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 53. 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 $A^T A\boldsymbol{x} = \mathbf{0}$ for some $\boldsymbol{x} \neq \mathbf{0}$. 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 $\boldsymbol{x} = \mathbf{0}$. A contradiction! □

Example 54.

(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 51. Note that computing P is more work than what we did in Example 51. 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 $\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$.

Armin Straub
straub@southalabama.edu

Example 55. (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 $A(A^TA)^{-1}A^T=\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}\begin{bmatrix}3&1\\1&3\end{bmatrix}^{-1}\begin{bmatrix}1&1&1\\1&-1&1\end{bmatrix}$
$=\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}\frac{1}{8}\begin{bmatrix}3&-1\\-1&3\end{bmatrix}\begin{bmatrix}1&1&1\\1&-1&1\end{bmatrix}=\frac{1}{8}\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}\begin{bmatrix}2&4&2\\2&-4&2\end{bmatrix}=\frac{1}{2}\begin{bmatrix}1&0&1\\0&2&0\\1&0&1\end{bmatrix}$.

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

For instance. For P as in Example 55, $P^2=\frac{1}{4}\begin{bmatrix}1&0&1\\0&2&0\\1&0&1\end{bmatrix}^2=\frac{1}{2}\begin{bmatrix}1&0&1\\0&2&0\\1&0&1\end{bmatrix}=P$.

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, $P^2=P$.]

Projecting onto 1-dimensional spaces

When we project onto a 1-dimensional space $\operatorname{span}\{\boldsymbol{w}\}$, 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 25 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!).

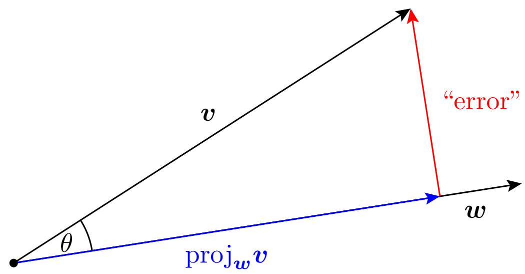
Two obvious cases. Observe that the cases $\theta=0$ and $\theta=90°$ are clearly true.

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 $\operatorname{proj}_{\boldsymbol{w}}\boldsymbol{v}$ of $\boldsymbol{v}$ onto $\boldsymbol{w}$:

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

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

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



Orthogonal bases

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

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

$\iff$ 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 57. 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 58. 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 59. 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 $\frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

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

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

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

Theorem 61. Suppose that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are nonzero and pairwise orthogonal. Then $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are linearly independent.

Proof. Suppose that $c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n = \boldsymbol{0}$. In order to show that $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are independent, we need to show that $c_1 = c_2 = \ldots = c_n = 0$.

Take the dot product of $\boldsymbol{v}_1$ with both sides:

$$\begin{aligned} 0 &= \boldsymbol{v}_1 \cdot (c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n) \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 + c_2 \boldsymbol{v}_1 \cdot \boldsymbol{v}_2 + \ldots + c_n \boldsymbol{v}_1 \cdot \boldsymbol{v}_n \\ &= c_1 \boldsymbol{v}_1 \cdot \boldsymbol{v}_1 = c_1 \|\boldsymbol{v}_1\|^2 \end{aligned}$$

But $\|\boldsymbol{v}_1\| \neq 0$ and hence $c_1 = 0$. Likewise, we find $c_2 = 0$, ..., $c_n = 0$. Hence, the vectors are independent. □

Comment. Note that this result is intuitively obvious: if the vectors were linearly dependent, then one of them could be written as a linear combination of the others. However, all these other vectors (and hence any combination of them) are orthogonal to it.

