Inner Product and Orthogonality

Inner Product

The notion of **inner product** is important in linear algebra in the sense that it provides a sensible notion of length and angle in a vector space. This seems very natural in the Euclidean space $\mathbb{R}^n$ through the concept of **dot product**. However, the inner product is much more general and can be extended to other non-Euclidean vector spaces. For this course, you are not required to understand the non-Euclidean examples. I just want to show you a glimpse of linear algebra in a more general setting in mathematics.

Definition. Let V be a vector space. An **inner product** on V is a function

$$\langle -, - \rangle \colon V \times V \to \mathbb{R}$$

such that for all vectors $u, v, w \in V$ and scalar $c \in \mathbb{R}$, it satisfies the axioms

1. $\langle u, v\rangle = \langle v, u\rangle$;
2. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$, and $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$;
3. $\langle cu, v\rangle = c\langle u, v\rangle + \langle u, cv\rangle$;
4. $\langle u, u\rangle \geq 0$, and $\langle u, u\rangle = 0$ if and only if $u = 0$.

Definition. In a Euclidean space $\mathbb{R}^n$, the **dot product** of two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined to be the function

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

In coordinates, if we write

$$\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \text{ and } \mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix},$$

then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \begin{bmatrix} u_1 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + \cdots u_n v_n.$$

The definitions in the remainder of this note will assume the Euclidean vector space $\mathbb{R}^n$, and the dot product as the natural inner product.

Lemma. *The dot product on $\mathbb{R}^n$ is an inner product.*

Exercise. Verify that the dot product satisfies the four axioms of inner products.

Example 1. Let $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$, and define the function

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T A \mathbf{v}^T$$

We will show that this function defines an inner product on $\mathbb{R}^2$.

Write the vectors

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \text{ and } \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix},$$

then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 7u_1v_1 + 2u_1v_2 + 2u_2v_1 + 4u_2v_2.$$

To verify axiom 1, we compute

$$\langle \mathbf{v}, \mathbf{u} \rangle = 7v_1u_1 + 2v_1u_2 + 2v_2u_1 + 4v_2u_2.$$

Hence, we can conclude that $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

To verify axiom 2, we see that

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle &= 7u_1(v_1 + w_1) + 2u_1(v_2 + w_2) + 2u_2(v_1 + w_1) + 4u_2(v_2 + w_2) \\ &= (7u_1v_1 + 2u_1v_2 + 2u_2v_1 + 4u_2v_2) + (7u_1w_1 + 2u_1w_2 + 2u_2w_1 + 4u_2w_2) \\ &= \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle. \end{aligned}$$

The other equality follows from axiom 1. Thus axiom 2 is verified.

To verify axiom 3, we see that

$$\begin{aligned} c\langle \mathbf{u}, \mathbf{v} \rangle &= c(7u_1v_1 + 2u_1v_2 + 2u_2v_1 + 4u_2v_2) \\ &= 7(cu_1)v_1 + 2(cu_1)v_2 + 2(cu_2)v_1 + 4(cu_2)v_2 = \langle c\mathbf{u}, \mathbf{v} \rangle. \end{aligned}$$

The other equality also follows from axiom 1. Thus axiom 3 is verified.

To verify axiom 4, notice that

$$\langle \mathbf{u}, \mathbf{u} \rangle = 7u_1u_1 + 2u_1u_2 + 2u_2u_1 + 4u_2u_2 = 7u_1^2 + 4u_1u_2 + 4u_2^2 = 3u_1^2 + 4(u_1 + u_2)^2.$$

If $\langle \mathbf{u}, \mathbf{u} \rangle = 0$, that means $u_1 = u_2 = 0$. The converse is clear. Hence, axiom 4 is verified.

