6 Orthogonality and Least Squares

6.1 Inner Product, Length, and Orthogonality

Definition If $\bar{u}$ and $\bar{v}$ are viewed as column vectors in $\mathbb{R}^n$, then the matrix product $\bar{u}^T\bar{v}$ is a 1×1 matrix (or just a scalar) called the **inner product** of $\bar{u}$ and $\bar{v}$. We often write it as $\bar{u} \cdot \bar{v}$, and call it the **dot product**.

More specifically, if

$$\bar{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \qquad \text{and} \qquad \bar{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix},$$

then

$$\bar{u}^T\bar{v} = \bar{u} \cdot \bar{v} = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Example Compute $\bar{u} \cdot \bar{v}$ and $\bar{v} \cdot \bar{u}$ for

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

Note that

$$\bar{u} \cdot \bar{v} = (2)(1) + (1)(-1) + (3)(2) = 7.$$

Likewise,

$$\bar{v} \cdot \bar{u} = (1)(2) + (-1)(1) + (2)(3) = 7.$$

We summarize some properties of the inner product in the following theorem.

Theorem Let $\bar{u}$, $\bar{v}$, and $\bar{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
(a) $\bar{u} \cdot \bar{v} = \bar{v} \cdot \bar{u}$;
(b) $(\bar{u} + \bar{v}) \cdot \bar{w} = \bar{u} \cdot \bar{w} + \bar{v} \cdot \bar{w}$;
(c) $(c\bar{u}) \cdot \bar{v} = c(\bar{u} \cdot \bar{v}) = \bar{u} \cdot (c\bar{v})$;
(d) $\bar{u} \cdot \bar{u} \geq 0$, and $\bar{u} \cdot \bar{u} = 0$ if and only if $\bar{u} = 0$.

We also get the following "linearity" property by combining (b) and (c) from above:

$$(c_1\bar{u}_1 + \cdots + c_p\bar{u}_p) \cdot \bar{w} = c_1(\bar{u}_1 \cdot \bar{w}) + \cdots + c_p(\bar{u}_p \cdot \bar{w}).$$

Definition The **length** (or **norm**) of a vector $\bar{v} \in \mathbb{R}^n$ with entries $v_1, \ldots, v_n$ is a nonnegative scalar, denoted $\|\bar{v}\|$, and defined by

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

Note If $\bar{v} \in \mathbb{R}^2$ or $\mathbb{R}^3$, then this definition coincides with the standard definition of distance coming from the Pythagorean theorem. We also have that $\|\bar{v}\|^2 = \bar{v} \cdot \bar{v}$.

Note For any scalar c, we have that $\|c\bar{v}\| = |c|\,\|\bar{v}\|$. That is, scalars scale the length of a vector as expected.

Definition A vector of length 1 is called a **unit vector**. If $\bar{v}$ is a vector, then we can **normalize** $\bar{v}$ by finding a unit vector $\bar{u}$ that points in the same direction as $\bar{v}$. Specifically

$$\bar{u} = \frac{\bar{v}}{\|\bar{v}\|},$$

i.e. you simply divide $\bar{v}$ by its length to get a unit vector that points in the same direction as $\bar{v}$.

Example Let $\bar{v} = (1, 2, 0, 2)$. Find a unit vector $\bar{u}$ in the same direction as $\bar{v}$.

$$\|\bar{v}\| = \sqrt{1^2 + 2^2 + 0^2 + 2^2} = 3,$$

so

$$\bar{u} = \frac{\bar{v}}{\|\bar{v}\|} = \left(\frac{1}{3}, \frac{2}{3}, 0, \frac{2}{3}\right).$$

We can use the definition of length to establish a distance formula between two vectors.