Orthogonal projections if we have an orthogonal basis

Lemma 62. (orthogonal projection if we have an orthogonal basis)
If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal, then the orthogonal projection of $\boldsymbol{w}$ onto $\mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is

$$\hat{\boldsymbol{w}} = \underbrace{\frac{\boldsymbol{w} \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1}_{\substack{\text{proj of } \boldsymbol{w} \\ \text{onto } \boldsymbol{v}_1}} + \ldots + \underbrace{\frac{\boldsymbol{w} \cdot \boldsymbol{v}_n}{\boldsymbol{v}_n \cdot \boldsymbol{v}_n} \boldsymbol{v}_n}_{\substack{\text{proj of } \boldsymbol{w} \\ \text{onto } \boldsymbol{v}_n}}.$$

Proof. It suffices to show that the error $\boldsymbol{w} - \hat{\boldsymbol{w}}$ is orthogonal to each $\boldsymbol{v}_i$. Indeed:

$$(\boldsymbol{w} - \hat{\boldsymbol{w}}) \cdot \boldsymbol{v}_i = \left(\boldsymbol{w} - \frac{\boldsymbol{w} \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 - \ldots - \frac{\boldsymbol{w} \cdot \boldsymbol{v}_n}{\boldsymbol{v}_n \cdot \boldsymbol{v}_n} \boldsymbol{v}_n \right) \cdot \boldsymbol{v}_i = \boldsymbol{w} \cdot \boldsymbol{v}_i - \frac{\boldsymbol{w} \cdot \boldsymbol{v}_i}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i} \boldsymbol{v}_i \cdot \boldsymbol{v}_i = 0.$$

□

Important consequence. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ is an orthogonal basis of V, and $\boldsymbol{w}$ is in V, then

$$\boldsymbol{w} = c_1 \boldsymbol{v}_1 + \ldots + c_n \boldsymbol{v}_n \qquad \text{with} \quad c_j = \frac{\boldsymbol{w} \cdot \boldsymbol{v}_j}{\boldsymbol{v}_j \cdot \boldsymbol{v}_j}.$$

If the $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are a basis, but not orthogonal, then we have to solve a system of equations to find the c_i. That is a lot more work than simply computing a few dot products.

Note. In other words, $\boldsymbol{w}$ decomposes as the sum of its projections onto each basis vector.

Note. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthonormal, then the denominators are all 1.

Armin Straub
straub@southalabama.edu

Example 63. What is the projection of $\begin{bmatrix}3\\7\\4\end{bmatrix}$ onto $W = \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ with $\boldsymbol{v}_1 = \begin{bmatrix}1\\-1\\0\end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}$?

Comment. We know how to do this using least squares. (Do it for practice!)
However, realizing that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are orthogonal makes things easier.
[Actually, here, it is obvious what the projection is going to be if we realized that W is the x-y-plane.]

Solution. (using orthogonality) Because $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are orthogonal, the projection is

$$\underbrace{\frac{\begin{bmatrix}3\\7\\4\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\0\end{bmatrix}}{\begin{bmatrix}1\\-1\\0\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\0\end{bmatrix}}\begin{bmatrix}1\\-1\\0\end{bmatrix}}_{\text{projection onto } \boldsymbol{v}_1} + \underbrace{\frac{\begin{bmatrix}3\\7\\4\end{bmatrix}\cdot\begin{bmatrix}1\\1\\0\end{bmatrix}}{\begin{bmatrix}1\\1\\0\end{bmatrix}\cdot\begin{bmatrix}1\\1\\0\end{bmatrix}}\begin{bmatrix}1\\1\\0\end{bmatrix}}_{\text{projection onto } \boldsymbol{v}_2} = \frac{-4}{2}\begin{bmatrix}1\\-1\\0\end{bmatrix} + \frac{10}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} = \begin{bmatrix}3\\7\\0\end{bmatrix}.$$

Important note. Note that, at this point, we can easily extend $\begin{bmatrix}1\\-1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}$ to an orthogonal basis of $\mathbb{R}^3$:
That is because the error $\begin{bmatrix}3\\7\\4\end{bmatrix} - \begin{bmatrix}3\\7\\0\end{bmatrix} = \begin{bmatrix}0\\0\\4\end{bmatrix}$ is orthogonal to both of the existing basis vectors.
Therefore $\begin{bmatrix}1\\-1\\0\end{bmatrix}, \begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\4\end{bmatrix}$ is an orthogonal basis of $\mathbb{R}^3$.
This observation underlies the Gram-Schmidt process, which we will discuss next class.