Remark (Symmetric Bilinear Form)**.** Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then the function given by

$$\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^T A\mathbf{v}$$

for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, defines an inner product on $\mathbb{R}^n$. Inner products on $\mathbb{R}^n$ defined in this way are called **symmetric bilinear form**. In fact, every inner product on $\mathbb{R}^n$ is a symmetric bilinear form. In particular, the standard dot product is defined with the identity matrix I, which is symmetric.

Definition. The **length** (or **norm**) of a vector $\mathbf{v} \in \mathbb{R}^n$, denoted by $\|\mathbf{v}\|$, is defined by

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

Remark. By the last axiom of the inner product, $\mathbf{v}\cdot\mathbf{v} \geq 0$, thus the length of $\mathbf{v}$ is always a non-negative real number, and the length is 0 if and only if $\mathbf{v}$ is the zero vector.

Definition. A vector with length 1 is called a **unit vector**. If $\mathbf{v} = \mathbf{0}$, then the vector

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{\sqrt{\mathbf{v}\cdot\mathbf{v}}}\mathbf{v}$$

is the **normalization** of $\mathbf{v}$.

Example 2. Consider the vector

$$\mathbf{v} = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}.$$

Its length is

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{2^2 + 2^2 + 1^2} = \sqrt{9} = 3,$$

and its normalization is

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{3}\mathbf{v} = \begin{bmatrix} 2/3 \\ 2/3 \\ 1/3 \end{bmatrix}.$$

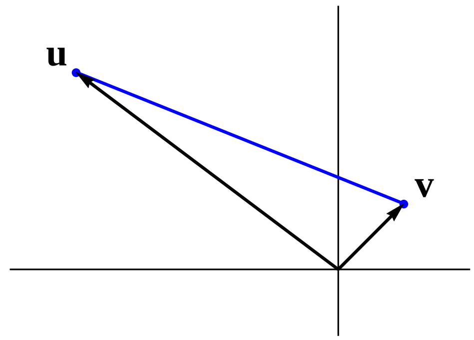
Definition. For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, we can define the **distance** between them by

$$\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|.$$

Example 3. Let $\mathbf{u} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, then the distance between them is

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{(-4-1)^2 + (3-1)^2} = \sqrt{29}.$$

This distance is demonstrated in the following figure.



Orthogonality

The notion of inner product allows us to introduce the notion of orthogonality, together with a rich family of properties in linear algebra.

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

Theorem 1 (Pythagorean)**.** *Two vectors are orthogonal if and only if* $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof. This well-known theorem has numerous different proofs. The linear-algebraic version looks like this. Notice that

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

The theorem follows from the fact that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\mathbf{u} \cdot \mathbf{v} = 0$. □

The following is an important concept involving orthogonality.

Definition. Let $W \subseteq R^n$ be a subspace. If a vector $\mathbf{x}$ is orthogonal to every vector $\mathbf{w} \in W$, we say that x is **orthogonal** to W. The **orthogonal complement** of W, denoted by $W^\perp$, is the collection of all vectors orthogonal to W, *i.e.*,

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

Lemma. *$W^\perp$ is a subspace of $\mathbb{R}^n$.*

Exercise. Verify that $W^\perp$ satisfies the axioms of a subspace.

Exercise. Let $W \subseteq \mathbb{R}^n$ be a subspace. Prove that $\dim W + \dim W^\perp = n$. [Hint: Use Rank-Nullity Theorem]

Theorem 2. *If* $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ *forms a basis of* W*, then* $\mathbf{x} \in W^\perp$ *if and only if* $\mathbf{x} \cdot \mathbf{w}_i = 0$ *for all integers* $1 \le i \le k$.

Proof. Let $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ be a basis of W. Assume that $\mathbf{x} \cdot \mathbf{w}_i = 0$. Let $\mathbf{w} \in W$ be arbitrary. Then $\mathbf{w}$ can be written as a linear combination

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

By the linearity of dot product, we have

$$\mathbf{x} \cdot \mathbf{w} = c_1\mathbf{x} \cdot \mathbf{w}_1 + \cdots + c_k\mathbf{x} \cdot \mathbf{w}_k = 0 + \cdots + 0 = 0.$$

Thus $x \in W^\perp$.

