Lecture 9: Orthogonality

Consider a vector space $V$ and a non-degenerate positive definite scalar product $\langle \cdot, \cdot \rangle$ on $V$. Two vectors $u, v \in V$ are said to be orthogonal relative to $\langle \cdot, \cdot \rangle$ if

$$\langle u, v \rangle = 0. \quad (1)$$

**Examples:**

- $V = \mathbb{R}^2$. The following vectors

  $$u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad (2)$$

  are orthogonal in $\mathbb{R}^2$ relative to the standard inner product

  $$\langle u, v \rangle = \sum_{i=1}^{2} u_i v_i = -1 + 1 = 0. \quad (3)$$

- $V = M_{2 \times 2}(\mathbb{R})$. The following matrices

  $$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad (4)$$

  are orthogonal in $M_{2 \times 2}(\mathbb{R})$ relative to the inner product

  $$\langle A, B \rangle = \text{Tr} \left( AB^T \right). \quad (5)$$

  In fact,

  $$\text{Tr} \left( AB^T \right) = \text{Tr} \left( \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) = \text{Tr} \left( \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right) = 0. \quad (6)$$

- $V = \mathbb{P}_2([-1, 1])$ (vector space of polynomials of degree at most two). The polynomials

  $$p_1(x) = x \quad \text{and} \quad p_2(x) = \frac{3}{2}x^2 - \frac{1}{2} \quad (7)$$

  are orthogonal with respect to the scalar product

  $$\langle p_1, p_2 \rangle = \int_{-1}^{1} p_1(x)p_2(x)dx. \quad (8)$$

  In fact,

  $$\langle p_1, p_2 \rangle = \int_{-1}^{1} p_1(x)p_2(x)dx = \int_{-1}^{1} \left( \frac{3}{2}x^3 - \frac{1}{2}x \right) dx = \left[ \frac{3}{8}x^4 - \frac{1}{4}x^2 \right]_{-1}^{1} = 0. \quad (9)$$
Orthogonal projections. Consider two vectors \( u \) and \( v \) in a vector space \( V \). The orthogonal projection of \( u \) onto \( v \) is defined as

\[
P_vu = \frac{\langle u, v \rangle}{\langle v, v \rangle} v = \frac{u}{\|v\|} \frac{v}{\|v\|},
\]

where \( \| \cdot \| \) is the norm induced by the scalar product. Clearly \( v/\|v\| \) is a vector with norm equal to one, i.e., a unit vector.

Examples:

- Let \( V = \mathbb{R}^2 \) and consider the following vectors
  \[
u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.
\]

  The projection of \( u \) onto \( v \) is
  \[
P_vu = \frac{\langle u, v \rangle}{\langle v, v \rangle} v = \frac{2}{4} \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.
\]

  Note that if we subtract \( \frac{\langle u, v \rangle}{\langle v, v \rangle} v \) from \( u \) we obtain a vector that is orthogonal to \( v \).

  \[
u - \frac{\langle u, v \rangle}{\langle v, v \rangle} v = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.
\]

- \( V = \mathbb{R}^3 \). Given three vectors \( v_1, v_2, v_3 \in \mathbb{R}^3 \) we can compute the orthogonal projection of any vector onto any other vector, e.g., the projection of \( v_2 \) onto \( v_1 \)

  \[
P_{v_2v_1} = \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1.
\]

  We can also compute the orthogonal projection one vector onto the span of the two vectors. For instance, the orthogonal projection of the vector \( v_2 \) onto the plane spanned by \( v_1 \) and \( v_3 \) (assuming \( v_1 \) and \( v_3 \) are orthogonal) is the vector

  \[
P_{Sv_2} = \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 + \frac{\langle v_2, v_3 \rangle}{\langle v_3, v_3 \rangle} v_3.
\]
This equation defines the orthogonal projection of a vector the plane \( S = \text{span}\{v_1, v_3\} \). If the vectors \( v_1 \) and \( v_3 \) are not orthogonal it is still possible to determine the projection of \( v_2 \) onto the plane. To this end it is sufficient to first calculate the component of \( v_2 \) normal to the plane by computing the orthogonal component of \( v_2 \) relative to \( v_1 \) and then computing the orthogonal component of the resulting vector relative to \( v_3 \).

