Lecture 14

Chapter 6 Matrix Diagonalization

6.1 Inner Product and Inner Product Space

6.2 Orthonormal Sets and Orthogonal Subspaces

6.3 The Gram-Schmidt Orthogonalization Process

6.1 Inner Product and Inner Product Space

Scalar products are useful not only in $\mathbf{R}^n$, but also in a wide variety of context. In this section, we add to the structure of a vector space by defining a scalar or inner product.

Inner Product

Definition 1. Let V be a vector space. An operation,

$$\langle \cdot,\cdot \rangle: V \times V \to \mathbf{R},$$

is said to be an **inner product [内积]** on V if it satisfies

(1) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$ and $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$;

(2) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$ for all $\boldsymbol{x}$ and $\boldsymbol{y}$ in V;

(3) $\langle \alpha \boldsymbol{x} + \beta \boldsymbol{y}, \boldsymbol{z} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \beta \langle \boldsymbol{y}, \boldsymbol{z} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ and all scalars α, β.

The vector space V with an inner product defined on it is called the **inner product space [内积空间]**.

In case that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, we say $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal [正交]** to each other, denoted by $\boldsymbol{x} \perp \boldsymbol{y}$.

Example 1. (Inner Products for Vector Space $\mathbf{R}^n$)

- **Standard inner product [标准内积] on $\mathbf{R}^n$**

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^T \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i \,.$$

Notice that
$$\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \boldsymbol{x}^T \boldsymbol{x} = \sum_{i=1}^{n} x_i^2 \geq 0.$$

- **Weighted inner product [带权内积]** on $\mathbf{R}^n$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_w = \sum_{i=1}^{n} w_i x_i y_i \,,$$

where $w_i > 0,\ i = 1,2,\dots,n$, are called the **weights [权重].**

Exercise: Check that $\langle \cdot,\cdot \rangle, \langle \cdot,\cdot \rangle_w$ defined above are inner products on $\mathbf{R}^n$.

Example 2. (Inner Products for Vector Space $\mathbf{R}^{m\times n}$)

- **Frobenius inner product:** $A, B \in \mathbf{R}^{m\times n}$,

$$\langle A, B\rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}\,.$$

- **weighted Frobenius inner product**

$$\langle A, B\rangle_w = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}a_{ij}b_{ij}$$

where $w_{ij} > 0$, $i = 1,2,\dots,m$, $j = 1,2,\dots,n$.

Exercise: Check that $\langle\cdot,\cdot\rangle, \langle\cdot,\cdot\rangle_w$ are inner products on $\mathbf{R}^{m\times n}$.

Example 3. (Inner Product for Vector Space $C[a,b]$)

$$f, g \in [a,b], \qquad \langle f, g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x.$$

Check that $\langle\cdot,\cdot\rangle$ is an inner product on $C[a,b]$.

Proof of condition (1) positivity: $\langle f, f\rangle = \int_a^b f^2(x)\mathrm{d}x \geq 0.$

- If $f(x) = 0$, we have $\langle f, f\rangle = 0$.
- If $\langle f, f\rangle = 0$, we will also have $f(x) = 0$ for all $x \in [a,b]$. If not so, suppose that $x_0 \in (a,b)$ is a point where $f(x_0) \neq 0$, then there exists a δ-neighborhood, say $U(x_0,\delta)$, $\delta > 0$ such that

$$f^2(x) \geq \frac{1}{2}f^2(x_0) > 0, \qquad x \in U(x_0,\delta).$$

$$\Longrightarrow \quad \langle f, f\rangle = \int_a^b f^2(x)\,\mathrm{d}x \geq \int_{x_0-\delta}^{x_0+\delta} f^2(x)\,\mathrm{d}x \geq f^2(x_0)\delta > 0,$$

contradiction

The similar conclusions hold for $x_0 = a$ or $x_0 = b$.

Conditions (2) and (3) can also be verified directly by properties of definite integral on $C[a, b]$.

weighted inner product on $C[a, b]$

$$\langle f, g \rangle_w = \int_a^b w(x) f(x) g(x) \, \mathrm{d}x,$$

$0 < w(x) \in C[a, b]$ is called a **weight function**.

Norm

Definition 2. Let V be a vector space. The **length [长度]** or **norm [范数]** of vectors in V is a non-negative real-valued function $\rho: V \to \mathbf{R}_+$ that satisfies

(1) $\rho(\lambda \boldsymbol{x}) = |\lambda| \cdot \rho(\boldsymbol{x})$, (positive homogeneity)

(2) $\rho(\boldsymbol{x} + \boldsymbol{y}) \leq \rho(\boldsymbol{x}) + \rho(\boldsymbol{y})$, (triangle inequality);

(3) $\rho(\boldsymbol{x}) \geq 0$ and $\rho(\boldsymbol{x}) = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$ (positivity),

where λ is a scalar, $\boldsymbol{x}, \boldsymbol{y} \in V$. The vector space V with a norm defined on it is said to be a **normed linear space [赋范线性空间]**.

