Chapter 5: Orthogonality

April 30, 2009

Week 13-14

1 Inner product

Geometric concepts of length, distance, angle, and orthogonality, which are well-known in $\mathbb{R}^2$ and $\mathbb{R}^3$, can be defined in $\mathbb{R}^n$. These concepts provide powerful geometric tools to solve many applied problems such as the least-squares problem.

Given two straight lines $\ell_1 : y = a_1 x$ and $\ell_2 : y = a_2 x$. We know that ℓ_1 and ℓ_2 are perpendicular, written $\ell_1 \perp \ell_2$, if and only if $a_1 a_2 = -1$. Note that the directions of the straight lines ℓ_1 and ℓ_2 can be described respectively by the nonzero vectors

$$\boldsymbol{u} = \begin{bmatrix} 1 \\ a_1 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} = \begin{bmatrix} 1 \\ a_2 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}.$$

Then $a_1 a_2 = -1$ is equivalent to $1 + a_1 a_2 = 0$. Let $\boldsymbol{u} \cdot \boldsymbol{v} := u_1 v_1 + u_2 v_2$. Thus

$$\ell_1 \perp \ell_2 \Longleftrightarrow \boldsymbol{u} \cdot \boldsymbol{v} = 0.$$

The **inner product** (or **dot product**) of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ is the number

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle := \boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + \cdots + u_n v_n,$$

where

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}.$$

The transpose of $\boldsymbol{u}$ is the row vector $\boldsymbol{u}^T = [u_1, \ldots, u_n]$. The matrix product $\boldsymbol{u}^T \boldsymbol{v}$ is a 1×1 matrix, and

$$\boldsymbol{u}^T \boldsymbol{v} = [\boldsymbol{u} \cdot \boldsymbol{v}].$$

If we identify any 1×1 matrix $[c]$ to its entry c, the inner product can be written as the matrix multiplication

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

Proposition 1.1. *For vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ in $\mathbb{R}^n$ and scalar c,*

(1) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$,

(2) $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$,

(3) $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v})$,

(4) $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$ *and* $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ *if and only if* $\boldsymbol{u} = \mathbf{0}$.

The **length** (or **norm**) of a vector $\boldsymbol{v}$ in $\mathbb{R}^n$ is the nonnegative number

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

It is clear that for any vector $\boldsymbol{v}$ in $\mathbb{R}^n$ and any scalar c,

$$\|c\boldsymbol{v}\| = |c| \, \|\boldsymbol{v}\|.$$

The **distance** $d(\boldsymbol{u}, \boldsymbol{v})$ between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ is the length of the vector $\boldsymbol{u} - \boldsymbol{v}$, i.e.,

$$d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|.$$

Theorem 1.2. *Two vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in* $\mathbb{R}^n$ *are orthogonal if and only if*

$$\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2. \tag{1.1}$$

Proof. By linearity of inner product, we have

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

It is clear that (1.1) is valid if and only if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$. □

The vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are called **orthogonal** if

$$\boldsymbol{u} \cdot \boldsymbol{v} = 0.$$

Lemma 1.3. *For nonzero vectors* $\boldsymbol{u}, \boldsymbol{v}$ *of* $\mathbb{R}^n$,

$$-1 \leq \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|} \leq 1.$$

Proof. Consider the vectors $\boldsymbol{w}(t) = \boldsymbol{u} + t\boldsymbol{v}$, where t is a real variable. Then

$$y(t) := (\boldsymbol{u} + t\boldsymbol{v}) \cdot (\boldsymbol{u} + t\boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{u} + 2t \, \boldsymbol{u} \cdot \boldsymbol{v} + t^2 \boldsymbol{v} \cdot \boldsymbol{v} \geq 0.$$

Thus quadratic function $y = y(t)$ is above the t-axis, and the equation $\boldsymbol{u} \cdot \boldsymbol{u}) + 2t\boldsymbol{u} \cdot \boldsymbol{v} + t^2 \boldsymbol{v} \cdot \boldsymbol{v} = 0$ has at mots one root. Therefore the discriminant

$$\Delta := b^2 - 4ac = 4(\boldsymbol{u} \cdot \boldsymbol{v})^2 - 4(\boldsymbol{u} \cdot \boldsymbol{u})(\boldsymbol{v} \cdot \boldsymbol{v}) \leq 0.$$

This inequality is equivalent to $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$. □

The **angle** θ between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ is defined by

$$\cos\theta = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\|\boldsymbol{u}\| \, \|\boldsymbol{v}\|}.$$

Let S be a nonempty subset of $\mathbb{R}^n$. A vector $\boldsymbol{z}$ is said to be **orthogonal** (or **perpendicular**) to S, written

$$\boldsymbol{z} \perp S,$$

if $\boldsymbol{z}$ is orthogonal to every vector in S, that is, $\langle \boldsymbol{z}, \boldsymbol{v} \rangle = 0$ for all $\boldsymbol{v} \in S$. The set of all vectors that are orthogonal to S is a subspace of $\mathbb{R}^n$, and is denoted by $S^\perp$. If W is a subspace of $\mathbb{R}^n$, the subspace $W^\perp$ is called the **orthogonal complement** of W in $\mathbb{R}^n$.