We can also construct an orthogonal set of vector by transforming the given set of linearly independent vectors \( \{v_1, v_2, v_3\} \) as follows

\[
\begin{align*}
    u_1 &= v_1, \\
    u_2 &= v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1, \\
    u_3 &= v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2.
\end{align*}
\]

This procedure is known as Gram-Schmidt orthogonalization, and it allows us to transform any set of linearly independent vectors into an orthogonal one. Such set of orthogonal vectors can be then normalized.

The previous example suggests that we can transform any basis \( \{v_1, \ldots, v_n\} \) of a \( n \)-dimensional vector space \( V \) into an orthonormal basis\(^1\) by using the Gram-Schmidt procedure. In fact, we can first compute

\[
\begin{align*}
    u_1 &= v_1, \\
    u_2 &= v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1, \\
    u_3 &= v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2, \\
    & \vdots \\
    u_n &= v_n - \frac{\langle v_n, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \cdots - \frac{\langle v_n, u_{n-1} \rangle}{\langle u_{n-1}, u_{n-1} \rangle} u_{n-1}.
\end{align*}
\]

and then normalize the vectors \( \{u_1, \ldots, u_n\} \) to obtain the orthonormal basis

\[
\left\{ \frac{u_1}{\|u_1\|}, \ldots, \frac{u_n}{\|u_n\|} \right\}.
\]

Alternatively, we can normalize each vector \( u_i \) right after we compute it. This reduces the number of calculations in the Gram-Schmidt procedure as we can write

\[
\begin{align*}
    u_1 &= v_1, \\
    \hat{u}_2 &= \frac{u_2}{\|u_2\|}, \\
    \hat{u}_3 &= \frac{u_3}{\|u_3\|}, \\
    & \vdots
\end{align*}
\]

\(^1\)Note that the orthogonal basis we obtain from the Gram-Schmidt procedure is not unique. In fact a reordering of the vectors \( \{v_1, \ldots, v_n\} \) yields a different orthogonal basis at the end of the procedure.
It is straightforward to show that
\[ \langle u_i, u_j \rangle = \delta_{ij} \| u_j \|^2, \] (17)
where \( \delta_{ij} \) is the Kronecker delta function\(^2\). For example,
\[ \langle u_1, u_2 \rangle = \left( v_1, v_2 - \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 \right) = \langle v_1, v_2 \rangle - \frac{\langle v_2, v_1 \rangle}{\langle v_1, v_1 \rangle} \langle v_1, v_1 \rangle = 0. \] (19)

**Example:** Let us use the Gram-Schmidt procedure to orthogonalize the following vectors in \( \mathbb{R}^2 \)
\[ v_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \]

We have
\[ u_1 = v_1, \quad u_2 = v_2 - \frac{\langle v_2, u_1 \rangle}{\| u_1 \|^2} u_1. \]

The norm of \( u_1 = v_1 \) is
\[ \| u_1 \|^2 = \| v_1 \|^2 = \langle v_1, v_1 \rangle = 2^2 + 1^2 = 5. \] (20)
This implies that
\[ u_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{4}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 - \frac{8}{5} \\ 2 - \frac{4}{5} \end{bmatrix} = \begin{bmatrix} -\frac{3}{5} \\ \frac{6}{5} \end{bmatrix} \]

Note that \( u_1 \) and \( u_2 \) are orthogonal. In fact,
\[ \langle u_1, u_2 \rangle = 2 \times \left( -\frac{3}{5} \right) + \frac{6}{5} = 0. \] (21)

The norm of \( u_2 \) is
\[ \| u_2 \| = \sqrt{\langle u_2, u_2 \rangle} = \sqrt{\frac{9}{25} + \frac{36}{25}} = \frac{3\sqrt{5}}{5}. \] (22)

This means that
\[ \begin{bmatrix} u_1 \\ \| u_1 \|^2 \| u_2 \| \end{bmatrix} = \begin{bmatrix} \frac{\sqrt{5}}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \\ \frac{\sqrt{5}}{5} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \end{bmatrix} \] (23)

\(^2\)The Kronecker delta is defined as:
\[ \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \] (18)
is an orthonormal basis of \( \mathbb{R}^2 \). Note that \( u_2/\|u_2\| \) can be obtained by rotating \( u_1/\|u_1\| \) by 90 degrees counterclockwise.

