Chapter 5

Orthogonality

5.1 The Dot Product in $\mathbb{R}^n$

Definition 5.1.1 (Dot Product)**.** *The **dot product** of two vectors* $\mathbf{x}, \mathbf{y}$ *in* $\mathbb{R}^n$ *is*

$$\mathbf{x}^T\mathbf{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \cdots + x_n y_n.$$

Other notations for the dot product: $\mathbf{x} \cdot \mathbf{y}$ or $\langle \mathbf{x}, \mathbf{y} \rangle$.

Properties: For any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, and $\alpha, \beta \in \mathbb{R}$.

1. Non-negative: $\mathbf{x}^T\mathbf{x} \geq 0$ with the equality holding iff $\mathbf{x} = \mathbf{0}$;
2. Symmetric: $\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$;
3. Bilinear: $(\alpha\mathbf{x} + \beta\mathbf{y})^T\mathbf{z} = \alpha\mathbf{x}^T\mathbf{z} + \beta\mathbf{y}^T\mathbf{z}$.

Definition 5.1.2 (Length)**.** *The Euclidean **length** of a vector* $\mathbf{x} \in \mathbb{R}^n$ *is*

$$\|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^T\mathbf{x}}.$$

Properties: For any $\mathbf{x} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$.

1. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$.
2. $\|\mathbf{x}\| \geq \mathbf{0}$ with equality holding iff $\mathbf{x} = \mathbf{0}$.

Definition 5.1.3 (Unit Vector)**.** *A vector* $\mathbf{u}$ *is called a* ***unit vector*** *if* $\|\mathbf{u}\| = 1$.

Remark. For any nonzero vector $\mathbf{u}$, the vector $\hat{\mathbf{u}} = \frac{1}{\|\mathbf{u}\|}\mathbf{u}$ is a unit vector in the same direction of $\mathbf{u}$. We call $\hat{\mathbf{u}}$ the **normalization** of $\mathbf{u}$.

Definition 5.1.4 (Distance)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be vectors in* $\mathbb{R}^n$. *The Euclidean* ***distance*** *between* $\mathbf{x}$ *and* $\mathbf{y}$ *is defined to be the number* $\|\mathbf{x} - \mathbf{y}\|$.

Example 5.1.5. Let $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$, so $\mathbf{x}+\mathbf{y} = \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}$ and $\mathbf{x}-\mathbf{y} = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$.

$$\mathbf{x}^T\mathbf{y} = \begin{pmatrix} 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} = 1 = \begin{pmatrix} 2 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \mathbf{y}^T\mathbf{x}.$$

$$\mathbf{x}^T\mathbf{x} = \begin{pmatrix} 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = 1^2 + 1^2 + 2^2 = 6, \qquad \mathbf{y}^T\mathbf{y} = 2^2 + (-1)^2 + 0^2 = 5.$$

Thus, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{6}$, $\|\mathbf{y}\| = \sqrt{5}$ and $\|\mathbf{x}+\mathbf{y}\| = \sqrt{13}$. The distance between $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{x} - \mathbf{y}\| = 3$. ∞

Definition 5.1.6 (Angle)**.** *The* **angle** *between two nonzero vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$ *is defined by*

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, \qquad 0 \le \theta \le \pi.$$

Proof. The vectors $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{x}-\mathbf{y}$ may be used to form a triangle. By the law of cosine:

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta,$$

and hence it follows that

$$\begin{aligned}
\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta &= \frac{1}{2}\left(\|\mathbf{x}\|^2+\|\mathbf{y}\|^2-\|\mathbf{x}-\mathbf{y}\|^2\right)\\
&= \frac{1}{2}\left(\|\mathbf{x}\|^2+\|\mathbf{y}\|^2-(\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y})\right)\\
&= \frac{1}{2}\left(\|\mathbf{x}\|^2+\|\mathbf{y}\|^2-(\mathbf{x}^T\mathbf{x}-\mathbf{y}^T\mathbf{x}-\mathbf{x}^T\mathbf{y}+\mathbf{y}^T\mathbf{y})\right)\\
&= \frac{1}{2}\left(2\mathbf{x}^T\mathbf{y}\right)=\mathbf{x}^T\mathbf{y}
\end{aligned}$$

□

If $\mathbf{x}^T\mathbf{y}=0$, then either one of the vectors is $\mathbf{0}$ or $\cos\theta=0$. For $\cos\theta=0$, $\theta=\pi/2$. Specially, when $\theta=\pi/2$, $\|\mathbf{x}-\mathbf{y}\|^2=\|\mathbf{x}\|^2+\|\mathbf{y}\|^2$, which is *the Pythagorean law.*

Example 5.1.7. Let $\mathbf{x}=\begin{pmatrix}1 & 1 & 2\end{pmatrix}^T$, $\mathbf{y}=\begin{pmatrix}2 & -1 & 0\end{pmatrix}^T$ and $\mathbf{z}=\begin{pmatrix}2 & -4 & 1\end{pmatrix}^T$. The angle between $\mathbf{x}$ and $\mathbf{y}$ is

$$\arccos\frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}=\arccos\frac{1}{\sqrt{6}\cdot\sqrt{5}}.$$

The angle between $\mathbf{x}$ and $\mathbf{z}$ is a right angle, since

$$\arccos\frac{\mathbf{x}^T\mathbf{z}}{\|\mathbf{x}\|\|\mathbf{z}\|}=\arccos\frac{0}{\sqrt{6}\cdot\sqrt{21}}=\frac{\pi}{2}.$$

ↂ

Theorem 5.1.8 (Cauchy-Schwarz Inequality)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be vectors in* $\mathbb{R}^n$*. Then*

$$|\mathbf{x}^T\mathbf{y}|\le\|\mathbf{x}\|\|\mathbf{y}\|$$

with equality holding iff one of the vectors is $\mathbf{0}$ *or one vector is a multiple of the other.*

Proof. Many different ways. One method: start with $\|\hat{\mathbf{x}}-\hat{\mathbf{y}}\|^2\ge\mathbf{0}$. □

Definition 5.1.9. *The vectors* $\mathbf{x}$ *and* $\mathbf{y}$ *in* $\mathbb{R}^n$ *are said to be* ***orthogonal****/perpendicular if* $\mathbf{x}^T\mathbf{y}=0$*, which is denoted as* $\mathbf{x}\perp\mathbf{y}$*.*

1. Nonzero vectors $\mathbf{x} \perp \mathbf{y}$ iff the angle between them is $\pi/2$.
2. The zero vector of $\mathbb{R}^n$ is orthogonal to any vectors in $\mathbb{R}^n$.

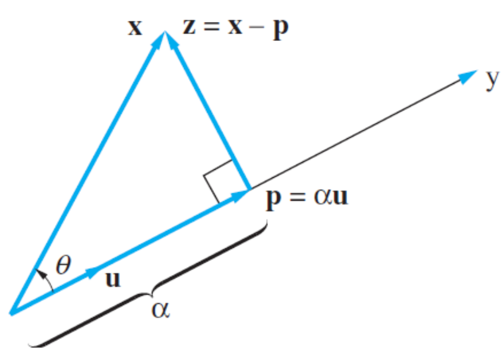
Definition 5.1.10 (Scalar and Vector Projection)**.** *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be vectors in* $\mathbb{R}^n$, $\mathbf{y} \neq \mathbf{0}$. *Besides,* $\mathbf{u}$ *is a unit vector in the direction of* $\mathbf{y}$.

*The **scalar projection** of* $\mathbf{x}$ *onto* $\mathbf{y}$ *is a scalar* $\alpha \in \mathbb{R}$ *given by*

$$\alpha = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}.$$

*The **vector projection** of a vector* $\mathbf{x}$ *onto* $\mathbf{y}$ *is the vector* $\mathbf{p}$ *so that* $(\mathbf{x} - \mathbf{p}) \perp \mathbf{y}$.

$$\mathbf{p} = \alpha\mathbf{u} = \alpha\frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\,\mathbf{y}.$$



$$\begin{aligned}\alpha &= \|\mathbf{x}\|\cos\theta \\ &= \frac{\|\mathbf{x}\|\,\|\mathbf{y}\|\cos\theta}{\|\mathbf{y}\|} \\ &= \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}\end{aligned}$$

Example 5.1.11. In Example 5.4.8, find the vector projection of $\mathbf{x}$ onto $\mathbf{y}$ and the vector projection of $\mathbf{y}$ onto $\mathbf{x}$. How about the vector projection of $\mathbf{x}$ onto $\mathbf{z}$?

$$\text{proj}_{\mathbf{y}}\mathbf{x} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|\|\mathbf{y}\|}\mathbf{y} = \frac{1}{5}\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \quad \text{proj}_{\mathbf{x}}\mathbf{y} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{x}\|}\mathbf{x} = \frac{1}{6}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

The vector projection of $\mathbf{x}$ onto $\mathbf{z}$ is $\mathbf{0}$, since $\mathbf{x} \perp \mathbf{z}$ and $\mathbf{x} \cdot \mathbf{z} = 0$.

