Orthogonal Sets

As we deepen our focus on geometry, we will focus on orthogonality.
Specifically, today we'll study the properties of **sets** of orthogonal vectors.
These can be very useful.

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, i.e.,

$$\mathbf{u}_i^T\mathbf{u}_j = 0 \text{ whenever } i \neq j.$$

Example. Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix}3\\1\\1\end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix}-1\\2\\1\end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix}-1/2\\-2\\7/2\end{bmatrix}.$$

Solution. Consider the three possible pairs of distinct vectors, namely, $\{\mathbf{u}_1, \mathbf{u}_2\}$, $\{\mathbf{u}_1, \mathbf{u}_3\}$, and $\{\mathbf{u}_2, \mathbf{u}_3\}$.

$$\mathbf{u}_1^T\mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$

$$\mathbf{u}_1^T\mathbf{u}_3 = 3(-1/2) + 1(-2) + 1(7/2) = 0$$

$$\mathbf{u}_2^T\mathbf{u}_3 = -1(-1/2) + 2(-2) + 1(7/2) = 0$$

Each pair of distinct vectors is orthogonal, and so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set.
In three space they describe three lines that say are **mutually perpendicular.**

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.HTML object>
```

Question Time! Q21.1

Theorem. If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

Proof. We will prove that there is no linear combination of the vectors in S with nonzero coefficients that yields the zero vector.

Our proof strategy will be:

we will show that for any linear combination of the vectors in S, if the combination is the zero vector, then all coefficients of the combination must be zero.

Specifically:

Assume $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \ldots, c_p$. Then:

$$\mathbf{0} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p$$

$$0 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)^T\mathbf{u}_1$$

$$= (c_1\mathbf{u}_1)^T\mathbf{u}_1 + (c_2\mathbf{u}_2)^T\mathbf{u}_1 \cdots + (c_p\mathbf{u}_p)^T\mathbf{u}_1$$

$$= c_1(\mathbf{u}_1^T\mathbf{u}_1) + c_2(\mathbf{u}_2^T\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p^T\mathbf{u}_1)$$

Because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$:

$$= c_1(\mathbf{u}_1^T\mathbf{u}_1)$$

Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1^T\mathbf{u}_1$ is not zero and so $c_1 = 0$.
We can use the same kind of reasoning to show that, $c_2, \ldots, c_p$ must be zero.

In other words, there is no nonzero combination of $\mathbf{u}_i$'s that yields the zero vector –so S is linearly independent.

Definition. An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

We have seen that for any subspace, there may be many different sets of vectors that can serve as a basis for W.

However an orthogonal basis is a particularly nice basis, because the weights (coordinates) of any point can be computed easily.

Theorem. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights of the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y}^T\mathbf{u}_j}{\mathbf{u}_j^T\mathbf{u}_j} \quad j = 1, \ldots, p$$

Proof. As we saw in the last proof, the orthogonality of $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ means that

$$\mathbf{y}^T\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_1\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)^T\mathbf{u}_1$$

$$= c_1(\mathbf{u}_1^T\mathbf{u}_1)$$

Since $\mathbf{u}_1^T\mathbf{u}_1$ is not zero, the equation above can be solved for c_1. To find any other c_j, compute $\mathbf{y}^T\mathbf{u}_j$ and solve for c_j.

Example. The set S which we saw earlier, ie,

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix},$$

is an orthogonal basis for $\mathbb{R}^3$.

Then, express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S (ie, in the basis S or in the coordinate system S).

Solution. Compute

$$\mathbf{y}^T\mathbf{u}_1 = 11, \quad \mathbf{y}^T\mathbf{u}_2 = -12, \quad \mathbf{y}^T\mathbf{u}_3 = -33,$$

$$\mathbf{u}_1^T\mathbf{u}_1 = 11, \quad \mathbf{u}_2^T\mathbf{u}_2 = 6, \quad \mathbf{u}_3^T\mathbf{u}_3 = 33/2$$

So

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y}^T\mathbf{u}_1}{\mathbf{u}_1^T\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}^T\mathbf{u}_2}{\mathbf{u}_2^T\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}^T\mathbf{u}_3}{\mathbf{u}_3^T\mathbf{u}_3}\mathbf{u}_3 \\ &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{33/2}\mathbf{u}_3 \end{aligned}$$

$$= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3.$$

Let's stop for a moment and think about how we would have done this if we had not known that the vectors $\mathbf{u}_1, \mathbf{u}_2$, and $\mathbf{u}_3$ form an orthogonal set.

We would have been looking for

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + c_3\mathbf{u}_3 = \mathbf{y}$$

The way we would find c_1, c_2, c_3 in that case would be to solve the linear system

$$[\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3] \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \mathbf{y}$$

which would have been much more trouble than what we did.
Instead, because the basis is an orthogonal basis, each coefficient c_1 can be found separately, and simply.

An Orthogonal Projection

Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors:

- one that is a multiple of $\mathbf{u}$, and
- one that is orthogonal to $\mathbf{u}$.

In other words, we wish to write:

$$\mathbf{y} = \mathbf{\hat{y}} + \mathbf{z}$$

where $\mathbf{\hat{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$.

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.HTML object>
```

