Part VI

Inner Product Spaces

Section 27

The Dot Product in $\mathbb{R}^n$

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is the dot product of two vectors? Under what conditions is the dot product defined?
- How do we find the angle between two nonzero vectors in $\mathbb{R}^n$?
- How does the dot product tell us if two vectors are orthogonal?
- How do we define the length of a vector in any dimension and how can the dot product be used to calculate the length?
- How do we define the distance between two vectors?
- What is the orthogonal projection of a vector $\mathbf{u}$ in the direction of the vector $\mathbf{v}$ and how do we find it?
- What is the orthogonal complement of a subspace W of $\mathbb{R}^n$?

Application: Hidden Figures in Computer Graphics

In video games, the speed at which a computer can render changing graphics views is vitally important. To increase a computer's ability to render a scene, programs often try to identify those parts of the images a viewer could see and those parts the viewer could not see. For example, in a scene involving buildings, the viewer could not see any images blocked by a solid building. In the mathematical world, this can be visualized by graphing surfaces. In Figure 27.1 we see a crude image of a house made up of small polygons (this is how programs generally represent surfaces). On the left in Figure 27.1 we see all of the polygons that are needed to construct the entire surface, even those polygons that lie behind others which we could not see if the surface was solid. On the right in Figure 27.1 we have hidden the parts of the polygons that we cannot see from our view. We

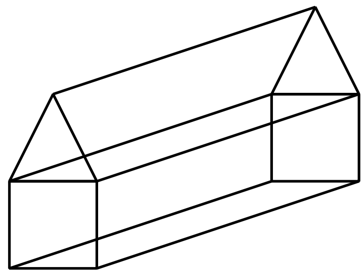

Figure 27.1: Images of a house.

also see this idea in mathematics when we graph surfaces. Figure 27.2 shows the graph of the surface defined by $f(x, y) = \sqrt{4 - x^2}$ that is made up of polygons. At left we see all of the polygons and at right only those parts that would be visible from our viewing perspective. By eliminating

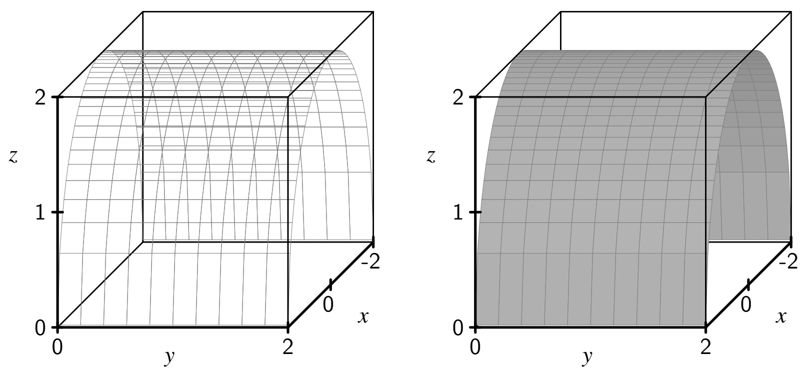

Figure 27.2: Graphs of $f(x, y) = \sqrt{4 - x^2}$.

the parts of the polygons we cannot see from our viewing perspective, the computer program can more quickly render the viewing image. Later in this section we will explore one method for how programs remove the hidden portions of images. This process involves the dot product of vectors.

Introduction

Orthogonality, a concept which generalizes the idea of perpendicularity, is an important concept in linear algebra. We use the dot product to define orthogonality and more generally angles between vectors in $\mathbb{R}^n$ for any dimension n. The dot product has many applications, e.g., finding components of forces acting in different directions in physics and engineering. The dot product is also an example of a larger concept, *inner products*, that we will discuss a bit later. We introduce and investigate dot products in this section.

We will illustrate the dot product in $\mathbb{R}^2$, but the process we go through will translate to any dimension. Recall that we can represent the vector $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ as the directed line segment (or arrow) from the origin to the point (v_1, v_2) in $\mathbb{R}^2$, as illustrated in Figure 27.3. Using the Pythagorean Theorem we can then define the length (or magnitude or norm) of the vector $\mathbf{v}$ in $\mathbb{R}^2$ as

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2}.$$

We can also write this norm as

$$\sqrt{v_1 v_1 + v_2 v_2}.$$

The expression under the square root is an important one and we extend it and give it a special name.

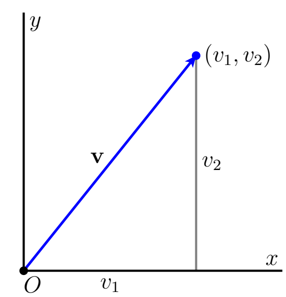

Figure 27.3: A vector in $\mathbb{R}^2$ from the origin to a point.

If $\mathbf{u} = [u_1\ u_2]^\mathsf{T}$ and $\mathbf{v} = [v_1\ v_2]^\mathsf{T}$ are vectors in $\mathbb{R}^2$, then we call the expression $u_1v_1 + u_2v_2$ the *dot product* of $\mathbf{u}$ and $\mathbf{v}$, and denote it as $\mathbf{u} \cdot \mathbf{v}$. With this idea in mind, we can rewrite the norm of the vector $\mathbf{v}$ as

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

The definition of the dot product translates naturally to $\mathbb{R}^n$ (see Exercise 5 in Section 5).

Definition 27.1. Let $\mathbf{u} = [u_1\ u_2\ \cdots\ u_n]$ and $\mathbf{v} = [v_1\ v_2\ \cdots\ v_n]$ be vectors in $\mathbb{R}^n$. The **dot product** (or **scalar product**) of $\mathbf{u}$ and $\mathbf{v}$ is the scalar

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \sum_{i=1}^{n} u_i v_i.$$

The dot product then allows us to define the norm (or magnitude or length) of any vector in $\mathbb{R}^n$.

Definition 27.2. The **norm** $||\mathbf{v}||$ of the vector $\mathbf{v} \in \mathbb{R}^n$ is

$$||\mathbf{v}|| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

We also use the words **magnitude** or **length** as alternatives for the word norm. We can equivalently write the norm of the vector $\mathbf{v} = [v_1\ v_2\ \cdots\ v_n]^\mathsf{T}$ as

$$||\mathbf{v}|| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

We can also realize the dot product as a matrix product. If $\mathbf{u} = [u_1\ u_2\ \cdots\ u_n]^\mathsf{T}$ and $\mathbf{v} = [v_1\ v_2\ \cdots\ v_n]^\mathsf{T}$, then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\mathsf{T}\mathbf{v}.$$[1]

IMPORTANT NOTE: The dot product is only defined between two vectors with the *same number of components*.

Preview Activity 27.1.

(1) Find $\mathbf{u} \cdot \mathbf{v}$ if $\mathbf{u} = [2\ 3\ \ -1\ 4]^\mathsf{T}$ and $\mathbf{v} = [4\ 6\ 7\ \ -5]^\mathsf{T}$ in $\mathbb{R}^4$.

(2) The dot product satisfies some useful properties as given in the next theorem.

Theorem 27.3. *Let* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *be vectors in* $\mathbb{R}^n$, *and let* c *be a scalar. Then*

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$ *(the dot product is* commutative*),*

(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$ *(the dot product* distributes over vector addition*),*

(c) $(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$*,*

(d) $\mathbf{u} \cdot \mathbf{u} \geq 0$ *with equality if and only if* $\mathbf{u} = \mathbf{0}$*,*

(e) $||c\mathbf{u}|| = |c|||\mathbf{u}||$*.*

Verification of some of these properties is left to the exercises. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^5$ with $\mathbf{u} \cdot \mathbf{v} = -1$, $||\mathbf{u}|| = 2$ and $||\mathbf{v}|| = 3$. Use the properties of the dot product given in Theorem 27.3 to find each of the following.

(a) $\mathbf{u} \cdot 2\mathbf{v}$

(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{v}$

(c) $(2\mathbf{u} + 4\mathbf{v}) \cdot (\mathbf{u} - 7\mathbf{v})$

(3) At times we will want to find vectors in the direction of a given vector that have a certain magnitude. Let $\mathbf{u} = [2\ 2\ 1]^\mathsf{T}$ in $\mathbb{R}^3$.

(a) What is $||\mathbf{u}||$?

(b) Show that $\left|\left|\frac{1}{||\mathbf{u}||}\mathbf{u}\right|\right| = 1$.

(c) Vectors with magnitude 1 are important and are given a special name.

Definition 27.4. A vector $\mathbf{v}$ in $\mathbb{R}^n$ is a **unit vector** if $||\mathbf{v}|| = 1$.

We can use unit vectors to find vectors of a given length in the direction of a given vector. Let c be a positive scalar and $\mathbf{v}$ a vector in $\mathbb{R}^n$. Use properties from Theorem 27.3 to show that the magnitude of the vector $c\frac{\mathbf{v}}{||\mathbf{v}||}$ is c.

[1]Technically, $\mathbf{u}^\mathsf{T}\mathbf{v}$ is a 1×1 matrix and not a scalar, but we usually think of 1×1 matrices as scalars.

The Distance Between Vectors

Finding optimal solutions to systems is an important problem in applied mathematics. It is often the case that we cannot find an exact solution that satisfies certain constraints, so we look instead for the "best" solution that satisfies the constraints. An example of this is fitting a least squares line to a set of data. As we will see, the dot product (and inner products in general) will allow us to find "best" solutions to certain types of problems, where we measure accuracy using the notion of a distance between vectors. Geometrically, we can represent a vector $\mathbf{u}$ as a directed line segment from the origin to the point defined by $\mathbf{u}$. If we have two vectors $\mathbf{u}$ and $\mathbf{v}$, we can think of the length of the difference $\mathbf{u} - \mathbf{v}$ as a measure of how far apart the two vectors are from each other. It is natural, then to define the distance between vectors as follows.

Definition 27.5. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. The **distance** between $\mathbf{u}$ and $\mathbf{v}$ is the length of the difference $\mathbf{u} - \mathbf{v}$ or

$$||\mathbf{u} - \mathbf{v}||.$$

As Figure 27.4 illustrates, if vectors $\mathbf{u}$ and $\mathbf{v}$ emanate from the same initial point, and P and Q are the terminal points of $\mathbf{u}$ and $\mathbf{v}$, respectively, then the difference $||\mathbf{u} - \mathbf{v}||$ is the standard Euclidean distance between the points P and Q.

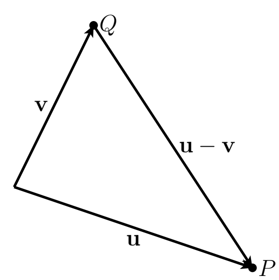

Figure 27.4: $||\mathbf{u} - \mathbf{v}||$.

The Angle Between Two Vectors

Determining a "best" solution to a problem often involves finding a solution that minimizes a distance. We generally accomplish a minimization through orthogonality – which depends on the angle between vectors. Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, we position the vectors so that they emanate from the same initial point. If the vectors are nonzero, then they determine a plane in $\mathbb{R}^n$. In that plane there are two angles that these vectors create. We will define the angle between the vectors to be the smaller of these two angles. The dot product will tell us how to find the angle between vectors. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ and θ the angle between them as illustrated in Figure 27.5. Using the Law of Cosines, we have

$$||\mathbf{u} - \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2 - 2||\mathbf{u}||\ ||\mathbf{v}||\ \cos(\theta)\,.$$

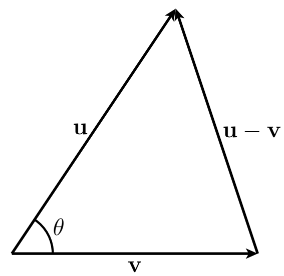

Figure 27.5: The angle between $\mathbf{u}$ and $\mathbf{v}$

By rearranging, we obtain

$$\begin{aligned} ||\mathbf{u}||\ ||\mathbf{v}||\ \cos(\theta) &= \frac{1}{2}\left(||\mathbf{u}||^2 + ||\mathbf{v}||^2 - ||\mathbf{u}-\mathbf{v}||^2\right) \\ &= \frac{1}{2}(||\mathbf{u}||^2 + ||\mathbf{v}||^2 - (\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})) \\ &= \frac{1}{2}(||\mathbf{u}||^2 + ||\mathbf{v}||^2 - \mathbf{u}\cdot\mathbf{u} + 2\mathbf{u}\cdot\mathbf{v} - \mathbf{v}\cdot\mathbf{v}) \\ &= \mathbf{u}\cdot\mathbf{v}\,. \end{aligned}$$

So the angle θ between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ satisfies the equation

$$\cos(\theta) = \frac{\mathbf{u}\cdot\mathbf{v}}{||\mathbf{u}||\ ||\mathbf{v}||}. \tag{27.1}$$

Of particular interest to us will be the situation where vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* (perpendicular).[2] Intuitively, we think of two vectors as orthogonal if the angle between them is 90°.

Activity 27.1.

(a) The vectors $\mathbf{e}_1 = [1\ 0]^\mathsf{T}$ and $\mathbf{e}_2 = [0\ 1]^\mathsf{T}$ are perpendicular in $\mathbb{R}^2$. What is $\mathbf{e}_1 \cdot \mathbf{e}_2$?

(b) Now let $\mathbf{u}$ and $\mathbf{v}$ be any vectors in $\mathbb{R}^n$.

i. Suppose the angle between nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ is 90°. What does Equation (27.1) tell us about $\mathbf{u}\cdot\mathbf{v}$?

ii. Now suppose that $\mathbf{u}\cdot\mathbf{v} = 0$. What does Equation (27.1) tell us about the angle between $\mathbf{u}$ and $\mathbf{v}$? Why?

iii. Explain why the following definition makes sense.

Definition 27.6. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if $\mathbf{u}\cdot\mathbf{v} = 0$.

iv. According to Definition 27.6, to which vectors is $\mathbf{0}$ orthogonal? Does this make sense to you intuitively? Explain.

Activity 27.2.

[2] We use the term orthogonal instead of perpendicular because we will be able to extend this idea to situations where we normally don't think of objects as being perpendicular.

(a) Find the angle between the two vectors $\mathbf{u} = [1\ 3\ \ -2\ 5]^\mathsf{T}$ and $\mathbf{v} = [5\ 2\ 3\ \ -1]^\mathsf{T}$.

(b) Find, if possible, two non-parallel vectors orthogonal to $\mathbf{u} = \begin{bmatrix} 0 \\ 3 \\ -2 \\ 1 \end{bmatrix}$.

Orthogonal Projections

When running a sprint, the racers may be aided or slowed by the wind. The wind assistance is a measure of the wind speed that is helping push the runners down the track. It is much easier to run a very fast race if the wind is blowing hard in the direction of the race. So that world records aren't dependent on the weather conditions, times are only recorded as record times if the wind aiding the runners is less than or equal to 2 meters per second. Wind speed for a race is recorded by a wind gauge that is set up close to the track. It is important to note, however, that weather is not always as cooperative as we might like. The wind does not always blow exactly in the direction of the track, so the gauge must account for the angle the wind makes with the track. If the wind is blowing in the direction of the vector $\mathbf{u}$ in Figure 27.6 and the track is in the direction of the vector $\mathbf{v}$ in Figure 27.6, then only part of the total wind vector is actually working to help the runners. This part is called the orthogonal projection of the vector $\mathbf{u}$ onto the vector $\mathbf{v}$ and is denoted $\text{proj}_{\mathbf{v}}\mathbf{u}$. The next activity shows how to find this projection.

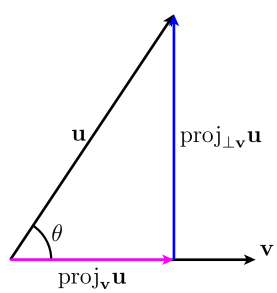

Figure 27.6: The orthogonal projection of $\mathbf{u}$ onto $\mathbf{v}$.

Activity 27.3. Since the orthogonal projection $\text{proj}_{\mathbf{v}}\mathbf{u}$ is in the direction of $\mathbf{v}$, there exists a constant c such that $\text{proj}_{\mathbf{v}}\mathbf{u} = c\mathbf{v}$. If we solve for c, we can find $\text{proj}_{\mathbf{v}}\mathbf{u}$.

(a) The wind component that acts perpendicular to the direction of $\mathbf{v}$ is called the projection of $\mathbf{u}$ orthogonal to $\mathbf{v}$ and is denoted $\text{proj}_{\perp\mathbf{v}}\mathbf{u}$ as shown in Figure 27.6. Write an equation that involves $\text{proj}_{\mathbf{v}}\mathbf{u}$, $\text{proj}_{\perp\mathbf{v}}\mathbf{u}$, and $\mathbf{u}$. Then solve that equation for $\text{proj}_{\perp\mathbf{v}}\mathbf{u}$.

(b) Given that $\mathbf{v}$ and $\text{proj}_{\perp\mathbf{v}}\mathbf{u}$ are orthogonal, what does that tell us about $\mathbf{v}\cdot\text{proj}_{\perp\mathbf{v}}\mathbf{u}$? Combine this fact with the result of part (a) and that $\text{proj}_{\mathbf{v}}\mathbf{u} = c\mathbf{v}$ to obtain an equation involving $\mathbf{v}$, $\mathbf{u}$, and c.

(c) Solve for c using the equation you found in the previous step.

(d) Use your value of c to identify $\text{proj}_{\mathbf{v}}\mathbf{u}$.

To summarize:

Definition 27.7. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$ with $\mathbf{v} \neq \mathbf{0}$.

(1) The **orthogonal projection** of $\mathbf{u}$ onto $\mathbf{v}$ is the vector

$$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{v}||^2}\mathbf{v}. \tag{27.2}$$

(2) The **projection** of $\mathbf{u}$ orthogonal to $\mathbf{v}$ is the vector

$$\text{proj}_{\perp \mathbf{v}}\mathbf{u} = \mathbf{u} - \text{proj}_{\mathbf{v}}\mathbf{u}.$$

Activity 27.4. Let $\mathbf{u} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 6 \\ -10 \end{bmatrix}$. Find $\text{proj}_{\mathbf{v}}\mathbf{u}$ and $\text{proj}_{\perp \mathbf{v}}\mathbf{u}$ and draw a picture to illustrate.

Orthogonal Complements

In Activity 27.1 we defined two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ to be orthogonal (or perpendicular) if $\mathbf{u} \cdot \mathbf{v} = 0$. With this in mind we can define the orthogonal complement of a subspace of $\mathbb{R}^n$.

Definition 27.8. Let W be a subspace of $\mathbb{R}^n$ for some $n \geq 1$. The **orthogonal complement** of W is the set

$$W^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}.$$

Preview Activity 27.2. Let $W = \text{Span}\{[1 \ \ -1]^{\mathsf{T}}\}$ in $\mathbb{R}^2$. Completely describe all vectors in $W^{\perp}$ both algebraically and geometrically.

The orthogonal projection of a vector $\mathbf{u}$ onto a vector $\mathbf{v}$ is really a projection of the vector $\mathbf{u}$ onto the vector space $\text{Span}\{\mathbf{v}\}$. The vector $\text{proj}_{\mathbf{v}}\mathbf{u}$ is the best approximation to $\mathbf{u}$ of all the vectors in $\text{Span}\{\mathbf{v}\}$ in the sense that $\text{proj}_{\mathbf{v}}\mathbf{u}$ is the closest to $\mathbf{u}$ among all vectors in $\text{Span}\{\mathbf{v}\}$, as we will prove later. Another familiar example where we see this type of behavior is when we look at planes in 3-space. Remember that a plane through the origin in $\mathbb{R}^3$ is a two dimensional subspace of $\mathbb{R}^3$. We define a plane through the origin to be the set of all vectors in $\mathbb{R}^3$ that are orthogonal to a given vector (called a *normal* vector). For example, to find the equation of the plane through the origin in $\mathbb{R}^3$ orthogonal to the normal vector $\mathbf{n} = [1 \ 2 \ \ -1]^{\mathsf{T}}$, we seek all the vectors $\mathbf{v} = [x \ y \ z]^{\mathsf{T}}$ such that

$$\mathbf{v} \cdot \mathbf{n} = \mathbf{0}.$$

This gives us the equation

$$x + 2y - z = 0$$

as the equation of this plane.

There is a more general idea here as defined in Preview Activity 27.2. If we have a set S of vectors in $\mathbb{R}^n$, we let $S^{\perp}$ (read as "S perp", called the *orthogonal complement* of S) be the set of all vectors in $\mathbb{R}^n$ that are orthogonal to every vector in S. In our plane example, the set S is $\{\mathbf{n}\}$ and $S^{\perp}$ is the plane with equation $x + 2y - z = 0$.

Activity 27.5. We have seen another example of orthogonal complements. Let A be an $m \times n$ matrix with rows $\mathbf{r}_1$, $\mathbf{r}_2$, ..., $\mathbf{r}_m$ in order. Consider the three spaces Nul A, Row A, and Col A related to A, where Row $A = \text{Span}\{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m\}$ (that is, Row A is the span of the rows of A). Let $\mathbf{x}$ be a vector in Row A.

(a) What does it mean for $\mathbf{x}$ to be in Row A?

(b) Now let $\mathbf{y}$ be a vector in Nul A. Use the result of part (a) and the fact that $A\mathbf{y} = \mathbf{0}$ to explain why $\mathbf{x} \cdot \mathbf{y} = 0$. (Hint: Calculate $A\mathbf{y}$ using scalar products of rows of A with $\mathbf{y}$.) Explain how this verifies (Row $A)^{\perp}$ = Nul A.

(c) Use A^{T} in place of A in the result of the previous part to show (Col $A)^{\perp}$ = Nul A^{T}.

The activity proves the following theorem:

Theorem 27.9. *Let A be an $m \times n$ matrix. Then*

$$(\textit{Row } A)^{\perp} = \textit{Nul } A \textit{ and } (\textit{Col } A)^{\perp} = \textit{Nul } A^{\mathsf{T}}.$$

To show that a vector is in the orthogonal complement of a subspace, it is not necessary to demonstrate that the vector is orthogonal to every vector in the subspace. If we have a basis for the subspace, it suffices to show that the vector is orthogonal to every vector in that basis for the subspace, as the next theorem demonstrates.

Theorem 27.10. *Let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be a basis for a subspace W of $\mathbb{R}^n$. A vector $\mathbf{v}$ in $\mathbb{R}^n$ is orthogonal to every vector in W if and only if $\mathbf{v}$ is orthogonal to every vector in $\mathcal{B}$.*

Proof. Let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be a basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Our theorem is a biconditional, so we need to prove both implications. Since $\mathcal{B} \subset W$, it follows that if $\mathbf{v}$ is orthogonal to every vector in W, then $\mathbf{v}$ is orthogonal to every vector in $\mathcal{B}$. This proves the forward implication. Now we assume that $\mathbf{v}$ is orthogonal to every vector in $\mathcal{B}$ and show that $\mathbf{v}$ is orthogonal to every vector in W. Let $\mathbf{x}$ be a vector in W. Then

$$\mathbf{x} = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_m\mathbf{w}_m$$

for some scalars x_1, x_2, ..., x_m. Then

$$\begin{aligned}\mathbf{v} \cdot \mathbf{x} &= \mathbf{v} \cdot (x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_m\mathbf{w}_m) \\ &= x_1(\mathbf{v} \cdot \mathbf{w}_1) + x_2(\mathbf{v} \cdot \mathbf{w}_2) + \cdots + x_m(\mathbf{v} \cdot \mathbf{w}_m) \\ &= 0.\end{aligned}$$

Thus, $\mathbf{v}$ is orthogonal to $\mathbf{x}$ and $\mathbf{v}$ is orthogonal to every vector in W. ∎

Activity 27.6. Let $W = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$. Find all vectors in $W^{\perp}$.

We will work more closely with projections and orthogonal complements in later sections.

Examples

What follows are worked examples that use the concepts from this section.

Example 27.11. Let ℓ be the line defined by the equation $ax + by + c = 0$ with in $\mathbb{R}^2$ and let $P = (x_0, y_0)$ be a point in the plane. In this example we will learn how to find the distance from P to ℓ.