Definition Given two vectors $\bar{u}, \bar{v} \in \mathbb{R}^n$, the **distance between** $\bar{u}$ **and** $\bar{v}$, written $\text{dist}(\bar{u}, \bar{v})$, is the length of the vector $\bar{u} - \bar{v}$. Specifically,

$$\text{dist}(\bar{u}, \bar{v}) = \|\bar{u} - \bar{v}\|.$$

Note In $\mathbb{R}^2$ and $\mathbb{R}^3$, this coincides with the usual definition of distance between two points.

Example Compute the distance between $\bar{u} = (2, 4)$ and $\bar{v} = (-1, 6)$.

$$\text{dist}(\bar{u}, \bar{v}) = \|\bar{u} - \bar{v}\| = \|(2, 4) - (-1, 6)\| = \|(3, -2)\| = \sqrt{3^2 + (-2)^2} = \sqrt{13}.$$

Note Supposing $\bar{u} = (u_1, \ldots, u_n)$ and $\bar{v} = (v_1, \ldots, v_n)$, we can write the distance formula a bit more explicitly, as follows:

$$\begin{aligned} \text{dist}(\bar{u}, \bar{v}) &= \sqrt{(\bar{u} - \bar{v}) \cdot (\bar{u} - \bar{v})} \\ &= \sqrt{(u_1 - v_1)^2 + \cdots + (u_n - v_n)^2}. \end{aligned}$$

Definition Two vectors $\bar{u}$ and $\bar{v}$ in $\mathbb{R}^n$ are **orthogonal** to each other if $\bar{u} \cdot \bar{v} = 0$.

You should think of orthogonality in $\mathbb{R}^n$ as the analogue of perpendicularity in $\mathbb{R}^2$.

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

Definition Let W be a subspace of $\mathbb{R}^n$. If a vector $\bar{z}$ is orthogonal to *every* vector in W, we say that $\bar{z}$ is **orthogonal to** W. The set of all vectors that are orthogonal to W is called the **orthogonal complement** of W, and is denoted $W^\perp$ ("W perp").

Facts Two facts about orthogonal complements include the following:
(i) A vector $\bar{x}$ is in $W^\perp$ if and only if $\bar{x}$ is orthogonal to every vector in a set that spans W.
(ii) $W^\perp$ is a subspace of $\mathbb{R}^n$.

The last theorem in this section is one that relates row spaces, column spaces, and null spaces of a matrix A via the notion of orthogonal complements.

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

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

6.2 Orthogonal Sets

Definition A set of vectors $\{\bar{u}_1, \ldots, \bar{u}_p\} \subset \mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, i.e. if $\bar{u}_i \cdot \bar{u}_j = 0$ whenever $i \neq j$.

Example The easiest example of an orthogonal set in $\mathbb{R}^n$ is the standard basis $\{\bar{e}_1, \ldots, \bar{e}_n\}$. Notice, for example that

$$\bar{e}_1 \cdot \bar{e}_2 = (1)(0) + (0)(1) + 0^2 + \cdots + 0^2 = 0.$$

Example We show that $\{\bar{u}_1, \bar{u}_2, \bar{u}_3\}$ is an orthogonal set, where

$$\bar{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \qquad \bar{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \qquad \bar{u}_3 = \begin{bmatrix} -\frac{1}{2} \\ -2 \\ \frac{7}{2} \end{bmatrix}.$$

We see that

$$\begin{aligned} \bar{u}_1 \cdot \bar{u}_2 &= (3)(-1) + (1)(2) + (1)(1) = 0 \\ \bar{u}_1 \cdot \bar{u}_3 &= (3)\left(-\frac{1}{2}\right) + (1)(-2) + (1)\left(\frac{7}{2}\right) = 0 \\ \bar{u}_2 \cdot \bar{u}_3 &= (-1)\left(-\frac{1}{2}\right) + (2)(-2) + (1)\left(\frac{7}{2}\right) = 0. \end{aligned}$$

As is perhaps suggested by the first example above, we have the following theorem.

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

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

In general, orthogonal bases for vector spaces are much easier to work with than other bases. The following theorem explains why.

