Chapter 4

Orthogonality

4.1 Orthogonality of the Four Subspaces

4.1.1 Definition of Orthogonality of Subspaces

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

4.1.2 Definition of Orthogonal Complements

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

4.1.3 Fundamental Theorem of Linear Algebra, Part 2

- $N(A)$ is the orthogonal complement of the row space $C(A^{\mathrm{T}})$, which is in $\mathbf{R}^n$.
- $N(A^{\mathrm{T}})$ is the orthogonal complement of the column space $C(A)$, which is in $\mathbf{R}^m$.

4.1.4 Combining Bases from Subspaces

- Any n independent vectors in $\mathbf{R}^n$ must span $\mathbf{R}^n$. So they are a basis.
- Any n vectors that span $\mathbf{R}^n$ must be indepentent. So they are a basis.
- If the n columns of A are independent, they span $\mathbf{R}^n$. So $A\mathbf{x} = \mathbf{b}$ is solvable.
- If the n columns of A span $\mathbf{R}^n$, they are independent. So $A\mathbf{x} = \mathbf{b}$ has only one solution.

4.2 Projections

4.2.1 Projection Onto a Line

$\forall\, n \in \mathbb{N}^+$, a line through the origin in the direction of $\mathbf{a} = (a_1, a_2, \cdots, a_n)$. Along the line, we want the point $\mathbf{p}$ closest to $\mathbf{b} = (b_1, b_2, \cdots, b_n)$. The line from $\mathbf{b}$ to $\mathbf{p}$ is perpendicular to the vector $\mathbf{a}$. Let $\mathbf{e} = \mathbf{b} - \mathbf{p}$ as the error. The projection $\mathbf{p}$ will be some multiple of $\mathbf{a}$. Let $\mathbf{p} = \widehat{\mathbf{x}}\mathbf{a}$. Then

$$\mathbf{a} \cdot \mathbf{e} = \mathbf{a} \cdot (\mathbf{b} - \mathbf{p}) = \mathbf{a} \cdot (\mathbf{b} - \widehat{\mathbf{x}}\mathbf{a}) = \mathbf{a} \cdot \mathbf{b} - \widehat{\mathbf{x}}\mathbf{a} \cdot \mathbf{a} = 0.$$

Therefore,

$$\widehat{\mathbf{x}} = \frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{a} \cdot \mathbf{a}} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}.$$

The projection of $\mathbf{b}$ onto the line through $\mathbf{a}$ is the vector

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

Define the projection matrix P:

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

There are two special cases:

- If $\mathbf{b} = \mathbf{a}$ then $\widehat{\mathbf{x}} = 1$. The projection of $\mathbf{a}$ onto $\mathbf{a}$ is itself. $P\mathbf{a} = \mathbf{a}$.
- If $\mathbf{b}$ is perpendicular to $\mathbf{a}$ then $\mathbf{a}^{\mathrm{T}}\mathbf{b} = 0$. The projection is $\mathbf{p} = \mathbf{0}$.

4.2.2 Projection Onto a Subspace

$\forall\, n$ linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ in $\mathbf{R}^m$ where $m, n \in \mathbb{N}^+$. We want to find the combination $\mathbf{p} = \widehat{x}_1\mathbf{a}_1 + \widehat{x}_2\mathbf{a}_2 + \cdots + \widehat{x}_n\mathbf{a}_n$ closest to a given vector $\mathbf{b}$. Let

$$A = \begin{bmatrix} | & \cdots & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & \cdots & | \end{bmatrix}, \qquad A^{\mathrm{T}} = \begin{bmatrix} — & \mathbf{a}_1^{\mathrm{T}} & — \\ \vdots & \vdots & \vdots \\ — & \mathbf{a}_n^{\mathrm{T}} & — \end{bmatrix}, \qquad \widehat{\mathbf{x}} = \begin{bmatrix} \widehat{x}_1 \\ \vdots \\ \widehat{x}_n \end{bmatrix}.$$

Then the projection $\mathbf{p} = A\widehat{\mathbf{x}}$; the error vector $\mathbf{e} = \mathbf{b} - \mathbf{p} = \mathbf{b} - A\widehat{\mathbf{x}}$ is perpendicular to the subspace. Therefore,

$$\begin{bmatrix} — & \mathbf{a}_1^{\mathrm{T}} & — \\ \vdots & \vdots & \vdots \\ — & \mathbf{a}_n^{\mathrm{T}} & — \end{bmatrix} (\mathbf{b} - A\widehat{\mathbf{x}}) = A^{\mathrm{T}}(\mathbf{b} - A\widehat{\mathbf{x}}) = \mathbf{0} \quad \Rightarrow \quad A^{\mathrm{T}}A\widehat{\mathbf{x}} = A^{\mathrm{T}}\mathbf{b}.$$

Now,

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

In addition,

$$P^2 = P.$$

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

4.3 Least Squares Approximations

Suppose that there are n points $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ where $n \in \mathbb{N}^+$. We want to find the closest line $b = C + Dt$ to these points. Let

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} C \\ D \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

We compute $\widehat{\mathbf{x}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\mathbf{b}$ to get the parameters of the line.