(a) Show that $\mathbf{n} = [a\ b]^{\mathsf{T}}$ is orthogonal to the line ℓ. That is, $\mathbf{n}$ is orthogonal to any vector on the line ℓ.

(b) Let $Q = (x_1, y_1)$ be any point on line ℓ. Draw a representative picture of P, $\mathbf{n}$ with its initial point at P, along with Q and ℓ. Explain how to use a projection to determine the distance from P to ℓ.

(c) Use the idea from part (b) to show that the distance d from P to ℓ satisfies

$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}. \tag{27.3}$$

(d) Use Equation (27.3) to find the distance from the point $(3, 4)$ to the line $y = 2x + 1$.

Example Solution.

(a) Any vector on the line ℓ is a vector between two points on the line. Let $Q = (x_1, y_1)$ and $R = (x_2, y_2)$ be points on the line ℓ. Then $\mathbf{u} = \overrightarrow{QR} = [x_2 - x_1\ y_2 - y_1]^{\mathsf{T}}$ is a vector on line ℓ. Since Q and R are on the line, we know that $ax_1 + by_1 + c = 0$ and $ax_2 + by_2 + c = 0$. So $-c = ax_1 + by_1 = ax_2 + by_2$ and

$$0 = a(x_2 - x_1) + b(y_2 - y_1) = [a\ b]^{\mathsf{T}}\mathbf{u}.$$

Thus, $\mathbf{n} = [a\ b]^{\mathsf{T}}$ is orthogonal to every vector on the line ℓ.

(b) A picture of the situation is shown in Figure 27.7. If $\mathbf{v} = \overrightarrow{PQ}$, then the distance from point P to line ℓ is given by $||\text{proj}_{\mathbf{n}}\mathbf{v}||$.

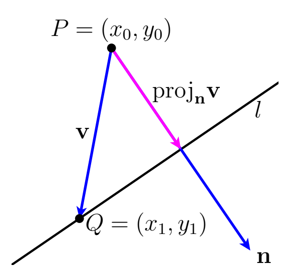

Figure 27.7: Distance from a point to a line.

(c) Recall that $\mathbf{n} = [a\ b]^{\mathsf{T}}$ and $\mathbf{v} = \overrightarrow{PQ} = [x_1 - x_0\ y_1 - y_0]^{\mathsf{T}}$. Since $ax_1 + by_1 + c = 0$, we have

$$\begin{aligned} \text{proj}_{\mathbf{n}}\mathbf{v} &= \frac{\mathbf{v}\cdot\mathbf{v}}{||\mathbf{n}||^2}\mathbf{n} \\ &= \frac{a(x_1-x_0)+b(y_1-y_0)}{a^2+b^2}[a\ b]^{\mathsf{T}} \\ &= \frac{ax_1+by_1-ax_0-by_0}{a^2+b^2}[a\ b]^{\mathsf{T}} \\ &= \frac{-c-ax_0-by_0}{a^2+b^2}[a\ b]^{\mathsf{T}}. \end{aligned}$$

So

$$||\text{proj}_{\mathbf{n}}\mathbf{v}|| = \frac{|ax_0+by_0+c|}{a^2+b^2}\sqrt{a^2+b^2} = \frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}}.$$

(d) Here we have $P = (3, 4)$, and the equation of our line is $2x - y + 1 = 0$. So $a = 2$, $b = -1$, and $c = -1$. Thus, the distance from P to the line is

$$\frac{|ax_0+by_0+c|}{\sqrt{a^2+b^2}} = \frac{|2(3)-(4)+1|}{\sqrt{4+1}} = \frac{3}{\sqrt{5}}.$$

Example 27.12. Let a, b, and c be scalars with $a \neq 0$, and let

$$W = \{ax + by + cz = 0 : x, y, z \in \mathbb{R}\}.$$

(a) Find two vectors that span W, showing that W is a subspace of $\mathbb{R}^3$. (In fact, W is a plane through the origin in $\mathbb{R}^3$.)

(b) Find a vector $\mathbf{n}$ that is orthogonal to the two vectors you found in part (a).

(c) Explain why $\{\mathbf{n}\}$ is a basis for $W^{\perp}$.

Example Solution.

(a) The coefficient matrix for the system $ax + by + cz = 0$ is $[a\ b\ c]^{\mathsf{T}}$. The first column is a pivot column and the others are not. So y and z are free variables and

$$[x\ y\ z]^{\mathsf{T}} = \left[-\frac{b}{a}y - \frac{c}{a}z, y, z\right]^{\mathsf{T}} = y\left[-\frac{b}{a}\ 1\ 0\right]^{\mathsf{T}} + z\left[-\frac{c}{a}\ 0\ 1\right]^{\mathsf{T}}.$$

So $W = \text{Span}\left\{\left[-\frac{b}{a}\ 1\ 0\right]^{\mathsf{T}}, \left[-\frac{c}{a}\ 0\ 1\right]^{\mathsf{T}}\right\}$.

(b) If we let $\mathbf{n} = [a\ b\ c]^{\mathsf{T}}$, then

$$\begin{aligned} \mathbf{n}\cdot\left[-\frac{b}{a}\ 1\ 0\right]^{\mathsf{T}} &= -b + b = 0 \\ \mathbf{n}\cdot\left[-\frac{c}{a}\ 0\ 1\right]^{\mathsf{T}} &= -c + c = 0. \end{aligned}$$

Thus, $[a\ b\ c]^{\mathsf{T}}$ is orthogonal to both $\left[-\frac{b}{a}\ 1\ 0\right]^{\mathsf{T}}$ and $\left[-\frac{c}{a}\ 0\ 1\right]^{\mathsf{T}}$.

(c) Let $\mathbf{u} = \left[-\frac{b}{a} \ 1 \ 0\right]^{\mathsf{T}}$ and $\mathbf{v} = \left[-\frac{c}{a} \ 0 \ 1\right]^{\mathsf{T}}$. Every vector in W has the form $x\mathbf{u} + y\mathbf{v}$ for some scalars x and y, and

$$\mathbf{n} \cdot (x\mathbf{u} + y\mathbf{v}) = x(\mathbf{n} \cdot \mathbf{u}) + y(\mathbf{n} \cdot \mathbf{v}) = 0.$$

So $\mathbf{n} \in W^{\perp}$.

Now we need to verify that $\{\mathbf{n}\}$ spans $W^{\perp}$. Let $\mathbf{w} = [w_1 \ w_2 \ w_3]^{\mathsf{T}}$ be in $W^{\perp}$. Then $\mathbf{w} \cdot \mathbf{z} = 0$ for every $\mathbf{z} \in W$. In particular, $\mathbf{w} \cdot \mathbf{u} = 0$ or $-\frac{b}{a}w_1 + w_2 = 0$, and $\mathbf{w} \cdot \mathbf{v} = 0$ or $-\frac{c}{a}w_1 + w_3 = 0$. Equivalently, we have $w_2 = \frac{b}{a}w_1$ and $w_3 = \frac{c}{a}w_1$. So

$$\begin{aligned}
\mathbf{w} &= [w_1 \ w_2 \ w_3]^{\mathsf{T}} \\
&= \left[w_1 \ \frac{b}{a}w_1 \ \frac{c}{a}w_1\right]^{\mathsf{T}} \\
&= \frac{1}{a}[a \ b \ c]^{\mathsf{T}} w_1 \\
&= \frac{w_1}{a}\mathbf{n}.
\end{aligned}$$

So every vector in $W^{\perp}$ is a multiple of $\mathbf{n}$, and $\{\mathbf{n}\}$ spans $W^{\perp}$. We conclude that $\{\mathbf{n}\}$ is a basis for $W^{\perp}$. Thus, the vector $[a \ b \ c]^{\mathsf{T}}$ is a normal vector to the plane $ax + by + cz = 0$ if $a \neq 0$. The same reasoning works if at least one of a, b,or c is nonzero, so we can say in every case that $[a \ b \ c]^{\mathsf{T}}$ is a normal vector to the plane $ax + by + cz = 0$.

Summary

- The dot product of vectors $\mathbf{u} = [u_1 \ u_2 \ \cdots \ u_n]^{\mathsf{T}}$ and $\mathbf{v} = [v_1 \ v_2 \ \cdots \ v_n]^{\mathsf{T}}$ in $\mathbb{R}^n$ is the scalar

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \sum_{i=1}^{n} u_iv_i.$$

- The angle θ between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ satisfies the equation

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| \ ||\mathbf{v}||}$$

  and $0 \leq \theta \leq 180$.

- Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.
- The length, or norm, of the vector $\mathbf{u}$ can be found as $||\mathbf{u}|| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$.
- The distance between the vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ is $||\mathbf{u} - \mathbf{v}||$, which is the length of the difference $\mathbf{u} - \mathbf{v}$.
- Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$.
  - The orthogonal projection of $\mathbf{u}$ onto $\mathbf{v}$ is the vector

$$\text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{v}||^2}\mathbf{v}.$$

– The projection of $\mathbf{u}$ perpendicular to $\mathbf{v}$ is the vector

$$\text{proj}_{\perp \mathbf{v}}\mathbf{u} = \mathbf{u} - \text{proj}_{\mathbf{v}}\mathbf{u}.$$

- The orthogonal complement of the subspace W of $\mathbb{R}^n$ is the set

$$W^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}.$$

Exercises

(1) For each of the following pairs of vectors, find $\mathbf{u} \cdot \mathbf{v}$, calculate the angle between $\mathbf{u}$ and $\mathbf{v}$, determine if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, find $||\mathbf{u}||$ and $||\mathbf{v}||$, calculate the distance between $\mathbf{u}$ and $\mathbf{v}$, and determine the orthogonal projection of $\mathbf{u}$ onto $\mathbf{v}$.

(a) $\mathbf{u} = [1\ 2]^{\mathsf{T}}$, $\mathbf{v} = [-2\ 1]^{\mathsf{T}}$

(b) $\mathbf{u} = [2\ \ -2]^{\mathsf{T}}$, $\mathbf{v} = [1\ \ -1]^{\mathsf{T}}$

(c) $\mathbf{u} = [2\ \ -1]^{\mathsf{T}}$, $\mathbf{v} = [1\ 3]^{\mathsf{T}}$

(d) $\mathbf{u} = [1\ 2\ 0]^{\mathsf{T}}$, $\mathbf{v} = [-2\ 1\ 1]^{\mathsf{T}}$

(e) $\mathbf{u} = [0\ 0\ 1]^{\mathsf{T}}$, $\mathbf{v} = [1\ 1\ 1]^{\mathsf{T}}$

(2) Given $\mathbf{u} = [2\ 1\ 2]^{\mathsf{T}}$, find a vector $\mathbf{v}$ so that the angle between $\mathbf{u}$ and $\mathbf{v}$ is 60° and the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$ has length 2.

(3) For which value(s) of h is the angle between $[1\ 1\ h]^{\mathsf{T}}$ and $[1\ 2\ 1]^{\mathsf{T}}$ equal to 60°?

(4) Let $A = [a_{ij}]$ be a $k \times m$ matrix with rows $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$, and let $B = [\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_n]$ be an $m \times n$ matrix with columns $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$. Show that we can write the matrix product AB in a shorthand way as $AB = [\mathbf{r}_i \cdot \mathbf{b}_j]$.

(5) Let A be an $m \times n$, $\mathbf{u}$ a vector in $\mathbb{R}^n$ and $\mathbf{v}$ a vector in $\mathbb{R}^m$. Show that

$$A\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^{\mathsf{T}}\mathbf{v}.$$

(6) Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$. Show that

(a) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = (\mathbf{u} \cdot \mathbf{w}) + (\mathbf{v} \cdot \mathbf{w})$ (the dot product *distributes over vector addition*)

(b) If c is an arbitrary constant, then $(c\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (c\mathbf{v}) = c(\mathbf{u} \cdot \mathbf{v})$

(7) The Pythagorean Theorem states that if a and b are the lengths of the legs of a right triangle whose hypotenuse has length c, then $a^2 + b^2 = c^2$. If we think of the legs as defining vectors $\mathbf{u}$ and $\mathbf{v}$, then the hypotenuse is the vector $\mathbf{u}+\mathbf{v}$ and we can restate the Pythagorean Theorem as

$$||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2.$$

In this exercise we show that this result holds in any dimension.

(a) Let $\mathbf{u}$ and $\mathbf{v}$ be orthogonal vectors in $\mathbb{R}^n$. Show that $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$. (Hint: Rewrite $||\mathbf{u} + \mathbf{v}||^2$ using the dot product.)

(b) Must it be true that if $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$ with $||\mathbf{u}+\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal? If not, provide a counterexample. If true, verify the statement.

(8) The Cauchy-Schwarz inequality,

$$|\mathbf{u}\cdot\mathbf{v}| \leq ||\mathbf{u}||\,||\mathbf{v}|| \tag{27.4}$$

for any vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, is considered one of the most important inequalities in mathematics. We verify the Cauchy-Schwarz inequality in this exercise. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$.

(a) Explain why the inequality (27.4) is true if either $\mathbf{u}$ or $\mathbf{v}$ is the zero vector. As a consequence, we assume that $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors for the remainder of this exercise.

(b) Let $\mathbf{w} = \text{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{v}}{||\mathbf{v}||^2}\mathbf{v}$ and let $\mathbf{z} = \mathbf{u}-\mathbf{w}$. We know that $\mathbf{w}\cdot\mathbf{z} = 0$. Use Exercise 7. of this section to show that

$$||\mathbf{u}||^2 \geq ||\mathbf{w}||^2.$$

(c) Now show that $||\mathbf{w}||^2 = \frac{|\mathbf{u}\cdot\mathbf{v}|^2}{||\mathbf{v}||^2}$.

(d) Combine parts (b) and (c) to explain why equation (27.4) is true.

(9) Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Then $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}+\mathbf{v}$ form a triangle. We should then expect that the length of any one side of the triangle is smaller than the sum of the lengths of the other sides (since the straight line distance is the shortest distance between two points). In other words, we expect that

$$||\mathbf{u}+\mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||. \tag{27.5}$$

Equation (27.5) is called the *Triangle Inequality*. Use the Cauchy-Schwarz inequality (Exercise 8) to prove the triangle inequality.

(10) Let W be a subspace of $\mathbb{R}^n$ for some n. Show that $W^\perp$ is also a subspace of $\mathbb{R}^n$.

(11) Let W be a subspace of $\mathbb{R}^n$. Show that W is a subspace of $(W^\perp)^\perp$.

(12) If W is a subspace of $\mathbb{R}^n$ for some n, what is $W \cap W^\perp$? Verify your answer.

(13) Suppose $W_1 \subseteq W_2$ are two subspaces of $\mathbb{R}^n$. Show that $W_2^\perp \subseteq W_1^\perp$.

(14) Label each of the following statements as True or False. Provide justification for your response.

(a) **True/False** The dot product is defined between any two vectors.

(b) **True/False** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then $\mathbf{u}\cdot\mathbf{v}$ is another vector in $\mathbb{R}^n$.

(c) **True/False** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then $\mathbf{u}\cdot\mathbf{v}$ is always non-negative.

(d) **True/False** If $\mathbf{v}$ is a vector in $\mathbb{R}^n$, then $\mathbf{v}\cdot\mathbf{v}$ is never negative.

(e) **True/False** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$ and $\mathbf{u}\cdot\mathbf{v} = 0$, then $\mathbf{u} = \mathbf{v} = \mathbf{0}$.

(f) **True/False** If $\mathbf{v}$ is a vector in $\mathbb{R}^n$ and $\mathbf{v} \cdot \mathbf{v} = 0$, then $\mathbf{v} = \mathbf{0}$.

(g) **True/False** The norm of the sum of vectors is the sum of the norms of the vectors.

(h) **True/False** If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then $\text{proj}_{\mathbf{v}}\mathbf{u}$ is a vector in the same direction as $\mathbf{u}$.

(i) **True/False** The only subspace W of $\mathbb{R}^n$ for which $W^{\perp} = \{\mathbf{0}\}$ is $W = \mathbb{R}^n$.

(j) **True/False** If a vector $\mathbf{u}$ is orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$, then $\mathbf{u}$ is also orthogonal to $\mathbf{v}_1 + \mathbf{v}_2$.

(k) **True/False** If a vector $\mathbf{u}$ is orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$, then $\mathbf{u}$ is also orthogonal to all linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.

(l) **True/False** If $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v}$ are parallel, then the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$ equals $\mathbf{v}$.

(m) **True/False** If $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v}$ are orthogonal, then the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$ equals $\mathbf{v}$.

(n) **True/False** For any vector $\mathbf{v}$ and $\mathbf{u} \neq \mathbf{0}$, $||\text{proj}_{\mathbf{u}}\mathbf{v}|| \leq ||\mathbf{v}||$.

(o) **True/False** Given an $m \times n$ matrix, $\dim(\text{Row } A) + \dim(\text{Row } A)^{\perp} = n$.

(p) **True/False** If A is a square matrix, then the columns of A are orthogonal to the vectors in Nul A.

(q) **True/False** The vectors in the null space of an $m \times n$ matrix are orthogonal to vectors in the row space of A.

Project: Back-Face Culling

To identify hidden polygons in a surface, we will utilize a technique called *back face culling*. This involves identifying which polygons are back facing and which are front facing relative to the viewer's perspective. The first step is to assign a direction to each polygon in a surface.

Project Activity 27.1. Consider the polygon $ABCD$ in Figure 27.8. Since a polygon is flat, every vector in the polygon is perpendicular to a fixed vector (which we call a *normal vector* to the polygon). A normal vector $\mathbf{n}$ for the polygon $ABCD$ in Figure 27.8 is shown. In this activity we learn how to find a normal vector to a polygon.

Let $\mathbf{x} = [x_1 \ x_2 \ x_3]^{\mathsf{T}}$ and $\mathbf{y} = [y_1 \ y_2 \ y_3]^{\mathsf{T}}$ be two vectors in $\mathbb{R}^3$. If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, then $\mathbf{x}$ and $\mathbf{y}$ determine a polygon as shown in Figure 27.8. Our goal is to find a vector $\mathbf{n}$ that is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$. Let $\mathbf{w} = [w_1 \ w_2 \ w_3]^{\mathsf{T}}$ be another vector in $\mathbb{R}^3$ and let $C = \begin{bmatrix} \mathbf{w}^{\mathsf{T}} \\ \mathbf{x}^{\mathsf{T}} \\ \mathbf{y}^{\mathsf{T}} \end{bmatrix}$ be the matrix whose rows are $\mathbf{w}$, $\mathbf{x}$, and $\mathbf{y}$. Let C_{ij} be the ijth cofactor of C, that is C_{ij} is $(-1)^{i+j}$ times the determinant of the submatrix of C obtained by deleting the ith row and jth column of C. Now define the vector $\mathbf{x} \times \mathbf{y}$ as follows:

$$\mathbf{x} \times \mathbf{y} = C_{11}\mathbf{e}_1 + C_{12}\mathbf{e}_2 + C_{13}\mathbf{e}_3.$$

The vector $\mathbf{x} \times \mathbf{y}$ is called the *cross product* of the vectors $\mathbf{x}$ and $\mathbf{y}$. (Note that the cross product is only defined for vectors in $\mathbb{R}^3$.) We will show that $\mathbf{x} \times \mathbf{y}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$, making $\mathbf{x} \times \mathbf{y}$ a normal vector to the polygon defined by $\mathbf{x}$ and $\mathbf{y}$.

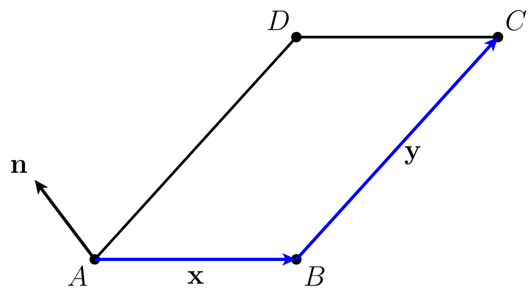

Figure 27.8: Normal vector to a polygon.

(a) Show that

$$\mathbf{x} \times \mathbf{y} = \begin{bmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{bmatrix}.$$

(b) Use a cofactor expansion of C along the first row and properties of the dot product to show that

$$\det(C) = \mathbf{w} \cdot (\mathbf{x} \times \mathbf{y}).$$

(c) Use the result of part (b) and properties of the determinant to calculate $\mathbf{x} \cdot (\mathbf{x} \times \mathbf{y})$ and $\mathbf{y} \cdot (\mathbf{x} \times \mathbf{y})$. Explain why $\mathbf{x} \times \mathbf{y}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$ and is therefore a normal vector to the polygon determined by $\mathbf{x}$ and $\mathbf{y}$.

Project Activity 27.1 shows how we can find a normal vector to a parallelogram – take two vectors $\mathbf{x}$ and $\mathbf{y}$ between the vertices of the parallelogram and calculate their cross products. Such a normal vector can define a direction for the parallelogram. There is still a problem, however.

Project Activity 27.2. Let $\mathbf{x} = [x_1 \ x_2 \ x_3]^{\mathsf{T}}$ and $\mathbf{y} = [y_1 \ y_2 \ y_3]^{\mathsf{T}}$ be any vectors in $\mathbb{R}^3$. There is a relationship between $\mathbf{x} \times \mathbf{y}$ and $\mathbf{y} \times \mathbf{x}$. Find and verify this relationship.

Project Activity 27.2 shows that the cross product is anticommutative, so we get different directions if we switch the order in which we calculate the cross product. To fix a direction, we establish the convention that we always label the vertices of our parallelogram in the counterclockwise direction as shown in Figure 27.8. This way we always use $\mathbf{x}$ as the vector from vertex A to vertex B rather than the reverse. With this convention established, we can now define the direction of a parallelogram as the direction of its normal vector.

Once we have a normal vector established for each polygon, we can now determine which polygons are back-face and which are front-face. Figure 27.9 at left provides the gist of the idea, where we represent the polygons with line segments to illustrate. If the viewer's eye is at point P and views the figures, the normal vectors of the visible polygons point in a direction toward the viewer (front-face) and the normal vectors of the polygons hidden from the viewer point away from the viewer (back-face). What remains is to determine an effective computational way to identify the front and back facing polygons.

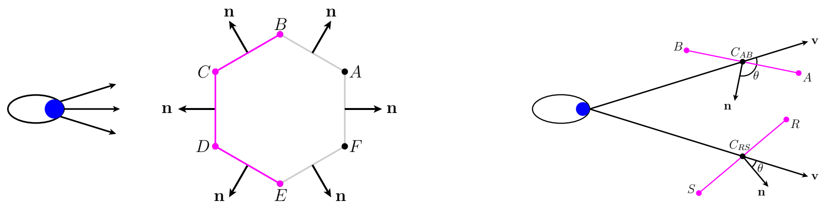

Figure 27.9: Left: Hidden faces. Right: Back face culling.

Project Activity 27.3. Consider the situation as depicted at right in Figure 27.9. Assume that AB and RS are polygons (rendered one dimensionally here) with normal vectors $\mathbf{n}$ at their centers as shown. The viewer's eye is at point P and the viewer's line of vision to the centers C_{AB} and C_{RS} are indicated by the vectors $\mathbf{v}$. Each vector $\mathbf{v}$ makes an angle θ with the normal to the polygon.

(a) What can be said about the angle θ for a front-facing polygon? What must be true about $\mathbf{v} \cdot \mathbf{n}$ for a front-facing polygon? Why?

(b) What can be said about the angle θ for a back-facing polygon? What must be true about $\mathbf{v} \cdot \mathbf{n}$ for a back-facing polygon? Why?

(c) The dot product then provides us with a simple computational tool for identifying back-facing polygons (assuming we have already calculated all of the normal vectors). We can then create an algorithm to cull the back-facing polygons. Assuming that we the viewpoint P and the coordinates of the polygons of the surface, complete the pseudo-code for a back-face culling algorithm:

for all polygons on the surface **do**
  calculate the normal vector $\mathbf{n}$ using the _______ product for the current polygon
  calculate the center C of the current polygon
  calculate the viewing vector ______________
    if ______________ **then**
      render the current polygon
    end if
end for

As a final comment, back-face culling generally reduces the number of polygons to be rendered by half. This algorithm is not perfect and does not always do what we want it to do (e.g., it may not remove all parts of a polygon that we don't see), so there are other algorithms to use in concert with back-face culling to correctly render objects.