Example 5.1.12.

$$A\mathbf{x} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \mathbf{b} \quad \text{has no solution.}$$

Find a $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest vector to $\mathbf{b}$ for all $\mathbf{x} \in \mathbb{R}^2$.

Idea: $A\hat{\mathbf{x}}$ should be the vector projection of $\mathbf{b}$ onto $A\mathbf{x}$ (or $(1,2)^T$).

$$A\hat{\mathbf{x}} = \frac{(1,2)\mathbf{b}}{\|(1,2)^T\|^2}\begin{pmatrix}1\\2\end{pmatrix} = \frac{7}{5}\begin{pmatrix}1\\2\end{pmatrix}.$$

Solve $A\hat{\mathbf{x}} = \frac{7}{5}\begin{pmatrix}1\\2\end{pmatrix}$, we have $\hat{\mathbf{x}} = \begin{pmatrix}5/7\\0\end{pmatrix} + x_2\begin{pmatrix}-2\\1\end{pmatrix}$, where $x_2 \in \mathbb{R}$. ꝏ

Example 5.1.13. Find the distance from a point $(2,0,0)^T$ to the plane $x+2y+2z=0$.

A **normal vector** to the plane is $\mathbf{n} = (1,2,2)'$. The plane passes through the origin. Define a vector $\mathbf{v}$ from the origin to the point $(2,0,0)^T$. Then the distance from the point to the plane is simply the absolute value of the scalar projection of $\mathbf{v}$ onto $\mathbf{n}$.

$$d = \frac{|\mathbf{v}^T\mathbf{n}|}{\|\mathbf{n}\|} = \frac{2}{3}.$$

ꝏ

Exercise 5.1.14. Find the point on the line $y = 2x+1$ that is closest to the point $(5,2)$. (Ans: $(1.4, 3.8)$) ✓

5.2 Orthogonal Subspaces

Definition 5.2.1 (Orthogonal Subspaces)**.** *Two subspaces X and Y of $\mathbb{R}^n$ are said to be* ***orthogonal*** *if every $\mathbf{x} \in X$ is orthogonal to every $\mathbf{y} \in Y$, i.e.*

$$\mathbf{x}^T\mathbf{y} = \mathbf{0}, \qquad \forall\ \mathbf{x} \in X, \quad \forall\ \mathbf{y} \in Y.$$

If subspaces X and Y are orthogonal, we write $X \perp Y$.

Theorem 5.2.2. *If $X \perp Y$, then $X \cap Y = \{\mathbf{0}\}$.*

Definition 5.2.3 (Orthogonal Complement)**.** *Let Y be a subspace of $\mathbb{R}^n$. The **orthogonal complement** of Y is defined as*

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

Example 5.2.4. Let X be the subspace of $\mathbb{R}^3$ spanned by $\mathbf{e}_1$, and let Y be the subspace of $\mathbb{R}^3$ spanned by $\mathbf{e}_2$. If $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, these vectors must be of the form $\mathbf{x} = (x_1, 0, 0)^T$ and $\mathbf{y} = (0, y_2, 0)^T$. Thus,

$$\mathbf{x} \cdot \mathbf{y} = x_1 \cdot 0 + 0 \cdot y_2 + 0 \cdot 0 = 0, \quad \forall \ x_1, y_2 \in \mathbb{R}.$$

The subspaces X and Y are orthogonal, $X \perp Y$, but they are not orthogonal complements. Indeed,

$$X^\perp = \operatorname{span}\{\mathbf{e}_2, \mathbf{e}_3\}, \quad Y^\perp = \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_3\}.$$

ꝏ

Theorem 5.2.5. *Let S be a subspace of $\mathbb{R}^n$. Then*

1. *$S^\perp$ is a subspace of $\mathbb{R}^n$*
2. *$S^\perp \perp S$. In fact, $S^\perp$ is the largest one among all subspaces of $\mathbb{R}^n$ that are orthogonal to S.*
3. $\dim S + \dim S^\perp = n$
4. $\mathbb{R}^n = S \oplus S^\perp$
5. *$(S^\perp)^\perp = S$.Moreover, S and $S^\perp$ are said to be orthogonal complement of each other.*

Proof. (1)-(2): Exercises. (3)-(4) Given in Theorem 5.2.14.

(5). For any vector $\mathbf{v} \in S$, we have $\mathbf{v}^T\mathbf{w} = 0$, $\forall\, \mathbf{w} \in S^{\perp}$. So $\mathbf{v} \in (S^{\perp})^{\perp}$ and $S \subseteq (S^{\perp})^{\perp}$. On the other hand, by part (1) and (3), both S and $S^{\perp}$ are subspaces of $\mathbb{R}^n$, thus,

$$\dim S + \dim S^{\perp} = n \quad \text{and} \quad \dim S^{\perp} + \dim(S^{\perp})^{\perp} = n.$$

$$\to \quad \dim S^{\perp} = n - \dim S = n - \dim(S^{\perp})^{\perp} \quad \to \quad \dim S = \dim(S^{\perp})^{\perp}.$$

Hence, $S = (S^{\perp})^{\perp}$.

$\square$

5.2.1 Sum and Direct Sum

Definition 5.2.6 (Sum)**.** *Let U and V be subspaces of the vector space W. The **sum** of U and V is defined as*

$$U + V = \{\mathbf{u} + \mathbf{v} \in W |\ \mathbf{u} \in U, \mathbf{v} \in V\}.$$

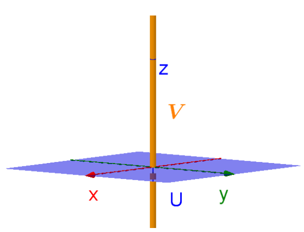
Theorem 5.2.7. *Let U and V be subspaces of W. Then $U + V$ is a subspace of W.*

Example 5.2.8. Consider the following examples of sum of subspaces

$U = \{(x, y, 0)^T | x, y \in \mathbb{R}\}$

$V = \{(0, 0, z)^T | z \in \mathbb{R}\}$

$U + V = \mathbb{R}^3$

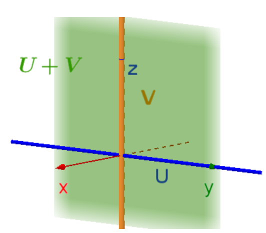


$U = \{(0, y, 0)^T | y \in \mathbb{R}\}$

$V = \{(0, 0, z)^T | z \in \mathbb{R}\}$

$U + V =$

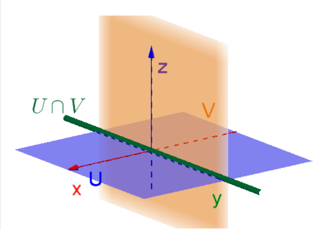
$\{(0, y, z)^T | y, z \in \mathbb{R}\}$



$U = \{(x, y, 0)^T | x, y \in \mathbb{R}\}$

$V = \{(0, y, z)^T | y, z \in \mathbb{R}\}$

$U + V = \mathbb{R}^3$



Theorem 5.2.9. *Let U and V be subspaces of a vector space W. Then*

$$\dim(U + V) \leq \dim U + \dim V.$$

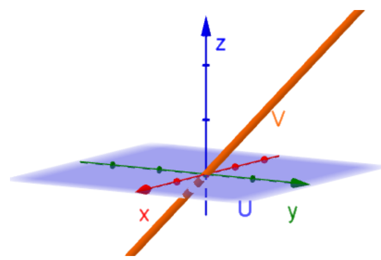
Proof. Consider the dimension theorem: $\dim U + \dim V - \dim(U \cap V) = \dim(U + V)$. While the proof of dimension theorem has been given in Chap 3, LA I. $\square$

Definition 5.2.10 (Direct Sum)**.** *Let U and V be subspaces of the vector space W and $U + V = S$. If every $\mathbf{z} \in S$ can be written uniquely as $\mathbf{z} = \mathbf{u} + \mathbf{v}$ for some $\mathbf{u} \in U$ and $\mathbf{v} \in V$, then we call the sum $U + V$ a **direct sum** and write $U \oplus V$.*

Theorem 5.2.11. *Let U and V be subspaces of the vector space W. Then $U + V$ is a direct sum if and only if $\dim(U + V) = \dim U + \dim V$, and $U \cap V = \{\mathbf{0}\}$.*

Example 5.2.12.

$U = \{(x, y, 0)^T | x, y \in \mathbb{R}\}$
$V = \{(0, a, a)^T | a \in \mathbb{R}\}$
Then $U \oplus V = \mathbb{R}^3$.



By Definition 5.2.10:
$(x, y, z)^T = (x, y - z, 0)^T + (0, z, z)^T$
where
$(x, y - z, 0)^T \in U$ and $(0, z, z)^T \in V$.

By Theorem 5.2.11:
Since $U + V = \mathbb{R}^3$ and $U \cap V = \{\mathbf{0}\}$, $U \oplus V = \mathbb{R}^3$.

