Overview

Last time

- we defined the dot product on $\mathbb{R}^n$;
- we recalled that the word "orthogonal" describes a relationship between two vectors in $\mathbb{R}^n$;
- we extended the definition of the word "orthogonal" to describe a relationship between a vector and a subspace;
- we defined the *orthogonal complement* $W^\perp$ of the the subspace W to be the subspace consisting of all the vectors orthogonal to W.

Today we'll extend the definition of the word "orthogonal" yet again. We'll also see how orthogonality can determine a particularly useful basis for a vector space.

From Lay, §6.2

Definition of an orthogonal set

Definition

A set $S \subset \mathbb{R}^n$ is *orthogonal* if its elements are pairwise orthogonal.

Example 1

Let $U = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix}.$$

To show that U is an orthogonal set we need to show that $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0, \mathbf{u}_1 \cdot \mathbf{u}_3 = 0$ and $\mathbf{u}_2 \cdot \mathbf{u}_3 = 0$.

Example 2

The set $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ where

$$\mathbf{w}_1 = \begin{bmatrix} 5 \\ -4 \\ 0 \\ 3 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} -4 \\ 1 \\ -3 \\ 8 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 3 \\ 3 \\ 5 \\ -1 \end{bmatrix}$$

is not an orthogonal set.

We note that $\mathbf{w}_1 \cdot \mathbf{w}_2 = 0$, $\mathbf{w}_1 \cdot \mathbf{w}_3 = 0$ but $\mathbf{w}_2 \cdot \mathbf{w}_3 = -32 \neq 0$.

Theorem (1)

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ *is an orthogonal set of* ***nonzero*** *vectors in* $\mathbb{R}^n$*, then* S *is a linearly independent set, and hence is a basis for the subspace spanned by* S*.*

Theorem (1)

*If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal set of **nonzero** vectors in $\mathbb{R}^n$, then S is a linearly independent set, and hence is a basis for the subspace spanned by S.*

Proof:
Suppose that $c_1, c_2, \ldots, c_k$ are scalars such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

Then

$$\begin{aligned} 0 &= \mathbf{0}\cdot\mathbf{v_1} = (\mathbf{c_1v_1} + \cdots + \mathbf{c_kv_k})\cdot\mathbf{v_1} \\ &= c_1(\mathbf{v}_1\cdot\mathbf{v}_1) + c_2(\mathbf{v}_2\cdot\mathbf{v}_1) + \cdots + c_k(\mathbf{v}_k\cdot\mathbf{v}_1) \\ &= c_1(\mathbf{v}_1\cdot\mathbf{v}_1) \end{aligned}$$

since $\mathbf{v}_1$ is orthogonal to $\mathbf{v}_2, \ldots, \mathbf{v}_k$.
Since $\mathbf{v}_1$ is nonzero, $\mathbf{v}_1\cdot\mathbf{v}_1$, and so $c_1 = 0$.
A similar argument shows that $c_2, \ldots, c_k$ must be zero.
Thus S is linearly independent. □

Definition

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

Definition

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

Example 3

Given $\begin{bmatrix}1\\2\\1\\0\end{bmatrix}, \begin{bmatrix}1\\-1\\1\\3\end{bmatrix}, \begin{bmatrix}2\\-1\\0\\-1\end{bmatrix}$, find a nonzero vector $\mathbf{x} = \begin{bmatrix}a\\b\\c\\d\end{bmatrix}$ so that the four vectors form an orthogonal set.

We are looking for a vector that satisfies the three conditions

$$\begin{bmatrix}a\\b\\c\\d\end{bmatrix} \cdot \begin{bmatrix}1\\2\\1\\0\end{bmatrix} = 0, \quad \begin{bmatrix}a\\b\\c\\d\end{bmatrix} \cdot \begin{bmatrix}1\\-1\\1\\3\end{bmatrix} = 0, \quad \begin{bmatrix}a\\b\\c\\d\end{bmatrix} \cdot \begin{bmatrix}2\\-1\\0\\-1\end{bmatrix} = 0$$

This gives a homogeneous system of three equations in the four variables a, b, c, d, which reduces the problem to one we already know how to solve.

An advantage of working with an orthogonal basis is that the coordinates of a vector with respect to that basis are easily determined.