Section 28

Orthogonal and Orthonormal Bases in $\mathbb{R}^n$

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is an orthogonal set in $\mathbb{R}^n$? What is one useful fact about orthogonal subsets of $\mathbb{R}^n$?
- What is an orthogonal basis for a subspace of $\mathbb{R}^n$? What is an orthonormal basis?
- How does orthogonality help us find the weights to write a vector as a linear combination of vectors in an orthogonal basis?
- What is an orthogonal matrix and why are orthogonal matrices useful?

Application: Rotations in 3D

An aircraft in flight, like a plane or the space shuttle, can perform three independent rotations: *roll*, *pitch*, and *yaw*. Roll is a rotation about the axis through the nose and tail of the aircraft, pitch is rotation moving the nose of the aircraft up or down through the axis from wingtip to wingtip, and yaw is the rotation when the nose of the aircraft turns left or right about the axis though the plane from top to bottom. These rotations take place in 3-space and the axes of the rotations change as the aircraft travels through space. To understand how aircraft maneuver, it is important to know about general rotations in space. These are more complicated than rotations in 2-space, and, as we will see later in this section, involve orthogonal sets.

Introduction

If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is a basis for a subspace W of $\mathbb{R}^n$, we know that any vector $\mathbf{w}$ in W can be written uniquely as a linear combination of the vectors in $\mathcal{B}$. In the past, the way we have found the coordinates of $\mathbf{x}$ with respect to $\mathcal{B}$, i.e. the weights needed to write a vector $\mathbf{x}$ as a linear combination of the elements in $\mathcal{B}$, has been to row reduce the matrix $[\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n \mid \mathbf{x}]$ to solve the corresponding system. This can be a cumbersome process, especially if we need to do it many times. For certain types of bases, namely the *orthogonal* and *orthonormal* bases, there is a much easier way to find the individual weights for this linear combination.

Recall that two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$. We can extend this idea to an entire set. For example, the standard basis $\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ for $\mathbb{R}^3$ has the property that any two distinct vectors in $\mathcal{S}$ are orthogonal to each other. The basis vectors in $\mathcal{S}$ make a very nice coordinate system for $\mathbb{R}^3$, where the basis vectors provide the directions for the coordinate axes. We could rotate this standard basis, or multiply any of the vectors in the basis by a nonzero constant, and retain a basis in which all distinct vectors are orthogonal to each other (e.g., $\{[2\ 0\ 0]^\mathsf{T}, [0\ 3\ 0]^\mathsf{T}, [0\ 0\ 1]^\mathsf{T}\}$). We define this idea of having all vectors be orthogonal to each other for sets, and then later for bases.

Definition 28.1. A non-empty subset S of $\mathbb{R}^n$ is **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$ for every pair of distinct vector $\mathbf{u}$ and $\mathbf{v}$ in S.

Preview Activity 28.1.

(1) Determine if the set $S = \{[1\ 2\ 1]^\mathsf{T}, [2\ \ -1\ 0]^\mathsf{T}\}$ is an orthogonal set.

(2) Orthogonal bases are especially important.

Definition 28.2. An **orthogonal basis** $\mathcal{B}$ for a subspace W of $\mathbb{R}^n$ is a basis of W that is also an orthogonal set.

Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_1 = [1\ 2\ 1]^\mathsf{T}$, $\mathbf{v}_2 = [2\ \ -1\ 0]^\mathsf{T}$, and $\mathbf{v}_3 = [1\ 2\ \ -5]^\mathsf{T}$.

(a) Explain why $\mathcal{B}$ is an orthogonal basis for $\mathbb{R}^3$.

(b) Suppose $\mathbf{x}$ has coordinates x_1, x_2, x_3 with respect to the basis $\mathcal{B}$, i.e.

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3\,.$$

Substitute for $\mathbf{x}$ in $\mathbf{x} \cdot \mathbf{v}_1$ and use the orthogonality property of the basis $\mathcal{B}$ to show that $x_1 = \frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}$. Then determine x_2 and x_3 similarly. Finally, calculate the values of x_1, x_2, and x_3 if $\mathbf{x} = [1\ 1\ 1]^\mathsf{T}$.

(c) Find components of $\mathbf{x} = [1\ 1\ 1]^\mathsf{T}$ by reducing the augmented matrix $[\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3 \mid \mathbf{x}]$. Does this result agree with your work from the previous part?

Orthogonal Sets

We defined orthogonal sets in $\mathbb{R}^n$ and bases of subspaces of $\mathbb{R}^n$ in Definitions 28.1 and 28.2. We saw that the standard basis in $\mathbb{R}^3$ is an orthogonal set and an orthogonal basis of $\mathbb{R}^3$, and there are many other examples as well.

Activity 28.1. Let $\mathbf{w}_1 = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{w}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$. In the same manner as in Preview Activity 28.1, we can show that the set $S_1 = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthogonal subset of $\mathbb{R}^3$.

(a) Is the set $S_2 = \left\{ \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \right\}$ an orthogonal subset of $\mathbb{R}^3$?

(b) Suppose a vector $\mathbf{v}$ is a vector so that $S_1 \cup \{\mathbf{v}\}$ is an orthogonal subset of $\mathbb{R}^3$. Then $\mathbf{w}_i \cdot \mathbf{v} = 0$ for each i. Explain why this implies that $\mathbf{v}$ is in Nul A, where $A = \begin{bmatrix} -2 & 1 & -1 \\ 0 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$.

(c) Assuming that the reduced row echelon form of the matrix A is I_3, explain why it is not possible to find a nonzero vector $\mathbf{v}$ so that $S_1 \cup \{\mathbf{v}\}$ is an orthogonal subset of $\mathbb{R}^3$.

The example from Activity 28.1 suggests that we can have three orthogonal nonzero vectors in $\mathbb{R}^3$, but no more. Orthogonal vectors are, in a sense, as far apart as they can be. So we might expect that there is no linear relationship between orthogonal vectors. The following theorem makes this clear.

Theorem 28.3. *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a set of nonzero orthogonal vectors in $\mathbb{R}^n$. Then the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ are linearly independent.*

Proof. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a set of nonzero orthogonal vectors in $\mathbb{R}^n$. To show that $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_m$ are linearly independent, assume that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m = \mathbf{0} \tag{28.1}$$

for some scalars x_1, x_2, ..., x_m. We will show that $x_i = 0$ for each i from 1 to m. Since the vectors in S are orthogonal to each other, we know that $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$. Fix an index k between 1 and m. We evaluate the dot product of both sides of (28.1) with $\mathbf{v}_k$ and simplify using the dot product properties:

$$\begin{aligned} \mathbf{v}_k \cdot (x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m) &= \mathbf{v}_k \cdot \mathbf{0} \\ (\mathbf{v}_k \cdot x_1\mathbf{v}_1) + (\mathbf{v}_k \cdot x_2\mathbf{v}_2) + \cdots + (\mathbf{v}_k \cdot x_m\mathbf{v}_m) &= 0 \\ x_1(\mathbf{v}_k \cdot \mathbf{v}_1) + x_2(\mathbf{v}_k \cdot \mathbf{v}_2) + \cdots + x_m(\mathbf{v}_k \cdot \mathbf{v}_m) &= 0. \end{aligned} \tag{28.2}$$

Now all of the dot products on the left are 0 except for $\mathbf{v}_k \cdot \mathbf{v}_k$, so (28.2) becomes

$$x_k(\mathbf{v}_k \cdot \mathbf{v}_k) = 0.$$

We assumed that $\mathbf{v}_k \neq \mathbf{0}$ and since $\mathbf{v}_k \cdot \mathbf{v}_k = ||\mathbf{v}_k||^2 \neq 0$, we conclude that $x_k = 0$. We chose k arbitrarily, so we have shown that $x_k = 0$ for each k between 1 and m. Therefore, the only solution to equation (28.1) is the trivial solution with $x_1 = x_2 = \cdots = x_m = 0$ and the set S is linearly independent. ■

Properties of Orthogonal Bases

Orthogonality is a useful and important property for a basis to have. In Preview Activity 28.1 we saw that if a vector $\mathbf{x}$ in the span of an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ could be written as a linear combination of the basis vectors as $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3$, then $x_1 = \frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}$. If we continued that same argument we could show that

$$\mathbf{x} = \left(\frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\right)\mathbf{v}_1 + \left(\frac{\mathbf{x}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\right)\mathbf{v}_2 + \left(\frac{\mathbf{x}\cdot\mathbf{v}_3}{\mathbf{v}_3\cdot\mathbf{v}_3}\right)\mathbf{v}_3.$$

We can apply this idea in general to see how the orthogonality of an orthogonal basis allows us to quickly and easily determine the weights to write a given vector as a linear combination of orthogonal basis vectors. To see why, let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{x}$ be any vector in W. We know that

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m$$

for some scalars x_1, x_2, …, x_m. Let $1 \le k \le m$. Then, using orthogonality of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$, we have

$$\mathbf{v}_k\cdot\mathbf{x} = x_1(\mathbf{v}_k\cdot\mathbf{v}_1) + x_2(\mathbf{v}_k\cdot\mathbf{v}_2) + \cdots + x_m(\mathbf{v}_k\cdot\mathbf{v}_m) = x_k\mathbf{v}_k\cdot\mathbf{v}_k.$$

So

$$x_k = \frac{\mathbf{x}\cdot\mathbf{v}_k}{\mathbf{v}_k\cdot\mathbf{v}_k}.$$

Thus, we can calculate each weight individually with two simple dot products. We summarize this discussion in the next theorem.

Theorem 28.4. *Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be an orthogonal basis for a subspace of $\mathbb{R}^n$. Let $\mathbf{x}$ be a vector in W. Then*

$$\mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{x}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 + \cdots + \frac{\mathbf{x}\cdot\mathbf{v}_m}{\mathbf{v}_m\cdot\mathbf{v}_m}\mathbf{v}_m. \tag{28.3}$$

Activity 28.2. Let $\mathbf{v}_1 = [1\ 0\ 1]^{\mathsf{T}}$, $\mathbf{v}_2 = [0\ 1\ 0]^{\mathsf{T}}$, and $\mathbf{v}_3 = [0\ 0 1]^{\mathsf{T}}$. The set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{R}^3$. Let $\mathbf{x} = [1\ 0\ 0]^{\mathsf{T}}$. Calculate

$$\frac{\mathbf{x}\cdot\mathbf{v}_1}{\mathbf{v}_1\cdot\mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{x}\cdot\mathbf{v}_2}{\mathbf{v}_2\cdot\mathbf{v}_2}\mathbf{v}_2 + \frac{\mathbf{x}\cdot\mathbf{v}_3}{\mathbf{v}_3\cdot\mathbf{v}_3}\mathbf{v}_3.$$

Compare to $\mathbf{x}$. Does this violate Theorem 28.4? Explain.

Orthonormal Bases

The decomposition (28.3) is even simpler if $\mathbf{v}_k\cdot\mathbf{v}_k = 1$ for each k, that is, if $\mathbf{v}_k$ is a unit vector for each k. In this case, the denominators are all 1 and we don't even need to consider them. We have a familiar example of such a basis for $\mathbb{R}^n$, namely the standard basis $\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$.

Recall that

$$\mathbf{v}\cdot\mathbf{v} = ||\mathbf{v}||^2,$$

so the condition $\mathbf{v}\cdot\mathbf{v} = 1$ implies that the vector $\mathbf{v}$ has norm 1. An orthogonal basis with this additional condition is a very nice basis and is given a special name.

Definition 28.5. An **orthonormal basis** $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ for a subspace W of $\mathbb{R}^n$ is an orthogonal basis such that $||\mathbf{u}_k|| = 1$ for $1 \leq k \leq m$.

In other words, an orthonormal basis is an orthogonal basis in which every basis vector is a unit vector. A good question to ask here is how we can construct an orthonormal basis from an orthogonal basis.

Activity 28.3.

(a) Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be orthogonal vectors. Explain how we can obtain unit vectors $\mathbf{u}_1$ in the direction of $\mathbf{v}_1$ and $\mathbf{u}_2$ in the direction of $\mathbf{v}_2$.

(b) Show that $\mathbf{u}_1$ and $\mathbf{u}_2$ from the previous part are orthogonal vectors.

(c) Use the ideas from this problem to construct an orthonormal basis for $\mathbb{R}^3$ from the orthogonal basis $S = \left\{ \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right\}$.

In general, we can construct an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ from an orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ by normalizing each vector in $\mathcal{B}$ (that is, dividing each vector by its norm).

Orthogonal Matrices

We have seen in the diagonalization process that we diagonalize a matrix A with a matrix P whose columns are linearly independent eigenvectors of A. In general, calculating the inverse of the matrix whose columns are eigenvectors of A in the diagonalization process can be time consuming, but if the columns form an orthonormal set, then the calculation is very straightforward.

Activity 28.4. Let $\mathbf{u}_1 = \frac{1}{3}[2\ 1\ 2]^{\mathsf{T}}$, $\mathbf{u}_2 = \frac{1}{3}[-2\ 2\ 1]^{\mathsf{T}}$, and $\mathbf{u}_3 = \frac{1}{3}[1\ 2\ -2]^{\mathsf{T}}$. It is not difficult to see that the set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis for $\mathbb{R}^3$. Let

$$A = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3] = \frac{1}{3}\begin{bmatrix} 2 & -2 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & -2 \end{bmatrix}.$$

(a) Use the definition of the matrix-matrix product to find the entries of the second row of the matrix product $A^{\mathsf{T}}A$. Why should you have expected the result? (Hint: How are the rows of A^{T} related to the columns of A?)

(b) With the result of part (a) in mind, what is the matrix product $A^{\mathsf{T}}A$? What does this tell us about the relationship between A^{T} and A^{-1}? Use technology to calculate A^{-1} and confirm your answer.

(c) Suppose P is an $n \times n$ matrix whose columns form an orthonormal basis for R^n. Explain why $P^{\mathsf{T}}P = I_n$.

The result of Activity 28.4 is that if the columns of a square matrix P form an orthonormal set, then $P^{-1} = P^{\mathsf{T}}$. This makes calculating P^{-1} very easy. Note, however, that this only works if the columns of P form an orthonormal basis for Col P. You should also note that if P is an $n \times n$ matrix satisfying $P^{\mathsf{T}}P = I_n$, then the columns of P must form an orthonormal set. Matrices like this appear quite often and are given a special name.

Definition 28.6. An **orthogonal** matrix is an $n \times n$ matrix P such that $P^{\mathsf{T}}P = I_n$.[1]

Activity 28.5. As a special case, we apply the result of Activity 28.4 to a 2×2 rotation matrix $P = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$.

(a) Show that the columns of P form an orthonormal set.

(b) Use the fact that $P^{-1} = P^{\mathsf{T}}$ to find P^{-1}. Explain how this shows that the inverse of a rotation matrix by an angle θ is just another rotation matrix but by the angle $-\theta$.

Orthogonal matrices are useful because they satisfy some special properties. For example, if P is an orthogonal $n \times n$ matrix and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

$$(P\mathbf{x}) \cdot (P\mathbf{y}) = (P\mathbf{x})^{\mathsf{T}}(P\mathbf{y}) = \mathbf{x}^{\mathsf{T}}P^{\mathsf{T}}P\mathbf{y} = \mathbf{x}^{\mathsf{T}}\mathbf{y} = \mathbf{x} \cdot \mathbf{y}.$$

This property tells us that the linear transformation T defined by $T(\mathbf{x}) = P\mathbf{x}$ preserves dot products and, hence, orthogonality. In addition,

$$||P\mathbf{x}||^2 = P\mathbf{x} \cdot P\mathbf{x} = \mathbf{x} \cdot \mathbf{x} = ||\mathbf{x}||^2,$$

so $||P\mathbf{x}|| = ||\mathbf{x}||$. This means that T preserves length. Such a transformation is called an *isometry* and it is convenient to work with functions that don't expand or contract things. Moreover, if $\mathbf{x}$ and $\mathbf{y}$ are nonzero vectors, then

$$\frac{P\mathbf{x} \cdot P\mathbf{y}}{||P\mathbf{x}|| \, ||P\mathbf{y}||} = \frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}|| \, ||\mathbf{y}||}.$$

Thus T also preserves angles. Transformations defined by orthogonal matrices are very well behaved transformations. To summarize,

Theorem 28.7. *Let P be an $n \times n$ orthogonal matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then*

(1) $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$,

(2) $||P\mathbf{x}|| = ||\mathbf{x}||$, *and*

(3) $\frac{P\mathbf{x} \cdot P\mathbf{y}}{||P\mathbf{x}|| \, ||P\mathbf{y}||} = \frac{\mathbf{x} \cdot \mathbf{y}}{||\mathbf{x}|| \, ||\mathbf{y}||}$ *if $\mathbf{x}$ and $\mathbf{y}$ are nonzero.*

We have discussed orthogonal and orthonormal bases for subspaces of $\mathbb{R}^n$ in this section. There are several questions that follow, such as

- Can we always find an orthogonal (or orthonormal) basis for any subspace of $\mathbb{R}^n$?
- Given a vector $\mathbf{v}$ in W, can we find an orthogonal basis of W that contain $\mathbf{v}$?
- Can we extend the concept of orthogonality to other vector spaces?

We will answer these questions in subsequent sections.

[1] It isn't clear why such matrices are called orthogonal since the columns are actually orthonormal, but that is the standard terminology in mathematics.

Examples

What follows are worked examples that use the concepts from this section.

Example 28.8. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where $\mathbf{v}_1 = [1\ 1\ -4]^\mathsf{T}$, $\mathbf{v}_2 = [2\ 2\ 1]^\mathsf{T}$, and $\mathbf{v}_3 = [1\ -1\ 0]^\mathsf{T}$.

(a) Show that S is an orthogonal set.

(b) Create an orthonormal set $S' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ from the vectors in S.

(c) Just by calculating dot products, write the vector $\mathbf{w} = [2\ 1\ \ -1]^\mathsf{T}$ as a linear combination of the vectors in S'.

Example Solution.

(a) Using the dot product formula, we see that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$, $\mathbf{v}_1 \cdot \mathbf{v}_3 = 0$, and $\mathbf{v}_2 \cdot \mathbf{v}_3 = 0$. Thus, the set S is an orthogonal set.

(b) To make an orthonormal set $S' = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ from S, we divide each vector in S by its magnitude. This gives us

$$\mathbf{u}_1 = \frac{1}{\sqrt{18}}[1\ 1\ \ -4]^\mathsf{T},\ \mathbf{u}_2 = \frac{1}{3}[2\ 2\ 1]^\mathsf{T},\ \text{ and }\ \mathbf{u}_3 = \frac{1}{\sqrt{2}}[1\ \ -1\ 0]^\mathsf{T}.$$

(c) Since S' is an orthonormal basis for $\mathbb{R}^3$, we know that

$$\begin{aligned}\mathbf{w} &= (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{w} \cdot \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{w} \cdot \mathbf{u}_3)\mathbf{u}_3 \\ &= \frac{7}{\sqrt{18}}[1\ 1\ \ -4]^\mathsf{T} + \frac{5}{3}[2\ 2\ 1]^\mathsf{T} + \frac{1}{\sqrt{2}}[1\ \ -1\ 0]^\mathsf{T}.\end{aligned}$$

Example 28.9. Let $\mathbf{u}_1 = \frac{1}{\sqrt{3}}[1\ 1\ 1]^\mathsf{T}$, $\mathbf{u}_2 = \frac{1}{\sqrt{2}}[1\ \ -1\ 0]^\mathsf{T}$, and $\mathbf{u}_3 = \frac{1}{\sqrt{6}}[1\ 1\ \ -2]^\mathsf{T}$. Let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

(a) Show that $\mathcal{B}$ is an orthonormal basis for $\mathbb{R}^3$.

(b) Let $\mathbf{w} = [1\ 2\ 1]^\mathsf{T}$. Find $[\mathbf{w}]_\mathcal{B}$.

(c) Calculate $||\mathbf{w}||$ and $||[\mathbf{w}]_\mathcal{B}||$. What do you notice?

(d) Show that the result of part (c) is true in general. That is, if $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$, and if $\mathbf{z} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$, then

$$||\mathbf{z}|| = \sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}.$$

Example Solution.

(a) Using the dot product formula, we see that $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ if $i \neq j$ and that $\mathbf{u}_i \cdot \mathbf{u}_i = 1$ for each i. Since orthogonal vectors are linearly independent, the set $\mathcal{B}$ is a linearly independent set with 3 vectors in a 3-dimensional space. It follows that $\mathcal{B}$ is an orthonormal basis for $\mathbb{R}^3$.

(b) Since $\mathcal{B}$ is an orthonormal basis for $\mathbb{R}^3$, we know that

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{w} \cdot \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{w} \cdot \mathbf{u}_3)\mathbf{u}_3.$$

Therefore,

$$[\mathbf{w}]_{\mathcal{B}} = [(\mathbf{w} \cdot \mathbf{u}_1)\ (\mathbf{w} \cdot \mathbf{u}_2)\ (\mathbf{w} \cdot \mathbf{u}_3)]^{\mathsf{T}} = \left[\frac{4}{\sqrt{3}}\ \ -\frac{1}{\sqrt{2}}\ \frac{1}{\sqrt{6}}\right]^{\mathsf{T}}.$$

(c) Using the definition of the norm of a vector we have

$$||\mathbf{w}|| = \sqrt{1^2 + 2^2 + 1^2} = \sqrt{6}$$
$$||[\mathbf{w}]_{|CB}|| = \sqrt{\left(\frac{4}{\sqrt{3}}\right)^2 + \left(-\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{6}}\right)^2} = \sqrt{6}.$$

So in this case we have $||\mathbf{w}|| = ||[\mathbf{w}]_{\mathcal{B}}||$.

(d) Let $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$, and suppose that $\mathbf{z} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. Then

$$\begin{aligned} ||\mathbf{z}|| &= \sqrt{\mathbf{z} \cdot \mathbf{z}} \\ &= \sqrt{(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \cdot (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n)}. \end{aligned} \tag{28.4}$$

Since $\mathcal{S}$ is an orthonormal basis for $\mathbb{R}^n$, it follows that $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ if $i \neq j$ and $\mathbf{v}.\mathbf{v}_i = 1$. Expanding the dot product in (28.4), the only terms that won't be zero are the ones that involve $\mathbf{v}_i \cdot \mathbf{v}_i$. This leaves us with

$$\begin{aligned} ||\mathbf{z}|| &= \sqrt{(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \cdot (c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n)} \\ &= \sqrt{c_1c_1(\mathbf{v}_1 \cdot \mathbf{v}_1) + c_2c_2(\mathbf{v}_2 \cdot \mathbf{v}_2) + \cdots + c_nc_n(\mathbf{v}_n \cdot \mathbf{v}_n)} \\ &= \sqrt{c_1^2 + c_2^2 + \cdots + c_n^2}. \end{aligned}$$

Summary

- A subset S of $\mathbb{R}^n$ is an orthogonal set if $\mathbf{u} \cdot \mathbf{v} = 0$ for every pair of distinct vector $\mathbf{u}$ and $\mathbf{v}$ in S.
- Any orthogonal set of nonzero vectors is linearly independent.
- A basis $\mathcal{B}$ for a subspace W of $\mathbb{R}^n$ is an orthogonal basis if $\mathcal{B}$ is also an orthogonal set.
- An orthogonal basis $\mathcal{B}$ for a subspace W of $\mathbb{R}^n$ is an orthonormal basis if each vector in $\mathcal{B}$ has unit length.
- If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is an orthogonal basis for a subspace of $\mathbb{R}^n$ and $\mathbf{x}$ is any vector in W, then

$$\mathbf{x} = \sum_{i=1}^{m} c_i\mathbf{v}_i$$

where $c_i = \frac{\mathbf{x} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$.
- An $n \times n$ matrix P is an orthogonal matrix if $P^{\mathsf{T}}P = I_n$. Orthogonal matrices are important, in part, because the matrix transformations they define are isometries.

