Review. We can compute the orthogonal projection $\hat{\boldsymbol{b}}$ of $\boldsymbol{b}$ onto W as follows:

- Write $W = \operatorname{col}(A)$, where the columns of A are a basis of W.
  Then, $\hat{\boldsymbol{b}} = A\hat{\boldsymbol{x}}$ where $\hat{\boldsymbol{x}}$ is the least squares solution to $A\boldsymbol{x} = \boldsymbol{b}$ (i.e. $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$).

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 $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$ and hence:

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

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

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

Lemma 51. 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 $A^T A\boldsymbol{x} = \boldsymbol{0}$ for some $\boldsymbol{x} \neq \boldsymbol{0}$. Multiply both sides with $\boldsymbol{x}^T$ to get

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

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

Example 52.

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

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

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

Solution.

(a) 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 $(A^T A)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$.

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

(b) 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 Example 49. Note that computing P is more work than what we did in Example 49. However, after having computed P once, we can easily project many vectors onto W.

(c) 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 53. (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 $A(A^TA)^{-1}A^T = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}^{-1}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}$
$= \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\frac{1}{8}\begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = \frac{1}{8}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 4 & 2 \\ 2 & -4 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

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

For instance. For P as in Example 53, $P^2 = \frac{1}{4}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}^2 = \frac{1}{2}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix} = P$.

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 $\boldsymbol{b}$, 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, $P^2 = P$.]

Projecting onto 1-dimensional spaces

When we project onto a 1-dimensional space $\text{span}\{\boldsymbol{w}\}$, we usually just say that we are projecting onto $\boldsymbol{w}$.

> The (orthogonal) projection of $\boldsymbol{v}$ onto $\boldsymbol{w}$ is $\dfrac{\boldsymbol{w}\cdot\boldsymbol{v}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}$.

Why? Replace $\boldsymbol{b}$ with $\boldsymbol{v}$ and A with $\boldsymbol{w}$ in our general projection matrix formula to get $\boldsymbol{w}(\boldsymbol{w}^T\boldsymbol{w})^{-1}\boldsymbol{w}^T\boldsymbol{v}$, which equals $\dfrac{\boldsymbol{w}\cdot\boldsymbol{v}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}$ (note that $\boldsymbol{w}^T\boldsymbol{v} = \boldsymbol{w}\cdot\boldsymbol{v}$ and $\boldsymbol{w}^T\boldsymbol{w} = \|\boldsymbol{w}\|^2$ are scalars).

Comment. If you have taken Calculus 3, you have seen that formula before. Most likely, you were deriving it using angles at that time. Namely, the dot product has the following connection to angles:

> $\boldsymbol{v}\cdot\boldsymbol{w} = \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|\cos\theta$ where $\theta \in [0, \pi]$ is the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$

Why? You can derive this by repeating what we did, right after Definition 23 to show that $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal if and only if $\boldsymbol{v}\cdot\boldsymbol{w} = 0$. Just replace Pythagoras with the law of cosines ($c^2 = a^2 + b^2 - 2ab\cos\theta$ holds in any triangle!).

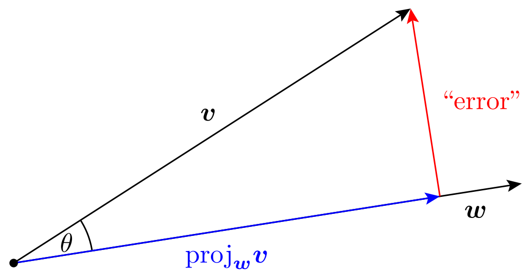
Two obvious cases. Observe that the cases $\theta = 0$ and $\theta = 90°$ are clearly true.

We will not discuss angles much further in this class. Just in case it is helpful, here is the typical argument given in Calculus 3 to determine the projection $\text{proj}_{\boldsymbol{w}}\boldsymbol{v}$ of $\boldsymbol{v}$ onto $\boldsymbol{w}$:

From the sketch, we see that "error" $= \boldsymbol{v} - \mathrm{proj}_{\boldsymbol{w}} \boldsymbol{v}$ and that this error is orthogonal to $\boldsymbol{w}$.

Basic trigonometry tells us that the length of $\mathrm{proj}_{\boldsymbol{w}} \boldsymbol{v}$ is $\|\boldsymbol{v}\| \cos\theta$. Hence:

$$\mathrm{proj}_{\boldsymbol{w}} \boldsymbol{v} = \underbrace{\|\boldsymbol{v}\| \cos\theta}_{\text{length}} \underbrace{\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}}_{\text{direction}} = \frac{\|\boldsymbol{v}\| \|\boldsymbol{w}\| \cos\theta}{\|\boldsymbol{w}\|} \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} = \left(\frac{\boldsymbol{v} \cdot \boldsymbol{w}}{\|\boldsymbol{w}\|^2}\right) \boldsymbol{w}$$



Orthogonal bases

Review. Vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are a **basis** for V.

$\Longleftrightarrow$ $V = \mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ and $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent.

$\Longleftrightarrow$ Any vector $\boldsymbol{w}$ in V can be written as $\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n$ in a unique way.

The latter is the practical reason why we care so much about bases!

V could be some abstract vector space (of polynomials or Fourier series), meaning that vectors are abstract objects and not just our usual column vectors. However, as soon as we pick a basis of V, then we can represent every (abstract) vector $\boldsymbol{w}$ by the (usual) column vector $(c_1, c_2, \ldots, c_n)^T$.

This means all of our results can be used, too, when working with these abstract spaces!

Definition 55. A basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of a vector space V is an **orthogonal basis** if the vectors are (pairwise) orthogonal. If, in addition, the basis vectors have length 1, then this is called an **orthonormal basis**.

Example 56. The standard basis $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ is an orthonormal basis for $\mathbb{R}^3$.

Example 57. Are the vectors $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ an orthogonal basis for $\mathbb{R}^3$? Is it orthonormal?

Solution. $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0$, $\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 0$.

So, this is an orthogonal basis.

On the other hand, the vectors do not all have length 1, so that this basis is not orthonormal.

Note. Orthogonal vectors are always linearly independent (see next class). Here, this certifies that the three vectors are linearly independent (and hence a basis for $\mathbb{R}^3$).

Normalize the vectors to produce an orthonormal basis.

Solution.

$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}} = \sqrt{2} \Longrightarrow$ normalized: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}} = \sqrt{2} \Longrightarrow$ normalized: $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ has length $\sqrt{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}} = 1 \Longrightarrow$ is already normalized: $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

The resulting orthonormal basis is $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

Armin Straub
straub@southalabama.edu