Proposition 1.4. *Let S be a nonempty subset of $\mathbb{R}^n$ and $W = \operatorname{Span} S$. Then*

(a) *$S^\perp$ is a subspace of $\mathbb{R}^n$.*

(b) *If $\boldsymbol{z}$ is orthogonal to S, then $\boldsymbol{z}$ is orthogonal to W, i.e.,*

$$S^\perp = W^\perp.$$

Proof. (1) Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be a vectors in $S^\perp$. Then for any $\boldsymbol{w}$ in S we have $\boldsymbol{u} \cdot \boldsymbol{w} = \boldsymbol{v} \cdot \boldsymbol{w} = 0$. Thus for any scalars c and d,

$$(c\boldsymbol{u} + d\boldsymbol{v}) \cdot \boldsymbol{w} = c\boldsymbol{u} \cdot \boldsymbol{w} + d\boldsymbol{v} \cdot \boldsymbol{w} = 0.$$

This means that $S^\perp$ is a subspace of $\mathbb{R}^n$.

(2) It clear that $W^\perp$ is contained in $S^\perp$, i.e., $W \subseteq S^\perp$, since $S \subseteq W$. Let $\boldsymbol{u}$ be a vector in $S^\perp$. For any $\boldsymbol{w}$ in $\operatorname{Span} S$ there exist vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ in S such that $\boldsymbol{w} = c_1\boldsymbol{w}_1 + \cdots + c_k\boldsymbol{w}_k$. Since $\boldsymbol{u} \cdot \boldsymbol{w}_1 = 0$, …, $\boldsymbol{u} \cdot \boldsymbol{w}_k = 0$, then $\boldsymbol{u} \cdot \boldsymbol{w} = c_1\boldsymbol{u} \cdot \boldsymbol{w}_1 + \cdots + c_k\boldsymbol{u} \cdot \boldsymbol{w}_k = 0$, i.e., $\boldsymbol{u} \perp \boldsymbol{w}$. This means that $\boldsymbol{u}$ is a vector in $W^\perp$. Therefore $S^\perp = W^\perp$. □

Example 1.1. Let $S = \{\boldsymbol{a}_1, \ldots \boldsymbol{a}_m\}$ be the set of row vectors of an $m \times n$ matrix A. Then the set of vectors that are perpendicular to S is the null space $\operatorname{Nul} A$, that is,

$$(\operatorname{Nul} A)^\perp = \operatorname{Row} A,$$

$$\dim \operatorname{Row} A + \dim \operatorname{Nul} A = n.$$

A set $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ of nonzero vectors in $\mathbb{R}^n$ is called an **orthogonal set** if every pair of distinct vectors from the set is orthogonal, i.e.,

$$\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0 \quad \text{for all} \quad i \neq j;$$

furthermore, if $\|\boldsymbol{u}_1\| = \cdots = \|\boldsymbol{u}_k\| = 1$, then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ is called an **orthonormal set**.

Example 1.2. The following set

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

is an orthogonal set of $\mathbb{R}^4$ but is not an orthonormal set.

Theorem 1.5. *Any orthogonal set $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ of nonzero vectors is a linear independent set and hence is a basis for* $\operatorname{Span} S$.

Proof. Consider

$$\boldsymbol{v} = c_1\boldsymbol{v}_1 + \cdots + c_p\boldsymbol{v}_p = 0.$$

Then for all $1 \leq i \leq p$,

$$\boldsymbol{v}_i \cdot \boldsymbol{v} = c_i\boldsymbol{v}_i \cdot \boldsymbol{v}_i = c_i\|\boldsymbol{v}_i\|^2 = 0$$

Since $\|\boldsymbol{v}_i\| \neq 0$, we must have $c_i = 0$. This means that $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ is a linearly independent set. □

Let W be a nonzero subspace of $\mathbb{R}^n$. A basis of W is called an **orthogonal basis** if it is an orthogonal set; if every vector of an orthogonal basis is a unit vector, the basis is called an **orthonormal basis**.

Example 1.3. The following set

$$\mathcal{B} = \left\{ \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \frac{1}{\sqrt{6}} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$$

is an orthonormal basis for $\mathbb{R}^3$.

