Chapter 5. Orthogonality

第五章包含兩個主題，內積空間和正交性，把線性代數和幾何做了個結合。第一節先由我們熟知的歐氏幾何空間開始，定義向量的內積、長度、夾角、正交性，和許多幾何問題的應用。第二節把正交性推廣到向量子空間，定義甚麼是正交的子空間。第三節介紹了正交子空間中的一個重要應用，解「線性最小平方和」的問題，對於大數據有興趣的同學要特別認真學。第四章將內積的定義推廣到一般的向量空間，在之上也可以定義長度、夾角、正交性等性質。第五章在正交性上加上了正規化，一個向量基底有了這兩個特性（正交性和正規化），基本上就有一堆好性質。

5.1 The Scalar Product in R^n

介紹歐氏幾何空間中的內積、長度、夾角、正交性，和許多幾何問題的應用。

- The **scalar product** of two $n\times 1$ matrices **x** and **y** is the 1×1 matrix $\mathbf{x}^T\mathbf{y}$, or simply regarded as a real number. That is, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, then

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + \ldots + x_ny_n$$

- A vector in $\boldsymbol{R}^2$ and $\boldsymbol{R}^3$ can be represented by directed line segment. The **Euclidean length** of a vector **x** in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$ can be defined in terms of the scalar product:

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} = \begin{cases} \sqrt{x_1^2 + x_2^2} & \text{if } \mathbf{x} \in R^2 \\ \sqrt{x_1^2 + x_2^2 + x_3^2} & \text{if } \mathbf{x} \in R^3 \end{cases}$$

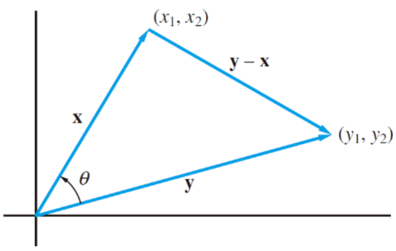


Figure 5.1.1.

- The angle between two vectors is defined as the angle θ between the line segments.
- The distance between the vectors is measured by the length of the vector joining the terminal point of **x** and the terminal point of **y**

- Definition: Let **x** and **y** be vectors in either R^2 or R^3. The **distance** between **x** and **y** is defined to be the number $||\mathbf{x} - \mathbf{y}||$.

- **Theorem 5.1.1**: If **x** and **y** are two nonzero vectors in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$ and θ is the angle between them, then $\mathbf{x}^T\mathbf{y} = ||\mathbf{x}||\ ||\mathbf{y}|| \cos\theta$

Proof: By the law of cosines, $||\mathbf{y} - \mathbf{x}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2 - 2\,||\mathbf{x}||\,||\mathbf{y}|| \cos\theta$ or

$$||\mathbf{x}||\,||\mathbf{y}|| \cos\theta = ½(||\mathbf{x}||^2 + ||\mathbf{y}||^2 - ||\mathbf{y} - \mathbf{x}||^2) = ½(||\mathbf{x}||^2 + ||\mathbf{y}||^2 - (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x}))$$
$$= ½(||\mathbf{x}||^2 + ||\mathbf{y}||^2 - (\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{x} - \mathbf{x}^T\mathbf{y} + \mathbf{x}^T\mathbf{x})) = ½(2\,\mathbf{x}^T\mathbf{y}) = \mathbf{x}^T\mathbf{y}$$

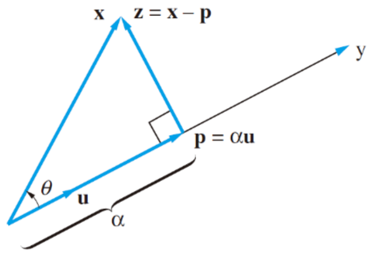
- If **x** is a nonzero vector, then we can form the **unit vector u** of **x** as $\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x}$

- If **x** and **y** are two nonzero vectors, **u** and **v** are the unit vectors of **x** and **y**, then the angle θ between **x** and **y** is

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \mathbf{u}^T\mathbf{v}$$

- **Corollary 5.1.2** If **x** and **y** are vectors in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$, then $|\mathbf{x}^T\mathbf{y}| \le \|\mathbf{x}\|\,\|\mathbf{y}\|$ with equality holding if and only if one of the vectors is 0 or one vector is a multiple of the other.

- The vectors **x** and **y** are in R^2 (or R^3) are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$.



- Let **x** and **y** be in either R^2 or R^3, then **x** can be represented as **p**+**z**, where **p** is in the direction of **y** and **z** is orthogonal to **p**.

- Let **u** = $(1/\|\mathbf{y}\|)\mathbf{y}$, thus **u** is a unit vector (length 1) in the direction of **y**. We wish to find α such that **p** = α**u** and is orthogonal to **z** = **x** – α**u**. Thus

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

α is called the **scalar projection** of **x** onto **y** and **p** is called the **vector projection** of **x** onto **y**:

$$\mathbf{p} = \alpha\mathbf{u} = \alpha\frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$$

- If **N** is a nonzero vector and P_0 is a fixed point, the set of points ***P*** such that is orthogonal to **N** forms a **plane** π in 3-space that passes through P_0. The vector **N** and the plane π are said to be **normal** to each other. A point $P = (x, y, z)$ will lie on π if and only if

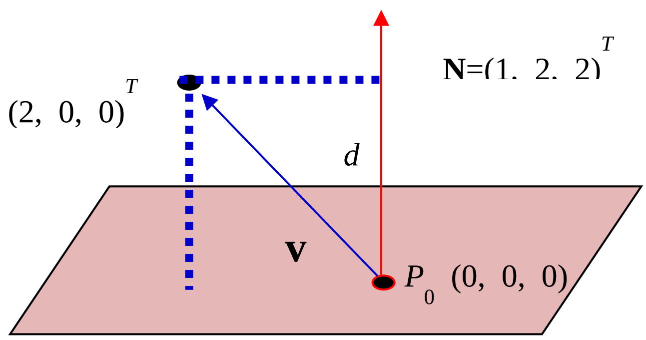
$$(\overrightarrow{P_0P})^T\mathbf{N} = 0$$

If $\mathbf{N} = (a, b, c)^T$ and $P_0 = (x_0, y_0, z_0)$, the above equation can be written as

$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$

Example: Find the equation of the plane that passes the point (2, -1, 3) and has the normal $(2, 3, 4)^T$.

Example: Compute the distance from (2, 0, 0) to x+2y+2z=0.



Example: Find the equation of the plane that passes the points (1, 1, 2), (2, 3, 3), and (3, -3, 3).