**Representation of vectors relative to orthonormal bases.** Let \( B_V = \{\hat{u}_1, \ldots, \hat{u}_n\} \) be an orthonormal basis of a \( n \)-dimensional vector space \( V \). Any vector \( v \in V \) can be represented relative to the basis \( B_V \) as

\[
v = x_1\hat{u}_1 + \cdots + x_n\hat{u}_n. \tag{24}\]

by projecting the vector \( v \) onto \( \hat{u}_i \) and taking into account the orthonormality conditions \( \langle u_i, u_j \rangle = \delta_{ij} \) yields

\[
\langle v, \hat{u}_j \rangle = \langle x_1\hat{u}_1 + \cdots + x_n\hat{u}_n, \hat{u}_j \rangle \\
= x_1\langle \hat{u}_1, \hat{u}_j \rangle + \cdots + x_n\langle \hat{u}_n, \hat{u}_j \rangle \\
= x_j\langle \hat{u}_j, \hat{u}_j \rangle \\
= x_j \tag{25}
\]

i.e., the \( j \)-the coordinate of \( v \) relative to \( B_V \) coincides with the projection of \( v \) onto \( \hat{u}_j \). On the other hand, if we consider an orthogonal basis \( \{u_1, \ldots, u_n\} \) we obtain

\[
v = y_1u_1 + \cdots + y_nu_n \Rightarrow x_j = \frac{\langle v, u_j \rangle}{\langle u_j, u_j \rangle}. \tag{26}\]

**Theorem 1.** Let \( B_V = \{\hat{u}_1, \ldots, \hat{u}_n\} \) be an orthonormal basis of a \( n \)-dimensional vector space \( V \). Then for any vector \( v \in V \) we have

\[
v = x_1\hat{u}_1 + \cdots + x_n\hat{u}_n \quad \text{and} \quad \|v\|^2 = \sum_{k=1}^{n} x_k^2. \tag{27}\]

**Orthogonal spaces.** Let \( S \) be a subspace of \( V \). We define

\[
S^\perp = \{v \in V : \langle v, w \rangle = 0 \quad \text{for all} \quad w \in S\}
\]

The orthogonal complement of \( S \) in \( V \). It can be shown that \( S^\perp \) is a vector subspace of \( V \). Any vector \( v \in V \) can be expressed as a sum of two vectors \( w_1 \in S \) and \( w_2 \in S^\perp \), i.e.,

\[
v = w_1 + w_2 \tag{28}\]

Equivalently, we say that \( V \) is the direct sum of \( S \) and \( S^\perp \), and write

\[
V = S \oplus S^\perp. \tag{29}\]

For example, any vector in \( V = \mathbb{R}^3 \) defines a one-dimensional vector subspace \( S \). The orthogonal complement of \( S \) in \( \mathbb{R}^3 \) is a plane orthogonal \( S \). Such plane is denoted by \( S^\perp \) in the following figure:
Next consider a $n \times n$ matrix $A$ and let \( \{v_1, \ldots, v_n\} \) be the columns of $A$. Denote by

$$\text{Col}(A) = \text{span}\{v_1, \ldots, v_n\}$$

the column space of $A$. We known that such space is a vector subspace of $\mathbb{R}^n$. The orthogonal complement of $\text{Col}(A)$ is

$$\text{Col}^\perp(A) = \{v \in \mathbb{R}^n : \langle v, w \rangle = 0 \text{ for all } w \in \text{Col}(A)\}.$$  

(31)

Let us write the condition $\langle v, w \rangle = 0$ more explicitly. An arbitrary vector in the column space $\text{Col}(A)$ can be written as

$$w = y_1 v_1 + \cdots + x_n v_n$$

(32)

Similarly, a vector $v \in \mathbb{R}^n$ can be written relative to the canonical basis as

$$v = x_1 e_1 + \cdots + x_n e_n$$

(33)

Taking the scalar product yields

$$\sum_{i,j=1}^{n} \langle e_i, v_j \rangle x_i y_j = \sum_{i,j=1}^{n} y_j A_{ji}^T x_i = 0$$

(34)

This can be written as

$$y^T (A^T x) = 0 \quad \text{where we defined} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$ 

(35)

Since the vector $y$ is arbitrary we obtain that $A^T x = 0$. This means that

$$\text{Col}^\perp(A) = N(A^T).$$

(36)

Similarly, it is straightforward to prove that

$$N^\perp(A) = \text{Col}(A^T)$$

(37)

In general, if $S$ is any vector subspace of $V$ then

$$V = S \oplus S^\perp \quad \text{which implies} \quad \text{dim}(V) = \text{dim}(S) + \text{dim}(S^\perp).$$

(38)