Theorem 1.6. *Let $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. Then every vector $\boldsymbol{y}$ in W can be expressed as*

$$\boldsymbol{y} = \left(\frac{\boldsymbol{v}_1 \cdot \boldsymbol{y}}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \right) \boldsymbol{v}_1 + \cdots + \left(\frac{\boldsymbol{v}_p \cdot \boldsymbol{y}}{\boldsymbol{v}_p \cdot \boldsymbol{v}_p} \right) \boldsymbol{v}_p.$$

If $\mathcal{B}$ is an orthonormal basis, then

$$\boldsymbol{y} = (\boldsymbol{v}_1 \cdot \boldsymbol{y})\boldsymbol{v}_1 + \cdots + (\boldsymbol{v}_p \cdot \boldsymbol{y})\boldsymbol{v}_p.$$

Proof. Let us write $\boldsymbol{y}$ as the linear combination

$$\boldsymbol{y} = \alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_p \boldsymbol{v}_p.$$

Then

$$\boldsymbol{v}_i \cdot \boldsymbol{y} = \alpha_1 \boldsymbol{v}_i \cdot \boldsymbol{v}_1 + \cdots + \alpha_i \boldsymbol{v}_i \cdot \boldsymbol{v}_i + \cdots + \alpha_p \boldsymbol{v}_i \cdot \boldsymbol{v}_p = \alpha_i \boldsymbol{v}_i \cdot \boldsymbol{v}_i, \quad 1 \le i \le p.$$

Thus

$$\alpha_i = \frac{\boldsymbol{v}_i \cdot \boldsymbol{y}}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i}, \quad 1 \le i \le p.$$

□

2 Orthogonal projection through a single vector

Let $\boldsymbol{v}$ be a nonzero vector in $\mathbb{R}^n$. Then any vector $\boldsymbol{y}$ in $\mathbb{R}^n$ can be decomposed into the form

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

where $\hat{\boldsymbol{y}}$ is parallel to $\boldsymbol{v}$ and $\boldsymbol{z}$ is orthogonal to $\boldsymbol{v}$. Since $\hat{\boldsymbol{y}}$ is parallel to $\boldsymbol{v}$, there is a scalar α such that

$$\hat{\boldsymbol{y}} = \alpha \boldsymbol{v}.$$

Then

$$\boldsymbol{z} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \alpha \boldsymbol{v}.$$

Since $\boldsymbol{z}$ is orthogonal to $\boldsymbol{v}$, we have

$$0 = \boldsymbol{v} \cdot \boldsymbol{z} = \boldsymbol{v} \cdot (\boldsymbol{y} - \alpha \boldsymbol{v}) = \boldsymbol{v} \cdot \boldsymbol{y} - \alpha \boldsymbol{v} \cdot \boldsymbol{v}.$$

Thus

$$\alpha = \frac{\boldsymbol{v} \cdot \boldsymbol{y}}{\boldsymbol{v} \cdot \boldsymbol{v}}.$$

Let us write

$$\hat{\boldsymbol{y}} = \mathrm{Proj}_{\boldsymbol{v}}(\boldsymbol{y}) = \left(\frac{\boldsymbol{v} \cdot \boldsymbol{y}}{\boldsymbol{v} \cdot \boldsymbol{v}} \right) \boldsymbol{v}, \quad \boldsymbol{y} \in \mathbb{R}^n;$$

the vector $\hat{\boldsymbol{y}}$ is called the **orthogonal projection** of $\boldsymbol{y}$ onto the direction $\boldsymbol{v}$. The projection

$$\mathrm{Proj}_{\boldsymbol{v}} : \mathbb{R}^n \to \mathbb{R}^n$$

is a linear transformation, called the **orthogonal projection** from $\mathbb{R}^n$ onto the line Span $\{\boldsymbol{v}\}$.

Note that $[\boldsymbol{v}\cdot\boldsymbol{y}] = \boldsymbol{v}^T\boldsymbol{y}$, and for any scalar c and vector $\boldsymbol{v}$,

$$c\boldsymbol{v} = c\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ \vdots \\ cv_n \end{bmatrix} = \begin{bmatrix} v_1 c \\ \vdots \\ v_n c \end{bmatrix} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}[c] = \boldsymbol{v}[c],$$

where $[c]$ is the 1×1 matrix with the only entry c. Then the orthogonal projection $\mathrm{Proj}_{\boldsymbol{v}}$ can be written as

$$\mathrm{Proj}_{\boldsymbol{v}}(\boldsymbol{y}) = \left(\frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)(\boldsymbol{v}\cdot\boldsymbol{y})\boldsymbol{v} = \left(\frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v}[\boldsymbol{v}\cdot\boldsymbol{y}] = \left(\frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v}\boldsymbol{v}^T\boldsymbol{y}.$$

This means that the standard matrix of $\mathrm{Proj}_{\boldsymbol{v}}$ is

$$\left(\frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v}\boldsymbol{v}^T.$$

Indeed, $\boldsymbol{v}$ is an $n\times 1$ matrix and $\boldsymbol{v}^T$ is a $1\times n$ matrix, the product $\boldsymbol{v}\boldsymbol{v}^T$ is an $n\times n$ matrix.