- Cross product in R^3
  - The cross product of **x** = (x1, x2, x3) and **y** = (y1, y2, y3) in R^3 is defined as

$$\boldsymbol{x} \times \boldsymbol{y} = \begin{bmatrix} i & j & k \\ x1 & x2 & x3 \\ y1 & y2 & y3 \end{bmatrix} = \begin{bmatrix} x2y3 - x3y2 \\ x1y3 - x3y1 \\ x1y2 - x2y1 \end{bmatrix}$$

  - The vector $\boldsymbol{x} \times \boldsymbol{y}$ is orthogonal to **x** and to **y**.
  - $\|\boldsymbol{x} \times \boldsymbol{y}\| = \|\boldsymbol{x}\|\|\boldsymbol{y}\| \sin\theta$

Proof:

- Orthogonality in R^n
  - If $\mathbf{x} \in R^n$ then the **Euclidean length** of **x** is defined by

$$\| \mathbf{x} \| = (\mathbf{x}^T \mathbf{x})^{1/2} = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$$

  - If **x** and **y** are two vectors in R^n, then the distance between **x** and **y** is **||y – x||**.
  - The Cauchy-Schwarz inequality holds in R^n: $-1 \le \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \le 1$
  - The angle θ between two vectors **x** and **y** in R^n is given by $\cos\theta = \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, 0 \le \theta \le \pi$
  - Two vectors **x** and **y** in R^n are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$ and often the symbol "⊥" is used to indicate orthogonality. If **x** and **y** are orthogonal, we will write $\mathbf{x} \perp \mathbf{y}$
  - If **x** and **y** are vectors in R^n and they are orthogonal, $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ (**Pythagorean Law**)

5.2 Orthogonal Subspaces

定義向量空間正交性，衍伸出互補空間的概念，定理 5.2.1 (**Fundamental Subspace Theorem**) 呼應了定理 3.6.5 的 **fundamental theorem of linear algebra**，基本上對於一個矩陣所帶出的四個子空間，給出了更多有用的性質。

- Let A be an $m \times n$ matrix and let $\mathbf{x} \in N(A)$, the null space of A.
  - $\mathbf{x}$ is orthogonal to the ith column vector of A^T for $i = 1, 2, \ldots, m$
  - $\mathbf{x}$ is orthogonal to any linear combination of the column vector of A^T
  - If $\mathbf{y}$ is any vector in <u>the column space of A^T</u>, then $\mathbf{x}^T\mathbf{y} = 0$
  - Each vector in $N(A)$ is orthogonal to every vector in the column space of A^T

- Two subspaces X and Y of R^n are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$ for <u>every</u> $\mathbf{x} \in X$ and <u>every</u> $\mathbf{y} \in Y$.
  If X and Y are orthogonal, we write $X \perp Y$.

- **Definition** Let Y be a subspace of R^n. The set of all vectors in R^n that are orthogonal to every vector in Y will be denoted $Y^\perp$. Thus

$$Y^\perp = \{\mathbf{x} \in R^n \mid \mathbf{x}^T\mathbf{y} = 0 \text{ for every } \mathbf{y} \in Y\}$$

  The set $Y^\perp$ is called the **orthogonal complement** of Y.
- Remarks:
  - If X and Y are orthogonal subspaces of R^n, then $X \cap Y = \{\mathbf{0}\}$.

  Proof: If $\mathbf{x} \in X \cap Y$ and $X \perp Y$, then $\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = 0 \Rightarrow \mathbf{x} = \mathbf{0}$
  - If Y is a subspace of R^n, then $Y^\perp$ is also a subspace of R^n.

  Proof: If $\mathbf{x} \in Y^\perp$ and α is a scalar, then for any $\mathbf{y} \in Y$, $(\alpha\mathbf{x})^T\mathbf{y} = \alpha(\mathbf{x}^T\mathbf{y}) = \alpha \cdot 0 = 0 \quad \Rightarrow \alpha\mathbf{x} \in Y^\perp$

  If $\mathbf{x}_1$ and $\mathbf{x}_2$ are elements of $Y^\perp$, then $(\mathbf{x}_1 + \mathbf{x}_2)^T\mathbf{y} = \mathbf{x}_1^T\mathbf{y} + \mathbf{x}_2^T\mathbf{y} = 0 + 0 = 0 \quad \Rightarrow (\mathbf{x}_1 + \mathbf{x}_2) \in Y^\perp$

  Therefore, $Y^\perp$ is a subspace of R^n.

- **Fundamental Subspaces**:
  - Let A be an $m \times n$ matrix, a vector $\mathbf{b} \in R^n$ is in the column space of A if and only if $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in R^n$.
  - If we think of A as a <u>linear transformation</u> mapping R^n into R^m, then the column space of A is the same as the range of A. Let A be an $m \times n$ matrix and $R(A)$ denote the range of A. Thus

$$R(A) = \{\mathbf{b} \in R^m \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in R^n\} = \text{the column space of } A$$

    The column space of A^T, $R(A^T)$, is a subspace of R^n:

$$R(A^T) = \{\mathbf{y} \in R^n \mid \mathbf{y} = A^T\mathbf{x} \text{ for some } \mathbf{x} \in R^m\} = \text{the column space of } A^T$$

  - The column space of A^T is essentially the same as the row space of A except that it consists of vectors in R^n ($n \times 1$ matrices) rather than n-tuples.

$$\mathbf{y} \in R(A^T) \text{ if and only if } \mathbf{y}^T \text{ is in the row space of } A$$

  - Since each vector in $N(A)$ is orthogonal to every vector in the column space of A^T (i.e., $R(A^T)$),

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

- **Theorem 5.2.1 (Fundamental Subspace Theorem)**: If A is an $m \times n$ matrix, then $N(A) = R(A^T)^{\perp}$ and $N(A^T) = R(A)^{\perp}$.

Proof: Since $N(A) \perp R(A^T) \Rightarrow N(A) \subset R(A^T)^{\perp}$. (1)

If $\mathbf{x} \in R(A^T)^{\perp}$, then $\mathbf{x}^T(A^T) = 0 \Rightarrow (A\mathbf{x})^T = \mathbf{0} \Rightarrow A\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} \in N(A)$ (2)

From (1) and (2), we proved $N(A) = R(A^T)^{\perp}$. Let $B = A^T$, then $N(A^T) = N(B) = R(B^T)^{\perp} = R(A)^{\perp}$.

- Let $\mathbf{x} \in N(A)$ (i.e., $A\mathbf{x} = 0$) and $\mathbf{y} \in R(A^T)$ (i.e., $\mathbf{y} = A^T\mathbf{z}$)
$\Rightarrow \mathbf{x}^T\mathbf{y} = \mathbf{x}^T(A^T\mathbf{z}) = (A\mathbf{x})^T\mathbf{z} = \mathbf{0}^T\mathbf{z} = 0$. So $N(A) \perp R(A^T)$?

Example: The <u>column space</u> of A consists of all vectors of the form:

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}, \quad \alpha\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \beta\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha \\ 2\alpha \end{bmatrix} = \alpha\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

If $\mathbf{x}$ is any vector in R^n, and $\mathbf{b} = A\mathbf{x}$, then $\mathbf{b}$ is in the column space of A.

$$\mathbf{b} = \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ 2x_1 \end{bmatrix} = x_1\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

The null space of A^T: $N(A^T)$ is of the form:

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2\beta \\ \beta \end{bmatrix} = \beta\begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

The range of A: $R(A)$ = the column space of A:

$$\because \begin{bmatrix} -2\beta & \beta \end{bmatrix}\begin{bmatrix} \alpha \\ 2\alpha \end{bmatrix} = 0 \Rightarrow \begin{bmatrix} -2\beta \\ \beta \end{bmatrix} \perp \begin{bmatrix} \alpha \\ 2\alpha \end{bmatrix},$$

$$\therefore N(A^T) = R(A)^{\perp}$$

- **Theorem 5.2.2**: If S is a subspace of R^n, then $\dim S + \dim S^{\perp} = n$. Furthermore, if $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ is a basis for S and $\{\mathbf{x}_{r+1}, \ldots, \mathbf{x}_n\}$ is a basis for $S^{\perp}$, then $\{\mathbf{x}_1, \ldots, \mathbf{x}_r, \mathbf{x}_{r+1}, \ldots, \mathbf{x}_n\}$ is a basis for R^n.