The norm of a vector $\boldsymbol{x} \in V$ is denoted by $||\boldsymbol{x}||$.

Remark. In particular, if V is an **inner product space** and the inner product defined on V is denoted by $\langle\cdot,\cdot\rangle$, then the **norm induced by inner product [内积诱导的范数]** is defined by

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}.$$

Theorem 1. (The Pythagorean law) Let $\boldsymbol{x}, \boldsymbol{y}$ be two **orthogonal** vectors in an inner product space V, and $\|\boldsymbol{x}\|$ is the norm induced by the inner product. Then

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^{\mathbf{2}} + \|\boldsymbol{y}\|^{\mathbf{2}}.$$

Proof. We have

$$\begin{aligned} \|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle \\ &= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle. \end{aligned}$$

Thus we get $\boldsymbol{x} \perp \boldsymbol{y}$ if and only if $\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

Example 4. Consider the inner product on $C[-1,1]$ defined by

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x)\, \mathrm{d}x. \qquad \textbf{(*)}$$

Let $f(x) = x$, $g(x) = 1 + x^2$. Calculate $\|f\|$, $\|g\|$ and $\|f + g\|$.

Solution.

$$\|f\|^2 = \langle f, f \rangle = \int_{-1}^{1} (x \cdot x)\, \mathrm{d}x = \frac{2}{3},$$

$$\|g\|^2 = \langle g, g \rangle = \int_{-1}^{1} (1 + x^2) \cdot (1 + x^2) \mathrm{d}x = \frac{56}{15},$$

$$\|f + g\|^2 = \langle f + g, f + g \rangle = \int_{-1}^{1} (1 + x + x^2) \cdot (1 + x + x^2) \mathrm{d}x = \frac{22}{5}.$$

We see that Pythagorean law holds $\|f + g\|^2 = \|f\|^2 + \|g\|^2$. It follows that $f(x)$ is orthogonal to $g(x)$ w.r.t the inner product defined in **(*)**.

Projection of vectors

Definition 3. Let $\boldsymbol{x}, \boldsymbol{y}$ be vectors in an inner product space V and $\boldsymbol{x} \neq \mathbf{0}$, then the **scalar projection [标量投影]** of $\boldsymbol{y}$ onto $\boldsymbol{x}$ is

$$\alpha = \frac{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\|\boldsymbol{x}\|}$$

and the **vector projection [向量投影]** of $\boldsymbol{y}$ onto $\boldsymbol{x}$ is

$$\boldsymbol{p} = \alpha \cdot \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \frac{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} \boldsymbol{x}.$$

Theorem 2. Let $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{p}$ be the vector projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$, then

(1) $(\boldsymbol{y}-\boldsymbol{p}) \perp \boldsymbol{p}$;

(2) $\boldsymbol{y}=\boldsymbol{p}$ if and only if $\boldsymbol{y}$ and $\boldsymbol{x}$ are linearly dependent.

Proof. (1) $\boldsymbol{p}$ is the projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$,

$$\boldsymbol{p}=\frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\boldsymbol{x},$$

then

$$\begin{aligned}\langle \boldsymbol{y}-\boldsymbol{p},\boldsymbol{p}\rangle &= \langle \boldsymbol{y},\boldsymbol{p}\rangle - \langle \boldsymbol{p},\boldsymbol{p}\rangle \\ &= \left\langle \boldsymbol{y},\frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\boldsymbol{x}\right\rangle - \left\langle \frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\boldsymbol{x},\frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\boldsymbol{x}\right\rangle \\ &= \frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\langle \boldsymbol{y},\boldsymbol{x}\rangle - \frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\cdot\frac{\langle \boldsymbol{y},\boldsymbol{x}\rangle}{\langle \boldsymbol{x},\boldsymbol{x}\rangle}\cdot\langle \boldsymbol{x},\boldsymbol{x}\rangle \\ &= 0.\end{aligned}$$

Theorem 2. Let $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{p}$ be the vector projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$, then

(1) $(\boldsymbol{y} - \boldsymbol{p}) \perp \boldsymbol{p}$;

(2) $\boldsymbol{y} = \boldsymbol{p}$ if and only if $\boldsymbol{y}$ and $\boldsymbol{x}$ are linearly dependent.

Proof. (2) If $\boldsymbol{y}$ and $\boldsymbol{x}$ are linearly dependent, there exists a scalar, say β, such that $\boldsymbol{y} = \beta\boldsymbol{x}$, then the vector projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$ can be calculated by

$$\boldsymbol{p} = \frac{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}\boldsymbol{x} = \beta\boldsymbol{x} = \boldsymbol{y}.$$

On the other hand, if $\boldsymbol{y} = \boldsymbol{p}$, it follows that

$$\boldsymbol{y} = \boldsymbol{p} = \frac{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}\boldsymbol{x} = \beta\boldsymbol{x}, \qquad \beta = \frac{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle},$$

implying that $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent.