Exercises

(1) Find an orthogonal basis for the subspace $W = \{[x\ y\ z] : 4x - 3z = 0\}$ of $\mathbb{R}^3$.

(2) Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be an orthogonal basis for $\mathbb{R}^n$ and, for some k between 1 and n, let $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. Show that $\{\mathbf{v}_{k+1}, \mathbf{v}_{k+2}, \ldots, \mathbf{v}_n\}$ is a basis for $W^\perp$.

(3) Let W be a subspace of $\mathbb{R}^n$ for some n, and let $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ be an orthogonal basis for W. Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$. and define $\mathbf{w}$ as

$$\mathbf{w} = \frac{\mathbf{x} \cdot \mathbf{w}_1}{\mathbf{w}_1 \cdot \mathbf{w}_1}\mathbf{w}_1 + \frac{\mathbf{x} \cdot \mathbf{w}_2}{\mathbf{w}_2 \cdot \mathbf{w}_2}\mathbf{w}_2 + \cdots + \frac{\mathbf{x} \cdot \mathbf{w}_k}{\mathbf{w}_k \cdot \mathbf{w}_k}\mathbf{w}_k.$$

(a) Explain why $\mathbf{w}$ is in W.

(b) Let $\mathbf{z} = \mathbf{x} - \mathbf{w}$. Show that $\mathbf{z}$ is in $W^\perp$.

(c) Explain why $\mathbf{x}$ can be written as a sum of vectors, one in W and one in $W^\perp$.

(d) Suppose $\mathbf{x} = \mathbf{w} + \mathbf{w}_1$ and $\mathbf{x} = \mathbf{u} + \mathbf{u}_1$, where $\mathbf{w}$ and $\mathbf{u}$ are in W and $\mathbf{w}_1$ and $\mathbf{u}_1$ are in $W^\perp$. Show that $\mathbf{w} = \mathbf{u}$ and $\mathbf{w}_1 = \mathbf{u}_1$, so that the representation of $\mathbf{x}$ as a sum of a vector in W and a vector in $W^\perp$ is unique.

(4) Use the result of problem (3.) above and that $W \cap W^\perp = \{\mathbf{0}\}$ to show that $\dim(W) + \dim(W^\perp) = n$ for a subspace W of $\mathbb{R}^n$. (See Exercise 12 in Section 22 for the definition of the sum of subspaces.)

(5) Let P be an $n \times n$ matrix. We showed that if P is an orthogonal matrix, then $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for any vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. Now we ask if the converse of this statement is true. That is, determine the validity of the following statement: if $(P\mathbf{x}) \cdot (P\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for any vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, then P is an orthogonal matrix? Verify your answer. (Hint: Consider $(P\mathbf{e}_i) \cdot (P\mathbf{e}_j)$ where $\mathbf{e}_t$ is the tth standard basis vector for $\mathbb{R}^n$.)

(6) In this exercise we completely describe the 2×2 orthogonal matrices. Let P be an orthogonal 2×2 matrix.

(a) The columns of P must be orthonormal vectors, so if we place the initial point of either of the columns of P at the origin, explain why its terminal point must have the form $(\cos(t), \sin(t))$ for some real number t.

(b) As a consequence of part (a), let $[\cos(\theta)\ \ \sin(\theta)]^{\mathsf{T}}$ be the first column of P. Let the second column of P be $[\cos(\alpha)\ \ \sin(\alpha)]^{\mathsf{T}}$. Since the columns of P are orthogonal, how must the angle α be related to θ?

(c) Use a trigonometric identity to explain why

$$P = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

So the only orthogonal matrices in $\mathbb{R}^2$ are the rotation matrices.

(7) Suppose A, B are orthogonal matrices of the same size.

(a) Show that AB is also an orthogonal matrix.

(b) Show that A^{T} is also an orthogonal matrix.

(c) Show that A^{-1} is also an orthogonal matrix.

(8) Label each of the following statements as True or False. Provide justification for your response.

(a) **True/False** Any orthogonal subset of $\mathbb{R}^n$ is linearly independent.

(b) **True/False** Every single vector set is an orthogonal set.

(c) **True/False** If S is an orthogonal set in $\mathbb{R}^n$ with exactly n nonzero vectors, then S is a basis for $\mathbb{R}^n$.

(d) **True/False** Every set of three linearly independent vectors in $\mathbb{R}^3$ is an orthogonal set.

(e) **True/False** If A and B are $n \times n$ orthogonal matrices, then $A + B$ must also be an orthogonal matrix.

(f) **True/False** If the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal set in $\mathbb{R}^n$, then so is the set $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2, \ldots, c_n\mathbf{v}_n\}$ for any scalars c_1, c_2, ..., c_n.

(g) **True/False** If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an orthogonal basis of $\mathbb{R}^n$, then so is $\{c_1\mathbf{v}_1$, $c_2\mathbf{v}_2$, ..., $c_n\mathbf{v}_n\}$ for any nonzero scalars c_1, c_2, ..., c_n.

(h) **True/False** If A is an $n \times n$ orthogonal matrix, the rows of A form an orthonormal basis of $\mathbb{R}^n$.

(i) **True/False** If A is an orthogonal matrix, any matrix obtained by interchanging columns of A is also an orthogonal matrix.

Project: Understanding Rotations in 3-Space

Recall that a counterclockwise rotation of 2-space around the origin by an angle θ is accomplished by left multiplication by the matrix $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$. Notice that the columns of this rotation matrix are orthonormal, so this rotation matrix is an orthogonal matrix. As the next activity shows, rotation matrices in 3D are also orthogonal matrices.

Project Activity 28.1. Let R be a rotation matrix in 3D. A rotation does not change lengths of vectors, nor does it change angles between vectors. Let $\mathbf{e}_1 = [1\ 0\ 0]^{\mathsf{T}}$, $\mathbf{e}_2 = [0\ 1\ 0]^{\mathsf{T}}$, and $\mathbf{e}_3 = [0\ 0\ 1]^{\mathsf{T}}$ be the standard unit vectors in $\mathbb{R}^3$.

(a) Explain why the columns of R form an orthonormal set. (Hint: How are $R\mathbf{e}_1$, $R\mathbf{e}_2$, and $R\mathbf{e}_3$ related to the columns of R?)

(b) Explain why R is an orthogonal matrix. What must be true about $\det(R)$? (Hint: What is R^{T} and what is $\det(R^{\mathsf{T}}R)$?)

By Project Activity 28.1 we know that the determinant of any rotation matrix is either 1 or -1. Having a determinant of 1 preserves orientation, and we will identify these rotations as being counterclockwise, and we will identify the others with determinant of -1 as being clockwise. We will set the convention that a rotation is always measured counterclockwise (as we did in $\mathbb{R}^2$), and so every rotation matrix will have determinant 1.

Returning to the counterclockwise rotation of 2-space around the origin by an angle θ determined by left multiplication by the matrix $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$, we can think of this rotation in 3-space as the rotation that keeps points in the xy plane in the xy plane, but rotates these points counterclockwise around the z axis. In other words, in the standard xyz coordinate system, with standard basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, our rotation matrix R has the property that $R\mathbf{e}_3 = \mathbf{e}_3$. Now $R\mathbf{e}_3$ is the third column of R, so the third column of R is $\mathbf{e}_3$. Similarly, $R\mathbf{e}_1$ is the first column of R and $R\mathbf{e}_2$ is the second column of R. Since R is a counterclockwise rotation of the xy plane space around the origin by an angle θ it follows that this rotation is given by the matrix

$$R_{\mathbf{e}_3}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{28.5}$$

In this notation in (28.5), the subscript gives the direction of the line fixed by the rotation and the angle provides the counterclockwise rotation in the plane perpendicular to this vector. This vector is called a *normal* vector for the rotation. Note also that the columns of $R_{\mathbf{e}_3}(\theta)$ form an orthogonal set such that each column vector has norm 1.

This idea describes a general rotation matrix $R_{\mathbf{n}}(\theta)$ in 3D by specifying a normal vector $\mathbf{n}$ and an angle θ. For example, with roll, a normal vector points from the tail of the aircraft to its tip. It is our goal to understand how we can determine an arbitrary rotation matrix of the form $R_{\mathbf{n}}(\theta)$. We can accomplish this by using the rotation around the z axis and change of basis matrices to find rotation matrices around other axes. Let $\mathcal{S} = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be the standard basis for $\mathbb{R}^3$

Project Activity 28.2. In this activity we see how to determine the rotation matrix around the x axis using the matrix $R_{\mathbf{e}_3}(\theta)$ and a change of basis.

(a) Define a new ordered basis $\mathcal{B}$ so that our axis of rotation is the third vector. So in this case the third vector in $\mathcal{B}$ will be $\mathbf{e}_1$. The other two vectors need to make $\mathcal{B}$ an orthonormal set. So we have plenty of choices. For example, we could set $\mathcal{B} = \{\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1\}$. Find the change of basis matrix $\underset{\mathcal{S}\leftarrow\mathcal{B}}{P}$ from $\mathcal{B}$ to $\mathcal{S}$.

(b) Use the change of basis matrix from part (a) to find the change of basis matrix $\underset{\mathcal{B}\leftarrow\mathcal{S}}{P}$ from $\mathcal{S}$ to $\mathcal{B}$.

(c) To find our rotation matrix around the x axis, we can first change basis from $\mathcal{S}$ to $\mathcal{B}$, then perform a rotation around the new z axis using (28.5), then changing basis back from $\mathcal{B}$ to $\mathcal{S}$. In other words,

$$R_{\mathbf{e}_1}(\theta) = \underset{\mathcal{S}\leftarrow\mathcal{B}}{P} R_{\mathbf{e}_3}(\theta) \underset{\mathcal{B}\leftarrow\mathcal{S}}{P}.$$

Find the entries of this matrix $R_{\mathbf{e}_1}(\theta)$.

IMPORTANT NOTE: We could have considered using $\mathcal{B}_1 = \{\mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_1\}$ in Project Activity 28.2

instead of $\mathcal{B} = \{\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1\}$. Then we would have

$$\underset{\mathcal{S}\leftarrow\mathcal{B}_1}{P} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

The difference between the two options is that, in the first we have $\det\left(\underset{\mathcal{S}\leftarrow\mathcal{B}_1}{P}\right) = -1$ while $\det\left(\underset{\mathcal{S}\leftarrow\mathcal{B}}{P}\right) = 1$ in the second. Using $\mathcal{B}_1$ will give clockwise rotations while $\mathcal{B}$ gives counterclockwise rotations (this is the difference between a left hand system and a right hand system). So it is important to ensure that our change of basis matrix is one with determinant 1.

Project Activity 28.3. Repeat Project Activity 28.3 to find the 3D rotation around the y axis.

We do one more example to illustrate the process before tackling the general case.

Project Activity 28.4. In this activity we find the rotation around the axis given by the line $x = y/2 = z$. This line is in the direction of the vector $\mathbf{n} = [1\ 2\ 1]^{\mathsf{T}}$. So we start with making a unit vector in the direction of $\mathbf{n}$ as the third vector in an ordered basis $\mathcal{B}$. The other two vectors need to make $\mathcal{B}$ an orthonormal set with $\det\left(\underset{\mathcal{S}\leftarrow\mathcal{B}}{P}\right) = 1$.

(a) Find a unit vector $\mathbf{w}$ in the direction of $\mathbf{n}$.

(b) Show that $[2\ \ -1\ 0]^{\mathsf{T}}$ is orthogonal to the vector $\mathbf{w}$ from part (a). Then find a unit vector $\mathbf{v}$ that is in the same direction as $[2\ \ -1\ 0]^{\mathsf{T}}$.

(c) Let $\mathbf{v}$ be as in the previous part. Now the trick is to find a third unit vector $\mathbf{u}$ so that $\mathcal{B} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is an orthonormal set. This can be done with the cross product. If $\mathbf{a} = [a_1\ a_2\ a_3]^{\mathsf{T}}$ and $\mathbf{b} = [b_1\ b_2\ b_3]^{\mathsf{T}}$, then the cross product $\mathbf{a} \times \mathbf{b}$ of $\mathbf{a}$ and $\mathbf{b}$ is the vector

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2)\,\mathbf{e}_1 - (a_1b_3 - a_3b_1)\,\mathbf{e}_2 + (a_1b_2 - a_2b_1)\,\mathbf{e}_3.$$

(You can check that $\{\mathbf{a} \times \mathbf{b}, \mathbf{a}, \mathbf{b}\}$ is an orthogonal set that gives the correct determinant for the change of basis matrix.) Use the cross product to find a unit vector $\mathbf{u}$ so that $\mathcal{B} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is an orthonormal set.

(d) Find the entries of the matrix $R_{\mathbf{w}}(\theta)$.

In the next activity we summarize the general process to find a 3D rotation matrix $R_{\mathbf{n}}(\theta)$ for any normal vector $\mathbf{n}$. There is a GeoGebra applet at https://www.geogebra.org/m/n9gbjhfx that allows you to visualize rotation matrices in 3D.

Project Activity 28.5. Let $\mathbf{n} = [n_1\ n_2\ n_3]^{\mathsf{T}}$ be a normal vector (nonzero) for our rotation. We need to create an orthonormal basis $\mathcal{B} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ where $\mathbf{w}$ is a unit vector in the direction of $\mathbf{n}$ so that the change of basis matrix $\underset{\mathcal{S}\leftarrow\mathcal{B}}{P}$ has determinant 1.

(a) Find, by inspection, a vector $\mathbf{y}$ that is orthogonal to $\mathbf{n}$. (Hint: You may need to consider some cases to ensure that $\mathbf{v}$ is not the zero vector.)

(b) Once we have a normal vector $\mathbf{n}$ and a vector $\mathbf{y}$ orthogonal to $\mathbf{n}$, the vector $\mathbf{z} = \mathbf{y} \times \mathbf{n}$ gives us an orthogonal set $\{\mathbf{z}, \mathbf{y}, \mathbf{n}\}$. We then normalize each vector to create our orthonormal basis $\mathcal{B} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$. Use this process to find the matrix that produces a 45° counterclockwise rotation around the normal vector $[1\ 0\ -1]^{\mathsf{T}}$.

Section 29

Inner Products

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is an inner product? What is an inner product space?
- What is an orthogonal set in an inner product space?
- What is an orthogonal basis for an inner product space?
- How do properties of orthogonality in R^n generalize to orthogonality in an inner product space?
- How do we find the coordinate vector for a vector in an inner product space relative to an orthogonal basis for the space?
- What is the projection of a vector orthogonal to a subspace and why are such orthogonal projections important?

Application: Fourier Series

In calculus, a Taylor polynomial for a function f is a polynomial approximation that fits f well around the center of the approximation. For this reason, Taylor polynomials are good *local* approximations, but they are not in general good global approximations. In particular, if a function f has periodic behavior it is impossible to model f well globally with polynomials that have infinite limits at infinity. For these kinds of functions, trigonometric polynomials are better choices. Trigonometric polynomials lead us to Fourier series, and we will investigate how inner products allow us to use trigonometric polynomials to model musical tones later in this section.

Introduction

We have seen that orthogonality in $\mathbb{R}^n$ is an important concept. We can extend the idea of orthogonality, as well as the notions of length and angles, to a variety of different vector spaces beyond just $\mathbb{R}^n$ as long as we have a product like a dot product. Such products are called *inner products*. Inner products lead us to many important ideas like Fourier series, wavelets, and others.

Recall that the dot product on $\mathbb{R}^n$ assigns to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ the scalar $\mathbf{u} \cdot \mathbf{v}$. Thus, the dot product defines a mapping from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$. Recall also that the dot product is commutative, distributes over vector addition, and respects scalar multiplication. Additionally, the dot product of a vector by itself is always non-negative and is equal to 0 only when the vector is the zero vector. There is nothing special about using $\mathbb{R}^n$ as the source for our vectors, and we can extend the notion of a dot product to any vector space using these properties.

Definition 29.1. An **inner product** $\langle\ ,\ \rangle$ on a vector space V is a mapping from $V \times V \to \mathbb{R}$ satisfying

(1) $\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$ for all $\mathbf{u}$ and $\mathbf{v}$ in V,

(2) $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V,

(3) $\langle c\mathbf{u}, \mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$ for all $\mathbf{u}$, $\mathbf{v}$ in V and all scalars c,

(4) $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ for all $\mathbf{u}$ in V and $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

An **inner product space** is a vector space on which an inner product is defined.

Preview Activity 29.1.

(1) Suppose we are given the mapping from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ defined by

$$\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 + 2u_2v_2$$

for $\mathbf{u} = [u_1\ u_2]^{\mathsf{T}}$ and $\mathbf{v} = [v_1\ v_2]^{\mathsf{T}}$ in $\mathbb{R}^2$. Check that this mapping satisfies all of the properties of an inner product.

(2)

(3) Now consider the mapping from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\mathbb{R}$ defined by

$$\langle \mathbf{u}, \mathbf{v}\rangle = 2u_1v_1 - 3u_2v_2$$

for $\mathbf{u} = [u_1\ u_2]^{\mathsf{T}}$ and $\mathbf{v} = [v_1\ v_2]^{\mathsf{T}}$ in $\mathbb{R}^2$. Show that this mapping does not satisfy the fourth property of an inner product.

(4) Finally, show that the mapping from $\mathbb{P}_1 \times \mathbb{P}_1 \to \mathbb{R}$ defined by

$$\langle a_0 + a_1t, b_0 + b_1t\rangle = a_0b_0 + a_1b_1$$

for $a_0 + a_1t, b_0 + b_1t$ in $\mathbb{P}_1$ is an inner product on $\mathbb{P}_1$.

Inner Product Spaces

As we saw in Definition 29.1, the idea of the dot product in $\mathbb{R}^n$ can be extended to define inner products on different types of vector spaces. Preview Activity 29.1 provides two examples of inner products. The examples below provide some important inner products on vector spaces. Verification of the following as inner products is left to the exercises.

- If a_1, a_2, ..., a_n are positive scalars, then

$$\langle [u_1\ u_2\ \cdots\ u_n]^{\mathsf{T}}, [v_1\ v_2\ \cdots\ v_n]^{\mathsf{T}} \rangle = a_1u_1v_1 + a_2u_2v_2 + \cdots + a_nu_nv_n$$

  defines an inner product on $\mathbb{R}^n$.

- Every invertible $n \times n$ matrix A defines an inner product on $\mathbb{R}^n$ by

$$\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v}).$$

- The definite integral defines an inner product:

$$\langle f, g \rangle = \int_a^b f(x)g(x)\,dx$$

  for $f, g \in C[a, b]$ (where $C[a, b]$ is the vector space of all continuous functions on the interval $[a, b]$ – that $C[a, b]$ is a vector space is left for Exercise 1.)

- We can use the trace of a matrix (see Definition 18.7) to define an inner product on matrix spaces. If A and B are in the space $\mathcal{M}_{n\times n}$ of $n \times n$ matrices with real entries, we define the product $\langle A, B \rangle$ as

$$\langle A, B \rangle = \text{trace}\left(AB^{\mathsf{T}}\right).$$

  This defines an inner inner product on the space $\mathcal{M}_{n\times n}$ called the *Frobenius* inner product.

Since we defined inner products using the properties of the dot product, we might wonder if inner products actually satisfy all of the other properties of the dot product. For example, $\mathbf{u} \cdot \mathbf{0} = 0$ for any vector $\mathbf{u}$ in $\mathbb{R}^n$; but is it true that $\langle \mathbf{u}, \mathbf{0} \rangle = 0$ in every inner product space? This property is not part of the definition of an inner product, so we need to verify it if true.

Activity 29.1. Let $\mathbf{u}$ be a vector in an inner product space V.

(a) Why is $\langle \mathbf{u}, \mathbf{0} \rangle = \langle \mathbf{u}, \mathbf{0} \rangle + \langle \mathbf{u}, \mathbf{0} \rangle$?

(b) How does the equation in part (a) show that $\langle \mathbf{u}, \mathbf{0} \rangle = 0$?

Activity 29.1 suggests that inner products share the defining properties of the dot product. Some properties of the inner product are given in the following theorem (the proofs of the remaining parts are left to the Exercises).

Theorem 29.2. *Let* $\langle\ ,\ \rangle$ *be an inner product on a vector space* V *and let* $\mathbf{u}, \mathbf{v}$*, and* $\mathbf{w}$ *be vectors in* V *and* c *a scalar. Then*

(1) $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$

(2) $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$

(3) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$

(4) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$

Inner products will allow us to extend ideas of orthogonality, lengths of vectors, and angles between vectors to any inner product space.

The Length of a Vector

We can use inner products to define the length of any vector in an inner product space and the distance between two vectors in an inner product space. The idea comes right from the relationship between lengths of vectors in $\mathbb{R}^n$ and the dot product (compare to Definition 27.2).

Definition 29.3. Let $\mathbf{v}$ be a vector in an inner product space V. The **length** of $\mathbf{v}$ is the real number

$$||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

The length of a vector in a vector space is also called *magnitude* or *norm*. Just as in $\mathbb{R}^n$ we can use the notion of length to define unit vectors in inner product spaces (compare to Definition 27.4).

Definition 29.4. A vector $\mathbf{v}$ in inner product space is a **unit vector** if $||\mathbf{v}|| = 1$.

We can find a unit vector in the direction of a nonzero vector $\mathbf{v}$ in an inner product space V by dividing by the norm of the vector. That is, the vector $\frac{\mathbf{v}}{||\mathbf{v}||}$ is a unit vector in the direction of the vector $\mathbf{v}$, provided that $\mathbf{v}$ is not zero.

We define the distance between vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V in the same way we defined distance in $\mathbb{R}^n$ (compare to Definition 27.5).

Definition 29.5. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in an inner product space V. The **distance between** $\mathbf{u}$ and $\mathbf{v}$ is the length of the difference $\mathbf{u} - \mathbf{v}$ or

$$||\mathbf{u} - \mathbf{v}||.$$

Activity 29.2. Find the indicated length or distance in the inner product space.

(a) Find the length of the vectors $\mathbf{u} = [1\ 3]^{\mathsf{T}}$ and $\mathbf{v} = [3\ 1]^{\mathsf{T}}$ using the inner product

$$\langle [u_1\ u_2]^{\mathsf{T}}, [v_1\ v_2]^{\mathsf{T}} \rangle = 2u_1v_1 + 3u_2v_2$$

in $\mathbb{R}^2$.

(b) Find the distance between the polynomials $p(t) = t + 1$ and $q(t) = t^2 - 1$ in $C[0, 1]$ using the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)\ dx$. (You may assume that $\langle\ ,\ \rangle$ defines an inner product on $C[0, 1]$, the space of continuous functions defined on the interval $[0, 1]$.

Orthogonality in Inner Product Spaces

We defined orthogonality in $\mathbb{R}^n$ using the dot product (see Definition 27.6) and the angle between vectors in $\mathbb{R}^n$. We can extend that idea to any inner product space.

We can define the angle between two vectors in an inner product space just as we did in $\mathbb{R}^n$. If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in an inner product space V, then the **angle θ between $\mathbf{u}$ and $\mathbf{v}$** is such that

$$\cos(\theta) = \frac{|\langle \mathbf{u}, \mathbf{v} \rangle|}{||\mathbf{u}|| \, ||\mathbf{v}||}.$$

and $0 \leq \theta \leq \pi$. This angle is well-defined due to the Cauchy-Schwarz inequality $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}|| \, ||\mathbf{v}||$ whose proof is left to the exercises.

With the angle between vectors in mind, we can define orthogonal vectors in an inner product space.

Definition 29.6. Vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V are **orthogonal** if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

Note that this defines the zero vector to be orthogonal to every vector.

Activity 29.3.

