Recall the formula for the orthogonal projection of a $\vec{y}$ in $\mathbb{R}^n$ onto a one dimensional subspace $L$ with basis vector $\vec{u}$:

$$\text{proj}_L \vec{y} = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}} \vec{u}.$$  

We extend this to the orthogonal projection of $\vec{y}$ onto a subspace $W$ to get a unique vector $\vec{p}$ in $W$ such that $\vec{y} - \vec{p}$ is orthogonal to $W$. [Draw the picture in $\mathbb{R}^3$.]

With an orthogonal basis for $W$, the formula for $\vec{p}$ extend that for the orthogonal projection onto a $L$.

**Theorem 8 (The Orthogonal Decomposition Theorem).** Let $W$ be a subspace of $\mathbb{R}^n$. Then each $\vec{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\vec{y} = \vec{p} + \vec{z},$$  

where $\vec{y}$ is in $W$ and $\vec{z}$ is in $W^\perp$. Additionally, if \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\} is an orthogonal basis for $W$, then

$$\vec{p} = \sum_{i=1}^{p} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i,$$

and $\vec{z} = \vec{y} - \vec{p}$.

**Proof.** The projection $\vec{p}$ is in $W$ and so it is a unique linear combination of an orthogonal basis \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\} of $W$ by Theorem 5:

$$\vec{p} = \sum_{i=1}^{p} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i.$$

Set $\vec{z} = \vec{y} - \vec{p}$.

Since $\vec{u}_1$ is orthogonal to $\vec{u}_2, \ldots, \vec{u}_p$, we have that

$$\vec{z} \cdot \vec{u}_1 = (\vec{y} - \vec{p}) \cdot \vec{u}_1 = \vec{y} \cdot \vec{u}_1 - \sum_{i=1}^{p} \left( \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i \right) \cdot \vec{u}_1 = \vec{y} \cdot \vec{u}_1 - \vec{y} \cdot \vec{u}_1 = 0.$$

Similarly, we obtain $\vec{z} \cdot \vec{u}_j = 0$ for all $j = 2, \ldots, p$.

Thus $\vec{z}$ is orthogonal to a spanning set of $W$, and so $\vec{z}$ belongs to $W^\perp$.

To get uniqueness of the decomposition $\vec{y} = \vec{p} + \vec{z}$, we suppose there is another decomposition $\vec{y} = \vec{q} + \vec{w}$ with $\vec{q} \in W$ and $\vec{w} \in W^\perp$.

Since both decompositions equal $\vec{y}$, we have that $\vec{p} - \vec{q} = \vec{w} - \vec{z}$.

Here $\vec{p} - \vec{q}$ is in $W$ while $\vec{w} - \vec{z}$ is in $W^\perp$.

You have it as a homework problem (#31 in §6.1) that there is only one vector that is in both $W$ and $W^\perp$, namely $\vec{0}$.

Thus $\vec{p} - \vec{q} = \vec{0}$ and $\vec{w} - \vec{z} = \vec{0}$, giving the uniqueness. □
Example. Find the orthogonal projection of
\[
\vec{y} = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 6 \end{bmatrix}
\]
onto the subspace \( W \) with orthogonal basis
\[
\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 3 \end{bmatrix}.
\]
We apply the formula from Theorem 8:
\[
\vec{p} = \vec{y} \cdot \vec{u}_1 \cdot \vec{u}_1 + \vec{y} \cdot \vec{u}_2 \cdot \vec{u}_2 + \vec{y} \cdot \vec{u}_3 \cdot \vec{u}_3
\]
\[
= \frac{4}{2} \vec{u}_1 - \frac{10}{10} \vec{u}_2 + \frac{18}{18} \vec{u}_3
\]
\[
= \begin{bmatrix} 0 \\ 4 \\ 2 \\ 4 \end{bmatrix}.
\]
How can we check this answer? Well, the difference
\[
\vec{z} = \vec{y} - \vec{p} = \begin{bmatrix} 1 \\ -1 \\ -2 \\ 2 \end{bmatrix}
\]
is supposedly in \( W^\perp \), and we can check that \( \vec{z} \) is orthogonal to each basis vector of \( W \).

Properties of Orthogonal Projections. We write \( \text{proj}_W \vec{y} \) for the orthogonal projection of \( \vec{y} \) onto the subspace \( W \).

If \( \vec{y} \) is in \( W \), then \( \text{proj}_W \vec{y} = \vec{y} \).

If \( \vec{y} \not\in W \), then \( \text{proj}_W \vec{y} \) is the best approximation of \( \vec{y} \) by vectors in \( W \), in the following sense.