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

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

are given by

$$c_j = \frac{\bar{y} \cdot \bar{u}_j}{\bar{u}_j \cdot \bar{u}_j}.$$

Example Express the vector

$$\bar{y} = \begin{bmatrix} 3 \\ \frac{1}{2} \\ -4 \end{bmatrix}$$

as a linear combination of the vectors used in the example above. We see that

$$\bar{y} \cdot \bar{u}_1 = (3)(3) + \left(\frac{1}{2}\right)(1) + (-4)(1) = \frac{11}{2} \qquad \bar{y} \cdot \bar{u}_2 = (3)(-1) + \left(\tfrac{1}{2}\right)(2) + (-4)(1) = -6 \qquad \bar{y} \cdot \bar{u}_3 = -\frac{33}{2}$$

$$\bar{u}_1 \cdot \bar{u}_1 = 3^2 + 1^2 + 1^2 = 11 \qquad \bar{u}_2 \cdot \bar{u}_2 = (-1)^2 + 2^2 + 1^2 = 6 \qquad \bar{u}_3 \cdot \bar{u}_3 = \frac{33}{2}.$$

It follows then that

$$\begin{aligned} \bar{y} &= \frac{\bar{y} \cdot \bar{u}_1}{\bar{u}_1 \cdot \bar{u}_1}\bar{u}_1 + \frac{\bar{y} \cdot \bar{u}_2}{\bar{u}_2 \cdot \bar{u}_2}\bar{u}_2 + \frac{\bar{y} \cdot \bar{u}_2}{\bar{u}_2 \cdot \bar{u}_2}\bar{u}_2 \\ &= \frac{1}{2}\bar{u}_1 - \bar{u}_2 - \bar{u}_3. \end{aligned}$$

Given a nonzero vector $\bar{u} \in \mathbb{R}^n$, it is often the case that we want to decompose another vector $\bar{y} \in \mathbb{R}^n$ into some multiple of $\bar{u}$ plus another vector that is orthogonal to $\bar{u}$. (A picture will make this idea a bit clearer.) In other words, we want to write

$$\bar{y} = \hat{y} + \bar{z},$$

where $\hat{y} = \alpha\bar{u}$ and $\bar{z}$ is orthogonal to $\bar{u}$.

Definition The vector $\hat{y}$ mentioned in the preceding discussion is called the **orthogonal projection of** $\bar{y}$ **onto** $\bar{u}$, and it is given explicitly as

$$\hat{y} = \frac{\bar{y}\cdot\bar{u}}{\bar{u}\cdot\bar{u}}\bar{u} \qquad \left(\text{so } \alpha = \frac{\bar{y}\cdot\bar{u}}{\bar{u}\cdot\bar{u}}\right).$$

Note The vector $\hat{y}$ is also sometimes denoted $\mathrm{proj}_L(\bar{y})$, where L is the space spanned by $\bar{u}$.

It follows from the equation above that

$$\bar{z} = \bar{y} - \hat{y}.$$

You may check that, in fact, $\hat{y}$ is orthogonal to $\bar{z}$ by taking the dot product $\hat{y}\cdot(\bar{y}-\hat{y})$.

Example Let $\bar{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\bar{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Find the orthogonal projection of $\bar{y}$ onto $\bar{u}$, then decompose $\bar{y}$ as the sum of two orthogonal vectors. We have first that

$$\begin{aligned} \hat{y} &= \frac{\bar{y}\cdot\bar{u}}{\bar{u}\cdot\bar{u}}\bar{u} \\ &= \frac{(7)(2)+(6)(1)}{2^2+1^2}\bar{u} \\ &= \frac{20}{5}\bar{u} \\ &= 4\bar{u} \\ &= \begin{bmatrix} 8 \\ 4 \end{bmatrix}. \end{aligned}$$

It follows that $\bar{z} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$. Hence

$$\bar{y} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix},$$

and we see that the vectors on the right hand side are orthogonal since

$$(8)(-1) + (4)(2) = 0.$$

Example Let $\bar{y}$ and $\bar{u}$ be as in the preceding example, with L the line spanned by $\bar{u}$. Then the distance from $\bar{y}$ to L is

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

Definition A set $\{\bar{u}_1, \ldots, \bar{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\bar{u}_1, \ldots, \bar{u}_p\}$ is an **orthonormal basis** for W.