(a) Find a nonzero vector in $\mathbb{R}^2$ orthogonal to the vector $\mathbf{u} = [3 \ 1]^{\mathsf{T}}$ using the inner product $\langle [u_1 \ u_2]^{\mathsf{T}}, [v_1 \ v_2]^{\mathsf{T}} \rangle = 2u_1v_1 + 3u_2v_2$.

(b) Determine if the vector $\mathbf{v} = \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}$ is orthogonal to the vector $\mathbf{w} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ using the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ on $\mathbb{R}^3$, where $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

(c) Find the angle between the two polynomials $p(t) = 1$ and $q(t) = t$ in $\mathbb{P}_1$ with inner product $\langle r(t), s(t) \rangle = \int_0^1 r(t)s(t) \, dt$.

Using orthogonality we can generalize the notions of orthogonal sets and bases, orthonormal bases and orthogonal complements we defined in $\mathbb{R}^n$ to all inner product spaces in a natural way.

Orthogonal and Orthonormal Bases in Inner Product Spaces

As we did in $\mathbb{R}^n$, we define an orthogonal set to be one in which all of the vectors in the set are orthogonal to each other (compare to Definition 28.1).

Definition 29.7. A subset S of an inner product space V for which $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ for all $\mathbf{u} \neq \mathbf{v}$ in S is called an **orthogonal set**.

As in $\mathbb{R}^n$, an orthogonal set of nonzero vectors is always linearly independent. The proof is similar to that of Theorem 28.3 and is left to the Exercises.

Theorem 29.8. *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a set of nonzero orthogonal vectors in an inner product space V. Then the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, ..., $\mathbf{v}_m$ are linearly independent.*

A basis that is also an orthogonal set is given a special name (compare to Definition 28.2).

Definition 29.9. An **orthogonal basis** $\mathcal{B}$ for a subspace W of an inner product space V is a basis of W that is also an orthogonal set.

Using the dot product in $\mathbb{R}^n$, we saw that the representation of a vector as a linear combination of vectors in an orthogonal or orthonormal basis was quite elegant. The same is true in any inner product space. To see this, let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be an orthogonal basis for a subspace W of an inner product space V and let $\mathbf{x}$ be any vector in W. We know that

$$\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_m\mathbf{v}_m$$

for some scalars x_1, x_2, ..., x_m. If $1 \leq k \leq m$, then, using inner product properties and the orthogonality of the vectors $\mathbf{v}_i$, we have

$$\langle \mathbf{v}_k, \mathbf{x}\rangle = x_1\langle \mathbf{v}_k, \mathbf{v}_1\rangle + x_2\langle \mathbf{v}_k, \mathbf{v}_2\rangle + \cdots + x_m\langle \mathbf{v}_k, \mathbf{v}_m\rangle = x_k\langle \mathbf{v}_k, \mathbf{v}_k\rangle.$$

So

$$x_k = \frac{\langle \mathbf{x}, \mathbf{v}_k\rangle}{\langle \mathbf{v}_k, \mathbf{v}_k\rangle}.$$

Thus, we can calculate each weight individually with two simple inner product calculations.

In other words, the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ of $\mathbf{x}$ in an inner product space V with orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is given by

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \frac{\langle \mathbf{x}, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle} \\ \frac{\langle \mathbf{x}, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle} \\ \vdots \\ \frac{\langle \mathbf{x}, \mathbf{v}_m\rangle}{\langle \mathbf{v}_m, \mathbf{v}_m\rangle} \end{bmatrix}.$$

We summarize this discussion in the next theorem (compare to Theorem 28.4).

Theorem 29.10. *Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be an orthogonal basis for a subspace of an inner product space V. Let $\mathbf{x}$ be a vector in W. Then*

$$\mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}\mathbf{v}_1 + \frac{\langle \mathbf{x}, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{x}, \mathbf{v}_m\rangle}{\langle \mathbf{v}_m, \mathbf{v}_m\rangle}\mathbf{v}_m. \tag{29.1}$$

Activity 29.4. Let $p_1(t) = 1 - t$, $p_2(t) = -2 + 4t + 4t^2$, and $p_3(t) = 7 - 41t + 40t^2$ be vectors in the inner product space $\mathbb{P}_2$ with inner product defined by $\langle p(t), q(t)\rangle = \int_0^1 p(t)q(t)\, dt$. Let $\mathcal{B} = \{p_1(t), p_2(t), p_3(t)\}$. You may assume that $\mathcal{B}$ is an orthogonal basis for $\mathbb{P}_2$. Let $z(t) = 4-2t^2$. Find the weight x_3 so that $z(t) = x_1p_1(t) + x_2p_2(t) + x_3p_3(t)$. Use technology as appropriate to evaluate any integrals.

The decomposition (29.1) is even simpler if $\langle \mathbf{v}_k, \mathbf{v}_k\rangle = 1$ for each k. Recall that

$$\langle \mathbf{v}, \mathbf{v}\rangle = ||\mathbf{v}||^2,$$

so the condition $\langle \mathbf{v}, \mathbf{v}\rangle = 1$ implies that the vector $\mathbf{v}$ has norm 1. As in $\mathbb{R}^n$, an orthogonal basis with this additional condition is given a special name (compare to Definition 28.5).

Definition 29.11. An **orthonormal basis** $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ for a subspace W of an inner product space V is an orthogonal basis such that $||\mathbf{v}_k|| = 1$ for $1 \leq k \leq m$.

If $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ is an orthonormal basis for a subspace W of an inner product space V and $\mathbf{x}$ is a vector in W, then (29.1) becomes

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{x}, \mathbf{v}_2 \rangle \mathbf{v}_2 + \cdots + \langle \mathbf{x}, \mathbf{v}_m \rangle \mathbf{v}_m. \tag{29.2}$$

A good question to ask here is how we can construct an orthonormal basis from an orthogonal basis.

Activity 29.5. Consider vectors from an inner product space V.

(a) Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be orthogonal vectors. Explain how we can obtain unit vectors $\mathbf{u}_1$ in the direction of $\mathbf{v}_1$ and $\mathbf{u}_2$ in the direction of $\mathbf{v}_2$.

(b) Show that $\mathbf{u}_1$ and $\mathbf{u}_2$ from the previous part are orthogonal vectors.

(c) Use the ideas from this problem to construct an orthonormal basis for the subspace

$$W = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 8 \\ 5 \\ -31 \\ -3 \end{bmatrix} \right\}$$

of the inner product space $\mathbb{R}^4$ with inner product

$$\langle [u_1\ u_2\ u_3\ u_4]^{\mathsf{T}}, [v_1\ v_2\ v_3\ v_4]^{\mathsf{T}} \rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3 + 5u_4v_4.$$

(Note that you need to check for orthogonality.)

Orthogonal Projections onto Subspaces

Preview Activity 29.2. Let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2\}$ be a basis for a subspace W of $\mathbb{R}^3$, where $\mathbf{w}_1 = [1\ 0\ 0]^{\mathsf{T}}$ and $\mathbf{w}_2 = [0\ 1\ 0]^{\mathsf{T}}$. Note that $\mathcal{B}$ is an orthonormal basis for W using the dot product as inner product. Let $\mathbf{v} = [1\ 2\ 1]^{\mathsf{T}}$. Notice also that W is the xy-plane and that $\mathbf{v}$ is not in W as illustrated in Figure 29.1.

(1) Find the orthogonal projection $\mathbf{u}_1$ of $\mathbf{v}$ onto $W_1 = \text{Span}\{\mathbf{w}_1\}$. (Hint: See Equation (27.2).)

(2) Find the orthogonal projection $\mathbf{u}_2$ of $\mathbf{v}$ onto $W_2 = \text{Span}\{\mathbf{w}_2\}$.

(3) Calculate the distance between $\mathbf{v}$ and $\mathbf{u}_1$ and $\mathbf{v}$ and $\mathbf{u}_2$. Which of $\mathbf{u}_1$ and $\mathbf{u}_2$ is closer to $\mathbf{v}$?

(4) Show that the vector $\frac{1}{2}[1\ 4\ 0]^{\mathsf{T}}$ is in W and find the distance between $\mathbf{v}$ and $\frac{1}{2}[1\ 4\ 0]^{\mathsf{T}}$.

(5) Part (4) shows that neither vector $\mathbf{u}_1 = \text{proj}_{\mathbf{w}_1}\mathbf{v}$ nor $\mathbf{u}_2 = \text{proj}_{\mathbf{w}_2}\mathbf{v}$ is the vector in W that is closest to $\mathbf{v}$. We should probably expect this since neither projection uses the fact that the other vector might contribute to the closest vector. Our goal is to find the linear combination $\mathbf{w}$ of $\mathbf{w}_1$ and $\mathbf{w}_2$ in W that makes $||\mathbf{w} - \mathbf{v}||$ the smallest. Letting $\mathbf{w} = a\mathbf{w}_1 + b\mathbf{w}_2$, we have that $||\mathbf{w} - \mathbf{v}|| = \sqrt{(a-1)^2 + (b-2)^2 + 1}$. Find the weights a and b that minimize this norm $||\mathbf{w} - \mathbf{v}||$.

(6) A picture of $\mathbf{w}_1$, $\mathbf{w}_2$, W, and $\mathbf{v}$ is shown in Figure 29.1. Draw in $\mathbf{u}_1$, $\mathbf{u}_2$, and the vector $\mathbf{w}$ you found in part (5). There is a specific relationship between this vector $\mathbf{w}$ and $\mathbf{u}_1$ and $\mathbf{u}_2$. Describe this relationship algebraically and illustrate it graphically in Figure 29.1.

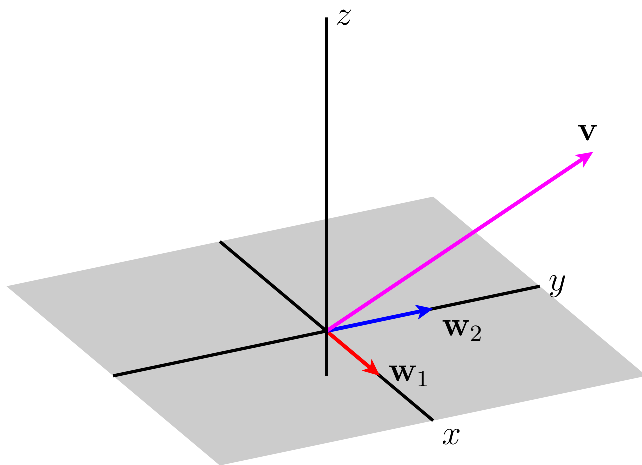

Figure 29.1: The space W and vectors $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{v}$

Preview Activity 29.2 gives an indication of how we can project a vector $\mathbf{v}$ in $\mathbb{R}^n$ onto a subspace W of $\mathbb{R}^n$. If we have an orthogonal basis for W, we can just add the orthogonal projections of $\mathbf{v}$ onto each basis vector. The resulting vector is called the orthogonal projection of $\mathbf{v}$ onto the subspace W. As we did with orthogonal projections onto vectors, we can also define the projection of $\mathbf{v}$ orthogonal to W. All of this can be done in the context of inner product spaces. Note that to make this all work out properly, we will need an orthogonal basis for W.

Definition 29.12. Let W be a subspace of an inner product space V and let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be an orthogonal basis for W. For a vector $\mathbf{v}$ in V, the **orthogonal projection of v onto** W is the vector

$$\text{proj}_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_m \rangle}{\langle \mathbf{w}_m, \mathbf{w}_m \rangle}\mathbf{w}_m.$$

The **projection of v orthogonal to** W is the vector

$$\text{proj}_{W^\perp} \mathbf{v} = \mathbf{v} - \text{proj}_W \mathbf{v}.$$

The notation $\text{proj}_{W^\perp} \mathbf{v}$ indicates that we expect this vector to be orthogonal to every vector in W.

Activity 29.6. Let $W = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ in $\mathbb{R}^4$, where $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 2 \end{bmatrix}$, and $\mathbf{w}_3 = \begin{bmatrix} 8 \\ 5 \\ -31 \\ -3 \end{bmatrix}$. Recall that we showed in Activity 29.5 that this was an orthogonal basis. Find the projection of the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ onto W using the inner product

$$\langle [u_1\ u_2\ u_3\ u_4]^{\mathsf{T}}, [v_1\ v_2\ v_3\ v_4]^{\mathsf{T}} \rangle = 2u_1v_1 + 3u_2v_2 + u_3v_3 + 5u_4v_4.$$

Show directly that $\text{proj}_{W^\perp}\mathbf{v}$ is orthogonal to the basis vectors for W.

Activity 29.6 indicates that the vector $\text{proj}_{W^\perp}\mathbf{v}$ is in fact orthogonal to every vector in W. To see that this is true in general, let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be an orthogonal basis for a subspace W of an inner product space V and let $\mathbf{v}$ be a vector in V. Let

$$\mathbf{w} = \text{proj}_W\mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_m \rangle}{\langle \mathbf{w}_m, \mathbf{w}_m \rangle}\mathbf{w}_m.$$

Then $\mathbf{v} - \mathbf{w}$ is the projection of $\mathbf{v}$ orthogonal to W. We will show that $\mathbf{v} - \mathbf{w}$ is orthogonal to every basis vector for W. Since $\mathcal{B}$ is an orthogonal basis for W, we know that $\mathbf{w}_i \cdot \mathbf{w}_j = 0$ for $i \neq j$. So if k is between 1 and m then

$$\begin{aligned}
\langle \mathbf{w}_k, \mathbf{v} - \mathbf{w} \rangle &= \langle \mathbf{w}_k, \mathbf{v} \rangle - \langle \mathbf{w}_k, \mathbf{w} \rangle \\
&= \langle \mathbf{w}_k, \mathbf{v} \rangle - \left[\langle \mathbf{w}_k, \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_m \rangle}{\langle \mathbf{w}_m, \mathbf{w}_m \rangle}\mathbf{w}_m \rangle \right] \\
&= \langle \mathbf{w}_k, \mathbf{v} \rangle - \left(\frac{\langle \mathbf{v}, \mathbf{w}_k \rangle}{\langle \mathbf{w}_k, \mathbf{w}_k \rangle} \right) \langle \mathbf{w}_k, \mathbf{w}_k \rangle \\
&= \langle \mathbf{w}_k, \mathbf{v} \rangle - \langle \mathbf{v}, \mathbf{w}_k \rangle \\
&= 0.
\end{aligned}$$

So the vector $\mathbf{v} - \mathbf{w}$ is orthogonal to every basis vector for W, and therefore to every vector in W (see Theorem 27.10). So, in fact, $\text{proj}_{W^\perp}\mathbf{v}$ is the projection of $\mathbf{v}$ onto the orthogonal complement of W, which will be defined shortly.

Best Approximations

In many situations we are interested in approximating a vector that is not in a subspace with a vector in the subspace (e.g., linear regression to fit a line to a set of data). In these cases we usually want to find the vector in the subspace that "best" approximates the given vector using a specified inner product. As we will soon see, the projection of a vector onto a subspace has the property that the projection is the "best" approximation over all vectors in the subspace in terms of the length. In other words, $\text{proj}_W\mathbf{v}$ is the vector in W closest to $\mathbf{v}$ and therefore the best approximation of $\mathbf{v}$ by a vector in W. To see that this is true in any inner product space, we first need a generalization of the Pythagorean Theorem that holds in inner product spaces.

Theorem 29.13 (Generalized Pythagorean Theorem)**.** *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be orthogonal vectors in an inner product space* V*. Then*

$$||\mathbf{u} - \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2.$$

Proof. Let $\mathbf{u}$ and $\mathbf{v}$ be orthogonal vectors in an inner product space V. Then

$$\begin{aligned}||\mathbf{u}-\mathbf{v}||^2 &= \langle \mathbf{u}-\mathbf{v}, \mathbf{u}-\mathbf{v}\rangle \\ &= \langle \mathbf{u},\mathbf{u}\rangle - 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\rangle \\ &= \langle \mathbf{u},\mathbf{u}\rangle - 2(0) + \langle \mathbf{v},\mathbf{v}\rangle \\ &= ||\mathbf{u}||^2 + ||\mathbf{v}||^2.\end{aligned}$$

■

Note that replacing $\mathbf{v}$ with $-\mathbf{v}$ in the theorem also shows that $||\mathbf{u}+\mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$ if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

Now we will prove that the projection of a vector $\mathbf{u}$ onto a subspace W of an inner product space V is the best approximation in W to the vector $\mathbf{u}$.

Theorem 29.14. *Let W be a subspace of an inner product space V and let $\mathbf{u}$ be a vector in V. Then*

$$||\mathbf{u} - proj_W \mathbf{u}|| < ||\mathbf{u}-\mathbf{x}||$$

for every vector $\mathbf{x}$ in W different from $proj_W \mathbf{u}$.

Proof. Let W be a subspace of an inner product space V and let $\mathbf{u}$ be a vector in V. Let $\mathbf{x}$ be a vector in W. Now

$$\mathbf{u}-\mathbf{x} = (\mathbf{u} - \text{proj}_W \mathbf{u}) + (\text{proj}_W \mathbf{u} - \mathbf{x}).$$

Since both $\text{proj}_W \mathbf{u}$ and $\mathbf{x}$ are in W, we know that $\text{proj}_W \mathbf{u} - \mathbf{x}$ is in W. Since $\text{proj}_{W^\perp} \mathbf{u} = \mathbf{u} - \text{proj}_W \mathbf{u}$ is orthogonal to every vector in W, we have that $\mathbf{u} - \text{proj}_W \mathbf{u}$ is orthogonal to $\text{proj}_W \mathbf{u} - \mathbf{x}$. We can now use the Generalized Pythagorean Theorem to conclude that

$$||\mathbf{u}-\mathbf{x}||^2 = ||\mathbf{u} - \text{proj}_W \mathbf{u}||^2 + ||\text{proj}_W \mathbf{u} - \mathbf{x}||^2.$$

Since $\mathbf{x} \neq \text{proj}_W \mathbf{u}$, it follows that $||\text{proj}_W \mathbf{u} - \mathbf{x}||^2 > 0$ and

$$||\mathbf{u}-\mathbf{x}||^2 > ||\mathbf{u} - \text{proj}_W \mathbf{u}||^2.$$

Since norms are nonnegative, we can conclude that $||\mathbf{u} - \text{proj}_W \mathbf{u}|| < ||\mathbf{u}-\mathbf{x}||$ as desired. ■

Theorem 29.14 shows that the distance from $\text{proj}_W \mathbf{v}$ to $\mathbf{v}$ is less than the distance from any other vector in W to $\mathbf{v}$. So $\text{proj}_W \mathbf{v}$ is the best approximation to $\mathbf{v}$ of all the vectors in W.

In $\mathbb{R}^n$ using the dot product as inner product, if $\mathbf{v} = [v_1\ v_2\ v_3\ \ldots\ v_n]^{\mathsf{T}}$ and $\text{proj}_W \mathbf{v} = [w_1\ w_2\ w_3\ \ldots\ w_n]^{\mathsf{T}}$, then the square of the error in approximating $\mathbf{v}$ by $\text{proj}_W \mathbf{v}$ is given by

$$||\mathbf{v} - \text{proj}_W \mathbf{v}||^2 = \sum_{i=1}^{n} (v_i - w_i)^2.$$

So $\text{proj}_W \mathbf{v}$ minimizes this sum of squares over all vectors in W. As a result, we call $\text{proj}_W \mathbf{v}$ the *least squares approximation* to $\mathbf{v}$.

Activity 29.7. The set $\mathcal{B} = \{1, t - \frac{1}{2}, t^3 - \frac{9}{10}t + \frac{1}{5}\}$ is an orthogonal basis for a subspace W of the inner product space $\mathbb{P}_3$ using the inner product $\langle p(t), q(t)\rangle = \int_0^1 p(t)q(t)\,dt$. Find the polynomial in W that is closest to the polynomial $r(t) = t^2$ and give a numeric estimate of how good this approximation is.

Orthogonal Complements

If we have a set of vectors S in an inner product space V, we can define the orthogonal complement of S as we did in $\mathbb{R}^n$ (see 27.8).

Definition 29.15. The **orthogonal complement** of a subset S of an inner product space V is the set

$$S^{\perp} = \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{u} \rangle = 0 \text{ for all } \mathbf{u} \in S\}.$$

As we saw in $\mathbb{R}^n$, to show that a vector is in the orthogonal complement of a subspace, it is enough to show that the vector is orthogonal to every vector in a basis for that subspace (see Theorem 27.10). The proof is left for the exercises.

Theorem 29.16. *Let $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ be a basis for a subspace W of an inner product space V. A vector $\mathbf{v}$ in V is orthogonal to every vector in W if and only if $\mathbf{v}$ is orthogonal to every vector in $\mathcal{B}$.*

Activity 29.8. Consider $\mathbb{P}_2$ with the inner product $\langle p(t), q(t) \rangle = \int_0^1 p(t)q(t)\, dt$.

(a) Find $\langle p(t), 1-t \rangle$ where $p(t) = a + bt + ct^2$ is in $\mathbb{P}_2$.

(b) Describe as best you can the orthogonal complement of Span$\{1-t\}$ in $\mathbb{P}_2$. Is $p(t) = 1 - 2t - 2t^2$ in this orthogonal complement? Is $p(t) = 1 + t - t^2$?

To conclude this section we will investigate an important connection between a subspace W and $W^{\perp}$.

Activity 29.9. Let V be an inner product space of dimension n, and let W be a subspace of V. Let $\mathbf{x}$ be any vector in V. We will demonstrate that $\mathbf{x}$ can be written uniquely as a sum of a vector in W and a vector in $W^{\perp}$.

(a) Explain why $\text{proj}_W \mathbf{x}$ is in W.

(b) Explain why $\text{proj}_{W^{\perp}} \mathbf{x}$ is in $W^{\perp}$.

(c) Explain why $\mathbf{x}$ can be written as a sum of vectors, one in W and one in $W^{\perp}$.

(d) Now we demonstrate the uniqueness of this decomposition. Suppose $\mathbf{x} = \mathbf{w} + \mathbf{w}_1$ and $\mathbf{x} = \mathbf{u} + \mathbf{u}_1$, where $\mathbf{w}$ and $\mathbf{u}$ are in W and $\mathbf{w}_1$ and $\mathbf{u}_1$ are in $W^{\perp}$. Show that $\mathbf{w} = \mathbf{u}$ and $\mathbf{w}_1 = \mathbf{u}_1$, so that the representation of $\mathbf{x}$ as a sum of a vector in W and a vector in $W^{\perp}$ is unique. (Hint: What is $W \cap W^{\perp}$?)

We summarize the result of Activity 29.9.

Theorem 29.17. *Let V be a finite dimensional inner product space, and let W be a subspace of V. Any vector in V can be written in a unique way as a sum of a vector in W and a vector in $W^{\perp}$.*

Theorem 29.17 is useful in many applications. For example, to compress an image using wavelets, we store the image as a collection of data, then rewrite the data using a succession of subspaces and their orthogonal complements. This new representation allows us to visualize the data in a way that compression is possible.

Examples

What follows are worked examples that use the concepts from this section.

Example 29.18. Let $V = \mathbb{P}_3$ be the inner product space with inner product

$$\langle p(t), q(t)\rangle = \int_{-1}^{1} p(t)q(t)\, dt.$$

Let $p_1(t) = 1 + t$, $p_2(t) = 1 - 3t$, $p_3(t) = 3t - 5t^3$, and $p_4(t) = 1 - 3t^2$.

(a) Show that the set $\mathcal{B} = \{p_1(t), p_2(t), p_3(t), p_4(t)\}$ is an orthogonal basis for V.

(b) Use 29.10 to write the polynomial $q(t) = t^2+t^3$ as a linear combination of the basis vectors in $\mathcal{B}$.

Example Solution. All calculations are done by hand or with a computer algebra system, so we leave those details to the reader.

(a) If we show that the set $\mathcal{B}$ is an orthogonal set, then Theorem 29.8 shows that $\mathcal{B}$ is linearly independent. Since $\dim(\mathbb{P}_3) = 4$, the linearly independent set $\mathcal{B}$ that contains four vectors must be a basis for $\mathbb{P}_3$.