That is, we are given $\mathbf{y}$ and $\mathbf{u}$, and asked to compute $\mathbf{z}$ and $\mathbf{\hat{y}}$.
To solve this, assume that we have some α, and with it we compute $\mathbf{y} - \alpha\mathbf{u} = \mathbf{y} - \mathbf{\hat{y}} = \mathbf{z}$.
We want $\mathbf{z}$ to be orthogonal to $\mathbf{u}$.
Now $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$ is orthogonal to $\mathbf{u}$ if and only if

$$0 = (\mathbf{y} - \alpha\mathbf{u})^T\mathbf{u}$$

$$= \mathbf{y}^T\mathbf{u} - (\alpha\mathbf{u})^T\mathbf{u}$$

$$= \mathbf{y}^T\mathbf{u} - \alpha(\mathbf{u}^T\mathbf{u})$$

That is, the solution in which $\mathbf{z}$ is orthogonal to $\mathbf{u}$ happens if and only if

$$\alpha = \frac{\mathbf{y}^T\mathbf{u}}{\mathbf{u}^T\mathbf{u}}$$

and since $\mathbf{\hat{y}} = \alpha\mathbf{u}$,

$$\mathbf{\hat{y}} = \frac{\mathbf{y}^T\mathbf{u}}{\mathbf{u}^T\mathbf{u}}\mathbf{u}.$$

The vector $\mathbf{\hat{y}}$ is called the **orthogonal projection of y onto u,** and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**.

```
<IPython.core.display.Javascript object>
```

<IPython.core.display.HTML object>

Now, note that if we had scaled $\mathbf{u}$ by any amount (ie, moved it to the right or the left), we would not have changed the location of $\hat{\mathbf{y}}$.

This can be seen as well by replacing $\mathbf{u}$ with $c\mathbf{u}$ and recomputing $\hat{\mathbf{y}}$:

$$\hat{\mathbf{y}} = \frac{\mathbf{y}^T c\mathbf{u}}{c\mathbf{u}^T c\mathbf{u}} c\mathbf{u} = \frac{\mathbf{y}^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \mathbf{u}.$$

Thus, the projection of $\mathbf{y}$ is determined by the *subspace* L that is spanned by $\mathbf{u}$ – in other words, the line through $\mathbf{u}$ and the origin.

Hence sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L \mathbf{y}$ and is called the **orthogonal projection of $\mathbf{y}$ onto L.**

Specifically:

$$\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y}^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \mathbf{u}$$

Example. Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$.

Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span$\{\mathbf{u}\}$, and one orthogonal to $\mathbf{u}$.

Solution. Compute

$$\mathbf{y}^T \mathbf{u} = \begin{bmatrix} 7 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$$

$$\mathbf{u}^T \mathbf{u} = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$$

The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}}{\mathbf{u}^T \mathbf{u}} \mathbf{u}$$

$$= \frac{40}{20} \mathbf{u} = 2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

And the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

So

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

$$\begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

Question Time! Q21.2

```
<IPython.core.display.Javascript object>
```

```
<IPython.core.display.HTML object>
```

The closest point.

Recall from geometry that given a line and a point P, the closest point on the line to P is given by the perpendicular from P to the line.

So this gives an important interpretation of $\hat{\mathbf{y}}$: it is **the closest point to $\mathbf{y}$ in the subspace L.**

The distance from $\mathbf{y}$ to L

The distance from $\mathbf{y}$ to L is the length of the perpendicular from $\mathbf{y}$ to its orthogonal projection on L, namely $\hat{\mathbf{y}}$.

This distance equals the length of $\mathbf{y} - \hat{\mathbf{y}}$.

In this example, the distance is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}.$$

A Geometric Interpretation

Earlier today, we saw that when we decompose a vector $\mathbf{y}$ into a linear combination of vectors $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ in a orthogonal set, we have

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

where

$$c_j = \frac{\mathbf{y}^T\mathbf{u}_j}{\mathbf{u}_j^T\mathbf{u}_j}$$

And just now we have seen that the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}$ is

$$\text{proj}_L\mathbf{y} = \frac{\mathbf{y}^T\mathbf{u}}{\mathbf{u}^T\mathbf{u}}\mathbf{u}.$$

So a decomposition like $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ is really decomposing $\mathbf{y}$ into **a sum of orthogonal projections onto one-dimensional subspaces.**