5.2.2 The Fundamental Subspaces Theorem

Theorem 5.2.13 (The Fundamental Subspaces Theorem)**.** *For any matrix A, we have*

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

Theorem 5.2.14. *If S is a subspace of $\mathbb{R}^n$, then $\dim S + \dim S^{\perp} = n$. Furthermore, if $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is a basis of S and $\{\mathbf{x}_{r+1}, \cdots, \mathbf{x}_n\}$ is a basis of $S^{\perp}$, then $\{\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{x}_{r+1}, \cdots, \mathbf{x}_n\}$ is a basis for $\mathbb{R}^n$. Namely,*

$$S \oplus S^{\perp} = \mathbb{R}^n.$$

Proof.

1. To prove $\dim S + \dim S^\perp = n$.

   If $S = \mathbb{R}^n$ or $S = \{\mathbf{0}\}$, the proof is trivial. Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be a basis for S, $0 < k < n$. Define an $n \times k$ matrix A whose columns are $\mathbf{v}_1, \cdots, \mathbf{v}_k$.

   $$A = \left[\begin{array}{c|c|c} \mathbf{v}_1 & \cdots & \mathbf{v}_k \end{array} \right]$$

   Then $S = \mathrm{Col}(A)$, which implies $S^\perp = \mathrm{Col}(A)^\perp = \mathrm{N}(A^T)$. Hence, $\dim S^\perp = \dim \mathrm{N}(A^T) = n - \mathrm{rank}(A^T) = n - k$. $\quad \to \quad \dim S + \dim S^\perp = k + n - k = n$.

2. To prove "furthermore" part.

   If $S = \mathbb{R}^n$ or $S = \{\mathbf{0}\}$, the proof is trivial. Otherwise, let $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ be a basis for S and let $\{\mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$ be a basis for $S^\perp$, since $\dim S + \dim S^\perp = n$. The set $\{\mathbf{v}_1, \cdots, \mathbf{v}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_n\}$ is linearly independent and contains n vectors. Therefore, it forms a basis of $\mathbb{R}^n$.

3. To prove $S \oplus S^\perp = \mathbb{R}^n$.

   $\dim S + \dim S^\perp = n$ and $S \cap S^\perp = \{\mathbf{0}\}$. Thus. $S \oplus S^\perp = \mathbb{R}^n$.

□

Theorem 5.2.15. *For any $m \times n$ matrix A, we have*

$$N(A) \oplus Col(A^T) = \mathbb{R}^n, \quad N(A^T) \oplus Col(A) = \mathbb{R}^m.$$

Corollary 5.2.16 (Orthogonal Decomposition Theorem)**.** *Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there exists a unique pair of vectors $\mathbf{p} \in W$ and $\mathbf{z} \in W^\perp$ such that*

$$\mathbf{v} = \mathbf{p} + \mathbf{z}.$$

Namely, any vector in $\mathbb{R}^n$ can be uniquely decomposed into the sum of two vectors, one from W and the other from $W^\perp$.

Example 5.2.17. Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix}$. Find the bases for $\mathrm{N}(A)$, $\mathrm{Col}(A^T)$, $\mathrm{N}(A^T)$, $\mathrm{Col}(A)$ and verify the fundamental subspaces theorem.

The reduced row echelon form of A is

$$A \xrightarrow[-3R_2+R_3\to R_3]{-R_2+R_1\to R_1} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \xrightarrow{-R_1+R_3\to R_3} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

$\mathrm{N}(A) = \mathrm{span}\{(1,1,-1)'\}$, $\mathrm{Col}(A^T) = \mathrm{Row}(A) = \mathrm{span}\{(1,0,1)', (0,1,1)'\}$. Clearly, $\mathrm{Col}(A^T) = \mathrm{N}(A)^\perp$. The reduced row echelon form of A^T is

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ 2 & 1 & 4 \end{bmatrix} \xrightarrow[-2R_1+R_3\to R_3]{-R_1+R_2\to R_2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{-R_2+R_3\to R_3} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

$\mathrm{N}(A^T) = \mathrm{span}\{(1,2,-1)'\}$, $\mathrm{Col}(A) = \mathrm{Row}(A^T) = \mathrm{span}\{(1,0,1)', (0,1,2)'\}$. Clearly, $\mathrm{Col}(A) = \mathrm{N}(A^T)^\perp$. ↂ

Example 5.2.18. Let S be the subspace of $\mathbb{R}^3$ spanned by $\mathbf{a}_1 = \begin{pmatrix} 1 & 0 & 2 \end{pmatrix}^T$ and $\mathbf{a}_2 = \begin{pmatrix} 0 & 1 & 5 \end{pmatrix}^T$. To find a basis for $S^\perp$, let $A = \left[\begin{array}{c|c} \mathbf{a}_1 & \mathbf{a}_2 \end{array}\right]$, then $\mathrm{Col}(A) = S$ which implies $S^\perp = \mathrm{Col}(A)^\perp = \mathrm{N}(A^T)$.

The solutions of $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 5 \end{bmatrix} \mathbf{x} = \mathbf{0}$ are of the form $\begin{pmatrix} -2\alpha \\ -5\alpha \\ \alpha \end{pmatrix}$ and $S^\perp = \mathrm{span}\left\{ \begin{pmatrix} -2 \\ -5 \\ 1 \end{pmatrix} \right\}$. ↂ

Theorem 5.2.19. *For any $m \times n$ matrix A, we have*

$$N(A^T A) = N(A), \quad N(AA^T) = N(A^T),$$

and

$$Col(A^T A) = Col(A^T), \quad Col(AA^T) = Col(A).$$

Notice that A^TA and AA^T are always square matrices for any matrix A.

Proof. We will prove $\mathrm{N}(A^TA) = \mathrm{N}(A)$ and $\mathrm{Col}(A^TA) = \mathrm{Col}(A^T)$. The rest are similar.

(a). To prove $\mathrm{N}(A) \subseteq \mathrm{N}(A^TA)$, that is, for all $\mathbf{x} \in \mathrm{N}(A)$, show that $\mathbf{x} \in \mathrm{N}(A^TA)$.

If $\mathbf{x} \in \mathrm{N}(A)$, then $A\mathbf{x} = \mathbf{0}$. $\rightarrow$ $A^TA\mathbf{x} = A^T\mathbf{0} = \mathbf{0}$ $\rightarrow$ $\mathbf{x} \in \mathrm{N}(A^TA)$.

(b). To prove $\mathrm{N}(A^TA) \subseteq \mathrm{N}(A)$, that is, for all $\mathbf{y} \in \mathrm{N}(A^TA)$, show that $\mathbf{y} \in \mathrm{N}(A)$.

If $\mathbf{y} \in \mathrm{N}(A^TA)$, then $(A^TA)\mathbf{y} = \mathbf{0}$. $\rightarrow$ $A^T(A\mathbf{y}) = \mathbf{0}$ $\rightarrow$ $A\mathbf{y} \in \mathrm{N}(A^T)$.

Meanwhile, $A\mathbf{y} \in \mathrm{Col}(A)$. Since $\mathrm{Col}(A) \cap \mathrm{N}(A^T) = \{\mathbf{0}\}$, so $A\mathbf{y} = \mathbf{0}$ $\rightarrow$ $\mathbf{y} \in \mathrm{N}(A)$.

Based on (a) and (b), $\mathrm{N}(A^TA) = \mathrm{N}(A)$.

(i). To prove $\mathrm{Col}(A^TA) \subseteq \mathrm{Col}(A^T)$.

For any $\mathbf{b} \in \mathrm{Col}(A^TA)$, there exists at least one $\mathbf{x} \in \mathbb{R}^n$ such that $A^TA\mathbf{x} = \mathbf{b}$.
$\rightarrow$ $A^T(A\mathbf{x}) = \mathbf{b}$ $\rightarrow$ $\mathbf{b} \in \mathrm{Col}(A^T)$.

(ii). To prove $\dim[\mathrm{Col}(A^TA)] = \dim[\mathrm{Col}(A^T)]$, use the rank-nullity theorem.

$$\begin{cases} \mathrm{N}(A^TA) = \mathrm{N}(A) \\ A^TA \text{ and } A \text{ both have } n \text{ columns.} \end{cases} \rightarrow \quad \mathrm{rank}(A^TA) = \mathrm{rank}(A) = \mathrm{rank}(A^T).$$

Based on (i) and (ii), $\mathrm{Col}(A^TA) = \mathrm{Col}(A^T)$. $\square$

Corollary 5.2.20. *Let A be an $m \times n$ matrix. Then*

1. *A^TA is nonsingular if and only if rank$(A) = n$,*
2. *AA^T is nonsingular if and only if rank$(A) = m$.*

Theorem 5.2.21. *Suppose A is an $m \times n$ matrix of rank r.*

(a). If $\{v_1, \dots, v_r\}$ is a basis of Col(A^T), then $\{Av_1, \dots, Av_r\}$ is a basis of Col(A).

(b). If $\{u_1, \dots, u_r\}$ is a basis of Col(A), then $\{A^Tu_1, \dots, A^Tu_r\}$ is a basis of Col(A^T).

Proof. Given $\text{rank}(A) = r$ and $A\mathbf{v}_i \in \text{Col}(A)$, only need to show $A\mathbf{v}_1, \cdots A\mathbf{v}_r$ are linearly independent.