Theorem (2)

Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ *be an orthogonal basis for a subspace* W *of* $\mathbb{R}^n$*, and let* $\mathbf{w}$ *be any vector in* W*. Then the unique scalars* $c_1, \ldots, c_k$ *such that*

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

are given by

$$c_i = \frac{\mathbf{w}\cdot\mathbf{v}_i}{\mathbf{v}_i\cdot\mathbf{v}_i} \quad \text{for } i = 1, \ldots, k.$$

Proof Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for W, we know that there are unique scalars $c_1, c_2, \ldots, c_k$ such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$.
To solve for c_1, we take the dot product of this linear combination with $\mathbf{v}_i$:

$$\begin{aligned} \mathbf{w}\cdot\mathbf{v}_1 &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k)\cdot\mathbf{v}_1 \\ &= c_1(\mathbf{v}_1\cdot\mathbf{v}_1) + \cdots + c_i(\mathbf{v}_i\cdot\mathbf{v}_1) + \cdots + c_k(\mathbf{v}_k\cdot\mathbf{v}_1) \\ &= c_1(\mathbf{v}_1\cdot\mathbf{v}_1) \end{aligned}$$

since $\mathbf{v}_j\cdot\mathbf{v}_1 = 0$ for $j \neq 1$.

Proof Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for W, we know that there are unique scalars $c_1, c_2, \ldots, c_k$ such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$.
To solve for c_1, we take the dot product of this linear combination with $\mathbf{v}_i$:

$$\begin{aligned} \mathbf{w}\cdot\mathbf{v}_1 &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k)\cdot\mathbf{v}_1 \\ &= c_1(\mathbf{v}_1\cdot\mathbf{v}_1) + \cdots + c_i(\mathbf{v}_i\cdot\mathbf{v}_1) + \cdots + c_k(\mathbf{v}_k\cdot\mathbf{v}_1) \\ &= c_1(\mathbf{v}_1\cdot\mathbf{v}_1) \end{aligned}$$

since $\mathbf{v}_j\cdot\mathbf{v}_1 = 0$ for $j \neq 1$.
Since $\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_1\cdot\mathbf{v}_1 \neq 0$. Dividing by $\mathbf{v}_1\cdot\mathbf{v}_1$, we obtain the desired result

$$c_1 = \frac{\mathbf{w}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}.$$

Similar results follow for $c = 2, \ldots, k$.

□

Example 4

Consider the orthogonal basis for $\mathbb{R}^3$:

$$\mathcal{U} = \left\{ \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \right\}.$$

Express $\mathbf{x} = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix}$ in $\mathcal{U}$ coordinates.

Example 4

Consider the orthogonal basis for $\mathbb{R}^3$:

$$\mathcal{U} = \left\{ \begin{bmatrix} 3 \\ -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix} \right\}.$$

Express $\mathbf{x} = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix}$ in $\mathcal{U}$ coordinates.

First, check that $\mathcal{U}$ really is an orthogonal basis for $\mathbb{R}^3$:

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = \mathbf{u}_1 \cdot \mathbf{u}_3 = \mathbf{u}_2 \cdot \mathbf{u}_3 = 0.$$

Hence the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, and since none of the vectors is the zero vector, the set is linearly independen a basis for $\mathbb{R}^3$.

Recall from Theorem (2) that the $\mathbf{u}_i$ coordinate of $\mathbf{x}$ is given by $\dfrac{\mathbf{x}\cdot\mathbf{v}_i}{\mathbf{v}_i\cdot\mathbf{v}_i}$.

Recall from Theorem (2) that the $\mathbf{u}_i$ coordinate of $\mathbf{x}$ is given by $\dfrac{\mathbf{x}\cdot\mathbf{v}_i}{\mathbf{v}_i\cdot\mathbf{v}_i}$. We compute

$$\mathbf{x}\cdot\mathbf{u}_1 = 6, \quad \mathbf{x}\cdot\mathbf{u}_2 = 13, \quad \mathbf{x}\cdot\mathbf{u}_3 = 2,$$
$$\mathbf{u}_1\cdot\mathbf{u}_1 = 18, \quad \mathbf{u}_2\cdot\mathbf{u}_2 = 9, \quad \mathbf{u}_3\cdot\mathbf{u}_3 = 18.$$

Hence

$$\begin{aligned}
\mathbf{x} &= \frac{\mathbf{x}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{x}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{x}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 \\
&= \frac{6}{18}\mathbf{u}_1 + \frac{13}{9}\mathbf{u}_2 + \frac{2}{18}\mathbf{u}_3 \\
&= \frac{1}{3}\mathbf{u}_1 + \frac{13}{9}\mathbf{u}_2 + \frac{1}{9}\mathbf{u}_3.
\end{aligned}$$

So $\mathbf{x} = \begin{bmatrix} \frac{1}{3} \\ \frac{13}{9} \\ \frac{1}{9} \end{bmatrix}_{\mathcal{U}}$.