To determine if the set $\mathcal{B}$ is an orthogonal set, we must calculate the inner products of pairs of distinct vectors in $\mathcal{B}$. Since $\langle 1+t, 1-3t\rangle = 0$, $\langle 1+t, 3t-5t^3\rangle = 0$, $\langle 1+t, 1-3t^2\rangle = 0$, $\langle 1-3t, 3t-5t^3\rangle = 0$, $\langle 1-3t, 1-3t^2\rangle = 0$, and $\langle 3t-5t^3, 1-3t^2\rangle = 0$, we conclude that $\mathcal{B}$ is an orthogonal basis for $\mathbb{P}_3$.

(b) We can write the polynomial $q(t) = 1 + t + t^2 + t^3$ as a linear combination of the basis vectors in $\mathcal{B}$ as follows:

$$q(t) = \frac{\langle q(t), p_1(t)\rangle}{\langle p_1(t), p_1(t)\rangle} p_1(t) + \frac{\langle q(t), p_2(t)\rangle}{\langle p_2(t), p_2(t)\rangle} p_2(t) + \frac{\langle q(t), p_3(t)\rangle}{\langle p_3(t), p_3(t)\rangle} p_3(t) + \frac{\langle q(t), p_4(t)\rangle}{\langle p_4(t), p_4(t)\rangle} p_4(t).$$

Now

$$\begin{array}{ccc} \langle q(t), p_1(t)\rangle = \frac{16}{15}, & \langle q(t), p_2(t)\rangle = -\frac{8}{15}, & \langle q(t), p_3(t)\rangle = -\frac{8}{35}, \\ \langle q(t), p_4(t)\rangle = -\frac{8}{15}, & \langle p_1(t), p_1(t)\rangle = \frac{8}{3}, & \langle p_2(t), p_2(t)\rangle = 8, \\ \langle p_3(t), p_3(t)\rangle = \frac{8}{7}, & \langle p_4(t), p_4(t)\rangle = \frac{8}{5} & \end{array}$$

so

$$\begin{aligned} q(t) &= \frac{\frac{16}{15}}{\frac{8}{3}} p_1(t) - \frac{\frac{8}{15}}{8} p_2(t) - \frac{\frac{8}{35}}{\frac{8}{7}} p_3(t) - \frac{\frac{8}{15}}{\frac{8}{5}} p_4(t) \\ & \frac{2}{5} p_1(t) - \frac{1}{15} p_2(t) - \frac{1}{5} p_3(t) - \frac{1}{3} p_4(t). \end{aligned}$$

Example 29.19. Let V be the inner product space $\mathbb{R}^4$ with inner product defined by

$$\langle [u_1\ u_2\ u_3\ u_4]^{\mathsf{T}}, [v_1\ v_2\ v_3\ v_4]^{\mathsf{T}}\rangle = u_1v_1 + 2u_2v_2 + 3u_3v_3 + 4u_4v_4.$$

(a) Let W be the plane spanned by $[-1\ 1\ 0\ 1]^{\mathsf{T}}$ and $[6\ 1\ 7\ 1]^{\mathsf{T}}$ in V. Find the vector in W that is closest to the vector $[2\ 0\ -1\ 3]^{\mathsf{T}}$. Exactly how close is your best approximation to the vector $[2\ 0\ -1\ 3]^{\mathsf{T}}$?

(b) Express the vector $[2\ 0\ -1\ 3]^{\mathsf{T}}$ as the sum of a vector in W and a vector orthogonal to W.

Example Solution.

(a) The vector we're looking for is the projection of $[2\ 0\ -1\ 3]^{\mathsf{T}}$ onto the plane. A spanning set for the plane is $\mathcal{B} = \{[-1\ 1\ 0\ 1]^{\mathsf{T}}, [6\ 1\ 7\ 1]^{\mathsf{T}}\}$. Neither vector in $\mathcal{B}$ is a scalar multiple of the other, so $\mathcal{B}$ is a basis for the plane. Since

$$\langle [-1\ 1\ 0\ 1]^{\mathsf{T}}, [6\ 1\ 7\ 1]^{\mathsf{T}} \rangle = -6 + 2 + 0 + 4 = 0,$$

the set $\mathcal{B}$ is an orthogonal basis for the plane.

The projection of the vector $\mathbf{v} = [2\ 0\ -1\ 3]^{\mathsf{T}}$ onto the plane spanned by $\mathbf{w}_1 = [-1\ 1\ 0\ 1]^{\mathsf{T}}$ and $\mathbf{w}_2 = [6\ 1\ 7\ 1]^{\mathsf{T}}$ is given by

$$\begin{aligned}\frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 &= \frac{10}{7}[-1\ 1\ 0\ 1]^{\mathsf{T}} + \frac{3}{189}[6\ 1\ 7\ 1]^{\mathsf{T}} \\ &= \frac{1}{189}[-252\ 273\ 21\ 273]\mathsf{T} \\ &= \left[-\frac{4}{3}\ \frac{13}{9}\ \frac{1}{9}\ \frac{13}{9}\right].\end{aligned}$$

To measure how close close $\left[-\frac{4}{3}\ \frac{13}{9}\ \frac{1}{9}\ \frac{13}{9}\right]$ is to $[2\ 0\ -1\ 3]^{\mathsf{T}}$, we calculate

$$\begin{aligned}\left|\left|\left[-\frac{4}{3}\ \frac{13}{9}\ \frac{1}{9}\ \frac{13}{9}\right] - [2\ 0\ -1\ 3]^{\mathsf{T}}\right|\right| &= \left|\left|\left[-\frac{10}{3}\ \frac{13}{9}\ \frac{10}{9}\ -\frac{14}{9}\right]\right|\right| \\ &= \sqrt{\frac{100}{9} + \frac{338}{81} + \frac{300}{81} + \frac{784}{81}} \\ &= \frac{1}{9}\sqrt{2322} \\ &\approx 5.35.\end{aligned}$$

(b) If $\mathbf{v} = [2\ 0\ -1\ 3]^{\mathsf{T}}$, then $\text{proj}_W \mathbf{v}$ is in W and

$$\text{proj}_{W^\perp}\mathbf{v} = \mathbf{v} - \text{proj}_W \mathbf{v} = \left[-\frac{10}{3}\ \frac{13}{9}\ \frac{10}{9}\ -\frac{14}{9}\right]$$

is in $W^\perp$, and $\mathbf{v} = \text{proj}_W \mathbf{v} + \text{proj}_{W^\perp}\mathbf{v}$.

Summary

- An inner product $\langle\ ,\ \rangle$ on a vector space V is a mapping from $V \times V \to \mathbb{R}$ satisfying

  (1) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$ for all $\mathbf{u}$ and $\mathbf{v}$ in V,

(2) $\langle \mathbf{u}+\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle + \langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V,

(3) $\langle c\mathbf{u}, \mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$ for all $\mathbf{u}$, $\mathbf{v}$ in V and $c \in \mathbb{R}$,

(4) $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ for all $\mathbf{u}$ in V and $\langle \mathbf{u}, \mathbf{u}\rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

- An inner product space is a pair V, $\langle\ ,\ \rangle$ where V is a vector space and $\langle\ ,\ \rangle$ is an inner product on V.
- The length of a vector $\mathbf{v}$ in an inner product space V is defined to be the real number $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$.
- The distance between two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V is the scalar $||\mathbf{u}-\mathbf{v}||$.
- The angle θ between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the angle which satisfies $0 \leq \theta \leq \pi$ and
$$\cos(\theta) = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{||\mathbf{u}||||\mathbf{v}||}.$$
- Two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V are orthogonal if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.
- A subset S of an inner product space is an orthogonal set if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$ for all $\mathbf{u} \neq \mathbf{v}$ in S.
- A basis for a subspace of an inner product space is an orthogonal basis if the basis is also an orthogonal set.
- Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be an orthogonal basis for a subspace of an inner product space V. Let $\mathbf{x}$ be a vector in W. Then
$$\mathbf{x} = \sum_{i=1}^{m} c_i \mathbf{v}_i,$$
where
$$c_i = \frac{\langle \mathbf{x}, \mathbf{v}_i\rangle}{\langle \mathbf{v}_i, \mathbf{v}_i\rangle}$$
for each i.
- An orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ for a subspace W of an inner product space V is an orthonormal basis if $||\mathbf{v}_k|| = 1$ for each k from 1 to m.
- If $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is an orthogonal basis for V and $\mathbf{x} \in V$, then
$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} \frac{\langle \mathbf{x}, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle} \\ \frac{\langle \mathbf{x}, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle} \\ \vdots \\ \frac{\langle \mathbf{x}, \mathbf{w}_m\rangle}{\langle \mathbf{w}_m, \mathbf{w}_m\rangle} \end{bmatrix}.$$
- The projection of the vector $\mathbf{v}$ in an inner product space V onto a subspace W of V is the vector
$$\text{proj}_W \mathbf{v} = \frac{\langle \mathbf{v}, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle}\mathbf{w}_2 + \cdots + \frac{\langle \mathbf{v}, \mathbf{w}_m\rangle}{\langle \mathbf{w}_m, \mathbf{w}_m\rangle}\mathbf{w}_m,$$
where $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is an orthogonal basis of W. Projections are important in that $\text{proj}_W \mathbf{v}$ is the best approximation of the vector $\mathbf{v}$ by a vector in W in the least squares sense.

- With W as in (a), the projection of $\mathbf{v}$ orthogonal to W is the vector

$$\text{proj}_{W^\perp}\mathbf{v} = \mathbf{v} - \text{proj}_W\mathbf{v}.$$

The norm of $\text{proj}_{W^\perp}\mathbf{v}$ provides a measure of how well $\text{proj}_W\mathbf{v}$ approximates the vector $\mathbf{v}$.

- The orthogonal complement of a subset S of an inner product space V is the set

$$S^\perp = \{\mathbf{v} \in V : \langle \mathbf{v}, \mathbf{u}\rangle = 0 \text{ for all } \mathbf{u} \in S\}.$$

Exercises

(1) Let $C[a,b]$ be the set of all continuous real valued functions on the interval $[a,b]$. If f is in $C[a,b]$, we can extend f to a continuous function from $\mathbb{R}$ to $\mathbb{R}$ by letting F be the function defined by

$$F(x) = \begin{cases} f(a) & \text{if } x < a \\ f(x) & \text{if } a \leq x \leq b \\ f(b) & \text{if } b < x \end{cases}.$$

In this way we can view $C[a,b]$ as a subset of $\mathcal{F}$, the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$. Verify that $C[a,b]$ is a vector space.

(2) Use the definition of an inner product to determine which of the following defines an inner product on the indicated space. Verify your answers.

(a) $\langle \mathbf{u}, \mathbf{v}\rangle = u_1v_1 - u_2v_1 - u_1v_2 + 3u_2v_2$ for $\mathbf{u} = [u_1\ u_2]^\mathsf{T}$ and $\mathbf{v} = [v_1\ v_2]^\mathsf{T}$ in $\mathbb{R}^2$

(b) $\langle f, g\rangle = \int_a^b f(x)g(x)\, dx$ for $f, g \in C[a,b]$ (where $C[a,b]$ is the vector space of all continuous functions on the interval $[a,b]$)

(c) $\langle f, g\rangle = f'(0)g'(0)$ for $f, g \in D(-1,1)$ (where $D(a,b)$ is the vector space of all differentiable functions on the interval (a,b))

(d) $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$ for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and A an invertible $n \times n$ matrix

(3) We can sometimes visualize an inner product in $\mathbb{R}^2$ or $\mathbb{R}^3$ (or other spaces) by describing the unit circle S^1, where

$$S^1 = \{\mathbf{v} \in V : ||\mathbf{v}|| = 1\}$$

in that inner product space. For example, in the inner product space $\mathbb{R}^2$ with the dot product as inner product, the unit circle is just our standard unit circle. Inner products, however, can distort this familiar picture of the unit circle. Describe the points on the unit circle S^1 in the inner product space $\mathbb{R}^2$ with inner product $\langle [u_1\ u_2], [v_1\ v_2]\rangle = 2u_1v_1 + 3u_2v_2$ using the following steps.

(a) Let $\mathbf{x} = [x\ y] \in \mathbb{R}^2$. Set up an equation in x and y that is equivalent to the vector equation $||\mathbf{x}|| = 1$.

(b) Describe the graph of the equation you found in $\mathbb{R}^2$. It should have a familiar form. Draw a picture to illustrate. What do you think of calling this graph a "circle"?

(4) Define $\langle\ ,\ \rangle$ on $\mathbb{R}^2$ by $\langle [u_1\ u_2]^{\mathsf{T}}, [v_1\ v_2]^{\mathsf{T}} \rangle = 4u_1v_1 + 2u_2v_2$.

(a) Show that $\langle\ ,\ \rangle$ is an inner product.

(b) The inner product $\langle\ ,\ \rangle$ can be represented as a matrix transformation $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathsf{T}} A\mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are written as column vectors. Find a matrix A that represents this inner product.

(5) This exercise is a generalization of Exercise 4. Define $\langle\ ,\ \rangle$ on $\mathbb{R}^n$ by

$$\langle [u_1\ u_2\ \cdots\ u_n]^{\mathsf{T}}, [v_1\ v_2\ \cdots\ v_n]^{\mathsf{T}} \rangle = a_1u_1v_1 + a_2u_2v_2 + \cdots + a_nu_nv_n$$

for some positive scalars a_1, a_2, ..., a_n.

(a) Show that $\langle\ ,\ \rangle$ is an inner product.

(b) The inner product $\langle\ ,\ \rangle$ can be represented as a matrix transformation $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathsf{T}} A\mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are written as column vectors. Find a matrix A that represents this inner product.

(6) Is the sum of two inner products on an inner product space V an inner product on V? If yes, prove it. If no, provide a counterexample. (By the sum of inner products we mean a function $\langle\ ,\ \rangle$ satisfying

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle_1 + \langle \mathbf{u}, \mathbf{v} \rangle_2$$

for all $\mathbf{u}$ and $\mathbf{v}$ in V, where $\langle\ ,\ \rangle_1$ and $\langle\ ,\ \rangle_2$ are inner products on V.)

(7) (a) Does $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathsf{T}} A\mathbf{v}$ define an inner product on $\mathbb{R}^n$ for every $n \times n$ matrix A? Verify your answer.

(b) If your answer to part (a) is no, are there any types of matrices for which $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathsf{T}} A\mathbf{v}$ defines an inner product? (Hint: See Exercises 4 and 5..)

(8) The trace of an $n \times n$ matrix $A = [a_{ij}]$ has some useful properties.

(a) Show that $\text{trace}(A + B) = \text{trace}(A) + \text{trace}(B)$ for any $n \times n$ matrices A and B.

(b) Show that $\text{trace}(cA) = c\text{trace}(A)$ for any $n \times n$ matrix A and any scalar c.

(c) Show that $\text{trace}\left(A^{\mathsf{T}}\right) = \text{trace}(A)$ for any $n \times n$ matrix.

(9) Let V be an inner product space and $\mathbf{u}, \mathbf{v}$ be two vectors in V.

(a) Check that if $\mathbf{v} = \mathbf{0}$, the Cauchy-Schwarz inequality

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}||||\mathbf{v}||$$

holds.

(b) Assume $\mathbf{v} \neq \mathbf{0}$. Let $\lambda = \langle \mathbf{u}, \mathbf{v} \rangle / ||\mathbf{v}||^2$ and $\mathbf{w} = \mathbf{u} - \lambda \mathbf{v}$. Use the fact that $||\mathbf{w}||^2 \geq 0$ to conclude the Cauchy-Schwarz inequality in this case.

(10) The *Frobenius inner product* is defined as

$$\langle A, B \rangle = \text{trace}\left(AB^{\mathsf{T}}\right).$$

for $n \times n$ matrices A and B. Verify that $\langle A, B \rangle$ defines an inner product on $\mathcal{M}_{n \times n}$.

(11) Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $n \times n$ matrices.

(a) Show that if $n = 2$, then the Frobenius inner product (see Exercise 10) of A and B is

$$\langle A, B\rangle = a_{11}b_{11} + a_{12}b_{12} + a_{21}b_{21} + a_{22}b_{22}.$$

(b) Extend part (a) to the general case. That is, show that for an arbitrary n,

$$\langle A, B\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ij}.$$

(c) Compare the Frobenius inner product to the scalar product of two vectors.

(12) Let $\mathcal{B} = \{[1\ 1\ 1]^{\mathsf{T}}, [1\ -1\ 0]^{\mathsf{T}}\}$ and let $W = \text{Span}\ \mathcal{B}$ in $\mathbb{R}^3$.

(a) Show that $\mathcal{B}$ is an orthogonal basis for W, using the dot product as inner product.

(b) Explain why the vector $\mathbf{v} = [0\ 2\ 2]^{\mathsf{T}}$ is not in W.

(c) Find the vector in W that is closest to $\mathbf{v}$. How close is this vector to $\mathbf{v}$?

(13) Let $\mathbb{R}^3$ be the inner product space with inner product

$$\langle [u_1\ u_2\ u_3]^{\mathsf{T}}, [v_1\ v_2\ v_3]^{\mathsf{T}}\rangle = u_1v_1 + 2u_2v_2 + u_3v_3.$$

Let $\mathcal{B} = \{[1\ 1\ 1]^{\mathsf{T}}, [1\ -1\ 1]^{\mathsf{T}}\}$ and let $W = \text{Span}\ \mathcal{B}$ in $\mathbb{R}^3$.

(a) Show that $\mathcal{B}$ is an orthogonal basis for W, using the given inner product.

(b) Explain why the vector $\mathbf{v} = [0\ 2\ 2]^{\mathsf{T}}$ is not in W.

(c) Find the vector in W that is closest to $\mathbf{v}$. How close is this vector to $\mathbf{v}$?

(14) Let $\mathbb{P}_2$ be the inner product space with inner product

$$\langle p(t), q(t)\rangle = \int_0^1 p(t)q(t)\ dt.$$

Let $\mathcal{B} = \{1, 1 - 2t\}$ and let $W = \text{Span}\ \mathcal{B}$ in $\mathbb{P}_2$.

(a) Show that $\mathcal{B}$ is an orthogonal basis for W, using the given inner product.

(b) Explain why the polynomial $q(t) = t^2$ is not in W.

(c) Find the vector in W that is closest to $q(t)$. How close is this vector to $q(t)$?

(15) Prove the remaining properties of Theorem 29.2. That is, if $\langle\ ,\ \rangle$ is an inner product on a vector space V and $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in V and c is any scalar, then

(a) $\langle \mathbf{0}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{0}\rangle = 0$

(b) $\langle \mathbf{u}, c\mathbf{v}\rangle = c\langle \mathbf{u}, \mathbf{v}\rangle$

(c) $\langle \mathbf{v} + \mathbf{w}, \mathbf{u}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle + \langle \mathbf{w}, \mathbf{u}\rangle$

(d) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{u} - \mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{w}\rangle - \langle \mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{w}, \mathbf{u}\rangle - \langle \mathbf{w}, \mathbf{v}\rangle$

(16) Prove Theorem 29.16. (Hint: Refer to Theorem 27.10.)

(17) Prove Theorem 29.8. (Hint: Compare to Theorem 28.3.)

(18) Let V be a vector space with basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Define $\langle\ ,\ \rangle$ as follows:

$$\langle \mathbf{u}, \mathbf{w} \rangle = \sum_{i=1}^{n} u_i w_i$$

if $\mathbf{u} = \sum_{i=1}^{n} u_i \mathbf{v}_i$ and $\mathbf{w} = \sum_{i=1}^{n} w_i \mathbf{v}_i$ in V. (Since the representation of a vector as a linear combination of basis elements is unique, this mapping is well-defined.) Show that $\langle\ ,\ \rangle$ is an inner product on V and conclude that any finite dimensional vector space can be made into an inner product space.

(19) Label each of the following statements as True or False. Provide justification for your response.

(a) **True/False** The only inner product on $\mathbb{R}^n$ is the dot product.

(b) **True/False** If W is a subspace of an inner product space and a vector $\mathbf{v}$ is orthogonal to every vector in a basis of W, then $\mathbf{v}$ is in $W^\perp$.

(c) **True/False** If $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $\mathbb{R}^3$, then so is $\{c\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for any nonzero scalar c.

(d) **True/False** An inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ in an inner product space V results in another vector in V.

(e) **True/False** An inner product in an inner product space V is a function that maps pairs of vectors in V to the set of non-negative real numbers.

(f) **True/False** The vector space of all $n \times n$ matrices can be made into an inner product space.

(g) **True/False** Any non-zero multiple of an inner product on space V is also an inner product on V.

(h) **True/False** Every set of k non-zero orthogonal vectors in a vector space V of dimension k is a basis for V.

(i) **True/False** For any finite-dimensional inner product space V and a subspace W of V, W is a subspace of $(W^\perp)^\perp$.

(j) **True/False** If W is a subspace of an inner product space, then $W \cap W^\perp = \{\mathbf{0}\}$.

Project: Fourier Series and Musical Tones

Joseph Fourier first studied trigonometric polynomials to understand the flow of heat in metallic plates and rods. The resulting series, called Fourier series, now have applications in a variety of areas including electrical engineering, vibration analysis, acoustics, optics, signal processing, image processing, geology, quantum mechanics, and many more. For our purposes, we will focus on synthesized music.

Pure musical tones are periodic sine waves. Simple electronic circuits can be designed to generate alternating current. Alternating current is current that is periodic, and hence is described by a combination of $\sin(kx)$ and $\cos(kx)$ for integer values of k. To synthesize an instrument like a violin, we can project the instrument's tones onto trigonometric polynomials – and then we can produce them electronically. As we will see, these projections are least squares approximations onto certain vector spaces. The website http://www.falstad.com/fourier/ provides a tool for hearing sounds digitally created by certain functions. For example, you can listen to the sound generated by a sawtooth function f of the form

$$f(x) = \begin{cases} x & \text{if } -\pi < x \leq \pi, \\ f(x-2\pi), & \text{if } \pi < x, \\ f(x+2\pi), & \text{if } x \leq -\pi. \end{cases}$$

Try out some of the tones on this website (click on the Sound button to hear the tones). You can also alter the tones by clicking on any one of the white dots and moving it up or down. and play with the buttons. We will learn much about what this website does in this project.

Pure tones are periodic and so are modeled by trigonometric functions. In general, trigonometric polynomials can be used to produce good approximations to periodic phenomena. A *trigonometric polynomial* is an object of the form

$$c_0 + c_1\cos(x) + d_1\sin(x) + c_2\cos(2x) + d_2\sin(2x) + \cdots \\ + c_n\cos(nx) + d_n\sin(nx) + \cdots ,$$

where the c_i and d_j are real constants. With judicious choices of these constants, we can approximate periodic and other behavior with trigonometric polynomials. The first step for us will be to understand the relationships between the summands of these trigonometric polynomials in the inner product space $C[-\pi,\pi]$[1] of continuous functions from $[\pi,\pi]$ to $\mathbb{R}$ with the inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx. \tag{29.3}$$

Our first order of business is to verify that (29.3) is, in fact, an inner product.

Project Activity 29.1. Let $C[a,b]$ be the set of continuous real-valued functions on the interval $[a,b]$. In Exercise 1. in Section 29 we are asked to show that $C[a,b]$ is a vector space, while Exercise 2 in Section 29 asks us to show that $\langle f,g\rangle = \int_a^b f(x)g(x)\,dx$ defines an inner product on $C[a,b]$. However, 29.3 is slightly different than this inner product. Show that any positive scalar multiple of an inner product is an inner product, and conclude that (29.3) defines an inner product on $C[-\pi,\pi]$. (We will see why we introduce the factor of $\frac{1}{\pi}$ later.)

Now we return to our inner product space $C[-\pi,\pi]$ with inner product (29.3). Given a function g in $C[-\pi,\pi]$, we approximate g using only a finite number of the terms in a trigonometric polynomial. Let W_n be the subspace of $C[-\pi,\pi]$ spanned by the functions

$$1, \cos(x), \sin(x), \cos(2x), \sin(2x), \ldots, \cos(nx), \sin(nx).$$

One thing we need to know is the dimension of W_n.