It follows that the orthogonal projection $\mathrm{Proj}_{\boldsymbol{v}^\perp} : \mathbb{R}^n \to \mathbb{R}^n$ is given by

$$\boldsymbol{z} = \mathrm{Proj}_{\boldsymbol{v}^\perp}(\boldsymbol{y}) = \boldsymbol{y} - \mathrm{Proj}_{\boldsymbol{v}}(\boldsymbol{y}) = \left(I - \frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}\boldsymbol{v}^T\right)\boldsymbol{y}.$$

This means that the standard matrix of $\mathrm{Proj}_{\boldsymbol{v}^\perp}$ is

$$I - \left(\frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v}\boldsymbol{v}^T.$$

The vector $\boldsymbol{z}$ is called the **complement of $\boldsymbol{y}$ orthogonal to $\boldsymbol{v}$.**

Example 2.1. Find the orthogonal projection $\mathrm{Proj}_b v : \mathbb{R}^3 \to \mathbb{R}^3$ through the vector $\boldsymbol{v} = [1,1,1]^T$.

$$\begin{aligned} \mathrm{Proj}_{\boldsymbol{v}}(\boldsymbol{y}) &= \left(\frac{\boldsymbol{y}\cdot\boldsymbol{v}}{\boldsymbol{v}\cdot\boldsymbol{v}}\right)\boldsymbol{v} = \frac{y_1+y_2+y_3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ &= \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}. \end{aligned}$$

The matrix A of $\mathrm{Proj}_{\boldsymbol{v}}$ can be found in a different way as

$$A = \frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}\boldsymbol{v}^T = \frac{1}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}[1,1,1] = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Example 2.2. Find the linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$, i.e., the orthogonal projection from $\mathbb{R}^3$ to the plane $x_1+x_2+x_3 = 0$. This is to find the orthogonal projection from $\mathbb{R}^3$ to the subspace $\boldsymbol{v}^\perp$.

$$\begin{aligned} \mathrm{Proj}_{\boldsymbol{v}^\perp}\boldsymbol{y} &= \boldsymbol{y} - \mathrm{Proj}_{\boldsymbol{v}}(\boldsymbol{y}) = \left(I - \frac{1}{\boldsymbol{v}\cdot\boldsymbol{v}}\boldsymbol{v}\boldsymbol{v}^T\right)\boldsymbol{y} \\ &= \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}. \end{aligned}$$

Theorem 2.1. *Let U be an $m \times n$ matrix. Then the column vectors of U are orthonormal if and only if*

$$U^T U = I_n.$$

Proof. Write $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$, where $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n \in \mathbb{R}^m$. Then $\boldsymbol{u}_i^T \boldsymbol{u}_i = 1$ for all $1 \le i \le n$, and

$$\boldsymbol{u}_i^T \boldsymbol{u}_j = \boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0 \quad \text{for all} \quad i \neq j.$$

Thus the $n \times n$ matrix

$$U^T U = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_n^T \end{bmatrix} [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n] = \begin{bmatrix} \boldsymbol{u}_1^T\boldsymbol{u}_1 & \boldsymbol{u}_1^T\boldsymbol{u}_2 & \cdots & \boldsymbol{u}_1^T\boldsymbol{u}_n \\ \boldsymbol{u}_2^T\boldsymbol{u}_1 & \boldsymbol{u}_2^T\boldsymbol{u}_2 & \cdots & \boldsymbol{u}_2^T\boldsymbol{u}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_n^T\boldsymbol{u}_1 & \boldsymbol{u}_n^T\boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n^T\boldsymbol{u}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}_{n\times n}$$

□

Theorem 2.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by $T(\boldsymbol{x}) = U\boldsymbol{x}$. If the column vectors of U are orthonormal, then for any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^k$,*

(1) $\|T\boldsymbol{u}\| = \|\boldsymbol{u}\|$.

(2) $(T\boldsymbol{u}) \cdot (T\boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}$.