Finally, note that if $P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 1 \\ -3 & 2 & 1 \\ 0 & -1 & 4 \end{bmatrix}$, then

$$P^T P = \begin{bmatrix} 18 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 18 \end{bmatrix}.$$

The diagonal form is because the vectors form an orthogonal set, diagonal entries are the squares of the lengths of the vectors. □

Orthonormal sets

Definition

A set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an *orthonormal set* if it is an orthogonal set of unit vectors.

Orthonormal sets

Definition

A set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is an *orthonormal set* if it is an orthogonal set of unit vectors.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$.
When the vectors in an orthogonal set of nonzero vectors are *normalised* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set.

Recall that in the last example, when P was a matrix with orthogonal columns, $P^T P$ was diagonal. When the columns of a matrix are vectors in an orthonormal set, the situation is even nicer:

Recall that in the last example, when P was a matrix with orthogonal columns, $P^T P$ was diagonal. When the columns of a matrix are vectors in an orthonormal set, the situation is even nicer:

Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set in $\mathbb{R}^3$ and $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$. Then

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

Recall that in the last example, when P was a matrix with orthogonal columns, $P^T P$ was diagonal. When the columns of a matrix are vectors in an orthonormal set, the situation is even nicer:

Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set in $\mathbb{R}^3$ and $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$. Then

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

Hence

$$U^T U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Recall that in the last example, when P was a matrix with orthogonal columns, $P^T P$ was diagonal. When the columns of a matrix are vectors in an orthonormal set, the situation is even nicer:

Suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set in $\mathbb{R}^3$ and $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$. Then

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

Hence

$$U^T U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since U is a square matrix, the relation $U^T U = I$ implies that $U^T = U^{-1}$ and thus we also have $UU^T = I$.

In fact,

A **square** matrix U has orthonormal columns if and only if U is invertible with $U^{-1} = U^T$.

In fact,

A **square** matrix U has orthonormal columns if and only if U is invertible with $U^{-1} = U^T$.

Definition

A **square** matrix U which is invertible and such that $U^{-1} = U^T$ is called an **orthogonal matrix**.

It follows from the result above that an orthogonal matrix is a square matrix whose columns form an **orthonormal** set (not just an orthogonal set as the name might suggest).

More generally, we have the following result:

Theorem (3)

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

More generally, we have the following result:

Theorem (3)

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

We also have the following theorem

Theorem (4)

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

(1) $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

(2) $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$.

(3) $(U\mathbf{x})\cdot(U\mathbf{y}) = 0$ *if and only if* $\mathbf{x}\cdot\mathbf{y} = 0$.

Properties (1) and (3) say that if U has orthonormal columns then the linear transformation $\mathbf{x} \to U\mathbf{x}$ (from $\mathbb{R}^n$ to $\mathbb{R}^m$) preserves lengths and orthogonality.

Examples

Example 5

The 4×3 matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 3 & -1 \end{bmatrix}$$

has orthogonal columns and $A^T A$ equals

$$\begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & -1 & 1 & 3 \\ 2 & -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & 0 \\ 0 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 12 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Note that here the rows of A are NOT orthogonal. For example, if we take the dot product of the first two rows we get

$$\langle 1, 1, 2 \rangle \cdot \langle 2, -1, -1 \rangle = 2 - 1 - 2 = -1 \neq 0\,.$$

Now consider the new matrix where each column of A is normalised:

$$B = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{12} & 2/\sqrt{6} \\ 2/\sqrt{6} & -1/\sqrt{12} & -1/\sqrt{6} \\ 1/\sqrt{6} & 1/\sqrt{12} & 0 \\ 0 & 3/\sqrt{12} & -1/\sqrt{6} \end{bmatrix}.$$

Then

$$B^T B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Example 6

Determine a, b, c such that

$$\begin{bmatrix} a & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ b & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ c & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}$$

is an orthogonal matrix.

The given 2nd and 3rd columns are orthonormal.

So we need to satisfy:

(1) $a^2 + b^2 + c^2 = 1$,

(2) $a/\sqrt{2} + b/\sqrt{6} + c/\sqrt{3} = 0$ which is equivalent to

$$\sqrt{3}a + b + \sqrt{2}c = 0$$

(3) $-a/\sqrt{2} + b/\sqrt{6} + c/\sqrt{3} = 0$ which is equivalent to

$$-\sqrt{3}a + b + \sqrt{2}c = 0.$$

From (2) and (3) we get $a = 0, b = -\sqrt{2}c$.
Substituting in (1) we get $2c^2 + c^2 = 1$ that is $c^2 = \frac{1}{3}$ which gives $c = \pm\frac{1}{\sqrt{3}}$. Thus possible 1st columns are $\pm \begin{bmatrix} 0 \\ -\frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}$ (there are only two possibilities). □