Example 64. Express $\underbrace{\begin{bmatrix}3\\7\\4\end{bmatrix}}_{\boldsymbol{x}}$ in terms of the basis $\underbrace{\begin{bmatrix}1\\-1\\0\end{bmatrix}}_{\boldsymbol{v}_1}, \underbrace{\begin{bmatrix}1\\1\\0\end{bmatrix}}_{\boldsymbol{v}_2}, \underbrace{\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\boldsymbol{v}_3}$.

Solution. Because $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ is an orthogonal basis of $\mathbb{R}^3$, we get (much as in the previous example):

$$\begin{aligned}\begin{bmatrix}3\\7\\4\end{bmatrix} &= c_1\begin{bmatrix}1\\-1\\0\end{bmatrix} + c_2\begin{bmatrix}1\\1\\0\end{bmatrix} + c_3\begin{bmatrix}0\\0\\1\end{bmatrix}\\ &= \underbrace{\frac{\begin{bmatrix}3\\7\\4\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\0\end{bmatrix}}{\begin{bmatrix}1\\-1\\0\end{bmatrix}\cdot\begin{bmatrix}1\\-1\\0\end{bmatrix}}\begin{bmatrix}1\\-1\\0\end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_1} + \underbrace{\frac{\begin{bmatrix}3\\7\\4\end{bmatrix}\cdot\begin{bmatrix}1\\1\\0\end{bmatrix}}{\begin{bmatrix}1\\1\\0\end{bmatrix}\cdot\begin{bmatrix}1\\1\\0\end{bmatrix}}\begin{bmatrix}1\\1\\0\end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_2} + \underbrace{\frac{\begin{bmatrix}3\\7\\4\end{bmatrix}\cdot\begin{bmatrix}0\\0\\1\end{bmatrix}}{\begin{bmatrix}0\\0\\1\end{bmatrix}\cdot\begin{bmatrix}0\\0\\1\end{bmatrix}}\begin{bmatrix}0\\0\\1\end{bmatrix}}_{\text{projection of } \boldsymbol{x} \text{ onto } \boldsymbol{v}_3}\\ &= \frac{-4}{2}\begin{bmatrix}1\\-1\\0\end{bmatrix} + \frac{10}{2}\begin{bmatrix}1\\1\\0\end{bmatrix} + \frac{4}{1}\begin{bmatrix}0\\0\\1\end{bmatrix}\end{aligned}$$

Because we spelled out all the details this looks more involved than it is. We only computed 6 dot products!

Alternative. We could have solved $\begin{bmatrix}1 & 1 & 0\\-1 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}3\\7\\4\end{bmatrix}$ to also find $\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}-2\\5\\4\end{bmatrix}$.

The numbers are particularly easy here but in general, to find this solution, we have to go through the entire process of Gaussian elimination. On the other hand, if we have an orthogonal basis, the former approach requires less work, because it is just computing a few dot products.

Example 65. Express $\begin{bmatrix}3\\7\\4\end{bmatrix}$ in terms of the basis $\begin{bmatrix}1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix}$.

Solution. This is not an orthogonal basis, so we cannot proceed as in the previous example.
To write $\begin{bmatrix}3\\7\\4\end{bmatrix} = c_1\begin{bmatrix}1\\1\\0\end{bmatrix} + c_2\begin{bmatrix}0\\1\\1\end{bmatrix} + c_3\begin{bmatrix}1\\0\\1\end{bmatrix}$, we need to solve $\begin{bmatrix}1 & 0 & 1\\1 & 1 & 0\\0 & 1 & 1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}3\\7\\4\end{bmatrix}$.