Let $c_1 A\mathbf{v}_1 + \cdots c_r A\mathbf{v}_r = \mathbf{0}$, then $A(c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r) = \mathbf{0}$ and $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r \in \text{N}(A)$.

Meanwhile $\{v_1, \ldots, v_r\}$ is a basis of $\text{Col}(A^T)$, so $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r \in \text{Col}(A^T)$.

Because $\text{Col}(A^T) \cap \text{N}(A) = \{\mathbf{0}\}$, we have $c_1\mathbf{v}_1 + \cdots + c_r\mathbf{v}_r = \mathbf{0}$.

Since $\mathbf{v}_1, \cdots \mathbf{v}_r$ are linearly independent, then $c_1 = \cdots = c_r = 0$. Part (a) is proved. □

5.3 Least Squares Problems

Motivation

A standard technique in mathematical and statistical modeling is to find a least squares fit to a set of data points in the plane. The least squares curve is usually the graph of a standard type of function, such as a linear function, a polynomial, or a trigonometric polynomial. Since the data may include errors in measurement or experiment-related inaccuracies, we do not require the curve to pass through all the data points. Instead, squares of errors between the y values of the data points and the corresponding y values of the approximating curve are minimized.

A least squares problem can generally be formulated as an overdetermined linear system of equations.

Let A be an $m \times n$ matrix ($m \geq n$) and $\mathbf{b} \in \mathbb{R}^m$. Consider the linear system $A\mathbf{x} = \mathbf{b}$. If the system has no solution, it is called inconsistent. In this case, we wish to find "approximate" solutions, that is, to find a vector $\hat{\mathbf{x}}$ for which $A\hat{\mathbf{x}}$ is "closest" to $\mathbf{b}$.

Definition 5.3.1. *For any $\mathbf{x} \in \mathbb{R}^n$, let's define the **residual** as*

$$\mathbf{r}(\mathbf{x}) = \mathbf{b} - A\mathbf{x}.$$

Definition 5.3.2. *A **least squares solution** of the linear system $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that*

$$\|\mathbf{r}(\hat{\mathbf{x}})\| = \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{r}(\mathbf{x})\|.$$

Notice that minimizing $\|\mathbf{r}(\mathbf{x})\|$ *is equivalent to minimizing* $\|r(\mathbf{x})\|^2$.

Theorem 5.3.3. *Let* S *be a subspace of* $\mathbb{R}^m$. *For each* $\mathbf{b} \in \mathbb{R}^m$, *there exists a unique* $\mathbf{p} \in S$ *that is closest to* $\mathbf{b}$, *in the sense that*

$$\|\mathbf{b}-\mathbf{p}\| < \|\mathbf{b}-\mathbf{y}\|, \qquad \textit{for any } \mathbf{y} \in S,\ \mathbf{y} \neq \mathbf{p}.$$

In fact, $\mathbf{p}$ *is the unique element in* S *such that* $\mathbf{b}-\mathbf{p} \in S^\perp$. *The vector* $\mathbf{p}$ *is called the* ***projection*** *of* $\mathbf{b}$ *onto* S.

Proof. Since $S \oplus S^\perp = \mathbb{R}^m$, then each $\mathbf{b} \in \mathbb{R}^m$ can be decomposed uniquely as

$$\mathbf{b} = \mathbf{p} + \mathbf{z},$$

where $\mathbf{p} \in S$ and $\mathbf{z} \in S^\perp$. Let $\mathbf{y}$ be any vector in S and $\mathbf{y} \neq \mathbf{p}$, then

$$\|\mathbf{b}-\mathbf{y}\|^2 = \|(\mathbf{b}-\mathbf{p})+(\mathbf{p}-\mathbf{y})\|^2 = \|\mathbf{b}-\mathbf{p}\|^2+\|\mathbf{p}-\mathbf{y}\|^2+2(\mathbf{b}-\mathbf{p})^T(\mathbf{p}-\mathbf{y}) = \|\mathbf{b}-\mathbf{p}\|^2+\|\mathbf{p}-\mathbf{y}\|^2,$$

since $\mathbf{p}-\mathbf{y} \in S$ and $(\mathbf{b}-\mathbf{p}) \in S^\perp$. Since $\mathbf{y} \neq \mathbf{p}$, then $\|\mathbf{p}-\mathbf{y}\| > 0$ and

$$\|\mathbf{b}-\mathbf{y}\|^2 = \|\mathbf{b}-\mathbf{p}\|^2 + \|\mathbf{p}-\mathbf{y}\|^2 > \|\mathbf{b}-\mathbf{p}\|^2.$$

□

When it comes to a least squares problem, a vector $\hat{\mathbf{x}}$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{p} = A\hat{\mathbf{x}}$ is a vector in $\mathrm{Col}(A)$ that is closest to $\mathbf{b}$. Namely, $\mathbf{p} = A\hat{\mathbf{x}}$ is the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}(A)$, a subspace of $\mathbb{R}^m$.

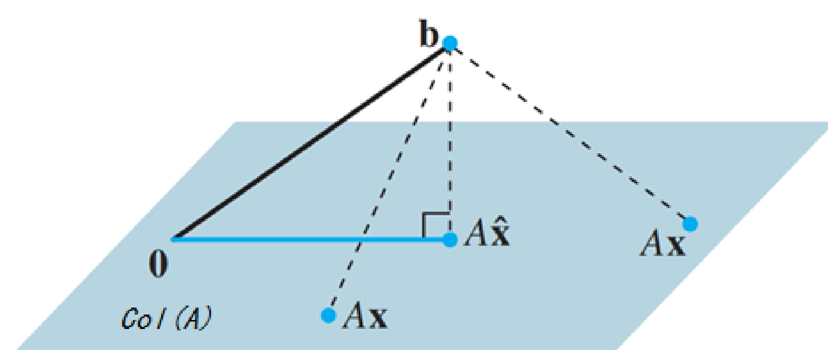


Figure 5.1: The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Theorem 5.3.4. *Let A be an $m \times n$ matrix with $m \geq n$. For each $\mathbf{b} \in \mathbb{R}^m$, there exists a a unique $\mathbf{p} = A\hat{\mathbf{x}} \in Col(A)$ that is closest to $\mathbf{b}$, in the sense that*

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| < \|\mathbf{b} - A\mathbf{x}\|, \qquad \text{for any } A\mathbf{x} \in Col(A),\ A\mathbf{x} \neq A\hat{\mathbf{x}}.$$

Furthermore, the residual $\mathbf{r}(\hat{\mathbf{x}}) = \mathbf{b} - A\hat{\mathbf{x}} \in Col(A)^{\perp} = N(A^T)$.

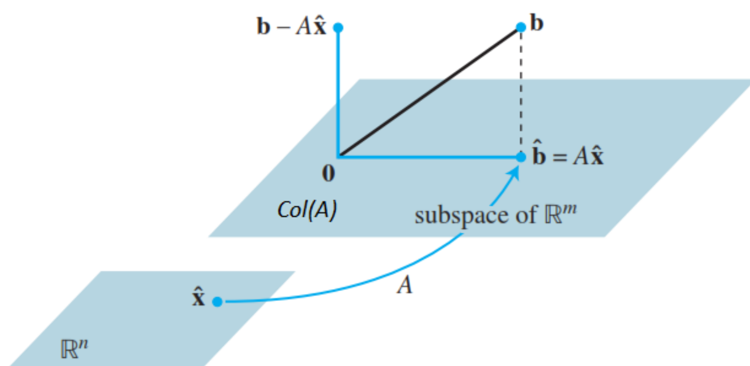


Figure 5.2: The least squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Question: How to find the orthogonal projection of a vector onto a subspace?

Theorem 5.3.5. *If A is an $m \times n$ matrix of rank n, the* ***normal equations***

$$A^T A\mathbf{x} = A^T\mathbf{b},$$

have a unique solution

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

and $\hat{\mathbf{x}}$ is the unique least squares solution of the system $A\mathbf{x} = \mathbf{b}$.

Proof. Since $\text{rank}(A) = n$, then $A^T A$ is invertible (*Prove as an exercise*).
By Theorem 5.3.4, $\mathbf{b} - A\hat{\mathbf{x}} \in \mathrm{N}(A^T)$,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0} \quad \leftrightarrow \quad A^T\mathbf{b} = A^T A\hat{\mathbf{x}} \quad \leftrightarrow \quad \hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}.$$

$\square$

Definition 5.3.6 (Projection Vector and Projection Matrix)**.**

The **projection vector** *is the orthogonal projection of* $\mathbf{b}$ *onto* $Col(A)$*:*

$$\mathbf{p} = A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$$

Among all vectors in $Col(A)$*,* $\mathbf{p}$ *is the closest one to* $\mathbf{b}$ *in the least squares sense.*

The **projection matrix***, or hat matrix, projects* $\mathbf{b}$ *onto* $Col(A)$*.*

$$P = A(A^TA)^{-1}A^T$$

Namely, the projection matrix maps any $\mathbf{b} \in \mathbb{R}^m$ *to its orthogonal projection on* $Col(A)$*.*

Example 5.3.7. Find the least squares solution of the system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}.$$

Determine the least square error $\|\mathbf{r}(\hat{\mathbf{x}})\|$ where $\mathbf{r}(\hat{\mathbf{x}}) = \mathbf{b} - A\hat{\mathbf{x}}$.