(3) $(T\boldsymbol{u}) \cdot (T\boldsymbol{v}) = 0$ *if and only if* $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

Proof. Note that $U^T U = I_n$. Then

$$(T\boldsymbol{u}) \cdot (T\boldsymbol{v}) = (U\boldsymbol{u}) \cdot (U\boldsymbol{v}) = (U\boldsymbol{u})^T U\boldsymbol{v} = \boldsymbol{u}^T U^T U \boldsymbol{v} = \boldsymbol{u}^T \boldsymbol{v} = \boldsymbol{u} \cdot \boldsymbol{v}.$$

□

Example 2.3. Find a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that the image of T is the plane $x_1 + x_2 + x_3 = 0$ and T preserves distance, that is,

$$d(T\boldsymbol{u}, T\boldsymbol{v}) = d(\boldsymbol{u}, \boldsymbol{v}) \quad \text{for} \quad \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^2.$$

It is clear that the vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

form an orthogonal basis for the plane $x_1 + x_2 + x_3 = 0$. The unit vectors

$$\boldsymbol{u}_1 = \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} \quad \text{and} \quad \boldsymbol{u}_2 = \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}$$

form an orthonormal basis. Thus T can be defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{2} & 1/\sqrt{6} \\ 0 & -2/\sqrt{6} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

For instance, for points $(1, 0)$ and $(1, 0)$, we have

$$T(1,0) = \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0 \right) \quad \text{and} \quad T(0,1) = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}} \right).$$

One easily verifies

$$d((1,0),(0,1)) = \sqrt{2} = d(T(1,0), T(0,1)).$$

Note that there are infinitely many such linear transformations.

Definition 2.3. An $n \times n$ matrix U is called an **orthogonal matrix** if

$$U^T U = I.$$

In other words, the column vectors of U form an orthonormal basis of $\mathbb{R}^n$.

Proposition 2.4. *Let A be an $n \times n$ matrix. Then A is orthogonal if and only if A^T is orthogonal. In other words, the column vectors of A are orthogonal if and only if the row vector vectors of A are orthonormal.*

Proof. The matrix A is orthogonal if and only if $A^T A = I$, that is, $A^{-1} = A^T$. Thus A is orthogonal if and only if $AA^T = I$, that is, $(A^T)^T A^T = I$. This means that the row vectors of A form an orthonormal basis for $\mathbb{R}^n$. $\square$

Note 1. If the column vectors of an $n \times n$ matrix A are orthogonal, then it is not necessary that the row vectors of A are orthogonal.

3 Orthogonal projection through a subspace

Given a subspace W of $\mathbb{R}^n$ and its orthogonal complement $W^\perp$; we wish to express every vector $\boldsymbol{y}$ in $\mathbb{R}^n$ as

$$\boldsymbol{y} = \boldsymbol{z}_1 + \boldsymbol{z}_2, \quad \text{where} \quad \boldsymbol{z}_1 \in W,\ \boldsymbol{z}_2 \in W^\perp.$$

If so, the vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are called the **orthogonal projections** of $\boldsymbol{y}$ onto W and $W^\perp$, respectively.

Theorem 3.1. *Let W be a nonzero subspace of $\mathbb{R}^n$. Then every vector $\boldsymbol{y}$ in $\mathbb{R}^n$ can be decomposed uniquely as*

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z}, \quad \textit{where} \quad \hat{\boldsymbol{y}} \in W,\ \boldsymbol{z} \in W^\perp.$$

Moreover, if $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$ is an orthogonal basis for W, then

$$\begin{aligned} \hat{\boldsymbol{y}} &= \mathrm{Proj}_W(\boldsymbol{y}) = \left(\frac{\boldsymbol{v}_1 \cdot \boldsymbol{y}}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\right)\boldsymbol{v}_1 + \left(\frac{\boldsymbol{v}_2 \cdot \boldsymbol{y}}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\right)\boldsymbol{v}_2 + \cdots + \left(\frac{\boldsymbol{v}_p \cdot \boldsymbol{y}}{\boldsymbol{v}_p \cdot \boldsymbol{v}_p}\right)\boldsymbol{v}_p & (3.1) \\ &= \left(\frac{1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1\boldsymbol{v}_1^T + \frac{1}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2\boldsymbol{v}_2^T + \cdots + \frac{1}{\boldsymbol{v}_p \cdot \boldsymbol{v}_p}\boldsymbol{v}_p\boldsymbol{v}_p^T\right)\boldsymbol{y}, & (3.2) \\ \boldsymbol{z} &= \boldsymbol{y} - \hat{\boldsymbol{y}}. & (3.3) \end{aligned}$$

In particular, if $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p\}$ is an orthonormal basis for W, then

$$\begin{aligned} \mathrm{Proj}_W(\boldsymbol{y}) &= (\boldsymbol{u}_1 \cdot \boldsymbol{y})\boldsymbol{u}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{y})\boldsymbol{u}_2 + \cdots + (\boldsymbol{u}_p \cdot \boldsymbol{y})\boldsymbol{u}_p & (3.4) \\ &= \left(\boldsymbol{u}_1\boldsymbol{u}_1^T + \boldsymbol{u}_2\boldsymbol{u}_2^T + \cdots + \boldsymbol{u}_p\boldsymbol{u}_p^T\right)\boldsymbol{y} & (3.5) \\ &= UU^T\boldsymbol{y}, & (3.6) \end{aligned}$$

where $U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p]$ is an $n \times p$ matrix.