Solving that system (do it!), we find $\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}3\\4\\0\end{bmatrix}$.

Armin Straub
straub@southalabama.edu

Review. If $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ are orthogonal, the orthogonal projection of $\boldsymbol{w}$ onto $\mathrm{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is

$$\hat{\boldsymbol{w}} = \frac{\boldsymbol{w} \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 + \ldots + \frac{\boldsymbol{w} \cdot \boldsymbol{v}_n}{\boldsymbol{v}_n \cdot \boldsymbol{v}_n} \boldsymbol{v}_n.$$

Example 66.

(a) Project $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ onto $W = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \right\}$.

(b) Express $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ in terms of the basis $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$.

Solution.

(a) The projection is $\frac{8}{6}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \frac{4}{5}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$. (Each coefficient is obtained as the quotient of two dot products.)

(b) $\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = \frac{8}{6}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + \frac{4}{5}\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + \frac{5}{30}\begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$

Gram–Schmidt

(Gram–Schmidt orthogonalization)

Given a basis $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots$ for W, we produce an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ for W as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1 \end{pmatrix}$
- $\boldsymbol{q}_3 = \boldsymbol{w}_3 - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} - \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2 \end{pmatrix}$
- $\boldsymbol{q}_4 = \ldots$

Note. Since $\boldsymbol{q}_1, \boldsymbol{q}_2$ are orthogonal, $\begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\} \end{pmatrix} = \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_1 \end{pmatrix} + \begin{pmatrix} \text{projection of} \\ \boldsymbol{w}_3 \text{ onto } \boldsymbol{q}_2 \end{pmatrix}$.

Important comment. When working numerically on a computer it actually saves time to compute an orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ by the same approach but always normalizing each $\boldsymbol{q}_i$ along the way. The reason this saves time is that now the projections onto $\boldsymbol{q}_i$ only require a single dot product (instead of two). This is called **Gram–Schmidt orthonormalization**. When working by hand, it is usually simpler to wait until the end to normalize (so as to avoid).

Note. When normalizing, the orthonormal basis $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ is the unique one (up to $\pm$ signs) with the property that $\mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_k\} = \mathrm{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ for all $k = 1, 2, \ldots$.

Armin Straub
straub@southalabama.edu

Example 67. Using Gram–Schmidt, find an orthogonal basis for $W = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$.

Solution. We already have the basis $\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{w}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ for W. However, that basis is not orthogonal.

We can construct an orthogonal basis $\boldsymbol{q}_1, \boldsymbol{q}_2$ for W as follows:

- $\boldsymbol{q}_1 = \boldsymbol{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$
- $\boldsymbol{q}_2 = \boldsymbol{w}_2 - \left(\begin{matrix} \text{projection of} \\ \boldsymbol{w}_2 \text{ onto } \boldsymbol{q}_1 \end{matrix} \right) = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix}$

  Note. $\boldsymbol{q}_2$ is the error of the projection of $\boldsymbol{w}_2$ onto $\boldsymbol{q}_1$. This guarantees that it is orthogonal to $\boldsymbol{q}_1$. On the other hand, since $\boldsymbol{q}_2$ is a combination of $\boldsymbol{w}_2$ and $\boldsymbol{q}_1$, we know that $\boldsymbol{q}_2$ actually is in W.

We have thus found the orthogonal basis $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{2}{3} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ for W (if we like, we can, of course, drop that $\frac{2}{3}$).

Important comment. By normalizing, we get an orthonormal basis for W: $\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

Practical comment. When implementing Gram–Schmidt on a computer, it is beneficial (slightly less work) to normalize each $\boldsymbol{q}_i$ during the Gram–Schmidt process. This typically introduces square roots, which is why normalizing at the end is usually preferable when working by hand.