Proof:

- **Definition**: If U and $\boldsymbol{V}$ are subspaces of a vector space W and each $\mathbf{w} \in W$ can be written uniquely as a sum $\mathbf{u}+\mathbf{v}$, where $\mathbf{u} \in U$ and $\mathbf{v} \in V$, then we say W is a **direct sum** of U and $\mathbf{V}$ and we write $W = U \oplus V$.
- **Theorem 5.2.3**: If S is a subspace of R^n, then $R^n = S \oplus S^{\perp}$

- **Theorem 5.2.4**: If S is a subspace of R^n, then $(S^{\perp})^{\perp} = S$

- If T is the orthogonal complement of a subspace S, then S is the orthogonal complement of T.
- From Theorem 5.2.1, $N(A)$ and $R(A^T)$ are orthogonal complements of each other and $N(A^T)$ and $R(A)$ are orthogonal complements, we can write $N(A)^{\perp} = R(A^T)$ and $N(A^T)^{\perp} = R(A)$

- **Corollary 5.2.5**: If A is an $m \times n$ matrix and $\mathbf{b} \in R^m$, then either there is a vector $\mathbf{x} \in R^n$ such that $A\mathbf{x} = \mathbf{b}$ or there is a vector $\mathbf{y} \in R^m$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$

  (1) $\mathbf{b} \in R(A)$

  (2) $\mathbf{b} \notin R(A) \Rightarrow \mathbf{b} \notin N(A^T)^{\perp}$

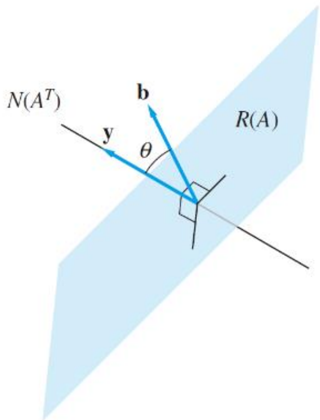


Figure 5.2.2.

Example: Find the bases for $N(A)$, $R(A^T)$, $N(A^T)$ and $R(A)$.

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A^T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ 2 & 1 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

(1) $N(A)$: $x_1 + x_3 = 0$ and $x_2 + x_3 = 0$

set $x_3 = \alpha$, then $x_1 = -\alpha$ and $x_2 = -\alpha$.

$$\therefore \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\alpha \\ -\alpha \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \Rightarrow N(A) = \mathrm{Span}(\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix})$$

(2) $R(A^T)$: Since (1, 0, 1) and (0, 1, 2) form a basis for the <u>row space</u> of A

$$\Rightarrow \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \text{ form a basis for the column space of } A^T$$

$$\Rightarrow R(A^T) = \mathrm{Span}(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix})$$

(3) $N(A^T)$: $x_1 + x_3 = 0$ and $x_2 + 2x_3 = 0$

set $x_3 = \alpha$, then $x_1 = -\alpha$ and $x_2 = -2\alpha$.

$$\therefore \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\alpha \\ -2\alpha \\ \alpha \end{bmatrix} = \alpha \begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix} \Rightarrow N(A^T) = \mathrm{Span}(\begin{bmatrix} -1 \\ -2 \\ 1 \end{bmatrix})$$

(4) $R(A)$ $\quad R(A) = \mathrm{Span}(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix})$

5.3 Least Squares Problems

線性最小平方和的問題，或是線性回歸問題，和線性系統很像，只是它的變數少於等式，所以只能求一個近似解，你們在中學或是微積分中應該有學過如何解。這節把這個問題寫成了矩陣的形式，利用幾何的概念（投影、正交），將這個問題的本質分析的清清楚楚，並利用 5.2 的定理來解這個問題，可以和之前所學得比較一下。

- Find a least squares curve (a linear function, a polynomial, a trigonometric polynomial, etc.) fit to a set of data points in the plane.
- The curve provides an <u>optimal approximation</u> in the sense that the sum of squares of errors between the *y* values of the data points and the corresponding *y* values of the approximating curve are **minimized**.
- A <u>least squares problems</u> can generally be formulated as an <u>overdetermined linear system of equations</u>. An over-determined system is one involving <u>more equations than unknowns</u>.
- An over-determined system is usually <u>inconsistent</u>.
- Given an $m \times n$ system $A\mathbf{x} = \mathbf{b}$ with $m > n$, we cannot expect in general to find a vector $\mathbf{x} \in R^n$ such that $A\mathbf{x}$ **<u>equals</u>** $\mathbf{b}$. We can look for a vector $\mathbf{x} \in R^n$ for which $A\mathbf{x}$ is **"closer"** to $\mathbf{b}$.

- Given a system of equations $A\mathbf{x} = \mathbf{b}$, where A is an $m\times n$ matrix with $m > n$ and $\mathbf{b} \in R^m$, for each $\mathbf{x} \in R^n$ we can form a **residual** $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$
- The distance between $\mathbf{b}$ and $A\mathbf{x}$ is given by $||\mathbf{b} - A\mathbf{x}|| = ||\mathbf{r(x)}||$
- we wish to find a vector $\mathbf{x} \in R^n$ for which $||r(\mathbf{x})||$ is minimized
- In fact, minimizing $||r(\mathbf{x})||$ is equivalent to minimizing $||r(\mathbf{x})||^2$
- A vector $\mathbf{x}^*$ that minimizing $||r(\mathbf{x})||^2$ is said to be a **least squares solution** to the system $A\mathbf{x} = \mathbf{b}$
- If $\mathbf{x}^*$ is a least squares solution to the system $A\mathbf{x} = \mathbf{b}$ and $\mathbf{p} = A\mathbf{x}^*$, then $\mathbf{p}$ is <u>a vector in the column space of **A** that is closest to **b**.</u>
- Such a closest vector $\mathbf{p}$ not only exists, but is <u>unique</u>

- **Theorem 5.3.1**: Let S be a subspace of R^m. For each $\mathbf{b} \in R^m$ there is a <u>unique</u> element $\mathbf{p}$ of S that is closest to $\mathbf{b}$, that is $||\ \mathbf{b} - \mathbf{y}\ || > ||\ \mathbf{b} - \mathbf{p}\ ||$ for any $\mathbf{y} \neq \mathbf{p}$ in S. Furthermore, a given vector $\mathbf{p}$ in S will be closest to a given vector $\mathbf{b} \in R^m$ if and only if $\mathbf{b} - \mathbf{p} \in S^{\perp}$.