[1]With suitable adjustments, we can work over any interval that is convenient, but for the sake of simplicity in this project, we will restrict ourselves to the interval $[-\pi,\pi]$.

Project Activity 29.2. We start with the initial case of W_1.

(a) Show directly that the functions 1, $\cos(x)$, and $\sin(x)$ are orthogonal.

(b) What is the dimension of W_1? Explain.

Now we need to see if what happened in Project Activity 29.2 happens in general. A few tables of integrals and some basic facts from trigonometry can help.

Project Activity 29.3. A table of integrals shows the following for $k \neq m$ (up to a constant):

$$\int \cos(mx)\cos(kx)\, dx = \frac{1}{2}\left(\frac{\sin((k-m)x)}{k-m} + \frac{\sin((k+m)x)}{k+m}\right) \tag{29.4}$$

$$\int \sin(mx)\sin(kx)\, dx = \frac{1}{2}\left(\frac{\sin((k-m)x)}{k-m} - \frac{\sin((k+m)x)}{k+m}\right) \tag{29.5}$$

$$\int \cos(mx)\sin(kx)\, dx = \frac{1}{2}\left(\frac{\cos((m-k)x)}{m-k} - \frac{\sin((m+k)x)}{m+k}\right) \tag{29.6}$$

$$\int \cos(mx)\sin(mx)\, dx = -\frac{1}{2m}\cos^2(mx) \tag{29.7}$$

(a) Use (29.4) to show that $\cos(mx)$ and $\cos(kx)$ are orthogonal in $C[-\pi,\pi]$ if $k \neq m$.

(b) Use (29.5) to show that $\sin(mx)$ and $\sin(kx)$ are orthogonal in $C[-\pi,\pi]$ if $k \neq m$.

(c) Use (29.6) to show that $\cos(mx)$ and $\sin(kx)$ are orthogonal in $C[-\pi,\pi]$ if $k \neq m$.

(d) Use (29.7) to show that $\cos(mx)$ and $\sin(mx)$ are orthogonal in $C[-\pi,\pi]$.

(e) What is $\dim(W_n)$? Explain.

Once we have an orthogonal basis for W_n, we might want to create an orthonormal basis for W_n. Throughout the remainder of this project, unless otherwise specified, you should use a table of integrals or any appropriate technological tool to find integrals for any functions you need.

Project Activity 29.4. Show that the set

$$\mathcal{B}_n = \left\{\frac{1}{\sqrt{2}}, \cos(x), \cos(2x), \ldots, \cos(nx), \sin(x), \sin(2x), \ldots, \sin(nx)\right\}$$

is an orthonormal basis for W_n. Use the fact that the norm of a vector $\mathbf{v}$ in an inner product space with inner product $\langle\ ,\ \rangle$ is defined to be $\sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$. (This is where the factor of $\frac{1}{\pi}$ will be helpful.)

Now we need to recall how to find the best approximation to a vector by a vector in a subspace, and apply that idea to approximate an arbitrary function g with a trigonometric polynomial in W_n. Recall that the best approximation of a function g in $C[-\pi,\pi]$ is the projection of g onto W_n. If we have an orthonormal basis $\{h_0, h_1, h_2, \ldots, h_{2n}\}$ of W_n, then the projection of g onto W_n is

$$\text{proj}_{W_n} g = \langle g, h_0\rangle h_0 + \langle g, h_1\rangle h_1 + \langle g, h_2\rangle h_2 + \ldots + \langle g, h_{2n}\rangle h_{2n}.$$

With this idea, we can find formulas for the coefficients when we project an arbitrary function onto W_n.

Project Activity 29.5. If g is an arbitrary function in $C[-\pi, \pi]$, we will write the projection of g onto W_n as

$$a_0\left(\frac{1}{\sqrt{2}}\right) + a_1\cos(x) + b_1\sin(x) + a_2\cos(2x) + b_2\sin(2x) + \cdots$$
$$+ a_n\cos(nx) + b_n\sin(nx).$$

The a_i and b_j are the *Fourier coefficients* for f. The expression $a_n\cos(nx) + b_n\sin(nx)$ is called the nth *harmonic* of g. The first harmonic is called the *fundamental frequency*. The human ear cannot hear tones whose frequencies exceed 20000 Hz, so we only hear finitely many harmonics (the projections onto W_n for some n).

(a) Show that
$$a_0 = \frac{1}{\sqrt{2}\pi}\int_{-\pi}^{\pi} g(x)\, dx. \tag{29.8}$$
Explain why $\frac{a_0}{\sqrt{2}}$ gives the average value of g on $[-\pi, \pi]$. You may want to go back and review average value from calculus. This is saying that the best constant approximation of g on $[-\pi, \pi]$ is its average value, which makes sense.

(b) Show that for $m \geq 1$,
$$a_m = \frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\cos(mx)\, dx. \tag{29.9}$$

(c) Show that for $m \geq 1$,
$$b_m = \frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\sin(mx)\, dx. \tag{29.10}$$

Let us return to the sawtooth function defined earlier and find its Fourier coefficients.

Project Activity 29.6. Let f be defined by $f(x) = x$ on $[-\pi, \pi]$ and repeated periodically afterwards with period 2π. Let p_n be the projection of f onto W_n.

(a) Evaluate the integrals to find the projection p_1.

(b) Use appropriate technology to find the projections p_{10}, p_{20}, and p_{30} for the sawtooth function f. Draw pictures of these approximations against f and explain what you see.

(c) Now we find formulas for all the Fourier coefficients. Use the fact that $x\cos(mx)$ is an odd function to explain why $a_m = 0$ for each m. Then show that $b_m = (-1)^{m+1}\frac{2}{m}$ for each m.

(d) Go back to the website `http://www.falstad.com/fourier/` and replay the sawtooth tone. Explain what the white buttons represent.

Project Activity 29.7. This activity is not connected to the idea of musical tones, so can be safely ignored if so desired. We conclude with a derivation of a very fascinating formula that you may have seen for $\displaystyle\sum_{n=1}^{\infty}\frac{1}{n^2}$. To do so, we need to analyze the error in approximating a function g with a function in W_n.

Let p_n be the projection of g onto W_n. Notice that p_n is also in W_{n+1}. It is beyond the scope of this project, but in "nice" situations we have $||g - p_n|| \to 0$ as $n \to \infty$. Now $g - p_n$ is orthogonal to p_n, so the Pythagorean theorem shows that

$$||g - p_n||^2 + ||p_n||^2 = ||g||^2.$$

Since $||g - p_n||^2 \to 0$ as $n \to \infty$, we can conclude that

$$\lim_{n\to\infty} ||p_n||^2 = ||g||^2. \tag{29.11}$$

We use these ideas to derive a formula for $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^2}$.

(a) Use the fact that $\mathcal{B}_n$ is an orthonormal basis to show that

$$||p_n||^2 = a_0^2 + a_1^2 + b_1^2 + \cdots + a_n^2 + b_n^2.$$

Conclude that

$$||g||^2 = a_0^2 + a_1^2 + b_1^2 + \cdots + a_n^2 + b_n^2 + \cdots . \tag{29.12}$$

(b) For the remainder of this activity, let f be the sawtooth function defined by $f(x) = x$ on $[-\pi, \pi]$ and repeated periodically afterwards. We determined the Fourier coefficients a_i and b_j of this function in Project Activity 29.6.

i. Show that

$$a_0^2 + a_1^2 + b_1^2 + \cdots + a_n^2 + b_n^2 + \cdots = 4\sum_{n=1}^{\infty} \frac{1}{n^2}.$$

ii. Calculate $||f||^2$ using the inner product and compare to (29.12) to find a surprising formula for $\sum_{n=1}^{\infty} \frac{1}{n^2}$.

Section 30

The Gram-Schmidt Process

Focus Questions

By the end of this section, you should be able to give precise and thorough answers to the questions listed below. You may want to keep these questions in mind to focus your thoughts as you complete the section.

- What is the Gram-Schmidt process and why is it useful?
- What is the QR-factorization of a matrix and why is it useful?

Application: Gaussian Quadrature

Since integration of functions is difficult, approximation techniques for definite integrals are very important. In calculus we are introduced to various methods, e.g., the midpoint, trapezoid, and Simpson's rule, for approximating definite integrals. These methods divide the interval of integration into subintervals and then use values of the integrand at the endpoints to approximate the integral. These are useful methods when approximating integrals from tabulated data, but there are better methods for other types of integrands. If we make judicious choices in the points we use to evaluate the integrand, we can obtain more accurate results with less work. One such method is Gaussian quadrature (which, for example, is widely used in solving problems of radiation heat transfer in direct integration of the equation of transfer of radiation over space), which we explore later in this section. This method utilizes the Gram-Schmidt process to produce orthogonal polynomials.

Introduction

We have seen that orthogonal bases make computations very convenient. So one question we might want to address is how we can create an orthogonal basis from any basis. We have already done this in the case that $W = \text{Span}\{\mathbf{w}_1\}$ is the span of a single vector in $\mathbb{R}^n$ – we can project the vector $\mathbf{v}$ in the direction of $\mathbf{w}_1$. Our goal is to generalize this to project a vector $\mathbf{v}$ onto an entire subspace in a useful way, and then use that projection to create a vector orthogonal to the subspace.

The Gram-Schmidt Process

To make best approximations by orthogonal projections onto subspaces we need to be able to find orthogonal bases for vector spaces. We now address the question of how to do that.

Preview Activity 30.1. Let $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in $\mathbb{R}^4$ (using the dot product), where $\mathbf{v}_1 = [1\ 1\ 1\ 1]^\mathsf{T}$, $\mathbf{v}_2 = [-1\ 4\ 4\ -1]^\mathsf{T}$, and $\mathbf{v}_3 = [2\ -1\ 1\ 0]^\mathsf{T}\}$. Our goal in this preview activity is to begin to understand how we can find an orthogonal set $\mathcal{B} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ in $\mathbb{R}^4$ so that Span $\mathcal{B} = W$. To begin, we could start by letting $\mathbf{w}_1 = \mathbf{v}_1$.

(1) Now we want to find a vector in W that is orthogonal to $\mathbf{w}_1$. Let $W_1 = \text{Span}\{\mathbf{w}_1\}$. Explain why $\mathbf{w}_2 = \text{proj}_{\perp W_1}\mathbf{v}_2$ is in W and is orthogonal to $\mathbf{w}_1$. Then calculate the vector $\mathbf{w}_2$.

(2) Next we need to find a third vector $\mathbf{w}_3$ that is in W and is orthogonal to both $\mathbf{w}_1$ and $\mathbf{w}_2$. Let $W_2 = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2\}$. Explain why $\mathbf{w}_3 = \text{proj}_{\perp W_2}\mathbf{v}_3$ is in W and is orthogonal to both $\mathbf{w}_1$ and $\mathbf{w}_2$. Then calculate the vector $\mathbf{w}_3$.

(3) Explain why the set $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is an orthogonal basis for W.

Preview Activity 30.1 shows the first steps of the Gram-Schmidt process to construct an orthogonal basis from any basis of a subspace in $\mathbb{R}^n$. Of course, there is nothing special about using the dot product in Preview Activity 30.1 – the same argument works in any inner product space. To understand why the process works in general, let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a basis for a subspace W of an inner product space V. Let $\mathbf{w}_1 = \mathbf{v}_1$ and let $W_1 = \text{Span}\{\mathbf{w}_1\}$. Since $\mathbf{w}_1 = \mathbf{v}_1$ we have that $W_1 = \text{Span}\{\mathbf{w}_1\} = \text{Span}\{\mathbf{v}_1\}$. Now consider the subspace

$$W_2 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$$

of W. The vectors $\mathbf{v}_1 = \mathbf{w}_1$ and $\mathbf{v}_2$ are possibly not orthogonal, but we know the orthogonal projection of $\mathbf{v}_2$ onto $W_1^\perp$ is orthogonal to $\mathbf{w}_1$. Let

$$\mathbf{w}_2 = \text{proj}_{\perp W_1}\mathbf{v}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\mathbf{w}_1.$$

Then $\{\mathbf{w}_1, \mathbf{w}_2\}$ is an orthogonal set. Note that $\mathbf{w}_1 = \mathbf{v}_1 \neq \mathbf{0}$, and the fact that $\mathbf{v}_2 \notin W_1$ implies that $\mathbf{w}_2 \neq \mathbf{0}$. So the set $\{\mathbf{w}_1, \mathbf{w}_2\}$ is linearly independent, being a set of non-zero orthogonal vectors. Now the question is whether $\text{Span}\{\mathbf{w}_1, \mathbf{w}_2\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Note that $\mathbf{w}_2$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$, so $\mathbf{w}_2$ is in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Since $\text{Span}\{\mathbf{w}_1, \mathbf{w}_2\}$ is a 2-dimensional subspace of the 2-dimensional space W_2, it must be true that $\text{Span}\{\mathbf{w}_1, \mathbf{w}_2\} = W_2 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$.

Now we take the next step, adding $\mathbf{v}_3$ into the mix. Let

$$W_3 = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{v}_3\}.$$

The vector

$$\mathbf{w}_3 = \text{proj}_{\perp W_2}\mathbf{v}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle}\mathbf{w}_2$$

is orthogonal to both $\mathbf{w}_1$ and $\mathbf{w}_2$ and, by construction, $\mathbf{w}_3$ is a linear combination of $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. So $\mathbf{w}_3$ is in W_3. The fact that $\mathbf{v}_3 \notin W_2$ implies that $\mathbf{w}_3 \neq \mathbf{0}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is a linearly independent set. Since Span$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is a 3-dimensional subspace of the 3-dimensional space W_3, we conclude that Span$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ equals Span$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

We continue inductively in this same manner. If we have constructed a set $\{\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, ..., $\mathbf{w}_{k-1}\}$ of $k-1$ orthogonal vectors such that

$$\text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_{k-1}\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{k-1}\},$$

then we let

$$\begin{aligned} \mathbf{w}_k &= \text{proj}_{\perp W_{k-1}} \mathbf{v}_k \\ &= \mathbf{v}_k - \frac{\langle \mathbf{v}_k, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_k, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_k, \mathbf{w}_{k-1} \rangle}{\langle \mathbf{w}_{k-1}, \mathbf{w}_{k-1} \rangle}\mathbf{w}_{k-1}, \end{aligned}$$

where

$$W_{k-1} = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_{k-1}\}.$$

We know that $\mathbf{w}_k$ is orthogonal to $\mathbf{w}_1$, $\mathbf{w}_2$, ..., $\mathbf{w}_{k-1}$. Since $\mathbf{w}_1$, $\mathbf{w}_2$, ..., $\mathbf{w}_{k-1}$, and $\mathbf{v}_k$ are all in W_k = Span$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ we see that $\mathbf{w}_k$ is also in W_k. Since $\mathbf{v}_k \notin W_{k-1}$ implies that $\mathbf{w}_k \neq \mathbf{0}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ is a linearly independent set. Then Span$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_k\}$ is a k-dimensional subspace of the k-dimensional space W_k, so it follows that

$$\text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_k\} = W_k = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_k\}.$$

This process will end when we have an orthogonal set $\{\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, ..., $\mathbf{w}_m\}$ with Span$\{\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, ..., $\mathbf{w}_m\}$ = W.

We summarize the process in the following theorem.

Theorem 30.1 (The Gram-Schmidt Process)**.** *Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a basis for a subspace W of an inner product space V. The set $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_m\}$ defined by*

- $\mathbf{w}_1 = \mathbf{v}_1$,
- $\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1$,
- $\mathbf{w}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2$,

  $\vdots$
- $\mathbf{w}_m = \mathbf{v}_m - \frac{\langle \mathbf{v}_m, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_m, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_m, \mathbf{w}_{m-1} \rangle}{\langle \mathbf{w}_{m-1}, \mathbf{w}_{m-1} \rangle}\mathbf{w}_{m-1}$.

is an orthogonal basis for W. Moreover,

$$\textit{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_k\} = \textit{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_k\}$$

for $1 \leq k \leq m$.

The Gram-Schmidt process builds an orthogonal basis $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_m\}$ for us from a given basis. To make an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots, \mathbf{u}_m\}$, all we need do is normalize each basis vector: that is, for each i, we let

$$\mathbf{u}_i = \frac{\mathbf{w}_i}{||\mathbf{w}_i||}.$$

Activity 30.1. Use the Gram-Schmidt process to find an orthogonal basis for W.

(a) $W = \text{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ using the dot product as the inner product

(b) $W = \text{Span}\{1, t, t^2\}$ in $\mathbb{P}_2$ using the inner product

$$\langle p(t), q(t) \rangle = \int_0^1 p(t)q(t)\, dt$$

The QR Factorization of a Matrix

There are several different factorizations, or decompositions, of matrices where each matrix is written as a product of certain types of matrices: LU decomposition using lower and upper triangular matrices (see Section 21), EVD (EigenVector Decomposition) decomposition using eigenvectors and diagonal matrices (see Section 18, and in this section we will introduce the QR decomposition using orthogonal matrix and upper triangular matrices. The QR factorization has applications to solving least squares problems and approximating eigenvalues of matrices.

Activity 30.2. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 2 \end{bmatrix}$.

(a) Find an orthonormal basis for Col A. Let $\mathcal{B}$ be the basis formed by these vectors. Let Q be the matrix whose columns are these orthonormal basis vectors.

(b) Write the columns of A as linear combinations of the columns of Q. That is, if $A = [\mathbf{a}_1\ \mathbf{a}_2]$, find $[\mathbf{a}_1]_{\mathcal{B}}$ and $[\mathbf{a}_2]_{\mathcal{B}}$. Let $R = [[\mathbf{a}_1]_{\mathcal{B}}\ [\mathbf{a}_2]_{\mathcal{B}}]$.

(c) Find the product QR and compare to Q.

Activity 30.2 contains the main ideas to find the QR factorization of a matrix. Let

$$A = [\mathbf{a}_1 \mid \mathbf{a}_2 \mid \mathbf{a}_3 \mid \cdots \mid \mathbf{a}_n]$$

be an $m \times n$ matrix with rank[1] n. We can use the Gram-Schmidt process to find an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ for Col A. Recall also that $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\}$ for any k between 1 and n. Let

$$Q = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \cdots\ \mathbf{u}_n].$$

If k is between 1 and n, then $\mathbf{a}_k$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ and

$$a_k = r_{k1}\mathbf{u}_1 + r_{k2}\mathbf{u}_2 + \cdots + r_{kk}\mathbf{u}_k$$

[1] Recall that the rank of a matrix A is the dimension of the column space of A.

for some scalars r_{k1}, r_{k2}, ..., r_{kk}. Then

$$Q \begin{bmatrix} r_{k1} \\ r_{k2} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{a}_k.$$

If we let $\mathbf{r}_k = \begin{bmatrix} r_{k1} \\ r_{k2} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ for k from 1 to n, then

$$A = [Q\mathbf{r}_1 \; Q\mathbf{r}_2 \; Q\mathbf{r}_3 \; \cdots \; Q\mathbf{r}_k].$$

This is the QR factorization of A into the product

$$A = QR$$

where the columns of Q form an orthonormal basis for Col A and

$$R = [\mathbf{r}_1 \; \mathbf{r}_2 \; \mathbf{r}_3 \; \cdots \; \mathbf{r}_n] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n-1} & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n-1} & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n-1} & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & r_{nn} \end{bmatrix}$$

is an upper triangular matrix. Note that Q is an $m \times n$ matrix and R is an $n \times n$ matrix.

Activity 30.3. Let $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix}$. Find the QR factorization of the following matrices or explain why the matrix does not have a QR factorization.

The QR factorization provides a widely used algorithm (the QR algorithm) for approximating all of the eigenvalues of a matrix. The computer system MATLAB utilizes four versions of the QR algorithm to approximate the eigenvalues of real symmetric matrices, eigenvalues of real nonsymmetric matrices, eigenvalues of pairs of complex matrices, and singular values of general matrices.

The algorithm works as follows.

- Start with an $n \times n$ matrix A. Let $A_1 = A$.

- Find the QR factorization for A_1 and write it as $A_1 = Q_1R_1$, where Q_1 is orthogonal and R_1 is upper triangular.
- Let $A_2 = Q_1^{-1}A_1Q_1 = Q_1^{\mathsf{T}}AQ_1 = R_1Q_1$. Find the QR factorization of A_2 and write it as $A_2 = Q_2R_2$.
- Let $A_3 = Q_2^{-1}A_2Q_2 = Q_2^{\mathsf{T}}AQ_2 = R_2Q_2$. Find the QR factorization of A_3 and write it as $A_3 = Q_3R_3$.
- Continue in this manner to obtain a sequence $\{A_k\}$ where $A_k = Q_kR_k$ and $A_{k+1} = R_kQ_k$.

Note that $A_{k+1} = Q_k^{-1}A_kQ_k$ and so all of the matrices A_k are similar to each other and therefore all have the same eigenvalues. We won't go into the details, but it can be shown that if the eigenvalues of A are real and have distinct absolute values, then the sequence $\{A_i\}$ converges to an upper triangular matrix with the eigenvalues of A as the diagonal entries. If some of the eigenvalues of A are complex, then the sequence $\{A_i\}$ converges to a block upper triangular matrix, where the diagonal blocks are either 1×1 (approximating the real eigenvalues of A) or 2×2 (which provide a pair of conjugate complex eigenvalues of A).

Examples

What follows are worked examples that use the concepts from this section.

Example 30.2. Let W be the subspace of $\mathbb{R}^4$ spanned by $\mathbf{w}_1 = [1\ 0\ 0\ 0]^{\mathsf{T}}$, $\mathbf{w}_2 = [1\ 1\ 1\ 0]^{\mathsf{T}}$, and $\mathbf{w}_3 = [1\ 2\ 0\ 1]^{\mathsf{T}}$.

(a) Use the Gram-Schmidt process to find an orthonormal basis for the subspace W of $\mathbb{R}^4$ spanned by $\mathbf{w}_1 = [1\ 0\ 0\ 0]^{\mathsf{T}}$, $\mathbf{w}_2 = [1\ 1\ 1\ 0]^{\mathsf{T}}$, and $\mathbf{w}_3 = [1\ 2\ 0\ 1]^{\mathsf{T}}$.

(b) Find a QR factorization of the matrix $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$.

Example Solution.

(a) First note that $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ are linearly independent. We let $\mathbf{v}_1 = \mathbf{w}_1$ and the Gram-Schmidt process gives us

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{w}_2 - \frac{\mathbf{w}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\ &= [1\ 1\ 1\ 0]^{\mathsf{T}} - \frac{1}{1}[1\ 0\ 0\ 0]^{\mathsf{T}} \\ &= [0\ 1\ 1\ 0]^{\mathsf{T}} \end{aligned}$$

and

$$\begin{aligned} \mathbf{v}_3 &= \mathbf{w}_3 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{w}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\ &= [1\ 2\ 0\ 1]^{\mathsf{T}} - \frac{1}{1}[1\ 0\ 0\ 0]^{\mathsf{T}} - \frac{2}{2}[0\ 1\ 1\ 0]^{\mathsf{T}} \\ &= [0\ 1\ -1\ 1]^{\mathsf{T}}. \end{aligned}$$

So $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for W. An orthonormal basis is found by dividing each vector by its magnitude, so

$$\{[1\ 0\ 0\ 0]^{\mathsf{T}}, \frac{1}{\sqrt{2}}[0\ 1\ 1\ 0]^{\mathsf{T}}, \frac{1}{\sqrt{3}}[0\ 1\ -1\ 1]^{\mathsf{T}}\}$$

is an orthonormal basis for W.