Example The easiest example of an orthonormal set (that is in fact also an orthonormal basis) is $\{\bar{e}_1, \ldots, \bar{e}_n\} \subset \mathbb{R}^n$.

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

The interesting thing about a matrix with orthonormal columns is that it preserves lengths and orthogonality, as follows.

Theorem Let U be an $m \times n$ matrix with orthonormal columns, and let $\bar{x}$ and $\bar{y}$ be in $\mathbb{R}^n$. Then
(a) $\|U\bar{x}\| = \|\bar{x}\|$;
(b) $(U\bar{x})\cdot(U\bar{y}) = \bar{x}\cdot\bar{y}$;
(c) $(U\bar{x})\cdot(U\bar{y}) = 0$ if and only if $\bar{x}\cdot\bar{y} = 0$.

6.3 Orthogonal Projections

We start this section with a theorem that says that if W is a subspace of $\mathbb{R}^n$, then any vector in $\bar{y} \in \mathbb{R}^n$ can be written in the form $\bar{y} = \hat{y} + \bar{z}$, where $\hat{y}$ is the projection of $\bar{y}$ onto W, and $\bar{z}$ is in $W^\perp$. Intuitively, you should think of $\hat{y}$ as the closest approximation to $\bar{y}$ subject to the restriction that you cannot leave W.

Theorem Let W be a subspace of $\mathbb{R}^n$. Then each $\bar{y} \in \mathbb{R}^n$ can be written *uniquely* in the form

$$\bar{y} = \hat{y} + \bar{z},$$

where $\hat{y} \in W$ and $\bar{z} \in W^\perp$. In fact, if $\{\bar{u}_1, \ldots, \bar{u}_p\}$ is an orthogonal basis for W, then

$$\hat{y} = \frac{\bar{y} \cdot \bar{u}_1}{\bar{u}_1 \cdot \bar{u}_1}\bar{u}_1 + \cdots + \frac{\bar{y} \cdot \bar{u}_p}{\bar{u}_p \cdot \bar{u}_p}\bar{u}_p.$$

You should think of each term in the sum as the projection of $\bar{y}$ onto the subspace L_i of W that is generated by $\bar{u}_i$. Taking the sum of all these projections, we end up with the total projection $\hat{y}$ of $\bar{y}$ onto W.

Example Let

$$\bar{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \qquad \bar{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \qquad \bar{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

You may check that $\{\bar{u}_1, \bar{u}_2\}$ is an orthogonal basis for $W = \text{span}\{\bar{u}_1, \bar{u}_2\}$ by computing $\bar{u}_1 \cdot \bar{u}_2$. We write $\bar{y}$ as a the sum of a vector in W and a vector orthogonal to W as follows:

$$\begin{aligned}
\hat{y} &= \frac{\bar{y} \cdot \bar{u}_1}{\bar{u}_1 \cdot \bar{u}_1}\bar{u}_1 + \frac{\bar{y} \cdot \bar{u}_2}{\bar{u}_2 \cdot \bar{u}_2}\bar{u}_2 \\
&= \frac{2 + 10 - 3}{2^2 + 5^2 + 1^2}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{-2 + 2 + 3}{2^2 + 1^2 + 1^2}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \\
&= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \\
&= \frac{3}{10}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} \frac{3}{5} - 1 \\ \frac{3}{2} + \frac{1}{2} \\ -\frac{3}{10} + \frac{1}{2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix}.
\end{aligned}$$