Solution: We have $\operatorname{rank}(A) = 2$, then

$$A^TA = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}, \quad A^T\mathbf{b} = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}.$$

Then

$$\begin{aligned} \hat{\mathbf{x}} &= (A^TA)^{-1}A^T\mathbf{b} \\ &= \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} 19 \\ 11 \end{pmatrix} \\ &= \frac{1}{84}\begin{pmatrix} 5 & -1 \\ -1 & 17 \end{pmatrix}\begin{pmatrix} 19 \\ 11 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{aligned}$$

So the least squares solution is $\hat{\mathbf{x}} = (1,2)^T$ and the least squares error is

$$\|\mathbf{r}(\hat{\mathbf{x}})\| = \|\mathbf{b} - A\hat{\mathbf{x}}\| = \left\| \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} - \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\| = \sqrt{84}.$$

ↂ

Example 5.3.8. Find all least squares solutions of the system $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{pmatrix} 4 & -8 \\ 0 & 0 \\ 1 & -2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}.$$

Can we still use $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$?

Solution: No, because

Least squares solutions: $\hat{\mathbf{x}} = \{(\frac{19}{17} + 2\alpha, \alpha)^T \,|\, \alpha \in \mathbb{R}\}$. Projection vector: $A\hat{\mathbf{x}} = \frac{19}{17}\left(\begin{smallmatrix} 4 \\ 0 \\ 1 \end{smallmatrix}\right)$.

ↂ

5.3.1 Data Fitting

Given a table of data

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

y_i: the i-th observed value of y

$\hat{y}_i = \beta_0 + \beta_1 x_i$: the i-th predicted value by a linear function $y = \beta_0 + \beta_1 x$.

$y_i - \hat{y}_i$: the residual related to the i-th data.

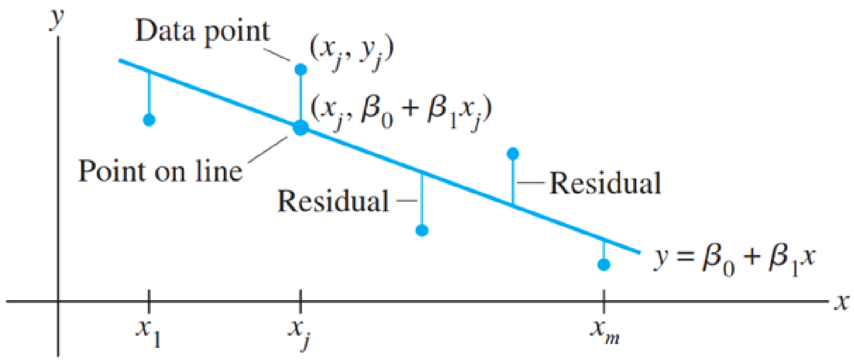


Figure 5.3: Fitting a linear line to observational data.

We wish to find a linear function of the form $y = \beta_0 + \beta_1 x$ which minimizes the sum of squares of the residuals, $\sum_{k=1}^{m} (y_k - \hat{y}_k)^2$.

The best least squares fit by **a linear function** $y = \beta_0 + \beta_1 x$: for all data, we have

$$\hat{\mathbf{y}} = \beta_0 + \beta_1 \mathbf{x}.$$

such that

$$\sum_{i=1}^{m} |\mathbf{y}_i - \hat{\mathbf{y}}_i|^2$$

is minimized over every choice of β_0, β_1. Namely, to find the least squares solution β_0, β_1 of the following system $A\mathbf{x} = \mathbf{b}$

$$\underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}}_{\mathbf{b}}.$$

Since the data are observed at distinct points, $x_i \neq x_i$ if $i \neq j$, then $\text{rank}(A) = 2$ and

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (A^T A)^{-1} A^T \mathbf{b}.$$

Example 5.3.9. Given the data

x	0	3	6
y	1	4	5

Find the best least squares fit by a linear function, the least square residual $\mathbf{r}(\hat{\mathbf{x}})$ and the error $\|\mathbf{r}(\hat{\mathbf{x}})\|$.

Solution: Let

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 4 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Then

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} = \begin{bmatrix} 3 & 9 \\ 9 & 45 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 42 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

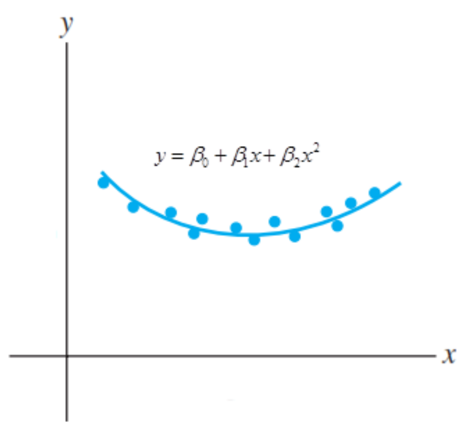
The best least squares fit by a linear function is $y = \frac{4}{3} + \frac{2}{3}x$.

The least square residual is $-\frac{1}{3}(1, -2, 1)^T$ and the error is $\sqrt{\frac{2}{3}}$. ∞

One can also fit the data by other types of functions in the least squares sense.

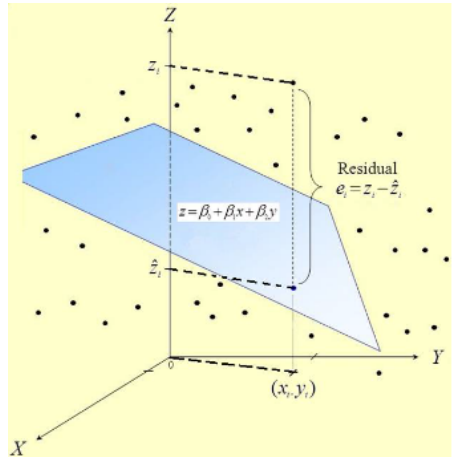
- The best least squares fit by **a quadratic function** $y = \beta_0 + \beta_1 x + \beta_2 x^2$, with

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$



- The best least squares fit by **a plane** in $\mathbb{R}^3$, $z = \beta_0 + \beta_1 x + \beta_2 y$, with

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & y_m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}$$



- The least squares fit by a higher degree polynomial, a circle, a trigonometric function, a hyperplane in $\mathbb{R}^n$...

Example 5.3.10. Find the best quadratic least squares fit to the data

x	0	1	2	3
y	3	2	4	4

Solution: Let

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix}^T, \quad \mathbf{b} = \begin{bmatrix} 3 & 2 & 4 & 4 \end{bmatrix}^T \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}.$$

Then the normal equations are $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$, i.e.

$$\begin{bmatrix} 4 & 6 & 14 \\ 6 & 14 & 36 \\ 14 & 36 & 98 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 13 \\ 22 \\ 54 \end{bmatrix} \quad \to \quad \hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{4}\begin{bmatrix} 11 \\ -1 \\ 1 \end{bmatrix}.$$

The best least squares fit by a quadratic function is $y = \frac{11}{4} - \frac{1}{4}x + \frac{1}{4}x^2$.

Exercise 5.3.11. Find the best least squares fitted plane to the data

(x, y)	$(0, 0)$	$(1, 0)$	$(0, 1)$	$(1, 1)$
z	2	3	5	7

(Ans: $z = 1.75 + 1.5x + 3.5y$.) ✓

5.4 Inner Product Spaces

5.4.1 Inner Products

Definition 5.4.1 (Inner Product and Inner Product Space)**.** *An* ***inner product*** *on a vector space V is a function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ with the following properties:*
For any $\mathbf{u}, \mathbf{v}$ and $\mathbf{w}$ in V, and any $\alpha, \beta \in \mathbb{R}$,

1. Non-negative: $\langle \mathbf{u}, \mathbf{u}\rangle \geq 0$ *with equality holding only if* $\mathbf{u} = \mathbf{0}$*;*

2. Symmetric: $\langle \mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{v}, \mathbf{u}\rangle$*;*

3. Bilinear: $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w}\rangle = \alpha\langle \mathbf{u}, \mathbf{w}\rangle + \beta\langle \mathbf{v}, \mathbf{w}\rangle$*.*

A vector space with an inner product is called an ***inner product space****.*

Example 5.4.2 ($\mathbb{R}^n$)**.** The dot product in $\mathbb{R}^n$ defined as

$$\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y} = x_1y_1 + \cdots + x_ny_n, \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

satisfies all properties as an inner product. Proofs were given in the Definition 5.1.1. ⚭

Example 5.4.3 ($\mathbb{R}^n$)**.** The weighted dot product in $\mathbb{R}^n$ can be defined as

$$\langle \mathbf{x}, \mathbf{y}\rangle = w_1x_1y_1 + w_2x_2y_2 + \cdots + w_nx_ny_n, \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$$

where $w_1, \cdots, w_n$ are positive weights. The weighted dot product is an inner product.