Proof. Suppose there are two decompositions for a vector $\boldsymbol{y}$, say

$$\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z} = \hat{\boldsymbol{y}}_1 + \boldsymbol{z}_1. \tag{3.7}$$

Then

$$\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_1 = \boldsymbol{z}_1 - \boldsymbol{z}$$

is a vector in both W and $W^{\perp}$. Thus it is orthogonal to itself. Hence it must be the zero vector. This proves the uniqueness of decomposition. Let us write

$$\hat{\boldsymbol{y}} = \alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_p \boldsymbol{v}_p.$$

Since $\boldsymbol{z} = \boldsymbol{y} - \hat{\boldsymbol{y}}$ is a vector in $W^{\perp}$, we have

$$0 = \boldsymbol{v}_i \cdot \boldsymbol{z} = \boldsymbol{v}_i \cdot \boldsymbol{y} - \alpha_i \boldsymbol{v}_i \cdot \hat{\boldsymbol{y}} = \boldsymbol{v}_i \cdot \boldsymbol{y} - \alpha_i \boldsymbol{v}_i \cdot \boldsymbol{v}_i, \quad 1 \le i \le p.$$

Thus

$$\alpha_i = \frac{\boldsymbol{v}_i \cdot \boldsymbol{y}}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i}, \quad 1 \le i \le p.$$

This shows the existence of $\hat{\boldsymbol{y}}$ and $\boldsymbol{z}$.

If $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p\}$ is an orthonormal basis, then $\boldsymbol{u}_i \cdot \boldsymbol{u}_i = 1$ for all $1 \le i \le p$. Thus

$$\hat{\boldsymbol{y}} = (\boldsymbol{u}_1 \cdot \boldsymbol{y})\boldsymbol{u}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{y})\boldsymbol{u}_2 + \cdots + (\boldsymbol{u}_p \cdot \boldsymbol{y})\boldsymbol{u}_p.$$

Since

$$U^T \boldsymbol{y} = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_p^T \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} \boldsymbol{u}_1^T \boldsymbol{y} \\ \boldsymbol{u}_2^T \boldsymbol{y} \\ \vdots \\ \boldsymbol{u}_p^T \boldsymbol{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{y} \\ \boldsymbol{u}_2 \cdot \boldsymbol{y} \\ \vdots \\ \boldsymbol{u}_p \cdot \boldsymbol{y} \end{bmatrix}.$$

Hence

$$UU^T \boldsymbol{y} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p] \begin{bmatrix} \boldsymbol{u}_1 \cdot \boldsymbol{y} \\ \boldsymbol{u}_2 \cdot \boldsymbol{y} \\ \vdots \\ \boldsymbol{u}_p \cdot \boldsymbol{y} \end{bmatrix} = (\boldsymbol{u}_1 \cdot \boldsymbol{y})\boldsymbol{u}_1 + (\boldsymbol{u}_2 \cdot \boldsymbol{y})\boldsymbol{u}_2 + \cdots + (\boldsymbol{u}_p \cdot \boldsymbol{y})\boldsymbol{u}_p = \hat{\boldsymbol{y}}.$$

$\square$

Example 3.1. Find the orthogonal projection $\mathrm{Proj}_W : \mathbb{R}^3 \to \mathbb{R}^3$, where W is the plane in $\mathbb{R}^3$ given by $x_1 + x_2 + x_3 = 0$.

Solution. We need an orthogonal basis for W. By inspection, the vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

form an orthogonal basis for W. Then

$$\begin{aligned} \mathrm{Proj}_W(\boldsymbol{y}) &= \left(\frac{\boldsymbol{v}_1 \cdot \boldsymbol{y}}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \right) \boldsymbol{v}_1 + \left(\frac{\boldsymbol{v}_2 \cdot \boldsymbol{y}}{\boldsymbol{v}_2 \cdot \boldsymbol{v} 2} \right) \boldsymbol{v}_2 \\ &= \frac{y_1 - y_2}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} + \frac{y_1 + y_2 - 2y_3}{6} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \\ &= \begin{bmatrix} (2/3)y_1 - (1/3)y_2 - (1/3)y_3 \\ -(1/3)y_1 + (2/3)y_2 - (1/3)y_3 \\ -(1/3)y_1 - (1/3)y_2 + (2/3)y_3 \end{bmatrix} \\ &= \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \end{aligned}$$

which confirms the formula found in the previous section in a different way.

Example 3.2. Let $W = \text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ be a subspace of $\mathbb{R}^3$, where

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Find the matrix of the orthogonal projection $\text{Proj}_W : \mathbb{R}^3 \to \mathbb{R}^3$.