Theorem 3. (Cauchy-Schwarz Inequality) Let V be an inner product space and $\boldsymbol{x}, \boldsymbol{y}$ be two vectors in V, then

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|.$$

The equality holds if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent.

Proof. If $\boldsymbol{x} = \boldsymbol{0}$, the conclusions of the theorem hold.

Suppose $\boldsymbol{x} \neq \boldsymbol{0}$. Let $\boldsymbol{p}$ be the vector projection of $\boldsymbol{y}$ onto $\boldsymbol{x}$. By **Theorem 2** and **Pythagorean law**, we have

$$\|\boldsymbol{p}\|^2 + \|\boldsymbol{y} - \boldsymbol{p}\|^2 = \|\boldsymbol{y}\|^2.$$

Notice that $$\|\boldsymbol{p}\|^2 = \langle \boldsymbol{p}, \boldsymbol{p} \rangle = \frac{(\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^2}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \frac{(\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^2}{\|\boldsymbol{x}\|^2},$$

we have $$\frac{(\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^2}{\|\boldsymbol{x}\|^2} = \|\boldsymbol{y}\|^2 - \|\boldsymbol{y} - \boldsymbol{p}\|^2.$$

Theorem 3. (Cauchy-Schwarz Inequality) Let V be an inner product space and $\boldsymbol{x}, \boldsymbol{y}$ be two vectors in V, then

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|.$$

The equality holds if and only if $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent.

Proof. (continue)

$$\frac{(\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^2}{\|\boldsymbol{x}\|^2} = \|\boldsymbol{y}\|^2 - \|\boldsymbol{y} - \boldsymbol{p}\|^2.$$

Since $\boldsymbol{x} \neq \mathbf{0}$, then

$$(\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^2 = \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2 \|\boldsymbol{y} - \boldsymbol{p}\|^2.$$

Therefore,

$$(\langle \boldsymbol{y}, \boldsymbol{x} \rangle)^2 \leq \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2, \qquad \text{or} \quad |\langle \boldsymbol{y}, \boldsymbol{x} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

It is clear that the equality holds if and only if $\boldsymbol{y} = \boldsymbol{p}$, which is equivalent to the fact that $\boldsymbol{y}$ and $\boldsymbol{x}$ are linearly dependent.

Angle between two vectors

One consequence of the Cauchy-Schwarz inequality is that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are nonzero vectors, then

$$-1 \leq \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \leq 1,$$

and hence there is a unique number $\theta \in [0, \pi]$ such that

$$\cos\theta = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}. \qquad (**)$$

Thus equation (**) can be used to define the **angle** θ between the two nonzero vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

6.2 Orthonormal Sets and Orthogonal Subspaces

In $\mathbf{R}^2$, it is generally more convenient to use the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ than to use some other basis, such as $\{(2,1)^T, (3,5)^T\}$. For example, it would be easier to find the coordinates of a vector with respect to the standard basis. The elements of the standard basis are **orthogonal unit** vectors.

Orthonormal Sets

Definition 1. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ be **nonzero vectors** in an inner product space V. The vector set $S = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ is said to be an **orthogonal set [正交集]** of vectors if $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$ holds for all $i \neq j$.

If moreover, $\boldsymbol{v}_i$, $i = 1, \ldots, n$ are all **unit** vectors, S is said to be an **orthonormal set [标准正交集或规范集].**

Example 1. The set $S = \{\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_5\}$ is an orthogonal set of vectors in $\mathbf{R}^n$, where $\mathbf{e}_\mathbf{i}$, $i = 1,3,5$, are vectors in a standard basis of $\mathbf{R}^n$. In fact, we have

$$\langle \mathbf{e}_\mathbf{i}, \mathbf{e}_\mathbf{j} \rangle = \delta_{ij},$$

where δ_{ij} is the Dirac delta function. S is also an orthonormal set.

Example 2. The set $S = \{(1,0,1)^T, (-1,0,1)^T\}$ is an orthogonal set of vectors in $\mathbf{R}^3$, since

$$\langle (1,0,1)^T, (-1,0,1)^T \rangle = 1 \cdot (-1) + 0 \cdot 0 + 1 \cdot 1 = 0.$$

If we denote

$$S' = \left\{ \frac{1}{\sqrt{2}} (1,0,1)^T, \frac{1}{\sqrt{2}} (-1,0,1)^T \right\},$$

then S' is an orthonormal set.

Theorem 1. If $S = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ is an orthogonal set in an inner product space V, then all vectors in S are linearly independent.

