Chapter 4

Orthogonality

4.1 Orthogonality

4.1.1 Definition of Orthogonality

1. **Definition of Orthogonal Vectors**

   $\forall$ two n-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$ where $n \in \mathbb{N}^+$, if

   $$\mathbf{u}^{\mathrm{T}}\mathbf{v} = \mathbf{u} \cdot \mathbf{v} = 0,$$

   then the vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

2. **Definition of Orthogonality of Subspaces**

   Two subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$ of a vector space are orthogonal if every vector $\mathbf{u}$ in $\boldsymbol{U}$ is perpendicular to every vector $\mathbf{v}$ in $\boldsymbol{V}$, which is also that $\mathbf{u}^{\mathrm{T}}\mathbf{v} = 0$ for all $\mathbf{u}$ in $\boldsymbol{U}$ and all $\mathbf{v}$ in $\boldsymbol{V}$.

3. **Definition of Orthogonal Complements**

   The orthogonal complements of a subspace $\boldsymbol{V}$ contains every vector that is perpendicular to $\boldsymbol{V}$. This orthogonal subspace is denoted by $\boldsymbol{V}^{\perp}$.

4.1.2 Properties of Orthogonal Subspaces $\square$

1. **Independence**

   $\forall$ two orthogonal subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$, assume that a set of vectors $S_u = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ is a basis of $\boldsymbol{U}$ and a set of vectors $S_v = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a basis of $\boldsymbol{V}$, where

$m, n \in \mathbb{N}^+$. Then the vectors in the set $S = S_u \cup S_v = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m, \mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ are linearly independent.

2. **Symmetry**

   $\forall$ subspaces $\boldsymbol{V}$, its orthogonal complement is $\boldsymbol{V}^\perp$. Then the orthogonal complement of $\boldsymbol{V}^\perp$ is $\boldsymbol{V}$, which means that $(\boldsymbol{V}^\perp)^\perp = \boldsymbol{V}$.

3. **Complementarity**

   $\forall$ subspaces $\boldsymbol{V}$, its orthogonal complement is $\boldsymbol{V}^\perp$. Assume that the vectors in $\boldsymbol{V}$ and $\boldsymbol{V}^\perp$ are $n-$dimensional, where $h \in \mathbb{N}^+$. Then if a set of vectors S_v is a basis of $\boldsymbol{V}$ and a set of vectors $S_{v^\perp}$ is a basis of $\boldsymbol{V}^\perp$, then the set of vectors $S = S_v \cup S_{v^\perp}$ is a basis of $\mathbf{R}^n$.

4.1.3 Fundamental Theorem of Linear Algebra, Part 2 $\square$

$\forall$ $m \times n$ matrix A where $m, n \in \mathbb{N}^+$,

1. $\boldsymbol{N}(A)$ is the orthogonal complement of $\boldsymbol{C}(A^\mathrm{T})$, which is in $\mathbf{R}^n$.

2. $\boldsymbol{C}(A^\mathrm{T})$ is the orthogonal complement of $\boldsymbol{N}(A)$, which is in $\mathbf{R}^n$.

3. $\boldsymbol{N}(A^\mathrm{T})$ is the orthogonal complement of $\boldsymbol{C}(A)$, which is in $\mathbf{R}^m$.

4. $\boldsymbol{C}(A)$ is the orthogonal complement of $\boldsymbol{N}(A^\mathrm{T})$, which is in $\mathbf{R}^m$.

4.1.4 The Decomposition of The Solution to $A\mathbf{x} = \mathbf{b}$ $\square$

$\forall$ $m \times n$ matrix A where $m, n \in \mathbb{N}^+$, $\forall$ vector $\mathbf{b} \in \boldsymbol{C}(A)$ and $\mathbf{b} \neq \mathbf{0}$. For the linear equations $A\mathbf{x} = \mathbf{b}$:

1. $\exists$ a unique $n-$dimensional vector $\mathbf{x}_r \in \boldsymbol{C}(A^\mathrm{T})$ such that $A\mathbf{x}_r = \mathbf{b}$.

2. $\forall$ solution $\mathbf{x}_c$ to $A\mathbf{x} = \mathbf{0}$ can be decomposed as $\mathbf{x}_c = \mathbf{x}_r + \mathbf{x}_n$, where $\mathbf{x}_n \in \boldsymbol{N}(A)$.