Proof: (1) Since $R^m = S \oplus S^{\perp}$, each $\mathbf{b} \in R^m$ can be expressed uniquely as a sum $\mathbf{b} = \mathbf{p} + \mathbf{z}$, where $\mathbf{p} \in S$ and $\mathbf{z} \in S^{\perp}$. If $\mathbf{y}$ is another element of S, then

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

Since $(\mathbf{p} - \mathbf{y}) \in S$ ($\because$ $\mathbf{p} \in S$ and $\mathbf{y} \in S$) and $(\mathbf{b} - \mathbf{p}) = \mathbf{z} \in S^{\perp}$. It follows from the Pythagorean Law

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

Therefore, $||\ \mathbf{b} - \mathbf{y}\ || > ||\ \mathbf{b} - \mathbf{p}\ ||$

(2) If $\mathbf{p} \in S$ and $\mathbf{b} - \mathbf{p} \in S^{\perp}$, then $\mathbf{p}$ is the element of S that is closest to $\mathbf{b}$.
Conversely, if $\mathbf{q} \in S$ and $\mathbf{b} - \mathbf{q} \notin S^{\perp}$, then $\mathbf{q} \neq \mathbf{p}$, and it follows from the preceding argument (with $\mathbf{q} = \mathbf{y}$) that $||\ \mathbf{b} - \mathbf{q}\ || > ||\ \mathbf{b} - \mathbf{p}\ ||$. Therefore, $\mathbf{p}$ will be closest to $\mathbf{b}$

- If **b** $\in S$, then we have **b** = **p** + **z**, where **p** $\in S$ and **z** $\in S^{\perp}$, since **b** = **b** + **0**, by the uniqueness of the direct sum representation, we have **p** = **b** and **z** = **0**
- A vector x* will be a solution to the least squares problem A**x** = **b** if and only if **p** = A is the vector in $R(A)$ that is closest to **b**. The vector **p** is said to be the projection of **b** onto $R(A)$.

(a) $\mathbf{b} \in R^2$ and A is a 2×1 matrix of rank 1. (b) $\mathbf{b} \in R^2$ and A is a 3×2 matrix of rank 2.

Figure 5.3.2.

- How to find the vector x* satisfying (1)?
- Since $r(\mathbf{x}^*) \in R(A)^{\perp}$, and from Theorem 5.2.1: $R(A)^{\perp} = N(A^T)$
- A vector x* will be a solution to the least squares problem $A\mathbf{x} = \mathbf{b}$ if and only if $r(\mathbf{x}^*) \in N(A^T)$ or equivalently, $\mathbf{0} = A^T r(\mathbf{x}^*) = A^T(\mathbf{b} - A)$

- To solve the least squares problem $A\mathbf{x} = \mathbf{b}$, we must solve
$$A^TA\mathbf{x} = A^T\mathbf{b} \tag{2}$$
- Eq. (2) represents an $n \times n$ system of equations (note: A^TA: $n \times n$). These equations are called **normal equations**.

- **Theorem 5.3.2**: If A is an $m \times n$ matrix of rank n, the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$ have a unique solution x*= $(A^TA)^{-1}A^T\mathbf{b}$ and x* is the unique least squares solution to the problem $A\mathbf{x} = \mathbf{b}$.

Proof: First, we have to show that A^TA is nonsingular

- Let **z** be the solution to $$A^TA\mathbf{x} = \mathbf{0} \tag{3}$$
- Then, $A\mathbf{z} \in N(A^T)$ (think of $A^T(A\mathbf{z}) = \mathbf{0}$)
- Clearly, $A\mathbf{z} \in R(A) = N(A^T)^{\perp}$ (think of A**z** as the projection of **z** by **A**)
  $\Rightarrow A\mathbf{z} \in N(A^T) \cap N(A^T)^{\perp} = \{\mathbf{0}\} \Rightarrow A\mathbf{z} = \mathbf{0}$

- If A has rank n, the column vectors of A are **linear independent** and thus $A\mathbf{x} = \mathbf{0}$ has only the trivial solution **0** $\Rightarrow \mathbf{z} = \mathbf{0}$
  $\Rightarrow$ Eq. (iii) $A^TA\mathbf{x} = \mathbf{0}$ has only the trivial solution
  $\Rightarrow$ By Theorem 1.5.2, A^TA is nonsingular
  $\Rightarrow$ $= (A^TA)^{-1}A^T\mathbf{b}$ is the unique solution to the normal equations and, consequently, the unique least squares solution to the problem $A\mathbf{x} = \mathbf{b}$

- The projection vector $\mathbf{p} = A\mathbf{x}^* = A(A^TA)^{-1}A^T\mathbf{b}$ is the element of $R(A)$ that is closest to **b** in the least squares sense. The matrix $P = A(A^TA)^{-1}A^T$ is called the **projection matrix**

Example: Find the least squares solution to the system

$$
\begin{aligned}
x_1 + \quad x_2 &= 3 \\
-2x_1 + 3x_2 &= 1 \\
2x_1 - \quad x_2 &= 2
\end{aligned}
$$

Example: Given the data. Find the best least squares fit to the data by a linear function.

x	0	3	6
y	1	4	5

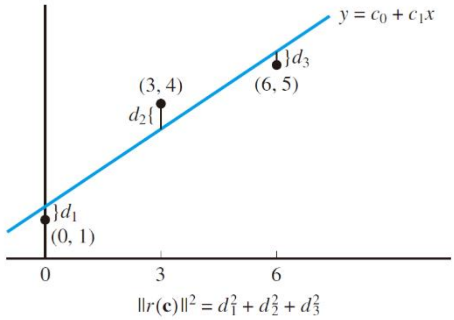


Figure 5.3.4.

5.4 Inner Product Spaces

把內積的概念一般化，和 3.1、3.2 的作法一樣，都是抽象代數的手法，藉由這些定義，多項式、可微分函數等其他向量空間也可以定義長度、角度、正交等幾何性質。

Definition

- An **inner product** on a vector space V is an operation on V that assigns to each pair of vectors **x** and **y** in V a real number **<x, y>** satisfying the following conditions:
  - I. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$.
  - II. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for all **x** and **y** in V.
  - III. $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$ for all **x**, **y**, **z** in V and all scalars α and β.
- A vector space V with an inner product is called an **inner product space**.

Example: The standard inner product for R^n is the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$

Given a vector **w** with positive entries $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i w_i$ where wi are referred to as weights.