Proof. Consider the linear equation

$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n} = \mathbf{0}.$$

Since $\boldsymbol{v_i} \neq \mathbf{0}, i = 1,2, \dots, n$, then we have

$$\begin{aligned}\langle \boldsymbol{v_i}, c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_n\boldsymbol{v_n}\rangle &= \langle \boldsymbol{v_i}, \mathbf{0}\rangle = 0 \\ &= c_1\langle \boldsymbol{v_i}, \boldsymbol{v_1}\rangle + c_2\langle \boldsymbol{v_i}, \boldsymbol{v_2}\rangle + \cdots + c_n\langle \boldsymbol{v_i}, \boldsymbol{v_n}\rangle \\ &= c_i\langle \boldsymbol{v_i}, \boldsymbol{v_i}\rangle = c_i\|\boldsymbol{v_i}\|^2\end{aligned}$$

This equation implies that c_i must be zero for all $i = 1, \dots, n$, and then $\boldsymbol{v_i}, i = 1, \dots, n$ are linearly independent.

Finding the coordinates w.r.t. basis E

Let V be an n-dim. vector space, $E = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ be a basis of V.

- Any vector $\boldsymbol{v} \in V$ can be uniquely written as

$$\boldsymbol{v} = c_1\boldsymbol{v_1} + \cdots + c_n\boldsymbol{v_n} = (\boldsymbol{v_1}, \dots, \boldsymbol{v_n})\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = (\boldsymbol{v_1}, \dots, \boldsymbol{v_n})\boldsymbol{c},$$

where $\boldsymbol{c} = [\boldsymbol{v}]_E \in \mathbf{R}^n$ is the **coordinate vector** of $\boldsymbol{v}$ w.r.t. basis E.

- In case of $V = \mathbf{R}^n$, in order to find the coordinate vector $\boldsymbol{c}$, we have to solve the linear system

$$A\boldsymbol{c} = \boldsymbol{v},$$

where $A = (\boldsymbol{v_1}, \dots, \boldsymbol{v_n})$. The coordinate vector can be calculated by

$$\boldsymbol{c} = A^{-1}\boldsymbol{v},$$

which is a very **exhausting** work.

Case of orthogonal basis

Theorem 2. Let $E = \{\boldsymbol{v_1}, \dots, \boldsymbol{v_n}\}$ be an orthogonal set and be a basis of inner product space V. Then for any $\boldsymbol{v} \in V$, the ith coordinate of $\boldsymbol{v}$ can be calculated by

$$c_i = \frac{\langle \boldsymbol{v_i}, \boldsymbol{v} \rangle}{\langle \boldsymbol{v_i}, \boldsymbol{v_i} \rangle} = \frac{\langle \boldsymbol{v_i}, \boldsymbol{v} \rangle}{\|\boldsymbol{v_i}\|^2}.$$

Proof. Since $\boldsymbol{v} \in V$, then $\boldsymbol{v} = c_1\boldsymbol{v_1} + \cdots + c_n\boldsymbol{v_n}$.

The vectors in E are orthogonal to each other so for $i \neq j$, $i, j = 1, \dots, n$,

$$\langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle = 0.$$

We then have for $i = 1, \dots, n$,

$$\langle \boldsymbol{v_i}, \boldsymbol{v} \rangle = \langle \boldsymbol{v_i}, c_1\boldsymbol{v_1} + \cdots + c_n\boldsymbol{v_n} \rangle = c_i\langle \boldsymbol{v_i}, \boldsymbol{v_i} \rangle,$$

finishing the proof of the theorem.

Example 3. Show that $E = \{\boldsymbol{v_1} = (1,1)^T, \boldsymbol{v_2} = (1,-1)^T\}$ is an orthogonal set and a basis of $\mathbf{R}^2$. Find the coordinate vector of $\boldsymbol{v} = (1,2)^T$ w.r.t. basis E.

Solution. The vectors $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ are orthogonal since

$$\langle \boldsymbol{v_1}, \boldsymbol{v_2} \rangle = (1,1)\begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1 \cdot 1 + 1 \cdot (-1) = 0.$$

By **Theorem 1**, the vectors in E are linearly independent, so E forms a basis of $\mathbf{R}^2$. By **Theorem 2**, the coordinates can be calculated by

$$c_1 = \frac{\langle \boldsymbol{v_1}, \boldsymbol{v} \rangle}{\langle \boldsymbol{v_1}, \boldsymbol{v_1} \rangle} = \frac{\langle (1,1)^T, (1,2)^T \rangle}{\langle (1,1)^T, (1,1)^T \rangle} = \frac{3}{2},$$

$$c_2 = \frac{\langle \boldsymbol{v_2}, \boldsymbol{v} \rangle}{\langle \boldsymbol{v_2}, \boldsymbol{v_2} \rangle} = \frac{\langle (1,-1)^T, (1,2)^T \rangle}{\langle (1,-1)^T, (1,-1)^T \rangle} = -\frac{1}{2}.$$

Therefore, the coordinate vector of $\boldsymbol{v}$ w.r.t basis E is $\boldsymbol{c} = \left(\frac{3}{2}, -\frac{1}{2}\right)^T$.

If moreover, the basis E in **Theorem 2** is an **orthonormal** set, we shall have

$$c_i = \langle \boldsymbol{v_i}, \boldsymbol{v} \rangle, \qquad i = 1, \dots, n.$$

As a result, it is very convenient to take an orthonormal set as a basis of the inner product space V.