4.2 Projections

4.2.1 Property of $A^{\mathrm{T}}A$ and AA^{T} $\square$

$\forall$ matrix A,

1. $A^{\mathrm{T}}A$ and AA^{T} are square and symmetric.

2. $A^{\mathrm{T}}A$ is invertible if and only if A has linearly independent columns.

3. AA^{T} is invertible if and only if A has linearly independent rows.

4.2.2 Definition of Projection

$\forall\ m, n \in \mathbb{N}^{+}$ such that $m > n$, assume that vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbf{R}^m$ and they are a basis of a subspace $\boldsymbol{V}$. $\forall$ vector $\mathbf{b} \in \mathbf{R}^m$ but $\mathbf{b} \notin \boldsymbol{V}$. $\exists$ a vector $\mathbf{p} \in \boldsymbol{V}$ and a vector $\mathbf{e} \in \boldsymbol{V}^{\perp}$ such that

$$\mathbf{p} + \mathbf{e} = \mathbf{b}.$$

$\mathbf{p}$ is called the projection onto $\boldsymbol{V}$ of $\mathbf{b}$; $\mathbf{e}$ is called the error vector. Assume that

$$\mathbf{p} = \widehat{x}_1\mathbf{a}_1 + \widehat{x}_2\mathbf{a}_2 + \cdots + \widehat{x}_n\mathbf{a}_n = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \vdots \\ \widehat{x}_n \end{bmatrix} = A\widehat{\mathbf{x}} = P\mathbf{b}.$$

P is called the projection matrix. It is clear that $\mathbf{p} \in \boldsymbol{C}(A)$ and $\mathbf{e} \in \boldsymbol{N}(A^{\mathrm{T}})$.

4.2.3 Existence and Uniqueness of Projection $\square$

$\forall\ m, n \in \mathbb{N}^{+}$ such that $m \geq n$, assume that vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbf{R}^m$ and they are a basis of a subspace $\boldsymbol{V}$. $\forall$ vector $\mathbf{b} \in \mathbf{R}^m$ but $\mathbf{b} \notin \boldsymbol{V}$. $\exists$ a vector $\mathbf{p} \in \boldsymbol{V}$ and a vector $\mathbf{e} \in \boldsymbol{V}^{\perp}$ such that

$$\mathbf{p} + \mathbf{e} = \mathbf{b}.$$

In addition, $\mathbf{p}$ and $\mathbf{e}$ are unique.

4.2.4 Process to Calculate Projection

Because $\mathbf{e} \in \boldsymbol{V}^{\perp}$, there is

$$\mathbf{a}_1^{\mathrm{T}}\mathbf{e} = \mathbf{a}_2^{\mathrm{T}}\mathbf{e} = \cdots = \mathbf{a}_n^{\mathrm{T}}\mathbf{e} = 0 \qquad \Rightarrow \qquad \begin{bmatrix} — & \mathbf{a}_1^{\mathrm{T}} & — \\ — & \mathbf{a}_2^{\mathrm{T}} & — \\ \vdots & \vdots & \vdots \\ — & \mathbf{a}_n^{\mathrm{T}} & — \end{bmatrix} \mathbf{e} = A^{\mathrm{T}}\mathbf{e} = \mathbf{0}.$$

Because $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - A\widehat{\mathbf{x}}$, there is

$$A^{\mathrm{T}}(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0} \qquad \Rightarrow \qquad A^{\mathrm{T}}A\widehat{\mathbf{x}} = A^{\mathrm{T}}\mathbf{b}.$$

Because $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ are linearly independent, which means that $\boldsymbol{N}(A)$ can only have $\mathbf{0}$, and $A^{\mathrm{T}}A$ is invertible. Thus,

$$\widehat{\mathbf{x}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b}, \qquad \mathbf{p} = A\widehat{\mathbf{x}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b}, \qquad P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}.$$

In addition,

$$P^2 = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} = A[(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}A](A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} = P.$$

Special Case: If $\boldsymbol{V}$ is a line and a vector $\mathbf{a} \in \boldsymbol{V}$, then

$$\widehat{\mathbf{x}} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}, \qquad \mathbf{p} = \mathbf{a}\widehat{\mathbf{x}} = \mathbf{a}\frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}, \qquad P = \frac{\mathbf{a}\mathbf{a}^{\mathrm{T}}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}.$$