⚭

Proof. For any $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ in $\mathbb{R}^n$, and any $\alpha, \beta \in \mathbb{R}$,

1. $\langle \mathbf{x}, \mathbf{x}\rangle = w_1x_1^2 + w_2x_2^2 + \cdots + w_nx_n^2 \geq 0$. Since w_i are positive, $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ only if all x_i are zeros, that is, $\mathbf{x} = \mathbf{0}$;

2. $\langle \mathbf{x}, \mathbf{y}\rangle = w_1x_1y_1 + w_2x_2y_2 + \cdots + w_nx_ny_n = \langle \mathbf{y}, \mathbf{x}\rangle$;

3. $$\begin{aligned}\langle \alpha\mathbf{x}+\beta\mathbf{y},\mathbf{z}\rangle &= w_1(\alpha x_1+\beta y_1)z_1+w_2(\alpha x_2+\beta y_2)z_2+\cdots+w_n(\alpha x_n+\beta y_n)z_n\\ &= \alpha(w_1x_1z_1+\cdots+w_nx_nz_n)+\beta(w_1y_1z_1+\cdots+w_ny_nz_n)\\ &= \alpha\langle\mathbf{x},\mathbf{z}\rangle+\beta\langle\mathbf{y},\mathbf{z}\rangle.\end{aligned}$$

Thus, the weighted dot product is an inner product on $\mathbb{R}^n$. □

It is possible to define many different inner products on a vector space.

Example 5.4.4 ($C[a,b]$)**.** An inner product on $C[a,b]$ can be defined as

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

If $w(x)$ is a positive continuous function on $[a,b]$, then

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

also defines an inner product on $C[a,b]$. The function $w(x)$ is called a weight function.

ↂ

Proof. Verify the weighted inner product: $\forall\, f,g$ and $h\in C[a,b]$, any $\alpha,\beta\in\mathbb{R}$,

1. Since $w(x)[f(x)]^2$ is nonnegative and continuous on $[a,b]$, then the area between the curve of $w(x)[f(x)]^2$ and x-axis on the interval $[a,b]$ should be nonnegative, i.e., $\langle f,f\rangle=\int_a^b w(x)[f(x)]^2\ \mathrm{d}x\ge 0$. Further, $\langle f,f\rangle=0$ only if $f(x)\equiv 0$;

2. $\langle f,g\rangle=\int_a^b w(x)f(x)g(x)\ \mathrm{d}x=\langle g,f\rangle$;

3. Because $f(x)$, $g(x)$ and $w(x)$ are continuous on $[a,b]$, then

$$\begin{aligned}\langle \alpha f(x)+\beta g(x),h(x)\rangle &= \int_a^b w(x)[\alpha f(x)+\beta g(x)]h(x)\ \mathrm{d}x\\ &= \int_a^b \alpha w(x)f(x)h(x)\ \mathrm{d}x+\int_a^b \beta w(x)g(x)h(x)\ \mathrm{d}x\\ &= \alpha\langle f,h\rangle+\beta\langle g,h\rangle.\end{aligned}$$

Thus, this is an inner product on $C[a,b]$. By letting $w(x)\equiv 1$, $\langle f,g\rangle=\int_a^b f(x)g(x)\ \mathrm{d}x$ is also an inner product on $C[a,b]$. □

Example 5.4.5 (P_n**). Let x_1, $x_2, \cdots, x_n$ be distinct real numbers. For any polynomials $p(x)$ and $q(x)$ in P_n, we may define an inner product by

$$\langle p, q\rangle = \sum_{i=1}^{n} p(x_i)q(x_i).$$

If $w(x)$ is a positive weight function, then

$$\langle p, q\rangle = \sum_{i=1}^{n} w(x_i)p(x_i)q(x_i)$$

also defines an inner product on P_n. ꝏ

5.4.2 Norms

The word *norm* in mathematics has its own meaning that is independent of an inner product and its use here should be justified.

Definition 5.4.6 (Norm and Normed Space)**.** *A* ***norm*** *on a vector space V is a function $\|\cdot\| : V \to \mathbb{R}$ with the following properties: For any $\mathbf{u}, \mathbf{v} \in V$ and any $\alpha \in \mathbb{R}$,*

1. $\|\mathbf{u}\| \geq 0$ *with equality holding only if* $\mathbf{u} = \mathbf{0}$*;*
2. $\|\alpha\mathbf{u}\| = |\alpha|\|\mathbf{u}\|$*;*
3. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$*.*

A vector space with a norm is called a ***normed space****.*
The ***distance*** *between $\mathbf{u}, \mathbf{v} \in V$ is $\|\mathbf{u} - \mathbf{v}\|$.*

Theorem 5.4.7 (Norm Induced by the Inner Product)**.** *If V is an inner product space, then the following equation defines a norm on V*

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}, \quad \textit{for all } \mathbf{v} \in V.$$

Example 5.4.8 ($C[-L, L]$)**.** Define an inner product on $C[-L, L]$ as

$$\langle f, g\rangle = \frac{1}{L}\int_{-L}^{L} f(x)g(x)\,\mathrm{d}x,$$

then define the norm by the inner product $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$. Let $m \in \mathbb{Z}^+$, then the followings are unit vectors in this inner product space:

$$
\begin{aligned}
\left\|\frac{1}{\sqrt{2}}\right\| &= \sqrt{\left\langle \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right\rangle} = \sqrt{\frac{1}{L}\int_{-L}^{L} \frac{1}{2}\,\mathrm{d}x} = 1 \\
\left\|\cos\frac{m\pi x}{L}\right\| &= \sqrt{\frac{1}{L}\int_{-L}^{L} \cos^2\frac{m\pi x}{L}\,\mathrm{d}x} = \sqrt{\frac{1}{2L}\int_{-L}^{L} \cos\frac{2m\pi x}{L} + 1\,\mathrm{d}x} = 1 \\
\left\|\sin\frac{m\pi x}{L}\right\| &= \sqrt{\frac{1}{L}\int_{-L}^{L} \sin^2\frac{m\pi x}{L}\,\mathrm{d}x} = 1.
\end{aligned}
$$

ↂ

It is possible to define many different norms on a given vector space.

Example 5.4.9 ($\mathbb{R}^n$)**.** In $\mathbb{R}^n$, for any $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$, we could define

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

Another important norm on $\mathbb{R}^n$ is the *infinity norm* defined by

$$\|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|$$

More generally, we could define a p-norm on $\mathbb{R}^n$ by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

for any real number $p \ge 1$. In particular, if $p = 2$, then $\|\cdot\|_2$ is the norm on $\mathbb{R}^n$ derived from the inner product. If $p \ne 2$, $\|\cdot\|_p$ does not correspond to any inner product. ↂ

Example 5.4.10. Consider $\mathbf{x} = (4, -5, 3)^T \in \mathbb{R}^3$. Compute $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$, and $\|\mathbf{x}\|_\infty$.

$$
\begin{aligned}
\|\mathbf{x}\|_1 &= |4| + |-5| + |3| = 12 \\
\|\mathbf{x}\|_2 &= \sqrt{16+25+9} = 5\sqrt{2} \\
\|\mathbf{x}\|_2 &= \max |4|, |-5|, |3| = 5
\end{aligned}
$$

ↂ

5.5 Orthonormal Sets

Definition 5.5.1 (Orthogonal Set). *A set $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of nonzero vectors in an inner product space is said to be **orthogonal** if $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ for every $i \neq j$.*

Theorem 5.5.2. *If $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is an orthogonal set, then it is linearly independent.*

Proof. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ are mutually orthogonal nonzero vectors and

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

For $1 \leq j \leq n$, take the inner product of $\mathbf{v}_j$ with both sides of the equation and we have

$$c_1\langle \mathbf{v}_1, \mathbf{v}_j \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_j \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_j \rangle = \langle \mathbf{0}, \mathbf{v}_j \rangle = 0$$

Therefore, $c_j\|\mathbf{v}_j\|^2 = 0$ and $c_j = 0$. Hence all scalars $c_1, c_2, \cdots, c_n$ must be 0. □

Example 5.5.3. The set $\left\{(1,1,1)^T, (2,1,-3)^T, (4,-5,1)^T\right\}$ is orthogonal in $\mathbb{R}^3$, since

$$(1,1,1)(2,1,-3)^T = 0, \quad (1,1,1)(4,-5,1)^T = 0,$$
$$(2,1,-3)(4,-5,1)^T = 0.$$

ↀ

Theorem 5.5.4. *Use a constant weight $w(x) = 1/L$ to define an inner product on $C[-L, L]$:*

$$\langle f, g \rangle = \frac{1}{L}\int_{-L}^{L} f(x)g(x)\ dx$$

Let $W_1 = \left\{ f \in C[-L, L] \,\middle|\, f(-x) = -f(x) \text{ for all } x \right\}$ be subspace of odd functions and

$$W_2 = \left\{ g \in C[-L, L] \,\middle|\, g(-x) = g(x) \text{ for all } x \right\} \text{ be subspace of even functions,}$$

then $W_1 \perp W_2$. That means $\langle f, g \rangle = 0$ for any $f \in W_1$ and $g \in W_2$.