**Theorem 9 (The Best Approximation Theorem).** For \( \vec{y} \) in \( \mathbb{R}^n \) and \( W \) a subspace of \( \mathbb{R}^n \), the projection \( \vec{p} = \text{prof}_W \vec{y} \) is the closest point on \( W \) to \( \vec{y} \), i.e.,
\[
\| \vec{y} - \vec{p} \| < \| \vec{y} - \vec{v} \|
\]
for all \( \vec{v} \) is \( W \) distinct from \( \vec{p} \).

**Proof.** Let \( \vec{v} \) be a vector in \( W \) different from \( \vec{p} \).
Because $\vec{v}$ and $\vec{p}$ belong to the subspace, so does $\vec{p} - \vec{v}$. This means that $\vec{p} - \vec{v}$ is orthogonal to $\vec{y} - \vec{p}$. [Draw the picture.]

We can write

$$\vec{y} - \vec{v} = (\vec{y} - \vec{p}) + (\vec{p} - \vec{v}).$$

The three vectors here are the sides of a right-angle triangle with $\vec{y} - \vec{v}$ as the hypothenuse. The Pythagorean Theorem gives

$$\|\vec{y} - \vec{v}\|^2 = \|\vec{y} - \vec{p}\|^2 + \|\vec{p} - \vec{v}\|^2.$$ 

Since $\vec{v} \neq \vec{p}$, we have $\|\vec{v} - \vec{p}\| > 0$, and the Pythagorean Theorem becomes

$$\|\vec{y} - \vec{v}\|^2 > \|\vec{y} - \vec{p}\|^2.$$ 

Taking square roots give the result. $\square$

We call the vector $\vec{p}$ in Theorem 9 the **best approximation to $\vec{y}$ by elements of $W$**. If we think of $\|\vec{y} - \vec{v}\|$ as the error of using $\vec{v}$ in place of $\vec{y}$, then the error is minimized when $\vec{v} = \vec{p}$.

Since the norm of a vector is determined by a sum of squares, we say that $\|\vec{y} - \vec{p}\|$ has the “least squares” error.

**Example (Continued).** The vector

$$\vec{p} = \begin{bmatrix} 0 \\ 4 \\ 2 \\ 4 \end{bmatrix}$$

in the subspace $W$ with orthogonal basis

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{u}_2 = \begin{bmatrix} 2 \\ -2 \\ 1 \\ -1 \end{bmatrix}, \quad \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 3 \end{bmatrix}$$

is the best approximation of

$$\vec{y} = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 6 \end{bmatrix}$$

by elements of $W$.

The error associated with the best approximation $\vec{p}$ is

$$\|\vec{y} - \vec{p}\| = \| [1 \quad -1 \quad -2 \quad 2]^T \| = \sqrt{10}.$$ 

For all $\vec{v}$ in $W$ different from $\vec{p}$ the error $\|\vec{y} - \vec{v}\| > \sqrt{10}$. / / /
We are seeing how useful an orthogonal basis is for computations. What is even better than an orthogonal basis? An orthonormal basis!

**Theorem 10.** If \( \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\} \) is an orthonormal basis of a subspace \( W \) of \( \mathbb{R}^n \), then for all \( \vec{y} \) in \( \mathbb{R}^n \), we have

\[
\text{proj}_W \vec{y} = \sum_{i=1}^{p} (\vec{y} \cdot \vec{u}_i) \vec{u}_i.
\]

Furthermore, if we set \( U = [\vec{u}_1 \, \vec{u}_2 \, \cdots \, \vec{u}_p] \), then

\[
\text{proj}_W \vec{y} = U U^T \vec{y}.
\]

**Proof.** Because \( \{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_p\} \) is an orthogonal basis we have by Theorem 8 that

\[
\text{proj}_W \vec{y} = \sum_{i=1}^{p} \frac{\vec{y} \cdot \vec{u}_i}{\vec{u}_i \cdot \vec{u}_i} \vec{u}_i.
\]

This simplifies to the desired formula because \( \vec{u}_i \cdot \vec{u}_i = 1 \) for all \( i = 1, 2, \ldots, p \).

The vector \( \text{proj}_W \vec{y} \) is linear combination of the columns of \( U \) where the weights are \( \vec{y} \cdot \vec{u}_i = \vec{u}_i^T \vec{y} \).

The weights are the entries of \( U^T \vec{y} \), and so the formula for \( \text{proj}_W \vec{y} \) becomes \( U U^T \vec{y} \). \( \square \)