

More on orthogonality

Example 52. (review) Find the least squares solution to $Ax = b$, where

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

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

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

Check. The error $A\hat{x} - b = \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ is indeed orthogonal to $\text{col}(A)$. 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{b} of a vector b onto a subspace W is the vector in W closest to b .

We can compute \hat{b} as follows:

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

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

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

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

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

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

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

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

Example 53.

- What is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto $W = \text{span} \left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right\}$?
- What is the matrix P for projecting onto $W = \text{span} \left\{ \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \right\}$?
- (once more)** Using P , what is the orthogonal projection of $\begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$ onto W ?
- Using P , what is the orthogonal projection of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ onto W ?

Solution.

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

In Example 52, we found that the system $A\mathbf{x} = \mathbf{b}$ has the least squares solution $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

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

Check. The error $\hat{\mathbf{b}} - \mathbf{b} = \begin{bmatrix} 2 \\ 4 \\ -8 \end{bmatrix}$ needs to be orthogonal to $\text{col}(A)$. 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 $W = \text{col}(A)$ 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 54. (extra)

- (a) What is the matrix P for projecting onto $W = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$?
- (b) Using the projection matrix, project $\begin{bmatrix} 2 \\ 3 \\ 3 \end{bmatrix}$ onto $W = \text{span} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

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 55. (extra)

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

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

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 56. If the columns of a matrix A are independent, then $A^T A$ is invertible.

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

$$\mathbf{x}^T A^T A \mathbf{x} = (A \mathbf{x})^T A \mathbf{x} = \|A \mathbf{x}\|^2 = 0,$$

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

Example 57. If P is a projection matrix, then what is P^2 ?

For instance. For P as in Example 54, $P^2 = \frac{1}{4} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}^2 = \frac{1}{2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} = P$.

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

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

Example 58. True or false? If P is the matrix for projecting onto W , then $W = \text{col}(P)$.

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 \mathbf{w} in W , we have $P \mathbf{w} = \mathbf{w}$ so that \mathbf{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. $\text{rank}(P) = \dim W$ (because, for any matrix, $\text{rank}(A) = \dim \text{col}(A)$)