(b) Technology shows that the reduced row echelon form of A is I_4, so the columns of A are linearly independent and A has rank 4. From part (a) we have an orthogonal basis for the span of the first three columns of A. To find a fourth vector to add so that the span is Col A, we apply the Gram-Schmidt process one more time with $\mathbf{w}_4 = [0\ 0\ 0\ 1]^{\mathsf{T}}$:

$$\begin{aligned}\mathbf{v}_4 &= \mathbf{w}_4 - \frac{\mathbf{w}_4 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{w}_4 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \frac{\mathbf{w}_4 \cdot \mathbf{v}_3}{\mathbf{v}_3 \cdot \mathbf{v}_3}\mathbf{v}_3 \\ &= [0\ 0\ 0\ 1]^{\mathsf{T}} - \frac{0}{1}[1\ 0\ 0\ 0]^{\mathsf{T}} - \frac{0}{2}[0\ 1\ 1\ 0]^{\mathsf{T}} - \frac{1}{3}[0\ 1\ -1\ 1]^{\mathsf{T}} \\ &= \frac{1}{3}[0\ -1\ 1\ 2]^{\mathsf{T}}.\end{aligned}$$

So $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ is an orthonormal basis for Col A, where

$$\begin{array}{rr}\mathbf{u}_1 = [1\ 0\ 0\ 0]^{\mathsf{T}} & \mathbf{u}_2 = \frac{\sqrt{2}}{2}[0\ 1\ 1\ 0]^{\mathsf{T}} \\ \mathbf{u}_3 = \frac{\sqrt{3}}{3}[0\ 1\ -1\ 1]^{\mathsf{T}} & \mathbf{u}_4 = \frac{\sqrt{6}}{6}[0\ -1\ 1\ 2]^{\mathsf{T}}.\end{array}$$

This makes

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

To find the matrix R, we write the columns of A in terms of the basis vectors $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, and $\mathbf{u}_4$. Technology shows that the reduced row echelon form of $[Q \mid A]$ is

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \sqrt{3} & \frac{\sqrt{3}}{3} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \frac{\sqrt{6}}{3} \end{array}\right].$$

So

$$R = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & \sqrt{2} & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} & \frac{\sqrt{3}}{3} \\ 0 & 0 & 0 & \frac{\sqrt{6}}{3} \end{bmatrix}.$$

Example 30.3. Consider the space $V = C[0, 1]$ with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\, dx$.

(a) Find the polynomial in $\mathbb{P}_2$ (considered as a subspace of V) that is closest to $h(x) = \sqrt{x}$. Use technology to calculate any required integrals. Draw a graph of your approximation against the graph of h.

(b) Provide a numeric measure of the error in approximating $\sqrt{x}$ by the polynomial you found in part (a).

Example Solution.

(a) Our job is to find $\text{proj}_{\mathbb{P}_2} h(x)$. To do this, we need an orthogonal basis of $\mathbb{P}_2$. We apply the Gram-Schmidt process to the standard basis $\{1, t, t^2\}$ of $\mathbb{P}_2$ to obtain an orthogonal basis $\{p_1(t), p_2(t), p_3(t)\}$ of $\mathbb{P}_2$. We start with $p_1(t) = 1$, then

$$\begin{aligned} p_2(t) &= t - \frac{\langle t, 1\rangle}{\langle 1, 1\rangle}(1) \\ &= t - \frac{1}{2} \end{aligned}$$

and

$$\begin{aligned} p_3(t) &= t^2 - \frac{\langle t^2, 1\rangle}{\langle 1, 1\rangle}(1) - \frac{\left\langle t^2, t - \frac{1}{2}\right\rangle}{\left\langle t - \frac{1}{2}, t - \frac{1}{2}\right\rangle}\left(t - \frac{1}{2}\right) \\ &= t^2 - \frac{1}{3} - \frac{\frac{1}{12}}{\frac{1}{12}}\left(t - \frac{1}{2}\right) \\ &= t^2 - t + \frac{1}{6}. \end{aligned}$$

Then

$$\begin{aligned} \text{proj}_{\mathbb{P}_2} \sqrt{x} &= \frac{\langle \sqrt{x}, 1\rangle}{\langle 1, 1\rangle}(1) + \frac{\left\langle \sqrt{x}, x - \frac{1}{2}\right\rangle}{\left\langle x - \frac{1}{2}, x - \frac{1}{2}\right\rangle}\left(x - \frac{1}{2}\right) \\ &\quad + \frac{\left\langle \sqrt{x}, x^2 - x + \frac{1}{12}\right\rangle}{\left\langle x^2 - x + \frac{1}{6}, x^2 - x + \frac{1}{6}\right\rangle}\left(x^2 - x + \frac{1}{6}\right) \\ &= \frac{\frac{2}{3}}{1}(1) + \frac{\frac{1}{15}}{\frac{1}{12}}\left(x - \frac{1}{2}\right) - \frac{\frac{1}{315}}{\frac{1}{180}}\left(x^2 - x + \frac{1}{6}\right) \\ &= \frac{2}{3} + \left(\frac{4}{5}\right)\left(x - \frac{1}{2}\right) - \frac{4}{7}\left(x^2 - x + \frac{1}{6}\right) \\ &= -\frac{4}{7}x^2 + \frac{48}{35}x + \frac{6}{35}. \end{aligned}$$

A graph of the approximation and h are shown in Figure 30.1

(b) The norm of $\text{proj}_{\mathbb{P}_2^\perp} \sqrt{x} = \sqrt{x} - \left(-\frac{4}{7}x^2 + \frac{48}{35}x + \frac{6}{35}\right)$ tells us how well our projection $-\frac{4}{7}x^2 + \frac{48}{35}x + \frac{6}{35}$ approximates $\sqrt{x}$. Technology shows that

$$||\text{proj}_{\mathbb{P}_2^\perp} \sqrt{x}|| = \sqrt{\langle \text{proj}_{\mathbb{P}_2^\perp} \sqrt{x}, \text{proj}_{\mathbb{P}_2^\perp} \sqrt{x}\rangle} = \frac{\sqrt{2}}{70} \approx 0.02.$$

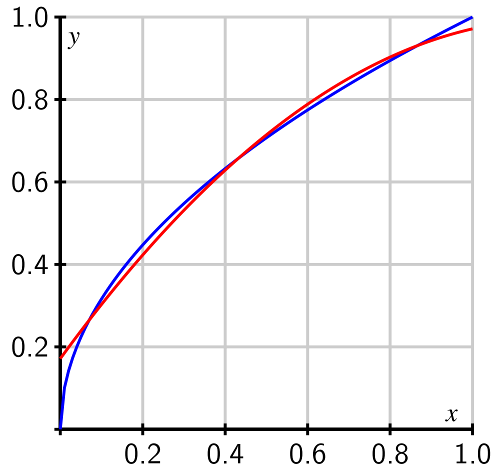

Figure 30.1: The graphs of $\sqrt{x}$ and $-\frac{4}{7}x^2 + \frac{48}{35}x + \frac{6}{35}$.

Summary

- The Gram-Schmidt process produces an orthogonal basis from any basis. It works as follows. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ be a basis for a subspace W of an inner product space V. The set $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_m\}$ defined by
  - $\mathbf{w}_1 = \mathbf{v}_1$,
  - $\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\mathbf{w}_1$,
  - $\mathbf{w}_3 = \mathbf{v}_3 - \frac{\langle \mathbf{v}_3, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_3, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle}\mathbf{w}_2$,

    $\vdots$
  - $\mathbf{w}_m = \mathbf{v}_m - \frac{\langle \mathbf{v}_m, \mathbf{w}_1\rangle}{\langle \mathbf{w}_1, \mathbf{w}_1\rangle}\mathbf{w}_1 - \frac{\langle \mathbf{v}_m, \mathbf{w}_2\rangle}{\langle \mathbf{w}_2, \mathbf{w}_2\rangle}\mathbf{w}_2 - \cdots - \frac{\langle \mathbf{v}_m, \mathbf{w}_{m-1}\rangle}{\langle \mathbf{w}_{m-1}, \mathbf{w}_{m-1}\rangle}\mathbf{w}_{m-1}$.

  is an orthogonal basis for W. Moreover,

  $$\text{Span}\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_k\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_k\}$$

  for each k between 1 and m.

- The QR factorization has applications to solving least squares problems and approximating eigenvalues of matrices. The QR factorization writes an $m \times n$ matrix with rank n as a product $A = QR$, where the columns of Q form an orthonormal basis for Col A and

  $$R = [\mathbf{r}_1 \mid \mathbf{r}_2 \mid \mathbf{r}_3 \mid \cdots \mid \mathbf{r}_n] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n-1} & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n-1} & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n-1} & r_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & r_{nn} \end{bmatrix}$$

  is an upper triangular matrix.

Exercises

(1) Let $\mathbf{w}_1 = [1\ 2\ 1]^\mathsf{T}$ and $\mathbf{w}_2 = [1\ -1\ 1]^\mathsf{T}\}$, and Let $W = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2\}$ in $\mathbb{R}^3$ with the dot product as inner product. Let $\mathbf{v} = [1\ 0\ 0]^\mathsf{T}$.

(a) Find $\text{proj}_W \mathbf{v}$ and $\text{proj}_{W^\perp} \mathbf{v}$

(b) Find the vector in W that is closest to $\mathbf{v}$. How close is this vector to $\mathbf{v}$?

(2) Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \right\}$ and let $W = \text{Span}(\mathcal{B})$ in $\mathbb{R}^4$. Find the best approximation in W to the vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \\ 1 \\ -1 \end{bmatrix}$ in W and give a numeric estimate of how good this approximation is.

(3) Let $V = C[0,1]$ with the inner product $\langle f(t), g(t) \rangle = \int_0^1 f(t)g(t)\,dt$. Let $W = \mathbb{P}_2$. Note that W is a subspace of V.

(a) Find the polynomial $q(t)$ in W that is closest to the function h defined by $h(t) = \frac{2}{1+t^2}$ in the least squares sense. That is, find the projection of $h(t)$ onto W. (Hint: Recall the work done in Activity 30.1.)

(b) Find the second order Taylor polynomial $P_2(t)$ for $h(t)$ centered at 0.

(c) Plot $h(t)$, $q(t)$, and $P_2(t)$ on the same set of axes. Which approximation provides a better fit to h on this interval. Why?

(4) In this exercise we determine the least-squares line (the line of best fit in the least squares sense) to a set of k data points (x_1, y_1), (x_2, y_2), ..., (x_k, y_k) in the plane. In this case, we want to fit a line of the form $f(x) = mx + b$ to the data. If the data were to lie on a line, then we would have a solution to the system

$$\begin{aligned} mx_1 + b &= y_1 \\ mx_2 + b &= y_2 \\ &\vdots \\ mx_k + b &= y_k \end{aligned}$$

This system can be written as

$$m\mathbf{w}_1 + b\mathbf{w}_2 = \mathbf{y},$$

where $\mathbf{w}_1 = [x_1\ x_2\ \cdots\ x_k]^\mathsf{T}$, $\mathbf{w}_2 = [1\ 1\ \cdots\ 1]^\mathsf{T}$, and $\mathbf{y} = [y_1\ y_2\ \cdots\ y_k]^\mathsf{T}$. If the data does not lie on a line, then the system won't have a solution. Instead, we want to minimize the square of the distance between $\mathbf{y}$ and a vector of the form $m\mathbf{w}_1 + b\mathbf{w}_2$. That is, minimize

$$||\mathbf{y} - (m\mathbf{w}_1 + b\mathbf{w}_2)||^2. \tag{30.1}$$

Rephrasing this in terms of projections, we are looking for the vector in $W = \text{Span}\{\mathbf{w}_1, \mathbf{w}_2\}$ that is closest to $\mathbf{y}$. In other words, the values of m and b will occur as the weights when we write $\text{proj}_W\mathbf{y}$ as a linear combination of $\mathbf{w}_1$ and $\mathbf{w}_2$. The one wrinkle in this problem is that we need an orthogonal basis for W to find this projection. Find the least squares line for the data points $(1, 2)$, $(2, 4)$ and $(3, 5)$ in $\mathbb{R}^2$.

(5) Each set S is linearly independent. Use the Gram-Schmidt process to create an orthogonal set of vectors with the same span as S. Then find an orthonormal basis for the same span.

(a) $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ -1 \\ 2 \end{bmatrix} \right\}$ in $\mathbb{R}^3$ using the dot product as the inner product

(b) $S = \{1+t, 1-t, t-t^2\}$ in $\mathbb{P}_2$ using the inner product

$$\langle p(t), q(t) \rangle = \int_{-1}^{1} p(t)q(t)\, dt$$

(c) $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$ in $\mathbb{R}^3$ using the inner product $\langle \mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u}) \cdot (A\mathbf{v})$, where $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

(d) $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 3 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 4 \\ -1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ in $\mathbb{R}^7$ with the weighted inner product

$$\begin{aligned} &\langle [u_1\ u_2\ u_3\ u_4\ u_5\ u_6\ u_7]^{\mathsf{T}}, [v_1\ v_2\ v_3\ v_4\ v_5\ v_6\ v_7]^{\mathsf{T}} \rangle \\ &\quad = u_1v_1 + u_2v_2 + u_3v_3 + 2u_4v_4 + 2u_5v_5 + u_6v_6 + u_7v_7 \end{aligned}$$

(6) Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a set of linearly dependent vectors in an inner product space V. What is the result if the Gram-Schmidt process is applied to the set S? Explain.

(7) Let $S = \{1, \cos(t), \sin(t)\}$.

(a) Show that S is a linearly independent set in $C[0, \pi]$.

(b) Use the Gram-Schmidt process to find an orthogonal basis from S using the inner product

$$\langle f(t), g(t) \rangle = \int_{0}^{\pi} f(t)g(t)\, dt$$

(8) The Legendre polynomials form an orthonormal basis for the infinite dimensional inner product space $\mathbb{P}$ of all polynomials using the inner product

$$\langle p(t), q(t) \rangle = \int_{-1}^{1} p(t)q(t)\, dt.$$

The Legendre polynomials have applications to differential equations, statistics, numerical analysis, and physics (e.g., they appear when solving Schrdinger equation in three dimensions for a central force). The Legendre polynomials are found by using the Gram-Schmidt process to find an orthogonal basis from the standard basis $\{1, t, t^2, \cdots\}$ for $\mathbb{P}$. Find the first four Legendre polynomials by creating a orthonormal basis from the set $\{1, t, t^2, t^3\}$.

(9) A fellow student wants to find a QR factorization for a 3×4 matrix. What would you tell this student and why?

(10) Label each of the following statements as True or False. Provide justification for your response. Throughout, let V be a vector space.

(a) **True/False** If $\{\mathbf{w}_1, \mathbf{w}_2\}$ is a basis for a subspace W of an inner product space V, then the vector $\frac{\mathbf{v}\cdot\mathbf{w}_1}{\mathbf{w}_1\cdot\mathbf{w}_1}\mathbf{w}_1 + \frac{\mathbf{v}\cdot\mathbf{w}_2}{\mathbf{w}_2\cdot\mathbf{w}_2}\mathbf{w}_2$ is the vector in W closest to $\mathbf{v}$.

(b) **True/False** If W is a subspace of an inner product space V, then the vector $\mathbf{v} - \text{proj}_W \mathbf{v}$ is orthogonal to every vector in W.

(c) **True/False** If $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$ are vectors in an inner product space V, then the Gram-Schmidt process constructs an orthogonal set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ with the same span as $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

(d) **True/False** Any set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of orthogonal vectors in an inner product space V can be extended to an orthogonal basis of V.

(e) **True/False** If A is an $n \times n$ matrix with $AA^{\mathsf{T}} = I_n$, then the rows of A form an orthogonal set.

(f) **True/False** Every nontrivial finite dimensional subspace of an inner product space has an orthogonal basis.

(g) **True/False** In any inner product space V, if W is a subspace satisfying $W^{\perp} = \{\mathbf{0}\}$, then $W = V$.

Project: Gaussian Quadrature and Legendre Polynomials

Simpson's rule is a reasonably accurate method for approximating definite integrals since it models the integrand on subintervals with quadratics. For that reason, Simpson's rule provides exact values for integrals of all polynomials of degree less than or equal to 2. In Gaussian quadrature, we will use a family of polynomials to determine points at which to evaluate an integral of the form $\int_{-1}^{1} f(t)\, dt$. By allowing ourselves to select evaluation points that are not uniformly distributed across the interval of integration, we will be able to approximate our integrals much more efficiently. The method is constructed so as to obtain exact values for as large of degree polynomial integrands as possible. As a result, if we can approximate our integrand well with polynomials, we can obtain very good approximations with Gaussian quadrature with minimal effort.

The Gaussian quadrature approximation has the form

$$\int_{-1}^{1} f(t)\, dt \approx w_1 f(t_1) + w_2 f(t_2) + \cdots + w_n f(t_n) = \sum_{i=1}^{n} w_i f(t_i), \tag{30.2}$$

where the w_i (weights) and the t_i (nodes) are points in the interval $[-1, 1]$[2]. Gaussian quadrature describes how to find the weights and the points in (30.2) to obtain suitable approximations. We begin to explore Gaussian quadrature with the simplest cases.

Project Activity 30.1. In this activity we find through direct calculation the node and weight with $n = 1$ so that

$$w_1 f(t_1) \approx \int_{-1}^{1} f(t)\, dt. \tag{30.3}$$

There are two unknowns in this situation (w_1 and t_1) and so we will need 2 equations to find these unknowns. Keep in mind that we want to have the approximation (30.3) be exact for as large of degree polynomials as possible.

(a) Assume equality in (30.3) if we choose $f(t) = 1$. Use the resulting equation to find w_1.

(b) Assume equality in (30.3) if we choose $f(t) = t$. Use the resulting equation to find t_1.

(c) Verify that (30.3) is in fact an equality for any linear polynomial of the form $f(t) = a_0 + a_1 t$, using the values of w_1 and t_1 you found

We do one more specific case before considering the general case.

Project Activity 30.2. In this problem we find through direct calculation the nodes and weights with $n = 2$ so that

$$w_1 f(t_1) + w_2 f(t_2) \approx \int_{-1}^{1} f(t)\, dt. \tag{30.4}$$

There are four unknowns in this situation (w_1, w_2 and t_1, t_2) and so we will need 4 equations to find these unknowns. Keep in mind that we want to have the approximation (30.4) be exact for as large of degree polynomials as possible. In this case we will use $f(t) = 1$, $f(t) = t$, $f(t) = t^2$, and $f(t) = t^3$.

(a) Assume equality in (30.4) if we choose $f(t) = 1$. This gives us an equation in w_1 and w_2. Find this equation.

(b) Assume equality in (30.4) if we choose $f(t) = t$. This gives us an equation in w_1, w_2 and t_1, t_2. Find this equation.

(c) Assume equality in (30.4) if we choose $f(t) = t^2$. This gives us an equation in w_1, w_2 and t_1, t_2. Find this equation.

(d) Assume equality in (30.4) if we choose $f(t) = t^3$. This gives us an equation in w_1, w_2 and t_1, t_2. Find this equation.

(e) Solve this system of 4 equations in 4 unknowns. You can do this by hand or with any other appropriate tool. Show that t_1 and t_2 are the roots of the polynomial $t^2 - \frac{1}{3}$.

(f) Verify that (30.4) is in fact an equality with the values of w_1, w_2 and t_1, t_2 you found for any polynomial of the form $f(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3$.

[2] As we will see later, integrations over $[a, b]$ can be converted to an integral over $[-1, 1]$ with a change of variable.

Other than solving a system of linear equations as in Project Activity 30.2, it might be reasonable to ask what the connection is between Gaussian quadrature and linear algebra. We explore that connection now.

In the general case, we want to find the weights and nodes to make the approximation exact for as large degree polynomials as possible. We have $2n$ unknowns w_1, w_2, ..., w_n and t_1, t_2, ..., t_n, so we need to impose $2n$ conditions to determine the unknowns. We will require equality for the $2n$ functions t^i for i from 0 to $2n-1$. This yields the equations

$$\begin{aligned}
w_1 \cdot 1 + w_2 \cdot 1 + \cdots + w_n \cdot 1 &= \int_{-1}^{1} 1 \, dt = t \Big|_{-1}^{1} = 2 \\
w_1 t_1 + w_2 t_2 + \cdots + w_n t_n &= \int_{-1}^{1} t \, dt = \frac{t^2}{2} \Big|_{-1}^{1} = 0 \\
w_1 t_1^2 + w_2 t_2^2 + \cdots + w_n t_n^2 &= \int_{-1}^{1} t^2 \, dt = \frac{t^3}{3} \Big|_{-1}^{1} = \frac{2}{3} \\
&\vdots \quad \vdots \\
w_1 t_1^{2n-1} + w_2 t_2^{2n-1} + \cdots + w_n t_n^{2n-1} &= \int_{-1}^{1} t^{2n-1} \, dt = \frac{t^{2n}}{2n} \Big|_{-1}^{1} = 0.
\end{aligned}$$

In the ith equation the right hand side is

$$\int_{-1}^{1} t^i \, dt = \frac{t^{i+1}}{i+1} \Big|_{-1}^{1} = \begin{cases} \frac{2}{i+1} & \text{if } i \text{ is even,} \\ 0 & \text{if } i \text{ is odd.} \end{cases}$$

Project Activity 30.3. It is inefficient to always solve these systems of equations to find the nodes and weights, especially since there is a more elegant way to find the nodes.

(a) Use appropriate technology to find the equations satisfied by the t_i for $n = 3$, $n = 4$, and $n = 5$.

(b) Now we will see the more elegant way to find the nodes. As we will show for some cases, the nodes can be found as roots of a set of orthogonal polynomials in $\mathbb{P}_n$ with the inner product $\langle f(t), g(t) \rangle = \int_{-1}^{1} f(t)g(t) \, dt$. Begin with the basis $\mathcal{S}_n = \{1, t, t^2, \ldots, t^n\}$ of $\mathbb{P}_n$. Use appropriate technology to find an orthogonal basis B_5 for $\mathbb{P}_5$ obtained by applying the Gram-Schmidt process to $\mathcal{S}_n$. The polynomials in this basis are called *Legendre polynomials*. Check that the nodes are roots of the Legendre polynomials by finding roots of these polynomials using any method. Explain why the t_i appear to be roots of the Legendre polynomials.

Although it would take us beyond the scope of this project to verify this fact, the nodes in the nth Gaussian quadrature approximation (30.2) are in fact the roots of the nth order Legendre polynomial. In other words, if $p_n(t)$ is the nth order Legendre polynomial, then t_1, t_2, ..., t_n are the roots of $p_n(t)$ in $[-1, 1]$. Gaussian quadrature as described in (30.2) using the polynomial $p_n(t)$ is exact if the integrand $f(t)$ is a polynomial of degree less than $2n$.

We can find the corresponding weights, w_i, using the formula[3]

$$w_i = \frac{2}{(1-t_i^2)(q_n'(t_i))^2}, \tag{30.5}$$

where $q_i(t)$ is the ith order Legendre polynomial scaled so that $q_i(1) = 1$.

Project Activity 30.4. Let us see now how good the integral estimates are with Gaussian quadrature method using an example. Use Gaussian quadrature with the indicated value of n to approximate $\displaystyle\int_{-1}^{1} e^t \cos(t)\, dt$. Be sure to explain how you found your nodes and weights (approximate the nodes and weights to 8 decimal places). Compare the approximations with the actual value of the integral. Use technology as appropriate to help with calculations.

(a) $n = 3$

(b) $n = 4$

(c) $n = 5$

Our Gaussian quadrature formula was derived for integrals on the interval $[-1, 1]$. We conclude by seeing how a definite integral on an interval can be converted to one on the interval $[-1, 1]$.

Project Activity 30.5. Consider the problem of approximating an integral of the form $I = \int_a^b g(x)\, dx$. Show that the change of variables $x = \frac{(b-a)t}{2} + \frac{a+b}{2}$, $f(t) = \frac{(b-a)g(x)}{2}$ reduces the integral I to the form $I = \int_{-1}^{1} f(t)\, dt$. (This change of variables can be derived by finding a linear function that maps the interval $[a, b]$ to the interval $[-1, 1]$.)

[3]Abramowitz, Milton; Stegun, Irene A., eds. (1972), 25.4, Integration, Handbook of Mathematical Functions (with Formulas, Graphs, and Mathematical Tables), Dover,