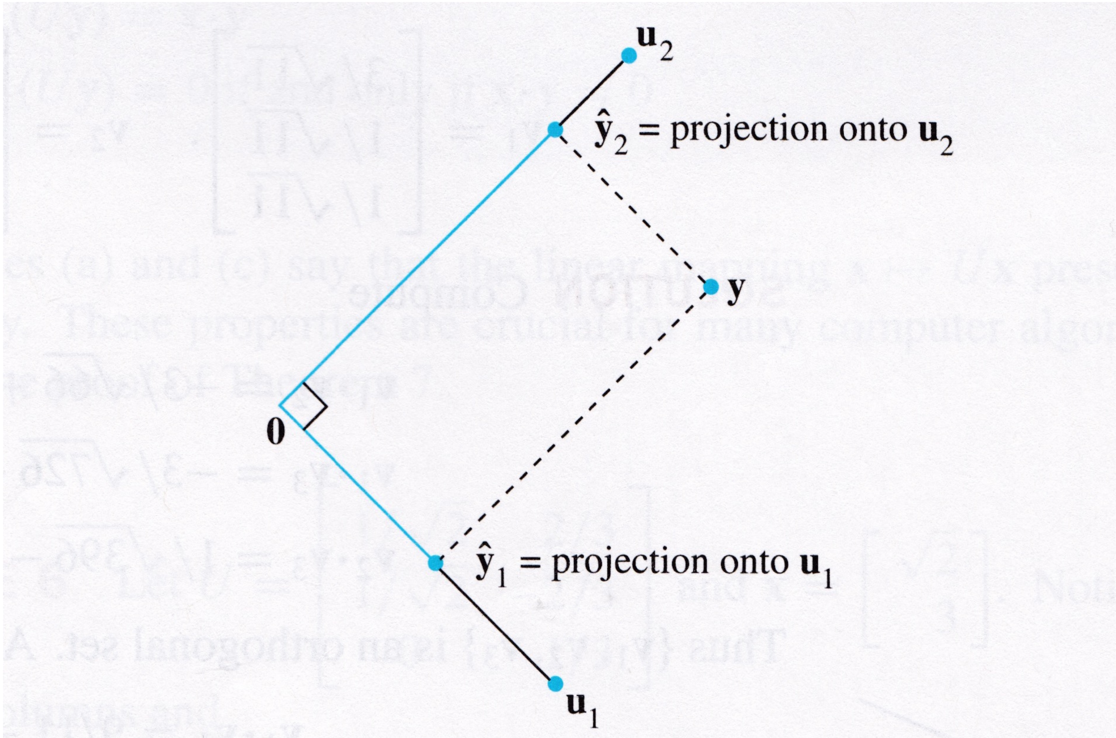
For example, let's take the case where $\mathbf{y} \in \mathbb{R}^2$. Let's say we are given $\mathbf{u}_1, \mathbf{u}_2$ such that $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2$, and so together they span $\mathbb{R}^2$.

Then $\mathbf{y}$ can be written in the form

$$\mathbf{y} = \frac{\mathbf{y}^T\mathbf{u}_1}{\mathbf{u}_1^T\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}^T\mathbf{u}_2}{\mathbf{u}_2^T\mathbf{u}_2}\mathbf{u}_2.$$

The first term is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_1$ and the second term is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_2$.

So this equation expresses $\mathbf{y}$ as the sum of its projections onto the (orthogonal) axes determined by $\mathbf{u}_1$ and $\mathbf{u}_2$.



Question Time! Q21.3

Orthonormal Sets

A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of **unit** vectors.

If W is the subspace spanned by such as a set, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal basis** for W since the set is automatically linearly independent.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal as well.

Pro tip: keep the terms clear in your head:

- **orthogonal** is (just) perpendicular, while
- **orthonormal** is perpendicular *and* unit length.

(You can see the word "normalized" inside "orthonormal").

Matrices with orthonormal columns are particularly important.

Theorem. A $m \times n$ matrix U has orthonormal columns if and only if $U^T U = I$.

Proof. Let us suppose that U has only three columns which are each vectors in $\mathbb{R}^m$ (but the proof will generalize to n columns).

Let $U = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3]$. Then:

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix}$$

The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \;\; \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \;\; \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0$$

The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \;\; \mathbf{u}_2^T\mathbf{u}_2 = 1, \;\; \mathbf{u}_3^T\mathbf{u}_3 = 1.$$

So $U^TU = I$.

Theorem. Let U by an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then:

1. $\|U\mathbf{x}\| = \|\mathbf{x}\|$.
2. $(U\mathbf{x})^T(U\mathbf{y}) = \mathbf{x}^T\mathbf{y}$.
3. $(U\mathbf{x})^T(U\mathbf{y}) = 0$ if and only if $\mathbf{x}^T\mathbf{y} = 0$.

Properties 1. and 3. say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality.

So, viewed as a linear operator, an orthonormal matrix is very special: the lengths of vectors, and therefore the **distances between points** is not changed by the action of U.

Example. Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Notice that U has orthonormal columns, and

$$U^TU = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Let's verify that $\|Ux\| = \|x\|$.

$$U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11}.$$

$$\|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}.$$

Orthonormal Square Matrices. Consider the case when U is square, and has orthonormal columns.

Then the fact that $U^TU = I$ implies that $U^{-1} = U^T$.

Then U is called an **orthogonal** matrix.

(Note that this terminology could be confusing; the columns of U are not just orthogonal but actually orthonormal.)