In **Section 6.3**, we shall study how to generate an **orthonormal basis** from a given basis of an inner product space V. (**Gram-Schmidt algorithm**)

Orthogonal Matrices

Definition 2. Let A be an $n \times n$ matrix. A is said to be an **orthogonal matrix [正交矩阵]** if and only if

$$A^T A = A A^T = I,$$

where I is the $n \times n$ identity matrix.

Example. Let $A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$. Rotation in $\mathbf{R}^2$

It is easy to calculate that

$$A^T A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I,$$

$$AA^T = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

Therefore, A is an orthogonal matrix.

Theorem 3. (Properties of Orthogonal Matrices) Let A be an $n \times n$ orthogonal matrix, then

(1)the column vectors of A form an orthonormal basis of $\mathbf{R}^n$;

(2) $A^T = A^{-1}$;

(3) $\langle A\boldsymbol{x}, A\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$;

(4) $\|A\boldsymbol{x}\| = \|\boldsymbol{x}\|$.

Proof. (1) By definition of orthogonal matrices, we have for $i, j = 1, \dots, n$

$$\boldsymbol{a}_i^T \boldsymbol{a}_j = \delta_{ij}.$$

Therefore the column vectors of A form an orthonormal basis of $\mathbf{R}^n$.

The statement (2) follows from the definition of inverse matrix.

To prove (3), by the definition of inner product in $\mathbf{R}^n$, we have

$$\langle A\boldsymbol{x}, A\boldsymbol{y}\rangle = (A\boldsymbol{x})^T(A\boldsymbol{y}) = \boldsymbol{x}^T\left(A^T A\right)\boldsymbol{y} = \boldsymbol{x}^T\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle.$$

The statement (4) follows from (3) by taking $\boldsymbol{x} = \boldsymbol{y}$.

Orthogonal Subspaces

Example. Let X and Y be two subspaces of $\mathbf{R}^3$, where

$$X = \mathrm{Span}\{\mathbf{e_1}\}, \qquad Y = \mathrm{Span}\{\mathbf{e_2}\}.$$

Any vector $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$,

$$\boldsymbol{x} = \alpha\mathbf{e_1}, \qquad \boldsymbol{y} = \beta\mathbf{e_2},$$

where α, β are scalars. We have

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \alpha\mathbf{e_1}, \beta\mathbf{e_2} \rangle = \alpha\beta\langle \mathbf{e_1}, \mathbf{e_2} \rangle = 0,$$

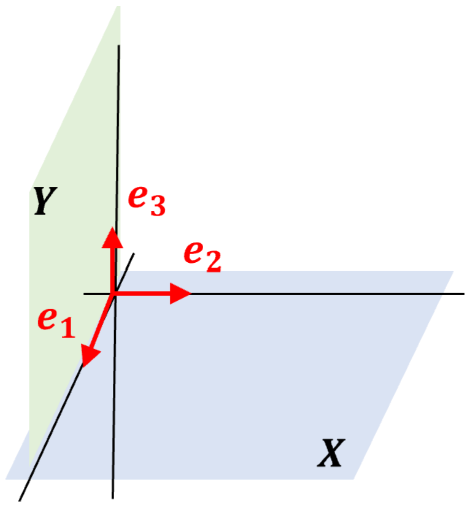
implying $\boldsymbol{x} \perp \boldsymbol{y}$, for all $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$.

$$X \perp Y$$

Definition 3. Let X and Y be two subspaces of $\mathbf{R}^n$. X and Y are said to be **orthogonal subspaces [正交子空间]** of $\mathbf{R}^n$ if and only if $\boldsymbol{x}^T\boldsymbol{y} = 0$ holds for all $\boldsymbol{x} \in X$ and $\boldsymbol{y} \in Y$.

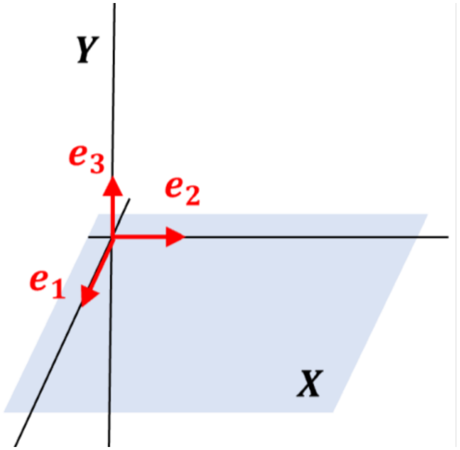
Notation $X \perp Y$ is used to state that X is orthogonal to Y.

Example.



(1) Let $X = \mathrm{Span}\{\mathbf{e_1}, \mathbf{e_2}\}$ and $Y = \mathrm{Span}\{\mathbf{e_1}, \mathbf{e_3}\}$ be two subspaces of $\mathbf{R}^3$.

Are X and Y orthogonal subspaces?