- Basic Properties of Inner Product Spaces
  - If **v** is a vector in an inner product space V, the **length** or **norm** of **v** is given by
  $$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$
  - Two vectors **u** and **v** are said to be **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$

- **Theorem 5.4.1:** If **u** and **v** are orthogonal vectors in an inner product space V, then
$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$$

Proof: $||\mathbf{u} + \mathbf{v}||^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} + \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle$

$= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle$

$= ||\mathbf{u}||^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + ||\mathbf{v}||^2$ (Note $\langle \mathbf{u}, \mathbf{v} \rangle = 0$)

$= ||\mathbf{u}||^2 + ||\mathbf{v}||^2$

Definition: If **u** and **v** are vectors in an inner product space *V* and $\mathbf{v} \neq \mathbf{0}$, then the **scalar projection** of **u** onto **v** are given by

$$\alpha = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|}$$

and the **vector projection** of **u** onto **v** is given by

$$\mathbf{p} = \alpha \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|} \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2} \mathbf{v} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$$

- Theorem: If $\mathbf{v} \neq \mathbf{0}$ and **p** is the vector projection of **u** onto **v**, then

  I. $\mathbf{u} - \mathbf{p}$ and **p** are orthogonal.

  II. $\mathbf{u} = \mathbf{p}$ if and only if **u** is a scalar multiple of **v.**

Proof: **I.** Since $\langle \mathbf{p}, \mathbf{p} \rangle = \langle \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v}, \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v} \rangle = \left(\frac{\alpha}{\|\mathbf{v}\|}\right)^2 \langle \mathbf{v}, \mathbf{v} \rangle = \alpha^2$ and

$$\langle \mathbf{u}, \mathbf{p} \rangle = \langle \mathbf{u}, \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v} \rangle = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \langle \mathbf{u}, \mathbf{v} \rangle = \frac{\langle \mathbf{u}, \mathbf{v} \rangle^2}{\|\mathbf{v}\|^2} = \alpha^2$$

$\Rightarrow \langle \mathbf{u} - \mathbf{p}, \mathbf{p} \rangle = \langle \mathbf{u}, \mathbf{p} \rangle - \langle \mathbf{p}, \mathbf{p} \rangle = \alpha^2 - \alpha^2 = 0$

$\Rightarrow$ $\mathbf{u} - \mathbf{p}$ and **p** are orthogonal.

II. If $\mathbf{u} = \beta\mathbf{v}$, then the vector projection of **u** onto **v** is given by

$$\mathbf{p} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v} = \frac{\langle \beta\mathbf{v}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v} = \beta\mathbf{v} = \mathbf{u}$$

If $\mathbf{u} = \mathbf{p} \Rightarrow$ $\mathbf{u} = \mathbf{p} = \alpha \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v} = \beta\mathbf{v}$, where $\beta = \frac{\alpha}{\|\mathbf{v}\|}$

- **Theorem 5.4.2** If **u** and **v** are any two vectors in an inner product space *V*, then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq ||\mathbf{u}||\ ||\mathbf{v}||$$

Equality holds if and only if **u** and **v** are **linearly dependent**.

- From the above theorem, if **u** and **v** are nonzero vectors, then $-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\ \|\mathbf{v}\|} \leq 1$

  and hence there is a unique angle $\theta \in [0, \pi]$ such that $\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}$

  this equation can be used to define the angle θ between two nonzero vectors **u** and **v.**

- **Definition** A vector space V is said to be a **normed linear space** if to each vector $\mathbf{v} \in V$ there is associated a real number $||\mathbf{v}||$ called the **norm** of $\mathbf{v}$, satisfying

  I. $||\mathbf{v}|| \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$

  II. $||\alpha\mathbf{v}|| = |\alpha|\ ||\mathbf{v}||$ for any scalar α

  III. $||\mathbf{v} + \mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$ for all $\mathbf{v}, \mathbf{w} \in V$ (**triangle inequality**)

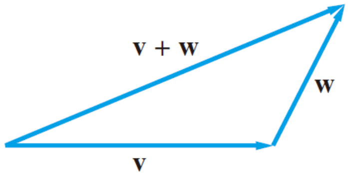


Figure 5.4.2.

- **Theorem 5.4.3**: If V is an **inner product space**, then the equation

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

  defines a **norm** on V.

Proof:

- It is easily seen that conditions **I** and **II** are satisfied. To show that condition **III** is satisfied

  $||\mathbf{u} + \mathbf{v}||^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + 2\underline{\langle \mathbf{u}, \mathbf{v} \rangle} + \langle \mathbf{v}, \mathbf{v} \rangle$

  $\leq ||\mathbf{u}||^2 + 2\,\underline{||\mathbf{u}||\ ||\mathbf{v}||} + ||\mathbf{v}||^2$ (from **The Cauchy-Schwarz Inequality**)

  $= (||\mathbf{u}|| + ||\mathbf{v}||)^2$

  Thus $||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}||$

- **Definitions**:
  - **1-norm:** $||\mathbf{x}||_1 = |x_1| + |x_2| + \dots + |x_n|$
  - **2-norm:** $||\mathbf{x}||_2 = (|x_1|^2 + |x_2|^2 + \dots + |x_n|^2)^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \sqrt{<\mathbf{x}, \mathbf{x}>}$
  - ***p*-norm:** $||\mathbf{x}||_p = (|x_1|^p + |x_2|^p + \dots + |x_n|^p)^{1/p} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$
  - **∞-norm (uniform norm, infinity norm):** $||\mathbf{x}||_\infty = \mathbf{max}(|x_1|, |x_2|, \dots, |x_n|)$

- A norm provides a way of measuring <u>the distance between two vectors</u>.

Example: Let $\mathbf{x} = (4, -5, 3)^T$ in R^3, then

$||\mathbf{x}||_1 =$

$||\mathbf{x}||_2 =$

$||\mathbf{x}||_\infty =$

- **Definition** Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in a normed linear space. The **distance** between $\mathbf{x}$ and $\mathbf{y}$ is defined to be the number $||\mathbf{y} - \mathbf{x}||$.

5.5 Orthonormal Sets

把正交性加上正規化，一個向量基底有了這兩個性質，在上面所有的計算都變得簡單，例如解最小平方和問題可以直接寫答案了，向量子空間的正交投影也變得容易了。

- In R^2, the elements of the standard basis {$\mathbf{e}_1$, $\mathbf{e}_2$} are <u>orthogonal unit vectors</u>.
- In working with an inner product space V, it is generally desirable to have a basis of <u>mutually orthogonal unit vectors</u>.
- Convenient in finding <u>coordinates</u> of vectors and <u>solving least square problems</u>.

- **Definition:** Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be nonzero vectors in an inner product space V. If $<\mathbf{v}_i, \mathbf{v}_j> = 0$ whenever $i \neq j$, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is said to be an **orthogonal set** of vectors.

Example: Is the set $\{(1, 1, 1)^T, (2, 1, -3)^T, (4, -5, 1)^T\}$ an orthogonal set in R^3 or not.

- Theorem 5.5.1: If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an **orthogonal set** of <u>nonzero</u> vectors in an inner product space V, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **linearly independent**.

Proof: Let $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n = \mathbf{0}$

$(\mathbf{v}_j)^T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) = 0$ for all j

$\Rightarrow$ $c_1<\mathbf{v}_j, \mathbf{v}_1> + c_2<\mathbf{v}_j, \mathbf{v}_2> + \ldots + c_n<\mathbf{v}_j, \mathbf{v}_n> = 0$

$\Rightarrow$ $c_j \|\mathbf{v}_j\|^2 = 0 \Rightarrow c_j = 0$ for all j

$\Rightarrow$ $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent.

