Chapter 4

Inner Product Space

在 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 中大家熟悉內積 (dot product) 的定義可以推廣到一般的 $\mathbb{R}^n$. 甚至可推廣到更一般的 vector space. 在一般的 vector space, 我們就不再用 dot product 而用所謂 *inner product* 來稱之. Inner product 可以幫助我們定義出 vector space 中許多重要的 subspaces. 具有 inner product 的 vector space 就稱為 *inner product space*. 在這一章中我們將介紹 inner product space 的性質.

4.1. Dot Product

在本節我們僅論及大家熟悉的內積性質在 $\mathbb{R}^n$ 的情況.

首先我們回顧在 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 中內積的定義. 若在 $\mathbb{R}^2$ 中 $\mathbf{u}=(a_1,a_2),\mathbf{v}=(b_1,b_2)$, 則 $\mathbf{u},\mathbf{v}$ 的內積 $\mathbf{u}\cdot\mathbf{v}$ 定義成 $\mathbf{u}\cdot\mathbf{v}=a_1b_1+a_2b_2$. 而在 $\mathbb{R}^3$ 中若 $\mathbf{u}=(a_1,a_2,a_3),\mathbf{v}=(b_1,b_2,b_3)$, 則 $\mathbf{u},\mathbf{v}$ 的內積 $\mathbf{u}\cdot\mathbf{v}$ 定義成 $\mathbf{u}\cdot\mathbf{v}=a_1b_1+a_2b_2+a_3b_3$. 由這定義我們很自然地可推廣到 $\mathbb{R}^n$ 中向量的內積如下:

Definition 4.1.1. 假設 $\mathbf{u}=(a_1,\ldots,a_n),\mathbf{v}=(b_1,\ldots,b_n)\in\mathbb{R}^n$. 則定義 $\mathbf{u},\mathbf{v}$ 的 *dot product* (*inner product*) 為

$$\mathbf{u}\cdot\mathbf{v}=a_1b_1+\cdots+a_nb_n=\sum_{i=1}^{n}a_ib_i.$$

向量的內積和向量的運算有一定的關係, 以下就是它們之間的關係

Proposition 4.1.2. 對任意 $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^n$, 我們有以下的性質:

(1) $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$.

(2) $\mathbf{u}\cdot\mathbf{u}\geq 0$ 且 $\mathbf{u}\cdot\mathbf{u}=0$ 若且唯若 $\mathbf{u}=\mathbf{0}$.

(3) 對任意 $r\in\mathbb{R}$ 皆有 $(r\mathbf{u})\cdot\mathbf{v}=\mathbf{u}\cdot(r\mathbf{v})=r(\mathbf{u}\cdot\mathbf{v})$.

(4) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$.

Proof. 這些性質在 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 大家應都了解, 在 $\mathbb{R}^n$ 上的證明其實也一樣. 我們假設 $\mathbf{u}=(a_1,\ldots,a_n),\mathbf{v}=(b_1,\ldots,b_n),\mathbf{w}=(c_1,\ldots,c_n)$.

(1) 依定義 $\mathbf{u}\cdot\mathbf{v}=\sum_{i=1}^n a_ib_i$ 由於每一項 a_ib_i 皆等於 b_ia_i (實數乘法交換率) 所以我們知道它們的和也相等, 也就是說 $\sum_{i=1}^n a_ib_i=\sum_{i=1}^n b_ia_i=\mathbf{v}\cdot\mathbf{u}$. 所以我們得 $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$.

(2) 依定義 $\mathbf{u}\cdot\mathbf{u}=\sum_{i=1}^n a_ia_i=\sum_{i=1}^n a_i^2$. 由於任一實數的平方皆大於等於 0, 即 $a_i^2\geq 0$, 故有 $\sum_{i=1}^n a_i^2\geq 0$, 而得證 $\mathbf{u}\cdot\mathbf{u}\geq 0$. 又上式中若 $\sum_{i=1}^n a_i^2=0$, 表示每一項 a_i^2 皆需等於 0, 故知對任意 $1\leq i\leq n$ 皆需有 $a_i=0$, 而得知 $\mathbf{u}=(a_1,\ldots,a_n)=(0,\ldots,0)=\mathbf{0}$. 反之若 $\mathbf{u}=(a_1,\ldots,a_n)=\mathbf{0}$ 表示對任意 $1\leq i\leq n$ 皆有 $a_i=0$, 故得 $\mathbf{u}\cdot\mathbf{u}=\sum_{i=1}^n a_ia_i=0$.

