begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$ is indeed orthogonal to both $\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$.

Example 56. (extra)

(a) What is the matrix P for projecting onto $W = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right\}$?

(b) Using the projection matrix, project $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$ onto $W = \operatorname{span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right\}$.

Solution.

(a) Choosing $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, the projection matrix P is $A(A^T A)^{-1} A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$

$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\frac{1}{8}\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 & 2 \\ 2 & -4 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

Comment. We can choose A in any way such that its columns are a basis for W. The final projection matrix will always be the same.

(b) The projection is $\frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix}$.

Check. The error $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 5 \\ 6 \\ 5 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$ is indeed orthogonal to W.

Example 57. (extra)

(a) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\right\}$?

(b) What is the orthogonal projection of $\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ onto $\operatorname{span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\}$?

Solution. (final answer only) The projections are $\left(\frac{11}{6}, \frac{1}{3}, \frac{7}{6}\right)^T$ and $\left(\frac{3}{2}, 0, \frac{3}{2}\right)^T$.

Lemma 58. If the columns of a matrix A are independent, then A^TA is invertible.

Proof. Assume A^TA is not invertible, so that $A^TA\boldsymbol{x} = \boldsymbol{0}$ for some $\boldsymbol{x} \neq \boldsymbol{0}$. Multiply both sides with $\boldsymbol{x}^T$ to get

$$\boldsymbol{x}^TA^TA\boldsymbol{x} = (A\boldsymbol{x})^TA\boldsymbol{x} = \|A\boldsymbol{x}\|^2 = 0,$$

which implies that $A\boldsymbol{x} = 0$. Since the columns of A are independent, this shows that $\boldsymbol{x} = \boldsymbol{0}$. A contradiction! □

Example 59. If P is a projection matrix, then what is P^2?

For instance. For P as in Example 56, $P^2 = \frac{1}{4}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}^2 = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} = P$.

Solution. Can you see why it is always true that $P^2 = P$?

[Recall that P projects a vector onto a space W (actually, $W = \operatorname{col}(P)$). Hence P^2 takes a vector $\boldsymbol{b}$, projects it onto W to get $\hat{\boldsymbol{b}}$, and then projects $\hat{\boldsymbol{b}}$ onto W again. But the projection of $\hat{\boldsymbol{b}}$ onto W is just $\hat{\boldsymbol{b}}$ (why?!), so that P^2 always has the exact same effect as P. Therefore, $P^2 = P$.]

Example 60. True or false? If P is the matrix for projecting onto W, then $W = \operatorname{col}(P)$.

Solution. True!

Why? The columns of P are the projections of the standard basis vectors and hence in W. On the other hand, for any vector $\boldsymbol{w}$ in W, we have $P\boldsymbol{w} = \boldsymbol{w}$ so that $\boldsymbol{w}$ is a combination of the columns of P.

[This may take several readings to digest but do read (or ask) until it makes sense!]

In particular. $\operatorname{rank}(P) = \dim W$ (because, for any matrix, $\operatorname{rank}(A) = \dim\operatorname{col}(A)$)