(2) Let $X = \mathrm{Span}\{\mathbf{e_1}, \mathbf{e_2}\}$ and $Y = \mathrm{Span}\{\mathbf{e_3}\}$ be two subspaces of $\mathbf{R}^3$.

Are X and Y orthogonal subspaces?

Theorem 4. Suppose that X and Y are two orthogonal subspaces of $\mathbf{R}^n$, say $X \perp Y$, then

$$X \cap Y = \{\mathbf{0}\}.$$

Definition 4. Let Y be a subspace of $\mathbf{R}^n$. The **orthogonal complement [正交补]** of Y, denoted by $Y^{\perp}$, is the set of all vectors in $\mathbf{R}^n$ which are orthogonal to every vector in Y. Thus

$$Y^{\perp} = \left\{ \boldsymbol{x} \in \mathbf{R}^n | \boldsymbol{x}^{\perp} \boldsymbol{y} = 0, \qquad \forall \boldsymbol{y} \in Y \right\}.$$

Example.

Let $Y = \mathrm{Span}\{\mathbf{e_3}\}$ be the subspace of $\mathbf{R}^3$. Then $Y^{\perp} = \mathrm{Span}\{\mathbf{e_1}, \mathbf{e_2}\}$.

Theorem 5. Let Y be a subspace of $\mathbf{R}^n$. Then $Y^{\perp}$ is a subspace of $\mathbf{R}^n$.

Fundamental Subspaces of a matrix

Given an $m \times n$ matrix A. Recall the **nullspace** of A

$$N(A) = \{\boldsymbol{x} \in \mathbf{R}^n | A\boldsymbol{x} = \mathbf{0}\}$$

is a subspace of $\mathbf{R}^n$.

Definition 5. Let A be an $m \times n$ matrix. The column space of A is called the **range [值域]** of A, denoted by $R(A)$. That is

$$R(A) = \text{Span}\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\} = \{\boldsymbol{b} \in \mathbf{R}^m | \boldsymbol{b} = A\boldsymbol{x}, \boldsymbol{x} \in \mathbf{R}^n\}.$$

Fundamental Subspaces of a matrix

Consider the linear system $A\boldsymbol{x} = \boldsymbol{0}$.

Let $\boldsymbol{a}(i,:)$ be the ith row vector of A, then

$$\boldsymbol{a}(i,:)\boldsymbol{x} = 0$$

holds for all vectors $\boldsymbol{x} \in N(A)$. Then

$$\langle \boldsymbol{a}^T(i,:), \boldsymbol{x} \rangle = 0, \qquad \forall \boldsymbol{x} \in N(A), \qquad \forall i = 1, \dots, m.$$

In other words, $\boldsymbol{a}^T(i,:)$, or the ith column vector of A^T, is orthogonal to vectors in $N(A)$:

$$N(A) \perp R(A^T).$$

Theorem 6. (Fundamental Subspaces Theorem) Let A be an $m \times n$ matrix, then

$$N(A) = R\left(A^T\right)^{\perp}, \qquad N\left(A^T\right) = R(A)^{\perp}.$$

Proof. We have known that $N(A) \perp R\left(A^T\right)$, that is $N(A) \subset R\left(A^T\right)^{\perp}$.

On the other hand, if $\boldsymbol{x} \in R\left(A^T\right)^{\perp}$ and $\boldsymbol{y} \in R(A^T)$,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0.$$

The above equality holds for all $\boldsymbol{y} \in R(A^T)$ which is defined as the column space of A^T, therefore it holds for $\boldsymbol{y} = \boldsymbol{a}^T(i,:),\ i = 1, \ldots, m$,

$$\left\langle \boldsymbol{x}, \boldsymbol{a}^T(i,:) \right\rangle = \boldsymbol{a}(i,:)\boldsymbol{x} = 0.$$

This gives

$$A\boldsymbol{x} = \mathbf{0}$$

and $\boldsymbol{x} \in N(A)$. We then have proved that $R\left(A^T\right)^{\perp} \subset N(A)$. The second statement can be proved similarly.

Example. Let $A = \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}$.

- The column space of A consists of all vectors of the form

$$\begin{pmatrix} \alpha \\ 2\alpha \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

- The nullspace of A^T consists of all vectors of the form $\beta(-2,1)^T$.

Since $(1,2)^T$ and $(-2,1)^T$ are orthogonal, it follows that every vector in $R(A)$ is orthogonal to every vector in $N(A^T)$. Therefore, $R(A) \perp N(A^T)$.

The same relationship holds between $R(A^T)$ and $N(A)$.

Projection of a vector onto a subspace

Theorem 1. Let S be a subspace of an inner product space V and let $\boldsymbol{x} \in V$. Let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_r}\}$ be an orthonormal basis for S. Let

$$\boldsymbol{p} = \sum_{i=1}^{r} \langle \boldsymbol{x}, \boldsymbol{v_i} \rangle \boldsymbol{v_i}\,,$$

then $\boldsymbol{p} - \boldsymbol{x} \in S^{\perp}$.