Solution. The vectors

$$\boldsymbol{u}_1 = \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix} \quad \text{and} \quad \boldsymbol{u}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$$

form an orthonormal basis of W. The standard matrix of the orthogonal projection Proj_W is

$$\begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} = \begin{bmatrix} 5/6 & -1/6 & 1/3 \\ -1/6 & 5/6 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}.$$

An alternative way to compute the orthogonal projection is

$$\begin{aligned} \text{Proj}_W(\boldsymbol{y}) &= \frac{y_1 + y_2 + y_3}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{y_1 - y_2}{2} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 5y_1 - y_2 + 2y_3 \\ -y_1 + 5y_2 + 2y_3 \\ 2y_1 + 2y_2 + 2y_3 \end{bmatrix} \\ &= \begin{bmatrix} 5/6 & -1/6 & 1/3 \\ -1/6 & 5/6 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}. \end{aligned}$$

Proposition 3.2. *Let W be a subspace of $\mathbb{R}^n$. Let $\hat{\boldsymbol{y}}$ be the orthogonal projection of a vector $\boldsymbol{y} \in \mathbb{R}^n$ in W. Then $\hat{\boldsymbol{y}}$ is the closest vector in W to $\boldsymbol{y}$, that is,*

$$\|\boldsymbol{y} - \hat{\boldsymbol{y}}\| < \|\boldsymbol{y} - \boldsymbol{w}\| \quad \textit{for all} \quad \boldsymbol{w} \in W, \ \boldsymbol{w} \neq \hat{\boldsymbol{y}}.$$

Proof. Since $\boldsymbol{y} - \hat{\boldsymbol{y}} \in W^\perp$ and $\hat{\boldsymbol{y}} - \boldsymbol{w} \in W$, the vectors $\boldsymbol{y} - \hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{y}} - \boldsymbol{w}$ are orthogonal. Then

$$\|\boldsymbol{y} - \boldsymbol{w}\|^2 = \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 + \|\hat{\boldsymbol{y}} - \boldsymbol{w}\|^2 > \|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 \quad \text{for all} \quad \boldsymbol{w} \neq \hat{\boldsymbol{y}}.$$

□

Example 3.3. Find the shortest distance from the point $(1, 2, 1, -3)$ in $\mathbb{R}^4$ to the plane defined by the equations $x_1 + x_2 - x_4 = 0$ and $x_2 - x_3 + x_4 = 0$.

Solution. Let W be the subspace defined by the given two equations. Then the orthogonal complement $W^\perp$ of W is the span of the two normal vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}.$$

Let $\boldsymbol{b}^T = [1, 2, 1, -3]$. Then the distance to be computed is to find the length of the orthogonal projection $\mathrm{Proj}_{W^\perp}\boldsymbol{b}$. Since

$$\mathrm{Proj}_{W^\perp}\boldsymbol{b} \;=\; \frac{6}{3}\begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix} - \frac{3}{3}\begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ -3 \end{bmatrix},$$

we have $d(\boldsymbol{b}, W) = \sqrt{15}$.

4 Gram-Schmidt process

For a vector subspace W of $\mathbb{R}^n$, the **Gram-Schmidt process** is an algorithm to construct an orthogonal basis for W from a given basis.

Example 4.1. Let W be the subspace spanned by the vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}.$$

Construct an orthogonal basis for W.

Solution. Set $\boldsymbol{w}_1 = \boldsymbol{v}_1$. We want to find another vector $\boldsymbol{w}_2$ orthogonal to $\boldsymbol{w}_1$. Note that $\boldsymbol{v}_2 - \mathrm{Proj}_{\boldsymbol{w}_1}\boldsymbol{v}_2$ is orthogonal to $\boldsymbol{w}_1$. So we set

$$\begin{aligned} \boldsymbol{w}_2 &= \boldsymbol{v}_2 - \mathrm{Proj}_{\boldsymbol{w}_1}\boldsymbol{v}_2 = \boldsymbol{v}_2 - \frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_2}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\boldsymbol{w}_1 \\ &= \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix} - \frac{12}{6}\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 3 \end{bmatrix}. \end{aligned}$$

Then the set $\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$ is an orthogonal basis for W.

Example 4.2. Let W be the subspace of $\mathbb{R}^4$ spanned by the vectors

$$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \ \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Construct an orthogonal basis for W.