The converse is clear. □

Example 4. Find the orthogonal complement of $W = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2\}$, where

$$\mathbf{w}_1 = \begin{bmatrix} 3 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \text{ and } \mathbf{w}_2 = \begin{bmatrix} 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

By the theorem above, to find all vectors $\mathbf{x} \in W^\perp$, we only need to make sure we find all $\mathbf{x} \in \mathbb{R}^4$ that satisfies

$$\mathbf{x} \cdot \mathbf{w}_1 = \mathbf{x} \cdot \mathbf{w}_2 = 0.$$

If we write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix},$$

we immediately obtain a system of linear equations

$$\begin{aligned} 3x_1 + x_3 + x_4 &= 0 \\ 2x_2 + 5x_3 + x_4 &= 0 \end{aligned}$$

A sequence of row operation gives all solutions to this linear system

$$\mathbf{x} = \begin{bmatrix} -x_3/3 - x_4/3 \\ -5x_3/2 - x_4/2 \\ x_3 \\ x_4 \end{bmatrix} = s \begin{bmatrix} 2 \\ 15 \\ -6 \\ 0 \end{bmatrix} + t \begin{bmatrix} 2 \\ 3 \\ 0 \\ -6 \end{bmatrix}$$

Exercise. Let A be an $m \times n$ matrix. Prove that $(\text{Col}A)^{\perp} = \text{Nul}A^T$. [Hint: Use a similar method as demonstrated in the previous example.]

Definition. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ in $\mathbb{R}^n$ is an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, *i.e.*, $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$. An **orthogonal basis** for a subspace W is a basis for W that is also an orthogonal set. An **orthonormal basis** for a subspace W is an orthogonal basis for W where each vector has length 1.

Example 5. The standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ forms an orthonormal basis for $\mathbb{R}^n$.

Example 6. Consider the following vectors in $\mathbb{R}^4$:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \ \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \ \mathbf{v}_4 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}.$$

One can easily verify that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ forms an orthogonal set. Furthermore, the matrix

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{bmatrix}$$

is row equivalent to the identity matrix. Thus the four vectors are linearly independent. It follows that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ forms an orthogonal basis for $\mathbb{R}^4$. Furthermore, if we normalize the vectors and obtain

$$\mathbf{u}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \ \mathbf{u}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \ \mathbf{u}_3 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}, \ \mathbf{u}_4 = \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix},$$

then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ forms an orthonormal basis for $\mathbb{R}^4$

Exercise. Verify that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in Example 6 forms an orthogonal basis for $\mathbb{R}^4$.

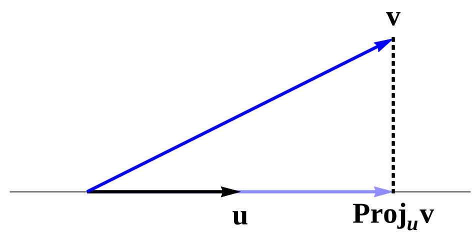
Definition. Let A be an $n \times n$ matrix. We say A is an **orthogonal matrix** if $A^T A = I$.

Theorem 3. *Let A be an $n \times n$ matrix. Then the followings are equivalent.*

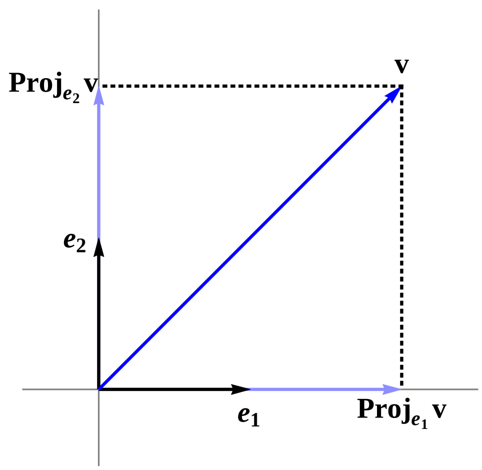
1. *A is an orthogonal matrix.*
2. *The column vectors of A are orthonormal.*
3. *The row vectors of A are orthonormal.*
4. *A preserves length, that is, $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$.*
5. *A preserves the dot product, which means $(A\mathbf{x}) \cdot (A\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.*

Orthogonal Projection

The idea of orthogonal projection is best depicted in the following figure.



The orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$ gives the component vector $\text{Proj}_{\mathbf{u}}\mathbf{v}$ of $\mathbf{v}$ in the direction of $\mathbf{u}$. This fact is best demonstrated in the case that $\mathbf{u}$ is one of the standard basis vectors.



As shown in the figure above, the lengths of the orthogonal projections in the $\mathbf{e}_1$ and $\mathbf{e}_2$ directions, respectively, give the coordinates of the vector $\mathbf{v}$ in the standard basis. On the other hand, each coordinate can be obtained by computing the dot product of $\mathbf{v}$ and the corresponding standard basis vector, *i.e.*,

$$\|\text{Proj}_{\mathbf{e}_1}\mathbf{v}\| = \mathbf{v} \cdot \mathbf{e}_1, \quad \text{and} \quad \|\text{Proj}_{\mathbf{e}_2}\mathbf{v}\| = \mathbf{v} \cdot \mathbf{e}_2.$$

However, the orthogonal projection of $\mathbf{v}$ in the $\mathbf{e}_1$ direction should not depend on the length of the vector we use to specify the direction. Hence, the validity of the observation above is based on the fact that $\mathbf{e}_1$ and $\mathbf{e}_2$ are "special" in some sense. The observation holds true precisely because the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are unit vectors.

To obtain a similar conclusion in the general setting, consider vectors $\mathbf{u}$ and $\mathbf{v}$ in the first figure. We first normalize $\mathbf{u}$ to get

$$\hat{\mathbf{u}} = \frac{\mathbf{u}}{\sqrt{\mathbf{u} \cdot \mathbf{u}}}$$

Now, this unit vector $\hat{\mathbf{u}}$ satisfies that

$$\|\text{Proj}_{\hat{\mathbf{u}}}\mathbf{v}\| = \mathbf{v} \cdot \hat{\mathbf{u}}.$$

Because $\mathbf{u}$ and $\hat{\mathbf{u}}$ are in the same direction, we have $\text{Proj}_{\hat{\mathbf{u}}}\mathbf{v} = \text{Proj}_{\mathbf{u}}\mathbf{v}$. Thus

$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \|\text{Proj}_{\hat{\mathbf{u}}}\mathbf{v}\|\,\hat{\mathbf{u}} = (\mathbf{v} \cdot \hat{\mathbf{u}})\,\hat{\mathbf{u}} = \left(\frac{\mathbf{v} \cdot \mathbf{u}}{\sqrt{\mathbf{u} \cdot \mathbf{u}}}\right)\frac{\mathbf{u}}{\sqrt{\mathbf{u} \cdot \mathbf{u}}} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\,\mathbf{u}.$$

Definition. Given vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, where $\mathbf{u} \neq \mathbf{0}$, then the **orthogonal projection** of $\mathbf{v}$ onto $\mathbf{u}$ is defined to be

$$\text{Proj}_{\mathbf{u}}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\,\mathbf{u}.$$

Exercise. Find the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$ in each case.

1. $\mathbf{u} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$
2. $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} -3 \\ 9 \end{bmatrix}$

Exercise. Find the distance from the point $(-3, 9)$ to the line $y = 2x$.