Theorem 2. Under the hypothesis of the previous theorem, $\boldsymbol{p}$ is the element of S that is closest to $\boldsymbol{x}$; that is

$$\|\boldsymbol{y} - \boldsymbol{x}\| > \|\boldsymbol{p} - \boldsymbol{x}\|,$$

for any $\boldsymbol{y} \neq \boldsymbol{p}$ in S.

The vector $\boldsymbol{p}$ defined above is said to be the **projection** of $\boldsymbol{x}$ onto S.

6.3 The Gram-Schmidt Orthogonalization Process

Construct an orthonormal basis from a given basis.

Let V be an n-dimensional inner product space.
Let $F = \{\boldsymbol{u_1}, \dots, \boldsymbol{u_n}\}$ be a given basis of V.

In general, the vectors $\boldsymbol{u_i}$, $i = 1, \dots, n$ are linearly independent, but they are not necessary to be orthogonal to each other or be unit vectors.

The **goal** of this section is to construct an **orthonormal** basis $E = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$ from F, such that

$$\mathrm{Span}\{\boldsymbol{v_1}, \dots, \boldsymbol{v_k}\} = \mathrm{Span}\{\boldsymbol{u_1}, \dots, \boldsymbol{u_k}\},$$

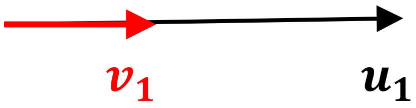
for $k = 1, \dots, n$.

Step 1. Take any vector in F, say $\boldsymbol{u_1}$, and let $\boldsymbol{q_1} = \boldsymbol{u_1}$,

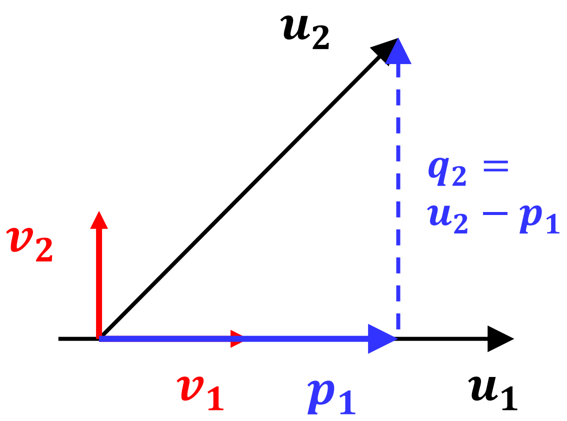
$$\boldsymbol{v_1} = \left(\frac{1}{\|\boldsymbol{q_1}\|}\right)\boldsymbol{q_1}.$$

Then $\boldsymbol{v_1}$ is a unit vector and

$$\mathrm{Span}\{\boldsymbol{v_1}\} = \mathrm{Span}\{\boldsymbol{u_1}\}.$$



Step 2. Since $\boldsymbol{u_2}$ and $\boldsymbol{u_1}$ are linearly independent, $\boldsymbol{u_2} \notin \text{Span}\{\boldsymbol{u_1}\}$.



Let $\boldsymbol{p_1}$ be the projection vector of $\boldsymbol{u_2}$ onto the vector $\boldsymbol{v_1}$:

$$\boldsymbol{p_1} = \langle \boldsymbol{u_2}, \boldsymbol{v_1} \rangle \boldsymbol{v_1}.$$

Let $\boldsymbol{q_2} = \boldsymbol{u_2} - \boldsymbol{p_1}$. It is easy to see that $\boldsymbol{v_1} \perp \boldsymbol{q_2}$. Therefore, if we take

$$\boldsymbol{v_2} = \left(\frac{1}{\|\boldsymbol{q_2}\|}\right)\boldsymbol{q_2},$$

then $\boldsymbol{v_2}$ is a unit vector and

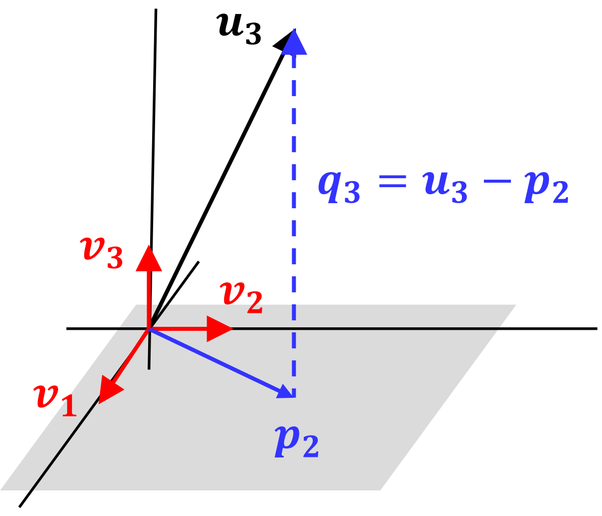
$$\text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\} = \text{Span}\{\boldsymbol{u_1}, \boldsymbol{u_2}\}.$$