Example 5.5.5. For the vector space $C[-L, L]$, if we define an inner product as

$$\langle f, g\rangle = \frac{1}{L}\int_{-L}^{L} f(x)g(x)\,\mathrm{d}x$$

then for any positive integer n, the following set is an orthogonal set.

$$\left\{\frac{1}{\sqrt{2}}, \cos\frac{\pi x}{L}, \sin\frac{\pi x}{L}, \cdots, \cos\frac{n\pi x}{L}, \sin\frac{n\pi x}{L}\right\}$$

For any positive integers m and n,

$$\begin{aligned}
\langle\frac{1}{\sqrt{2}}, \sin\frac{m\pi x}{L}\rangle &= \frac{1}{\sqrt{2}L}\int_{-L}^{L}\sin\frac{m\pi x}{L}\,\mathrm{d}x = 0\\
\langle\frac{1}{\sqrt{2}}, \cos\frac{n\pi x}{L}\rangle &= \frac{1}{\sqrt{2}L}\int_{-L}^{L}\cos\frac{n\pi x}{L}\,\mathrm{d}x = 0\\
\langle\cos\frac{n\pi x}{L}, \sin\frac{m\pi x}{L}\rangle &= \frac{1}{L}\int_{-L}^{L}\cos\frac{n\pi x}{L}\sin\frac{m\pi x}{L}\,\mathrm{d}x = 0.
\end{aligned}$$

For positive integers $m \neq n$,

$$\begin{aligned}
\langle\cos\frac{m\pi x}{L}, \cos\frac{n\pi x}{L}\rangle &= \frac{1}{L}\int_{-L}^{L}\cos\frac{m\pi x}{L}\cos\frac{n\pi x}{L}\,\mathrm{d}x\\
&= \frac{1}{2L}\int_{-L}^{L}\cos\frac{(m-n)\pi x}{L} + \cos\frac{(m+n)\pi x}{L}\,\mathrm{d}x\\
&= \frac{1}{2(m-n)\pi}\left[\sin\frac{(m-n)\pi x}{L}\right]_{-L}^{L} + \frac{1}{2(m+n)\pi}\left[\sin\frac{(m+n)\pi x}{L}\right]_{-L}^{L}\\
&= 0\\
\langle\sin\frac{n\pi x}{L}, \sin\frac{m\pi x}{L}\rangle &= \frac{1}{L}\int_{-L}^{L}\sin\frac{n\pi x}{L}\sin\frac{m\pi x}{L}\,\mathrm{d}x = 0.
\end{aligned}$$

ↂ

Definition 5.5.6 (Orthogonal Basis)**.** *A basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is called an **orthogonal basis** if it is an orthogonal set.*

Definition 5.5.7 (Orthonormal Set)**.** *A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ of nonzero vectors in an inner product space is said to be **orthonormal** if $\langle\mathbf{u}_i, \mathbf{u}_j\rangle = 0$ for every $i \neq j$ and $\|\mathbf{u}_i\| = 1$ for every i.*

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is an orthogonal set, then

$$\left\{\frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \ldots, \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|}\right\}$$

is an orthonormal set.

Definition 5.5.8 (Orthonormal Basis). *A basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is called an* ***orthonormal basis*** *if it is an orthonormal set.*

Example 5.5.9. Let $\mathbf{v}_1 = (1,1,1)^T$, $\mathbf{v}_2 = (2,1,-3)^T$, and $\mathbf{v}_3 = (4,-5,1)^T$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set in $\mathbb{R}^3$. To form an orthonormal set,

$$\begin{aligned}
\mathbf{u}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \frac{1}{\sqrt{3}}(1,1,1)^T \\
\mathbf{u}_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \frac{1}{\sqrt{14}}(2,1,-3)^T \\
\mathbf{u}_3 &= \frac{\mathbf{v}_3}{\|\mathbf{v}_3\|} = \frac{1}{\sqrt{42}}(4,-5,1)^T
\end{aligned}$$

Since $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis of $\mathbb{R}^3$, so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ form an orthonormal basis of $\mathbb{R}^3$. ∞

Example 5.5.10. Based on Example 5.4.8 and Example 5.5.5, the set in Example 5.5.5 is an orthonormal set. ∞

Theorem 5.5.11. *Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V. Then for any vector $\mathbf{v} \in V$,*

$$\mathbf{v} = \sum_{j=1}^{n} \langle \mathbf{v}, \mathbf{u}_j \rangle \mathbf{u}_j.$$

Proof. Assume that $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{u}_i$, then for $1 \le i \le n$, taking the inner product of $\mathbf{u}_j$ with both sides of the equation, we have

$$\langle \mathbf{v}, \mathbf{u}_j \rangle = c_1 \langle \mathbf{u}_1, \mathbf{u}_j \rangle + c_2 \langle \mathbf{u}_2, \mathbf{u}_j \rangle + \cdots + c_n \langle \mathbf{u}_n, \mathbf{u}_j \rangle = c_j \langle \mathbf{u}_j, \mathbf{u}_j \rangle = c_j.$$

□

Example 5.5.12. Express $\mathbf{x} = (1,2,3)^T$ as a linear combination of $\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)^T$, $\mathbf{u}_2 = \frac{1}{\sqrt{14}}(2,1,-3)^T$ and $\mathbf{u}_3 = \frac{1}{\sqrt{42}}(4,-5,1)^T$.

Solution: Note that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, then

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{6}{\sqrt{3}}\mathbf{u}_1 - \frac{5}{\sqrt{14}}\mathbf{u}_2 - \frac{3}{\sqrt{42}}\mathbf{u}_3,$$

since

$$\langle \mathbf{x}, \mathbf{u}_1 \rangle = \frac{1+2+3}{\sqrt{3}} = \frac{6}{\sqrt{3}}, \quad \langle \mathbf{x}, \mathbf{u}_2 \rangle = \frac{2+2-9}{\sqrt{14}} = -\frac{5}{\sqrt{14}}, \quad \langle \mathbf{x}, \mathbf{u}_3 \rangle = -\frac{3}{\sqrt{42}}.$$

ↀ

Theorem 5.5.13 (Parseval's Formula)**.** *If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for an inner product space V and $\mathbf{v} = \sum_{i=1}^n c_i \mathbf{u}_i$, then*

$$\|\mathbf{v}\|^2 = \sum_{i=1}^n c_i^2.$$

Exercise 5.5.14. Prove the Parseval's Formula. ✓

Theorem 5.5.15. *Let W be a subspace of an inner product space V and $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthonormal basis for W. Then for any $\mathbf{b} \in V$, the projection of $\mathbf{b}$ onto W is given by*

$$\mathbf{p} = \langle \mathbf{b}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{b}, \mathbf{u}_k \rangle \mathbf{u}_k$$

and $\mathbf{b} - \mathbf{p} \in W^\perp$.

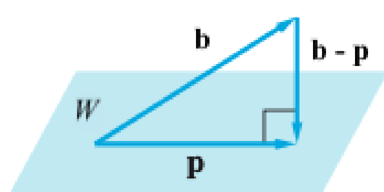


Figure 5.4: $\mathbf{p}$ is the orthogonal projecting vector of $\mathbf{b}$ onto W.

Example 5.5.16. Find the orthogonal projection of $\mathbf{y} = (9,0,0,0)^T$ onto the subspace W spanned by $\mathbf{u}_1 = \frac{1}{3}(2,2,1,0)^T$ and $\mathbf{u}_2 = \frac{1}{3}(-2,2,0,1)^T$.

Solution: Note that $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$, $\|\mathbf{u}_1\| = 1$ and $\|\mathbf{u}_2\| = 1$, so $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis of W. To find the projection of $\mathbf{y}$ onto W, we apply Theorem 5.5.15,

$$\text{proj}_W \mathbf{y} = \langle \mathbf{y}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{y}, \mathbf{u}_2 \rangle \mathbf{u}_2 = 18\mathbf{u}_1 - 18\mathbf{u}_2 = (8,0,2,-2)^T.$$

∞

Definition 5.5.17 (Orthogonal Matrix)**.** *An $n \times n$ matrix is called an* **orthogonal matrix** *if its column vectors form an orthonormal set in $\mathbb{R}^n$.*

Theorem 5.5.18. *For an $n \times n$ matrix Q, the following are equivalent:*

1. *Q is an orthogonal matrix,*
2. *$Q^TQ = I$, or equivalently, $Q^{-1} = Q^T$,*
3. *$\langle Q\mathbf{x}, Q\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,*
4. *$\|Q\mathbf{x}\| = \|\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^n$.*

From property (2) and (3), the linear transformation represented by an orthogonal matrix preserves the lengths of vectors and the angle between vectors.

Exercise 5.5.19. Prove Theorem 5.5.18. ✓

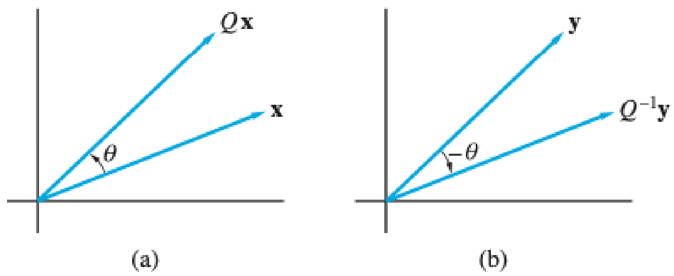
Example 5.5.20 (Rotation matrix)**.** For any fixed θ, the rotation matrix

$$Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

is orthogonal and

$$Q^{-1} = Q^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

The matrix Q can be viewed as a linear mapping from $\mathbb{R}^2$ onto $\mathbb{R}^2$ that rotates each vector anti-clockwisely by θ. Similarly, Q^{-1} can be regarded as a rotation by angle $-\theta$.