4.3 Orthonormal Bases and Gram-Schmidt

4.3.1 Definition about Orthonormal Concepts

1. **Orthonormal Vectors**

   $\forall$ m-dimensional vectors $\mathbf{q}_1, \cdots, \mathbf{q}_n$ where $m, n \in \mathbb{N}^+$, if

$$\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_j = \mathbf{q}_i \cdot \mathbf{q}_j = \begin{cases} 0 & \text{when } i \neq j \quad \text{(orthogonal vectors)} \\ 1 & \text{when } i = j \quad \text{(unit vectors : } \|\mathbf{q}_i\| = 1) \end{cases}$$

   where $1 \leq i, j \leq n$ and $i, j \in \mathbb{N}^+$, then the vectors $\mathbf{q}_1, \cdots, \mathbf{q}_n$ are orthonormal.

2. **Orthogonal Matrices**

   $\forall$ $m \times n$ matrix Q with orthonormal columns is called an orthogonal matrix, where $m, n \in \mathbb{N}^+$.

$$Q = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix},$$

   where $\mathbf{q}_1, \cdots, \mathbf{q}_n$ are m-dimensional orthonormal vectors.

4.3.2 Properties of Orthogonal Matrices $\square$

1. $\forall$ orthogonal matrix Q satisfies $Q^{\mathrm{T}}Q = I$.

2. $\forall$ square orthogonal matrix Q satisfies $Q^{\mathrm{T}} = Q^{-1}$.

3. $\forall$ $m \times n$ orthogonal matrix Q, where $m, n \in \mathbb{N}^+$

   (1) $\forall$ $n-$dimensional vector $\mathbf{x}$, there is $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.

   (2) $\forall$ $n-$dimensional vectors $\mathbf{x}$ and $\mathbf{y}$, there is $(Q\mathbf{x})^{\mathrm{T}}Q\mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{y}$.

4.3.3 Projections Using Orthonormal Bases: Q Replaces A

$\forall$ $m, n \in \mathbb{N}^+$ such that $m > n$, assume that orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n \in \mathbf{R}^m$ and they are a basis of a subspace $\boldsymbol{V}$. $\forall$ vector $\mathbf{b} \in \mathbf{R}^m$ but $\mathbf{b} \notin \boldsymbol{V}$. In order to obtain the projection onto $\boldsymbol{V}$ of $\mathbf{b}$, there is

$$\mathbf{p} = \widehat{x}_1\mathbf{q}_1 + \widehat{x}_2\mathbf{q}_2 + \cdots + \widehat{x}_n\mathbf{q}_n = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} \widehat{x}_1 \\ \widehat{x}_2 \\ \vdots \\ \widehat{x}_n \end{bmatrix} = Q\widehat{\mathbf{x}} = P\mathbf{b}.$$

Then

$$\widehat{\mathbf{x}} = (Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}}\mathbf{b} = Q^{\mathrm{T}}\mathbf{b},$$
$$\mathbf{p} = Q\widehat{\mathbf{x}} = QQ^{\mathrm{T}}\mathbf{b},$$
$$P = QQ^{\mathrm{T}}.$$

Specifically,

$$\begin{aligned}\mathbf{p} = QQ^{\mathrm{T}}\mathbf{b} &= \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} — & \mathbf{q}_1^{\mathrm{T}} & — \\ — & \mathbf{q}_2^{\mathrm{T}} & — \\ \vdots & \vdots & \vdots \\ — & \mathbf{q}_n^{\mathrm{T}} & — \end{bmatrix} \mathbf{b} \\ &= \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} — & \mathbf{q}_1^{\mathrm{T}}\mathbf{b} & — \\ — & \mathbf{q}_2^{\mathrm{T}}\mathbf{b} & — \\ \vdots & \vdots & \vdots \\ — & \mathbf{q}_n^{\mathrm{T}}\mathbf{b} & — \end{bmatrix} \\ &= (\mathbf{q}_1^{\mathrm{T}}\mathbf{b})\mathbf{q}_1 + (\mathbf{q}_2^{\mathrm{T}}\mathbf{b})\mathbf{q}_2 + \cdots + (\mathbf{q}_n^{\mathrm{T}}\mathbf{b})\mathbf{q}_n.\end{aligned}$$