Set $\boldsymbol{w}_1 = \boldsymbol{v}_1$. Let $W_1 = \mathrm{Span}\,\{\boldsymbol{w}_1\}$. To find a vector $\boldsymbol{w}_2$ in W that is orthogonal to W_1, we set

$$\begin{aligned} \boldsymbol{w}_2 &= \boldsymbol{v}_2 - \mathrm{Proj}_{W_1}\boldsymbol{v}_2 = \boldsymbol{v}_2 - \frac{\boldsymbol{v}_1 \cdot \boldsymbol{v}_2}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\boldsymbol{w}_1 \\ &= \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}. \end{aligned}$$

Let $W_2 = \text{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2\}$. To find a vector $\boldsymbol{w}_3$ in W that is orthogonal to W_2, we set

$$\begin{aligned}
\boldsymbol{w}_3 &= \boldsymbol{v}_3 - \text{Proj}_{W_2}\boldsymbol{v}_3 \\
&= \boldsymbol{v}_3 - \frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\boldsymbol{w}_1 - \frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\boldsymbol{w}_2 \\
&= \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2/4}{12/16} \cdot \frac{1}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix} \\
&= \begin{bmatrix} 1/3 \\ 1/3 \\ -2/3 \\ 0 \end{bmatrix}.
\end{aligned}$$

Then the set $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$ is an orthogonal basis for W.

Theorem 4.1 (Gram-Schmidt Process)**.** *Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p\}$ be a basis for a subspace W of $\mathbb{R}^n$. Set*

$$\begin{aligned}
\boldsymbol{w}_1 &= \boldsymbol{v}_1, \\
\boldsymbol{w}_2 &= \boldsymbol{v}_2 - \left(\frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_2}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\right)\boldsymbol{w}_1, \\
\boldsymbol{w}_3 &= \boldsymbol{v}_3 - \left(\frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\right)\boldsymbol{w}_1 - \left(\frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\right)\boldsymbol{w}_2, \\
&\vdots \\
\boldsymbol{w}_p &= \boldsymbol{v}_p - \left(\frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\right)\boldsymbol{w}_1 - \left(\frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\right)\boldsymbol{w}_2 - \cdots - \left(\frac{\boldsymbol{w}_{p-1} \cdot \boldsymbol{v}_p}{\boldsymbol{w}_{p-1} \cdot \boldsymbol{w}_{p-1}}\right)\boldsymbol{w}_{p-1}.
\end{aligned}$$

Then $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p\}$ is an orthogonal basis for W. Moreover, for any $1 \le k \le p$,

$$\text{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\} = \text{Span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}. \tag{4.1}$$

Proof. Let $W_k = \text{Span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$, $1 \le k \le p$. Set $\boldsymbol{w}_1 = \boldsymbol{v}_1$. Then

$$\boldsymbol{w}_k = \boldsymbol{v}_k - \text{Proj}_{W_{k-1}}\boldsymbol{v}_k, \quad 2 \le k \le p.$$

By the properties of orthogonal projection, the vector $\boldsymbol{w}_k$ is orthogonal to W_{k-1}. The identity (4.1) follows from the expression

$$\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{w}_1, \\
\boldsymbol{v}_2 &= \left(\frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_2}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\right)\boldsymbol{w}_1 + \boldsymbol{w}_2, \\
\boldsymbol{v}_3 &= \left(\frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\right)\boldsymbol{w}_1 + \left(\frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\right)\boldsymbol{w}_2 + \boldsymbol{w}_3, \\
&\vdots \\
\boldsymbol{v}_p &= \left(\frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1}\right)\boldsymbol{w}_1 + \left(\frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2}\right)\boldsymbol{w}_2 + \cdots + \left(\frac{\boldsymbol{w}_{p-1} \cdot \boldsymbol{v}_p}{\boldsymbol{w}_{p-1} \cdot \boldsymbol{w}_{p-1}}\right)\boldsymbol{w}_{p-1} + \boldsymbol{w}_p.
\end{aligned}$$

□

Note 2. The Gram-Schmidt process can be applied to any vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$, not necessarily linearly independent. In fact, if $\boldsymbol{v}_k$ is a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{k-1}$, then $\boldsymbol{v}_k \in W_{k-1}$. Thus $\mathrm{Proj}_{W_{k-1}} \boldsymbol{v}_k = \boldsymbol{v}_k$. Hence $\boldsymbol{w}_k = \mathbf{0}$. In this case, we assume that $\frac{\boldsymbol{w}_k \cdot \boldsymbol{v}_l}{\boldsymbol{w}_k \cdot \boldsymbol{w}_k} = 0$ in the expression of $\boldsymbol{w}_l$ for all $l > k$. Hence

$$[\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p] = [\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p] \begin{bmatrix} 1 & \frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_2}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} & \frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} & \cdots & \frac{\boldsymbol{w}_1 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_1 \cdot \boldsymbol{w}_1} \\ 0 & 1 & \frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_3}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2} & \cdots & \frac{\boldsymbol{w}_2 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_2 \cdot \boldsymbol{w}_2} \\ 0 & 0 & 1 & \cdots & \frac{\boldsymbol{w}_3 \cdot \boldsymbol{v}_p}{\boldsymbol{w}_3 \cdot \boldsymbol{w}_3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$