It is also useful to consider the orthogonal projection of a vector onto a subspace (not necessarily 1-dimensional). Let $W \subseteq \mathbb{R}^n$ be a subspace, and let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthonormal basis for the subspace W. Given a vector $\mathbf{v} \in \mathbb{R}^n$, its projection on the orthonormal basis vectors are

$$\text{Proj}_{\mathbf{u}_i}\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}_i}{\mathbf{u}_i \cdot \mathbf{u}_i}\,\mathbf{u}_i.$$

So the orthogonal projection of $\mathbf{v}$ onto the subspace W is the linear combination

$$\text{Proj}_W\mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\,\mathbf{u}_1 + \cdots + \frac{\mathbf{v} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k}\,\mathbf{u}_k.$$

Notice that the orthogonal projection of $\mathbf{v}$ onto $\mathbf{u}$ is the same with the orthogonal projection of $\mathbf{v}$ onto the 1-dimensional subspace W spanned by the vector $\mathbf{u}$, since W contains a unit vector, namely $\mathbf{u}/\|\mathbf{u}\|$, and it forms an orthonormal basis for W.

Example 7. Consider the subspace $W \subseteq \mathbb{R}^n$ given by

$$W = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_{k+1} = \cdots = x_n = 0\}.$$

Clearly, $\{e_1, \ldots, e_k\}$ gives an orthonormal basis for W. A straightforward computation will show that the orthogonal projection of any vector $\mathbf{x} = \{x_1, \ldots, x_n\} \in \mathbb{R}^n$ is

$$\text{Proj}_W\mathbf{x} = (x_1, \ldots, x_k, 0, \ldots, 0).$$

In the case that $n = 3$ and $k = 2$, this is the orthogonal projection onto the xy-plane.

Exercise. The following vectors in $\mathbb{R}^4$ form an orthogonal set.

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

Find the orthogonal projection of the vector

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

onto the subspace $W = \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

The Gram-Schmidt Process

When we compute orthogonal projection onto a subspace W, we need an orthonormal basis of this subspace. The **Gram-Schmidt process** provides an algorithm to find an orthonormal basis of a subspace.

Algorithm (Gram-Schmidt)**.** Given a subspace $W \subseteq \mathbb{R}^n$ of dimension k, the following procedure will provide an orthonormal basis for W.

- Find a set of basis $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$;
- Let $\mathbf{v}_1 = \mathbf{x}_1$;
- Let $\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_2)$;
- Let $\mathbf{v}_3 = \mathbf{x}_3 - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_3) - \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{x}_3)$;
- $\cdots\cdots\cdots\cdots$
- Let $\mathbf{v}_k = \mathbf{x}_k - \mathrm{Proj}_{\mathbf{v}_1}(\mathbf{x}_k) - \mathrm{Proj}_{\mathbf{v}_2}(\mathbf{x}_k) - \cdots - \mathrm{Proj}_{\mathbf{v}_{k-1}}(\mathbf{x}_k)$;
- Now let $\mathbf{u}_i = \dfrac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$.
- Then $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ forms an orthogonal basis for W, and $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ forms an orthonormal basis for W.

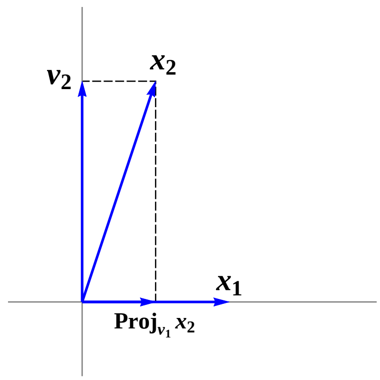
Example 8. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be vectors in $\mathbb{R}^2$ given by

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \text{ and } \mathbf{x}_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

By the Gram-Schmidt process, let

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \text{ and } \mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix} - \frac{2}{4}\begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}.$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms an orthogonal basis for $\mathbb{R}^2$. The geometric meaning for this process is the following figure.



To get an orthonormal basis for $\mathbb{R}^2$, we normalize the vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, and get

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ and } \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Then $\{\mathbf{u}_1, \mathbf{u}_2\}$ forms an orthonormal basis for $\mathbb{R}^2$. Notice that this orthonormal basis coincides with the standard basis of $\mathbb{R}^2$.