- **Definition:** An <u>**orthonormal**</u> **set** of vectors is an <u>**orthogonal**</u> set of <u>**unit vectors**</u>.

- The set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ will be orthonormal iff

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Given any orthogonal set of nonzero vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, it is possible to form an **orthonormal set** by defining

$$\mathbf{u}_i = \left(\frac{1}{\|\mathbf{v}_i\|} \right) \mathbf{v}_i, \quad \text{for } i = 1, 2, \cdots, n$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ will be an orthonormal set

- From Theorem 5.5.1, if $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthonormal set in an inner product space V

$\Rightarrow$ $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are **linearly independent**

$\Rightarrow$ B is a **basis** for a subspace $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of V

$\Rightarrow$ B is an **orthonormal basis** for S

- **Theorem 5.5.2**: Let {$\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_n$} be an **orthonormal basis** for an inner product space V. If

  $\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{u}_j$, then c_i = < **v**, $\mathbf{u}_i$ >

Proof:

$$\langle \mathbf{v},\ \mathbf{u}_i \rangle = \left\langle \sum_{j=1}^{n} c_j \mathbf{u}_j,\ \mathbf{u}_i \right\rangle = \sum_{j=1}^{n} c_j \left\langle \mathbf{u}_i,\ \mathbf{u}_j \right\rangle = \sum_{j=1}^{n} c_j\ \delta_{ji} = c_i$$

- Corollary 5.5.3: Let {$\mathbf{u}_1$, $\mathbf{u}_2$, ..., $\mathbf{u}_n$} be an orthonormal basis for an inner product space V. If

  $\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{u}_i$ and $\mathbf{v} = \sum_{i=1}^{n} b_i \mathbf{u}_i$. Then $\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} a_i\ b_i$

Proof:

- **Corollary 5.5.4**: If {u_1, u_2, ..., u_n} be an orthonormal basis for an inner product space V and

  $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{u}_i$ Then $\|\mathbf{v}\|^2 = \sum_{i=1}^{n} c_i^2$

- **Definition** An $n \times n$ matrix Q is said to be an **orthogonal matrix** if the **column vectors** of Q form an **orthonormal set** in R^n.

- **Theorem 5.5.5**: An $n \times n$ matrix Q is orthogonal if and only if $Q^TQ = I$. ($Q^{-1} = Q^T$)

Proof:

Example

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

- The matrix Q can be thought of as a linear transformation from R^2 to R^2 that has the effect of <u>rotating each vector by an angle θ</u> while leaving the length of the vector unchanged. Q^{-1} can be thought of as a <u>rotation by the angle $-\theta$</u>

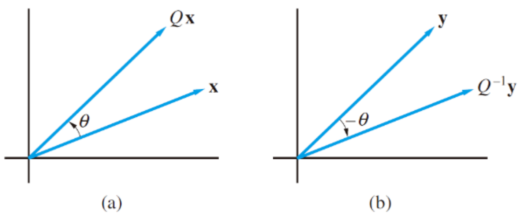


Figure 5.5.1.

- If Q is an $n \times n$ orthogonal matrix, then

(1) The column vectors of Q form an <u>orthonormal basis</u> for R^n.

(2) $Q^TQ = I$

(3) $Q^T = Q^{-1}$

(4) $\langle Q\mathbf{x}, Q\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$

(5) $||Q\mathbf{x}||_2 = ||\mathbf{x}||_2$

- A **permutation matrix** is a matrix formed from the <u>identity matrix</u> by **reordering its columns**.
    - A permutation matrix is an **orthogonal matrix**.
    - If P is the permutation matrix formed by reordering the columns of I in the order $(k_1, k_2, \ldots, k_n)$, then $P = (\mathbf{e}_{k1}, \mathbf{e}_{k2}, \ldots, \mathbf{e}_{kn})$. If A is an $m \times n$ matrix, then
    $$AP = (A\mathbf{e}_{k1}, A\mathbf{e}_{k2}, \ldots, A\mathbf{e}_{kn}) = (\mathbf{a}_{k1}, \mathbf{a}_{k2}, \ldots, \mathbf{a}_{kn})$$
    - <u>Post multiplication of A by P</u> reorders the columns of A in the order $(k_1, k_2, \ldots, k_n)$.

Example: Compute AP, PA, and P^{-1}.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix} \text{ and } P = \begin{array}{cc} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} & \begin{matrix} 2 \\ 3 \\ 1 \end{matrix} \\ \begin{matrix} 3 & 1 & 2 \end{matrix} & \end{array}$$

- **Theorem 5.5.6**: If the column vectors of A form an orthonormal set of vectors in R^m, then $A^TA = I$ and the solution to the least squares problem is $x^* = A^T\mathbf{b}$.

- **Theorem 5.5.7**: Let S be a subspace of an inner product space V and let $\mathbf{x} \in V$. Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be an orthonormal basis of S. If $\mathbf{p} = \sum_{i=1}^{n} c_i\mathbf{x}_i$, where $c_i = \langle \mathbf{x}, \mathbf{x}_i \rangle$ for each i then $\mathbf{p} - \mathbf{x} \in S^\perp$

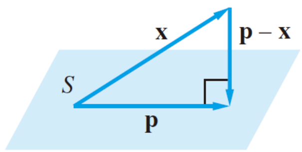


Figure 5.5.2.

Proof: (1) First, we will show that $(\mathbf{p} - \mathbf{x}) \perp \mathbf{x}_i$ for each i

$$\langle \mathbf{x}_i, \mathbf{p} - \mathbf{x} \rangle = \langle \mathbf{x}_i, \mathbf{p} \rangle - \langle \mathbf{x}_i, \mathbf{x} \rangle = \left\langle \mathbf{x}_i, \sum_{j=1}^{n} c_j \mathbf{x}_j \right\rangle - c_i \quad = \sum_{j=1}^{n} c_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - c_i = c_i - c_i = 0$$

So, $\mathbf{p} - \mathbf{x}$ is orthogonal to all $\mathbf{u}_i$'s

(2) If $\mathbf{y} \in S$, then $\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i$ and hence $\langle \mathbf{p} - \mathbf{x}, \mathbf{y} \rangle = \left\langle \mathbf{p} - \mathbf{x}, \sum_{i=1}^{n} \alpha_i \mathbf{x}_i \right\rangle = \sum_{i=1}^{n} \alpha_i \langle \mathbf{p} - \mathbf{x}, \mathbf{x}_i \rangle = 0$

- **Theorem 5.5.8**: Under the hypothesis of Theorem 5.5.7, $\mathbf{p}$ is the element of S that is closest to $\mathbf{x}$,

$$\|\mathbf{y} - \mathbf{x}\| \geq \|\mathbf{p} - \mathbf{x}\|, \qquad \text{for any } \mathbf{y} \neq \mathbf{p} \text{ in } S$$

Proof: From Theorem 5.5.7, $\mathbf{p} - \mathbf{x}$ is orthogonal to any vector in $\boldsymbol{S}$, thus $(\mathbf{p} - \mathbf{x}) \perp (\mathbf{y} - \mathbf{p})$.