Step 3. Since $\boldsymbol{u_3} \notin \text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$, we can take $\boldsymbol{p_2}$ as the projection vector of $\boldsymbol{u_3}$ onto the space $\text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\} = \text{Span}\{\boldsymbol{u_1}, \boldsymbol{u_2}\}$, that is

$$\boldsymbol{p_2} = \langle \boldsymbol{u_3}, \boldsymbol{v_1} \rangle \boldsymbol{v_1} + \langle \boldsymbol{u_3}, \boldsymbol{v_2} \rangle \boldsymbol{v_2}.$$

Let $$\boldsymbol{q_3} = \boldsymbol{u_3} - \boldsymbol{p_2},$$

and $$\boldsymbol{v_3} = \left(\frac{1}{\|\boldsymbol{q_3}\|} \right) \boldsymbol{q_3}.$$



We can show that $\boldsymbol{v_3}$ is a unit vector, $\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}$ are orthogonal to each other, and

$$\text{Span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\} = \text{Span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}.$$

Theorem (The Gram-Schmidt Orthogonalization Process) Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots \boldsymbol{u_n}\}$ be a given basis for the inner product space V. Let

$$\boldsymbol{v_1} = \left(\frac{1}{\|\boldsymbol{u_1}\|}\right)\boldsymbol{u_1}$$

and define $\boldsymbol{v_2}, \dots \boldsymbol{v_n}$ recursively by

$$\boldsymbol{v_{k+1}} = \left(\frac{1}{\|\boldsymbol{u_{k+1}} - \boldsymbol{p_k}\|}\right)(\boldsymbol{u_{k+1}} - \boldsymbol{p_k}) \;\text{ for } k = 1, \dots, n-1$$

where

$$\boldsymbol{p_k} = \langle \boldsymbol{u_{k+1}}, \boldsymbol{v_1}\rangle \boldsymbol{v_1} + \langle \boldsymbol{u_{k+1}}, \boldsymbol{v_2}\rangle \boldsymbol{v_2} + .. + \langle \boldsymbol{u_{k+1}}, \boldsymbol{v_k}\rangle \boldsymbol{v_k}$$

is the projection of $\boldsymbol{u_{k+1}}$ onto Span$\{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k}\}$. Then the set

$$\{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_n}\}$$

is an orthonormal basis for V.

Example. Given a basis of $\mathbf{R}^3$

$$F = \{\boldsymbol{u_1} = (1,1,1)^T, \boldsymbol{u_2} = (1,1,0)^T, \boldsymbol{u_3} = (1,0,0)^T\}.$$

Derive an orthonormal basis from basis F.

Step 1. Set $\boldsymbol{q_1} = \boldsymbol{u_1}$ and $\boldsymbol{v_1} = \left(\frac{1}{\|\boldsymbol{q_1}\|}\right)\boldsymbol{q_1} = \frac{1}{\sqrt{3}}(1,1,1)^T.$

Step 2. Let

$$\boldsymbol{p_1} = \langle \boldsymbol{u_2}, \boldsymbol{v_1} \rangle \boldsymbol{v_1} = \left\langle (1,1,0)^T, \frac{1}{\sqrt{3}}(1,1,1)^T \right\rangle \frac{1}{\sqrt{3}}(1,1,1)^T$$

$$= \frac{2}{3}(1,1,1)^T.$$

$$\boldsymbol{q_2} = \boldsymbol{u_2} - \boldsymbol{p_1} = (1,1,0)^T - \frac{2}{3}(1,1,1)^T = \frac{1}{3}(1,1,-2)^T.$$

$$\boldsymbol{v_2} = \left(\frac{1}{\|\boldsymbol{q_2}\|}\right)\boldsymbol{q_2} = \frac{1}{\sqrt{6}}(1,1,-2)^T.$$

Step 3. Set

$$\boldsymbol{p_2} = \langle \boldsymbol{u_3}, \boldsymbol{v_1} \rangle \boldsymbol{v_1} + \langle \boldsymbol{u_3}, \boldsymbol{v_2} \rangle \boldsymbol{v_2} = \frac{1}{2}(1,1,0)^T,$$

$$\boldsymbol{q_3} = \boldsymbol{u_3} - \boldsymbol{p_2} = \frac{1}{2}(1,-1,0)^T,$$

$$\boldsymbol{v_3} = \left(\frac{1}{\|\boldsymbol{q_3}\|}\right)\boldsymbol{q_3} = \frac{1}{\sqrt{2}}(1,-1,0)^T.$$

The set $E = \{\boldsymbol{v_1} = \frac{1}{\sqrt{3}}(1,1,1)^T, \boldsymbol{v_2} = \frac{1}{\sqrt{6}}(1,1,-2)^T, \boldsymbol{v_3} = \frac{1}{\sqrt{2}}(1,-1,0)^T\}$ is an orthonormal basis.

Review

- Inner Product and Inner Product Space
- Orthonormal Sets and Orthogonal Subspaces
- The Gram-Schmidt Orthogonalization Process

Preview

- Eigenvalues and Eigenvectors
- Diagonalization