Comment. There are, of course, many orthogonal bases $\boldsymbol{q}_1, \boldsymbol{q}_2$ for W. Up to the length of the vectors, ours is the unique one with the property that $\operatorname{span}\{\boldsymbol{q}_1\} = \operatorname{span}\{\boldsymbol{w}_1\}$ and $\operatorname{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\} = \operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$.

A matrix Q has orthonormal columns $\iff$ $Q^T Q = I$

Why? Let $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots$ be the columns of Q. By the way matrix multiplication works, the entries of $Q^T Q$ are dot products of these columns:

$$\begin{bmatrix} — & \boldsymbol{q}_1^T & — \\ — & \boldsymbol{q}_2^T & — \\ & \vdots & \end{bmatrix} \begin{bmatrix} | & | & \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots \\ | & | & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \ddots \end{bmatrix}$$

Hence, $Q^T Q = I$ if and only if the dot products $\boldsymbol{q}_i^T \boldsymbol{q}_j = 0$ (that is, the columns are orthogonal), for $i \neq j$, and $\boldsymbol{q}_i^T \boldsymbol{q}_i = 1$ (that is, the columns are normalized).

Example 68. $Q = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6} \\ 1/\sqrt{3} & -2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$ obtained from Example 67 satisfies $Q^T Q = I$.

Armin Straub
straub@southalabama.edu

The QR decomposition

Just like the LU decomposition encodes the steps of Gaussian elimination, the QR decomposition encodes the steps of Gram–Schmidt.

(QR decomposition) Every $m\times n$ matrix A of rank n can be decomposed as $A=QR$, where

- Q has orthonormal columns, $(m\times n)$
- R is upper triangular and invertible. $(n\times n)$

How to find Q and R?

- Gram–Schmidt orthonormalization on (columns of) A, to get (columns of) Q
- $R=Q^TA$
  Why? If $A=QR$, then $Q^TA=Q^TQR$ which simplifies to $R=Q^TA$ (since $Q^TQ=I$).

The decomposition $A=QR$ is unique if we require the diagonal entries of R to be positive (and this is exactly what happens when applying Gram–Schmidt).

Practical comment. Actually, no extra work is needed for computing R. All of its entries have been computed during Gram–Schmidt.

Variations. We can also arrange things so that Q is an $m\times m$ orthogonal matrix and R a $m\times n$ upper triangular matrix. This is a tiny bit more work (and not required for many applications): we need to complement "our" Q with additional orthonormal columns and add corresponding zero rows to R. For square matrices this makes no difference.

Example 69. Determine the QR decomposition of $A=\begin{bmatrix} 1 & 1\\ 1 & -1\\ 1 & 1 \end{bmatrix}$.

Solution. The first step is Gram–Schmidt orthonormalization on the columns of A. We then use the resulting orthonormal vectors (they need to be normalized!) as the columns of Q.

We already did Gram–Schmidt in Example 67: from that work, we have $Q=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6}\\ 1/\sqrt{3} & -2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}$.

Hence, $R=Q^TA=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3}\\ 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}\begin{bmatrix} 1 & 1\\ 1 & -1\\ 1 & 1 \end{bmatrix}=\begin{bmatrix} \sqrt{3} & 1/\sqrt{3}\\ 0 & 4/\sqrt{6} \end{bmatrix}$.

Comment. The entries of R have actually all been computed during Gram–Schmidt, so that, if we pay attention, we could immediately write down R (no extra work required). Looking back at Example 67, can you see this?

Check. Indeed, $QR=\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{6}\\ 1/\sqrt{3} & -2/\sqrt{6}\\ 1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}\begin{bmatrix} \sqrt{3} & 1/\sqrt{3}\\ 0 & 4/\sqrt{6} \end{bmatrix}=\begin{bmatrix} 1 & 1\\ 1 & -1\\ 1 & 1 \end{bmatrix}$ equals A.