From the Pythagorean Law, we can get

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

- **Theorem 5.5.9**: Let S be a nonzero subspace of R^m and let $\mathbf{b} \in R^m$, If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthonormal basis for S and $U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k)$, then the projection $\mathbf{p}$ of $\mathbf{b}$ onto S is given by

$$\mathbf{p} = UU^T\mathbf{b}$$

Proof: From Theorem 5.5.7, the projection $\mathbf{p}$ of $\mathbf{b}$ onto S is given by

$$\mathbf{p} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = U\mathbf{c}$$

Where $c_i = \langle \mathbf{b}, \mathbf{u}_i \rangle = \mathbf{u}_i^T\mathbf{b}$ and $\Rightarrow \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{b} \\ \mathbf{u}_2^T\mathbf{b} \\ \vdots \\ \mathbf{u}_k^T\mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \vdots \\ \mathbf{u}_k^T \end{bmatrix} \mathbf{b} = U^T\mathbf{b} \quad \Rightarrow \mathbf{p} = U\mathbf{c} = UU^T\mathbf{b}$

- The matrix UU^T is the <u>projection matrix</u> corresponding to the subspace S of R^m

  To project any vector $\mathbf{b} \in R^m$ onto S, we need only

  (1) find an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ for S

  (2) form and he matrix UU^T

  (3) multiply UU^T times $\mathbf{b}$

Chapter 6. Eigenvalues

第六章介紹了矩陣特徵值問題，由之前線性美好的世界來到了非線性的世界，用途之廣可能超乎你的想像，對於影像處理、人工智慧、機器學習、電腦視覺、訊號處理、資料探勘、資訊檢索、資料壓縮、科學計算、圖譜論、網路分析、控制理論、最佳化等，對這些領域有興趣的同學要好好學。第一節介紹基本定義和性值。第二節是以微分方程介紹這個問題的來源，我們會先跳過，交由工數來講。第三節介紹了矩陣的對角化，更深入的探討了矩陣特徵值的特性。第四章介紹了對稱矩陣的特徵值問題，我們在這邊會避開複數運算的定義，直接以實數的結果來看，所以內容會和課本有不同。第五章介紹奇異值分解，矩陣不再需要是正方形的，可以是任意長方形，都可以有奇異值分解，最重要的是它在 low rank approximation 的應用。

6.1 Eigenvalues and Eigenvectors

特徵值和特徵向量的定義。

- **Definition:** Let A be an $n \times n$ matrix. A scalar λ is said to be an **eigenvalue** or a **characteristic value** of A if there exists a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. The vector $\mathbf{x}$ is said to be an **eigenvector** or a **characteristic vector** belonging to λ.

Example:

$$\text{If } A = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

$$\text{then } A\mathbf{x} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 3\mathbf{x}$$

- If $A\mathbf{x} = \lambda\mathbf{x} \Rightarrow A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0} \Rightarrow (A - \lambda I)\mathbf{x} = \mathbf{0}$. λ is an eigenvalue of A if and only if (1) has a nontrivial solution. The set of solutions to (1) is $N(A - \lambda I)$, which is a subspace of R^n.
- If λ is an eigenvalue of A, then $N(A - \lambda I) \neq \{\mathbf{0}\}$ and any nonzero vector in $N(A - \lambda I)$ is an eigenvector belonging to λ. The subspace $N(A - \lambda I)$ is called the **eigenspace** corresponding to the eigenvalue λ.
- $(A - \lambda I)\mathbf{x} = \mathbf{0}$ will have a **nontrivial solution** if and only if $(A - \lambda I)$ is singular, or $\det(A - \lambda I) = 0$ which is called the **characteristic equation** for the matrix A.

Example: Find the eigenvalues and the corresponding eigenvectors of the matrix

$$A = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix} \qquad A = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix} \qquad A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

- If A is an $n \times n$ matrix with real entries, then the characteristic polynomial of A will have **real** coefficients. All of its **complex roots** must occur in **conjugate pairs**. That is, if $\lambda = a + bi$ ($b \neq 0$) is an eigenvalue of A, then $= a - bi$ must also be an eigenvalue of A.

- If $P(\lambda)$ is the characteristic polynomial of an $n \times n$ matrix A, then

$$P(\lambda) = \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix}$$

Expanding along the first column, we get

$$\det(A - \lambda I) = (a_{11} - \lambda)\det(M_{11}) + \sum_{i=2}^{n} a_{i1}(-1)^{i+1}\det(M_{i1})$$

where the minor M_{i1} does not contain the two diagonal elements $(a_{11} - \lambda)$ and $(a_{ii} - \lambda)$.
Expanding $\det(M_{11})$, we conclude that $(a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda)$ is the only term in the expansion of $\det(A - \lambda I)$ involving a product of more than $n - 2$ of the diagonal elements

the coefficient of λ^n is $(-1)^n$ the coefficient of $(-\lambda)^{n-1}$ is $\sum_{i=1}^{n} a_{ii}$

- If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A, then

$$P(\lambda) = (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \ldots (\lambda - \lambda_n) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_n - \lambda) \qquad (6)$$

$$p(0) = \det(A) = \lambda_1 \lambda_2 \ldots \lambda_n$$

The coefficient of $(-\lambda)^{n-1}$ is $\boxed{\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}}$

The sum of the diagonal elements of A is called the **trace** of A and is denoted by tr**(**A**)**

Example: For $A = \begin{bmatrix} 5 & -18 \\ 1 & -1 \end{bmatrix}$, the characteristic polynomial of A is $\begin{vmatrix} 5 - \lambda & -18 \\ 1 & -1 - \lambda \end{vmatrix} = \lambda^2 - 4\lambda + 13$

Hence the eigenvalues of A are $\lambda_1 = 2 + 3i$ and $\lambda_2 = 2 - 3i$. $\lambda_1 + \lambda_2 = 4 = \mathrm{tr}(A)$, $\lambda_1\lambda_2 = 13 = \det(A)$

- Theorem 6.1.1: Let A and B be $n \times n$ matrices. If B is similar to A ($B = S^{-1}AS$), then A and B have **the same characteristic polynomial** and consequently both have the **same eigenvalues.**

Example: Show T and S are similar and find their eigenvalues.

$$T = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \text{ and } S = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$$

6.3 Diagonalization

矩陣特徵值的性質和對角化。

- An $n \times n$ matrix A is said to be **diagonalizable** if there exists a nonsingular matrix X and a diagonal matrix D such that $X^{-1}AX = D$ (i.e., $A = XDX^{-1}$). We say that X **diagonalizes** A.

- **Theorem 6.3.2**: An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

Proof:

- **Theorem 6.3.1**: If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of an $n \times n$ matrix A with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, then $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are linearly independent.

Proof:

- Once we have a factorization $A = XDX^{-1}$, it is easy to compute powers of A.