Exercise. Use the Gram-Schmidt process to find an orthonormal basis of the column space of the following matrix.

$$\begin{bmatrix} 3 & -5 & 1 \\ 1 & 1 & 1 \\ -1 & 5 & -2 \\ 3 & -7 & 8 \end{bmatrix}$$

Least-Squares Problems

When we solve for a linear system, sometimes we run into the trouble that the system may not be consistent, which means there is no solution *exactly* solving the system. However, we sometimes only need an approximate solution. The **least-squares method** provides a way to solve for the best approximation of the solution.

Example 9. Given three points $(2, 3)$, $(3, 4)$, and $(4, 5)$, find a line $y = mx + b$ that best approximate these three points.

We can set up a system of linear equations to solve for m and b:

$$2 \cdot m + b = 3;$$
$$3 \cdot m + b = 4;$$
$$4 \cdot m + b = 5.$$

This system corresponds to the augmented matrix and reduced echelon form

$$\begin{bmatrix} 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, when $m = 1$, $b = 1$, the line $y = x + 1$ passes through these three given points.

Example 10. Given three points $(2, 3)$, $(3, 2)$, and $(4, 5)$, find a line $y = mx + b$ that best approximates these three points.

Once again, we set up a system of linear equations

$$2 \cdot m + b = 3;$$
$$3 \cdot m + b = 2;$$
$$4 \cdot m + b = 5.$$

The augmented matrix and reduced echelon form would be

$$\begin{bmatrix} 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Notice that the system is inconsistent. This makes sense because the three points given are not colinear. Now let's try to find a best linear approximation to these three points. Let

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}, \quad \mathbf{x}^* = \begin{bmatrix} m \\ b \end{bmatrix}$$

We know that $A\mathbf{x}^*$ gives a linear combination of the column vectors of A. So $A\mathbf{x}^* \in \mathrm{Col}(A)$. However, $\mathbf{b} \notin \mathrm{Col}(A)$, as the linear system $A\mathbf{x} = \mathbf{b}$ is inconsistent. The best approximation of $\mathbf{b}$ in the subspace $\mathrm{Col}(A) \subseteq \mathbb{R}^3$ would be the orthogonal projection

$$\mathbf{b}^* = \mathrm{Proj}_{\mathrm{Col}(A)}\mathbf{b}$$

This projection lands in the column space $\mathrm{Col}(A)$. Thus the system $A\mathbf{x}^* = \mathbf{b}^*$ is consistent. Now we need to solve for the least-squares solution $\mathbf{x}^*$. Because $\mathbf{b} - \mathbf{b}^*$ is orthogonal to $\mathrm{Col}(A)$, we may set

$$A^T(\mathbf{b} - A\mathbf{x}^*) = A^T(\mathbf{b} - \mathbf{b}^*) = 0.$$

Hence, we obtain the **normal equation**

$$A^T A\mathbf{x}^* = A^T\mathbf{b}.$$

By the above argument, the normal equation is always consistent, and we can always find a (not necessarily unique) least-square solution $\mathbf{x}^*$ to the system.

Now we go back to the example we are given.

$$A^T A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 29 & 9 \\ 9 & 3 \end{bmatrix}, \text{ and } A^T\mathbf{b} = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 32 \\ 10 \end{bmatrix}.$$

The least-square system has augmented matrix and reduced echelon form

$$\begin{bmatrix} 29 & 9 & 32 \\ 9 & 3 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/3 \end{bmatrix}$$

Thus, when $m = 1$, $b = 1/3$, the line $y = x + 1/3$ is the best approximation to these three points.

Exercise. Find the trigonometric function of the form $f(x) = c_0 + c_1 \sin^2(x) + c_2 \cos^3(x)$ that best fits the data points $(0, 0), (\pi/2, 1), (\pi, 2), (3\pi/2, 3)$.