Using the equation $\bar{z} = \bar{y} - \hat{y}$, we see

$$\bar{z} = \begin{bmatrix} 1 + \frac{2}{5} \\ 2 - 2 \\ 3 - \frac{1}{5} \end{bmatrix} = \begin{bmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{bmatrix},$$

so

$$\begin{aligned}
\bar{y} &= \hat{y} + \bar{z} \\
&= \begin{bmatrix} -\frac{2}{5} \\ 2 \\ \frac{1}{5} \end{bmatrix} + \begin{bmatrix} \frac{7}{5} \\ 0 \\ \frac{14}{5} \end{bmatrix}.
\end{aligned}$$

Since $\hat{y}$ is a combination of $\bar{u}_1$ and $\bar{u}_2$ by construction, it is clear that $\hat{y} \in W$. To see that $\bar{z} \in W^\perp$, you can check that

$$(c_1\bar{u}_1 + c_2\bar{u}_2) \cdot \bar{z} = c_1(\bar{u}_1 \cdot \bar{z}) + c_2(\bar{u}_2 \cdot \bar{z}) = c_1(0) + c_2(0) = 0.$$

The following theorem is a formalization of the remark made earlier that you should think of $\hat{y}$ as the closest approximation to $\bar{y}$ that lives in W.

Theorem (The Best Approximation Theorem)
Let W be a subspace of $\mathbb{R}^n$, let $\bar{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{y}$ be the orthogonal projection of $\bar{y}$ onto W. Then $\hat{y}$ is the closest "point" of W to $\bar{y}$ in the sense that

$$\|\bar{y} - \hat{y}\| < \|\bar{y} - \bar{v}\|$$

for any $\bar{v} \in W$ such that $\bar{v} \neq \hat{y}$.

Drawing a picture of the situation is perhaps the best way to understand what's really going on here. Another way to phrase this is that the shortest distance between the "point" $\bar{y}$ and the "plane" W is given by $\|\bar{y} - \hat{y}\|$, where $\hat{y}$ is the orthogonal projection of $\bar{y}$ onto W.

Example Find the shortest distance from $\bar{y}$ to $W = \text{span}\{\bar{u}_1, \bar{u}_2\}$, where

$$\bar{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \qquad \bar{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \qquad \bar{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

First we find $\hat{y}$:

$$\begin{aligned} \hat{y} &= \frac{\bar{y} \cdot \bar{u}_1}{\bar{u}_1 \cdot \bar{u}_1}\bar{u}_1 + \frac{\bar{y} \cdot \bar{u}_2}{\bar{u}_2 \cdot \bar{u}_2}\bar{u}_2 \\ &= \frac{15}{30}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} + \frac{-21}{6}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{5}{2} - \frac{7}{2} \\ -1 - \frac{21}{3} \\ \frac{1}{2} + \frac{21}{6} \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}. \end{aligned}$$

Hence

$$\begin{aligned} \bar{y} - \hat{y} &= \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix} \end{aligned}$$

so that

$$\begin{aligned} \|\bar{y} - \hat{y}\| &= \sqrt{3^2 + 6^2} \\ &= \sqrt{45} \\ &= 3\sqrt{5}. \end{aligned}$$

Theorem If $\{\bar{u}_1, \ldots, \bar{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W\bar{y} = (\bar{y} \cdot \bar{u}_1)\bar{u}_1 + \cdots + (\bar{y} \cdot \bar{u}_p)\bar{u}_p,$$

and if $U = \begin{bmatrix} \bar{u}_1 & \bar{u}_2 & \ldots & \bar{u}_p \end{bmatrix}$, then

$$\text{proj}_W\bar{y} = UU^T\bar{y}$$

for all $\bar{y} \in \mathbb{R}^n$.

Proof The first equation follows because all of the denominators $\bar{u}_i \cdot \bar{u}_i$ are equal to 1 since the basis is ortho*normal*. The second equation follows because $\bar{y} \cdot \bar{u}_i = \bar{u}_i^T\bar{y}$ for all i.