- An $n \times n$ matrix A is said to be **defective** if A has fewer than n linearly independent eigenvectors. From Theorem 6.3.2, *a defective matrix is not diagonalizable*

Example: $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

Example:

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 4 & 0 \\ -3 & 6 & 2 \end{bmatrix}$$

- The matrix B has the effect of stretching two linearly independent vectors by a factor of 2
- The eigenvalue $\lambda = 2$ has *geometric multiplicity* 2 since the dim of the eigenspace $N(B - 2I)$ is 2
- The matrix A only stretches the vectors along the z axis by a factor of 2
- The eigenvalue $\lambda = 2$ has *algebraic multiplicity* 2, but dim $N(A - 2I) = 1$, *so its geometric multiplicity is only* 1

6.4 Symmetric Matrix

對稱矩陣 $A=A^T$ 的特徵值性質，好的不得了。

- **Theorem 6.4.1**: The eigenvalues of a symmetric matrix are all <u>real</u>. Furthermore, eigenvectors belonging to distinct eigenvalues are <u>orthogonal</u>.

Proof:

- **Theorem 6.4.4 (Spectral Theorem—Real Symmetric Matrices)** : If *A* is real and symmetric, then there exists an orthogonal matrix *U* that diagonalizes *A*.

Proof:

Example: find an orthogonal matrix *U* that diagonalizes *A*.

$$A=\begin{bmatrix} 0 & 2 & -1 \\ 2 & 3 & -2 \\ -1 & -2 & 0 \end{bmatrix}$$

- A matrix *A* is said to be **normal** if $AA^T = A^TA$.
- **Theorem 6.4.6**: A matrix *A* is normal if and only if *A* possesses a complete set of orthonormal eigenvectors.

Proof:

6.5 The Singular Value Decomposition

在對稱矩陣的特徵值基礎上，解紹奇異值分解，它的存在性，和 low rank approximation 的理論。

- For an $m \times n$ matrix A, with $m > n$, the matrix A can be factored into a product $U\Sigma V^T$, where U is an $m \times m$ <u>orthogonal</u> matrix, V is an $n \times n$ <u>orthogonal</u> matrix, $\mathbf{\Sigma}$ is an $m \times n$ matrix whose off diagonal entries are all 0's and whose diagonal entries satisfy $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.
- The σ_i's determined by this factorization are <u>unique</u> and are called the *singular values* of A.   The factorization $U\Sigma V^T$ is called the *singular value decomposition* (SVD) of A.

- **Theorem 6.5.1 (The SVD Theorem)** : If A is an $m \times n$ matrix, then A has a singular value decomposition.

Proof: A^TA is a <u>symmetric</u> $n \times n$ matrix (From Corollary 6.4.5): The eigenvalues of A^TA are all <u>real</u>.

A^TA has an orthogonal diagonal matrix V (V^T <u>A^TA</u> $V = D$). The eigenvalues must all be <u>nonnegative.</u>

Let λ be an eigenvalue of A^TA and **x** be an eigenvector belonging to λ (i.e., $A^TA\lambda = \lambda\mathbf{x}$)

Let the columns of V be ordered such that the corresponding eigenvalues satisfy $\lambda_1 \geq \lambda_2 \geq \ldots\lambda_n \geq 0$

The singular value of A are given by $\sigma_j = \sqrt{\lambda_j} \quad j = 1, 2, \ldots, n$

Let r denote the rank of $A \Rightarrow$ The matrix A^TA will also have rank r. Since A^TA is <u>symmetric</u>, its rank equals the number of nonzero eigenvalues:

- Let A be an $m \times n$ matrix with a singular value decomposition $U\Sigma V^T$
  - The singular values $\sigma_1, \sigma_2, \ldots, \sigma_n$ of A are <u>unique</u>; however the matrices U and V are not unique.
  - Since V diagonalizes A^TA, the $\mathbf{v}_j$'s are eigenvectors of A^TA**.**
  - Since $AA^T = U\Sigma V^T V\Sigma^T U^T = U\Sigma\Sigma^T U^T$, U diagonalizes AA^T and that the $\mathbf{u}_j$'s are eigenvectors of AA^T
  - Comparing the jth column of each side of $AV = U\Sigma$, we get $A\,\mathbf{v}_j = \sigma_j\,\mathbf{u}_j$, $j = 1, \ldots, n$

    Similarly, sinve $AV = U\Sigma$,

    $(AV)^T = V^TA^T = (U\Sigma)^T = \Sigma^TU^T \quad \Rightarrow A^T = V\Sigma^TU^T \Rightarrow A^TU = (V\Sigma^TU^T)U = V\Sigma^T$

    $A^T\,\mathbf{u}_j = \sigma_j\mathbf{v}_j$, $j = 1, \ldots, n$ and $A^T\,\mathbf{u}_j = \mathbf{0}$, $j = n+1, \ldots, m$

    The $\mathbf{v}_j$'s are called the *right singular vectors* of A, and the $\mathbf{u}_j$'s are called the *left singular vectors* of A.

- If A has rank r, then
  - $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ form an orthonormal basis for $R(A^T)$
  - $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n$ form an orthonormal basis for $N(A)$
  - $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ form an orthonormal basis for $R(A)$
  - $\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \ldots, \mathbf{u}_m$ form an orthonormal basis for $N(A^T)$
- The rank of the matrix A is equal to the number of nonzero singular values.
- In the case that A has rank $r < n$, if we set and , then

$$A = U_1 \Sigma_1 V_1^T$$

The factorization is called the *compact form of the singular value decomposition of A*

- If A is an $m \times n$ matrix with rank r and $0 < k < r$, we can use the singular value decomposition to find a matrix in $R^{m \times n}$ of rank k that is closest to A with respect to the Frobenius norm

  Let M be the set of all $m \times n$ matrices of rank k or less, it can be shown that there is a matrix X in M such that

$$\| A - X \|_F = \min_{S \in \mathcal{M}} \| A - S \|_F$$

- **Lemma 6.5.2**: If A is an $m \times n$ matrix and Q is an $m \times m$ orthogonal matrix, then $\|QA\|_F = \|A\|_F$

Proof:

- If A has singular value decomposition $U\Sigma V^T$, then $\| A \|_F = \| U\Sigma V^T \|_F = \| \Sigma V^T \|_F$. Since $\| \Sigma V^T \|_F = \| (\Sigma V^T)^T \|_F = \| V\Sigma^T \|_F = \| \Sigma^T \|_F$.

$$\boxed{\| A \|_F = \| \Sigma V^T \|_F = \| \Sigma^T \|_F = \left(\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2\right)^{1/2}}$$

- **Theorem 6.5.3**: Let $A = U\Sigma V^T$ be an $m \times n$ matrix and let M denote the set of all $m \times n$ matrices of rank k or less, where $0 < k <$ rank(A). If X is a matrix in M satisfying

$$\| A - X \|_F = \min_{S \in \mathcal{M}} \| A - S \|_F$$

then

$$\| A - X \|_F = \left(\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2\right)^{1/2}$$

In particular, if $A' = U\Sigma' V^T$, where

$$\Sigma' = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & O \\ & & \sigma_k & \\ & O & & O \end{bmatrix} = \begin{bmatrix} \Sigma_k & O \\ O & O \end{bmatrix}$$

$$\| A - A' \|_F = \left(\sigma_{k+1}^2 + \sigma_{k+2}^2 + \cdots + \sigma_n^2\right)^{1/2} = \min_{S \in \mathcal{M}} \| A - S \|_F$$