4.4 Orthonormal Bases and Gram-Schmidt

4.4.1 Definition of Orthonormal Vectors

The vectors $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$ are orthonormal if

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

A matrix with orthonormal columns is assigned the special letter Q.

4.4.2 Properties of Orthogonal Matrices

1. A matrix Q with orthonormal columns satisfies $Q^{\mathrm{T}}Q = I$:

$$Q^{\mathrm{T}}Q = \begin{bmatrix} — & \mathbf{q}_1^{\mathrm{T}} & — \\ — & \mathbf{q}_2^{\mathrm{T}} & — \\ \vdots & \vdots & \vdots \\ — & \mathbf{q}_n^{\mathrm{T}} & — \end{bmatrix} \begin{bmatrix} | & | & \cdots & | \\ \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \\ | & | & \cdots & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I.$$

2. When Q is square, $Q^{\mathrm{T}}Q = I$ means that $Q^{\mathrm{T}} = Q^{-1}$.

3. If Q has orthonormal columns, it leaves lengths unchanged: $\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for every vector $\mathbf{x}$. Q also preserves dot products: $(Q\mathbf{x})^{\mathrm{T}}(Q\mathbf{y}) = \mathbf{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\mathbf{y} = \mathbf{x}^{\mathrm{T}}\mathbf{y}$.

4.4.3 Projections Using Orthonormal Bases: Q Replaces A

Suppose that $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$ in $\mathbf{R}^m$ are n orthonormal vectors where $m, n \in \mathbb{N}^+$. Let

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

Therefore, $Q^{\mathrm{T}}Q = I$. We want to find the combination $\mathbf{p} = \widehat{x}_1\mathbf{q}_1 + \widehat{x}_2\mathbf{q}_2 + \cdots + \widehat{x}_n\mathbf{q}_n$ closest to a given vector $\mathbf{b}$. Then

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

$$\mathbf{p} = Q\widehat{\mathbf{x}} = QQ^{\mathrm{T}}\mathbf{b} = \begin{bmatrix} | & \cdots & | \\ \mathbf{q}_1 & \cdots & \mathbf{q}_n \\ | & \cdots & | \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{q}_n^{\mathrm{T}} \end{bmatrix} = \mathbf{q}_1(\mathbf{q}_1^{\mathrm{T}}\mathbf{b}) + \cdots + \mathbf{q}_n(\mathbf{q}_n^{\mathrm{T}}\mathbf{b}),$$

$$P = Q(Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}} = QQ^{\mathrm{T}}.$$

4.4.4 The Gram-Schmidt Process

$\forall$ n linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ in $\mathbf{R}^m$ where $m, n \in \mathbb{N}^+$. First of all, we intend to construct n orthogonal vectors $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_n$:

$$\begin{aligned}
\mathbf{A}_1 &= \mathbf{a}_1, \\
\mathbf{A}_2 &= \mathbf{a}_2 - \frac{\mathbf{A}_1^{\mathrm{T}}\mathbf{a}_2}{\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1}\mathbf{A}_1, \\
\mathbf{A}_3 &= \mathbf{a}_3 - \frac{\mathbf{A}_1^{\mathrm{T}}\mathbf{a}_3}{\mathbf{A}_1^{\mathrm{T}}\mathbf{A}_1}\mathbf{A}_1 - \frac{\mathbf{A}_2^{\mathrm{T}}\mathbf{a}_3}{\mathbf{A}_2^{\mathrm{T}}\mathbf{A}_2}\mathbf{A}_2, \\
&\vdots \\
\mathbf{A}_n &= \mathbf{a}_n - \sum_{i=1}^{n-1} \frac{\mathbf{A}_i^{\mathrm{T}}\mathbf{a}_n}{\mathbf{A}_i^{\mathrm{T}}\mathbf{A}_i}\mathbf{A}_i.
\end{aligned}$$

Then we get n orthonormal bases $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$:

$$\mathbf{q}_1 = \frac{\mathbf{A}_1}{\|\mathbf{A}_1\|}, \qquad \mathbf{q}_2 = \frac{\mathbf{A}_2}{\|\mathbf{A}_2\|}, \qquad \cdots, \qquad \mathbf{q}_n = \frac{\mathbf{A}_n}{\|\mathbf{A}_n\|}.$$

4.4.5 The Factorization $A = QR$

From $\forall\, n$ linearly independent vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ in $\mathbf{R}^m$ where $m, n \in \mathbb{N}^+$, 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,

$$A = QR.$$