Example 5.5.21 (Permutation matrix)**. Permutation matrix** is a matrix formed from the identity by reordering its columns. Permutation matrices are orthogonal. The following presents 3×3 permutation matrices.

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ & & 1 \\ & 1 & \end{pmatrix} \begin{pmatrix} & 1 & \\ 1 & & \\ & & 1 \end{pmatrix} \begin{pmatrix} & & 1 \\ 1 & & \\ & 1 & \end{pmatrix} \begin{pmatrix} & 1 & \\ & & 1 \\ 1 & & \end{pmatrix} \begin{pmatrix} & & 1 \\ & 1 & \\ 1 & & \end{pmatrix}$$

For example,

$$\text{if} \quad A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \text{then} \quad AP = \begin{pmatrix} 2 & 3 & 1 \\ 5 & 6 & 4 \end{pmatrix}.$$

5.6 Gram-Schmidt Orthogonalization Process

Since orthonormal sets have very nice properties, we wish to construct an orthonormal set out of any given set (for the same space). The method involves using projections to transform an ordinary basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}$ into an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$.

Example 5.6.1. The orthogonal projection of $(9,0,0,0)^T$ onto the subspace W spanned by $\mathbf{x}_1 = (-2,2,0,1)^T$ and $\mathbf{x}_2 = (0,4,1,1)^T$. Notice that $\{\mathbf{x}_1, \mathbf{x}_2\}$ is just an ordinary basis of W, Thm 5.5.15 cannot be appiled directly.

Idea: Transform $\{\mathbf{x}_1, \mathbf{x}_2\}$ to an orthonormal basis, then use Thm 5.5.15.

Theorem 5.6.2 (Gram-Schmidt Orthogonalization Process)**.** *Let W be a subspace of V. Suppose W has a basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$. Define*

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\langle \mathbf{x}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \frac{\langle \mathbf{x}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\langle \mathbf{x}_p, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 - \cdots - \frac{\langle \mathbf{x}_p, \mathbf{v}_{p-1} \rangle}{\langle \mathbf{v}_{p-1}, \mathbf{v}_{p-1} \rangle}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ form an orthogonal basis for W. In addition, let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1, \ \mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2, \ \cdots, \ \mathbf{u}_p = \frac{1}{\|\mathbf{v}_p\|}\mathbf{v}_p.$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_p\}$ forms an orthonormal basis for W.

$$W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$$

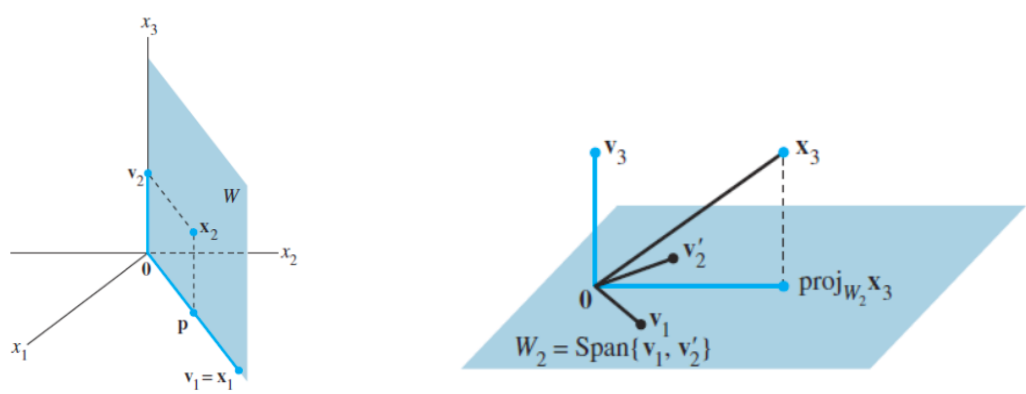


Figure 5.5: Left: to find $\mathbf{v}_2$ from $\mathbf{x}_2$; Right: to find $\mathbf{v}_3$ from $\mathbf{x}_3$

Example 5.6.3. Let $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_1\}$, where $\mathbf{x}_1 = (-2, 2, 0, 1)^T$ and $\mathbf{x}_2 = (0, 4, 1, 1)^T$. Construct an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W.

Solution: Let $\mathbf{v}_1 = \mathbf{x}_1$. Remove the component of $\mathbf{x}_2$ that is parallel to $\mathbf{x}_1$, and keep the orthogonal component as $\mathbf{v}_2$:

$$\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{\mathbf{v}_1}\mathbf{x}_2 = (0, 4, 1, 1)^T - \frac{\langle \mathbf{x}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\mathbf{v}_1 = (0, 4, 1, 1)^T - \mathbf{v}_1 = (2, 2, 1, 0)^T.$$

Normalize $\mathbf{v}_1$ and $\mathbf{v}_2$:

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{3}}(-2, 2, 0, 1)^T, \quad \mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{1}{\sqrt{3}}(2, 2, 1, 0)^T.$$

ↂ

Example 5.6.4. Let $\mathbf{x}_1 = (1, 1, 1, 1)^T$, $\mathbf{x}_2 = (0, 1, 1, 1)^T$ and $\mathbf{x}_3 = (0, 0, 1, 1)^T$. Construct an orthonormal basis for the subspace $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ of $\mathbb{R}^4$.

Solution: Let $\mathbf{v}_1 = \mathbf{x}_1$. Let

$$\mathbf{v}_2 = \mathbf{x}_2 - \text{proj}_{\mathbf{v}_1}\mathbf{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Thus, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace $W_2 = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$. Let

$$\begin{aligned} \mathbf{v}_3 &= \mathbf{x}_3 - \text{proj}_{W_2}\mathbf{x}_3 \\ &= \mathbf{x}_3 - \text{proj}_{\mathbf{v}_1}\mathbf{x}_3 - \text{proj}_{\mathbf{v}_2}\mathbf{x}_3 \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \underbrace{\frac{2}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}}_{\text{projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_1} - \underbrace{\frac{\frac{2}{4}}{\frac{12}{16}}\begin{pmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{pmatrix}}_{\text{projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_2} = \frac{1}{3}\begin{pmatrix} 0 \\ -2 \\ 1 \\ 1 \end{pmatrix}. \end{aligned}$$

Thus, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for W.

Normalizing $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, $\left\{ \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{12}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix} \right\}$ is an orthonormal basis of W.

ↂ

5.6.1 QR-Factorization**

Theorem 5.6.5. *If $A = [\mathbf{x}_1| \ldots | \mathbf{x}_n]$ is an $m \times n$ matrix of rank n, then A can be factored into a product QR, where Q is an $m \times n$ matrix with orthonormal columns and R is an upper triangular $n \times n$ matrix with positive diagonal entries:*

$$Q = \left[\ \mathbf{u}_1 \ \middle|\ \mathbf{u}_2 \ \middle|\ \cdots \ \middle|\ \mathbf{u}_n\ \right]_{m\times n}, \qquad R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix}_{n\times n}$$

1. *The set $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is the orthonormal basis obtained from $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ by the Gram-Schmidt Process.*

2. *In matrix R, the entry $r_{ij} = \langle \mathbf{u}_i, \mathbf{x}_j \rangle$, for $1 \le i \le j \le n$.*

Example 5.6.6. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then $Q = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}$ by Example 5.6.4. Since $r_{11} = \langle \mathbf{u}_1, \mathbf{x}_1 \rangle = 2$, $r_{12} = \langle \mathbf{u}_1, \mathbf{x}_2 \rangle = 2/3$, $r_{13} = \langle \mathbf{u}_1, \mathbf{x}_3 \rangle = 1$

$$r_{22} = \langle \mathbf{u}_2, \mathbf{x}_2 \rangle = \sqrt{3}/2, \quad r_{23} = \langle \mathbf{u}_2, \mathbf{x}_3 \rangle = 1/\sqrt{3}$$

$$r_{33} = \langle \mathbf{u}_3, \mathbf{x}_3 \rangle = \sqrt{2}/\sqrt{3}.$$

Then $R = \begin{bmatrix} 2 & 2/3 & 1 \\ 0 & \sqrt{3}/2 & 1/\sqrt{3} \\ 0 & 0 & \sqrt{2}/\sqrt{3} \end{bmatrix}$, and $A = QR$. ꝏ

Theorem 5.6.7 (QR in least-square problems)**.** *If A is an $m \times n$ matrix of rank n, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is given by $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$, where $A = QR$ be a QR-factorization of A. The solution $\hat{\mathbf{x}}$ can be obtained by using back substitution to solve $R\hat{\mathbf{x}} = Q^T\mathbf{b}$.*