4.3.4 Vectors in A Subspace with an Orthonormal Basis □

$\forall\ n \in \mathbb{N}^+$, assume that orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$ are a basis of a subspace $\boldsymbol{V}$. $\forall$ vector $\mathbf{v} \in \boldsymbol{V}$, there is

$$\mathbf{v} = (\mathbf{q}_1^{\mathrm{T}}\mathbf{v})\mathbf{q}_1 + (\mathbf{q}_2^{\mathrm{T}}\mathbf{v})\mathbf{q}_2 + \cdots + (\mathbf{q}_n^{\mathrm{T}}\mathbf{v})\mathbf{q}_n.$$

4.3.5 The Gram-Schmidt Process

$\forall\ m, n \in \mathbb{N}^+$ such that $m \geq n$, assume that vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbf{R}^m$ are a basis of a subspace $\boldsymbol{V}$. The Gram-Schmidt Process can be used to get an orthonormal basis of $\boldsymbol{V}$: $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$.

1. **Complement**

   Let

$$\mathbf{p}_i = \mathbf{a}_{n+1} = \mathbf{0}.$$

2. **Recursion**

   $\forall\ 1 \leq i \leq n$ and $i \in \mathbb{N}^+$,

$$\mathbf{e}_i = \mathbf{a}_i - \mathbf{p}_i, \qquad \mathbf{q}_i = \frac{\mathbf{e}_i}{\|\mathbf{e}_i\|},$$

$$\mathbf{p}_{i+1} = (\mathbf{q}_1^{\mathrm{T}}\mathbf{a}_{i+1})\mathbf{q}_1 + (\mathbf{q}_2^{\mathrm{T}}\mathbf{a}_{i+1})\mathbf{q}_2 + \cdots + (\mathbf{q}_i^{\mathrm{T}}\mathbf{a}_{i+1})\mathbf{q}_i.$$

3. **Result**

   The vectors $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$ is an orthonormal basis of $\boldsymbol{V}$.

4.3.6 The Factorization $A = QR$ $\square$

$\forall\ m, n \in \mathbb{N}^+$ such that $m \geq n$, assume that vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n \in \mathbf{R}^m$ are a basis of a subspace $\boldsymbol{V}$. Gram-Schmidt constructs orthonormal vectors $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$. The matrices with these columns satisfy $A = QR$. Then $R = Q^{\mathrm{T}}A$ is upper triangular because later $\mathbf{q}$'s are orthogonal to earlier $\mathbf{a}$'s.

$$A = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & \cdots & | \end{bmatrix}, \quad Q = \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix}, \quad R = \begin{bmatrix} \mathbf{q}_1^{\mathrm{T}}\mathbf{a}_1 & \mathbf{q}_1^{\mathrm{T}}\mathbf{a}_2 & \cdots & \mathbf{q}_1^{\mathrm{T}}\mathbf{a}_n \\ 0 & \mathbf{q}_2^{\mathrm{T}}\mathbf{a}_2 & \cdots & \mathbf{q}_2^{\mathrm{T}}\mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{q}_n^{\mathrm{T}}\mathbf{a}_n \end{bmatrix}.$$

Thus,

$$A = QR.$$

It is clear that R is invertible.

Special Case: For the projection, if $A = QR$, there is

$$A^{\mathrm{T}}A = (QR)^{\mathrm{T}}QR = R^{\mathrm{T}}Q^{\mathrm{T}}QR = R^{\mathrm{T}}(Q^{\mathrm{T}}Q)R = R^{\mathrm{T}}IR = R^{\mathrm{T}}R.$$

Then

$$\widehat{\mathbf{x}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b} = (R^{\mathrm{T}}R)^{-1}(QR)^{\mathrm{T}}\mathbf{b} = R^{-1}(R^{\mathrm{T}})^{-1}R^{\mathrm{T}}Q^{\mathrm{T}}\mathbf{b} = R^{-1}Q^{\mathrm{T}}\mathbf{b},$$

$$\mathbf{p} = A\widehat{\mathbf{x}} = QRR^{-1}Q^{\mathrm{T}}\mathbf{b} = QQ^{\mathrm{T}}\mathbf{b},$$

$$P = QQ^{\mathrm{T}}.$$