(3) $(r\mathbf{u})\cdot\mathbf{v}$ 這個符號表示 $r\mathbf{u}$ 這個向量與 $\mathbf{v}$ 的內積, 因 $r\mathbf{u}=(ra_1,\ldots,ra_n)$ 故由定義知 $(r\mathbf{u})\cdot\mathbf{v}=\sum_{i=1}^n(ra_i)b_i$. 又對所有的 $1\leq i\leq n$ 皆有 $(ra_i)b_i=r(a_ib_i)$ (實數乘法結合律) 故知 $\sum_{i=1}^n(ra_i)b_i=\sum_{i=1}^n r(a_ib_i)$ 再加上 $\sum_{i=1}^n r(a_ib_i)$ 中每一項皆有 r 可提出, 故由實數加法與乘法的分配律可知 $\sum_{i=1}^n r(a_ib_i)=r\sum_{i=1}^n a_ib_i=r(\mathbf{u}\cdot\mathbf{v})$, 而得證 $(r\mathbf{u})\cdot\mathbf{v}=r(\mathbf{u}\cdot\mathbf{v})$. 我們也可用同樣方法證得 $\mathbf{u}\cdot(r\mathbf{v})=r(\mathbf{u}\cdot\mathbf{v})$, 不過我們這裡可利用 (1) 知 $\mathbf{u}\cdot(r\mathbf{v})=(r\mathbf{v})\cdot\mathbf{u}$ 再利用剛才的結果得 $(r\mathbf{v})\cdot\mathbf{u}=r(\mathbf{v}\cdot\mathbf{u})$, 再利用一次 (1) 得到 $r(\mathbf{v}\cdot\mathbf{u})=r(\mathbf{u}\cdot\mathbf{v})$ 而得證 $\mathbf{u}\cdot(r\mathbf{v})=r(\mathbf{u}\cdot\mathbf{v})$.

(4) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})$ 這個符號表示 $\mathbf{u}$ 這個向量與 $\mathbf{v}+\mathbf{w}$ 的內積, 因 $\mathbf{v}+\mathbf{w}=(b_1+c_1,\ldots,b_n+c_n)$ 故由定義知 $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\sum_{i=1}^n a_i(b_i+c_i)$. 由實數加法與乘法的分配律知每一項 $a_i(b_i+c_i)$ 可表為 $a_ib_i+a_ic_i$, 也就是說 $\sum_{i=1}^n a_i(b_i+c_i)=\sum_{i=1}^n(a_ib_i+a_ic_i)$. 因為實數加法有交換率, 我們可以先將 a_ib_i 的部份先加在一起, 再將 a_ic_i 的部份加在一起, 再求它們之和, 故知

$$\sum_{i=1}^n(a_ib_i+a_ic_i)=\sum_{i=1}^n a_ib_i+\sum_{i=1}^n a_ic_i=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w},$$

依此得證 $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$. □

Proposition 4.1.2 (2) 告訴我們除了零向量 $\mathbf{0}$ 以外, 其餘向量 $\mathbf{v}$ 皆需符合 $\mathbf{v}\cdot\mathbf{v}>0$, 所以很自然地我們可依此定義向量的長度.

Definition 4.1.3. 令 $\mathbf{v}=(a_1,\ldots,a_n)\in\mathbb{R}^n$, 我們定義 $\mathbf{v}$ 的 *norm* (or *length*) 為

$$\|\mathbf{v}\|=\sqrt{\mathbf{v}\cdot\mathbf{v}}=\sqrt{a_1^2+a_2^2+\cdots+a_n^2}.$$

我們可以利用 Proposition 4.1.2 的處理一些有關於內積的性質, 而不必涉及內積的定義.

Lemma 4.1.4. 假設 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, 則 $\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+2\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2$.

Proof. 依定義 $\|\mathbf{u}+\mathbf{v}\|^2=(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})$, 再依 Proposition 4.1.2 (4) 可得

$$(\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v})=(\mathbf{u}+\mathbf{v})\cdot\mathbf{u}+(\mathbf{u}+\mathbf{v})\cdot\mathbf{v}=\mathbf{u}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}+\mathbf{v}\cdot\mathbf{v}.$$

最後再依 Proposition 4.1.2 (1) 的交換律知 $\mathbf{v}\cdot\mathbf{u}+\mathbf{u}\cdot\mathbf{v}=2\mathbf{u}\cdot\mathbf{v}$ 而得證本定理. □

Question 4.1. 試證明平行四邊形定理 *(parallelogram relation):* 平行四邊形兩對角線長的平方和等於四邊長的平方和, 即對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 皆有

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

再次強調一次 Lemma 4.1.4 僅用到內積的性質, 所以在一般的情形若我們不是利用 Definition 4.1.1 的方法定義內積 (當然此時長度的定義也跟著改變) 但所定義的內積仍保有 Proposition 4.1.2 中的性質, 我們依然可得到 Lemma 4.1.4 中的性質. Lemma 4.1.4 最常見的就是可以幫助我們推得所謂的「柯希、舒瓦茲」不等式.

Proposition 4.1.5 (Cauchy-Schwarz inequality)**.** 若 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, 則 $|\mathbf{u}\cdot\mathbf{v}|\leq\|\mathbf{u}\|\,\|\mathbf{v}\|$. 特別地當 $\mathbf{u},\mathbf{v}$ 皆不為零向量時, 等號成立若且唯若存在 $\lambda\in\mathbb{R}$ 使得 $\mathbf{v}=\lambda\mathbf{u}$.

Proof. 假設 $\mathbf{u}$ 和 $\mathbf{v}$ 中有一個為零向量, 即 $\mathbf{u}\cdot\mathbf{v}=0$ 且 $\|\mathbf{u}\|\,\|\mathbf{v}\|=0$, 故此不等式成立.

若 $\mathbf{u},\mathbf{v}$ 皆不為零向量, 考慮 $\mathbf{u}_0=\mathbf{u}/\|\mathbf{u}\|$ 且 $\mathbf{v}_0=\mathbf{v}/\|\mathbf{v}\|$. 此時

$$\|\mathbf{u}_0\|^2=\mathbf{u}_0\cdot\mathbf{u}_0=\frac{1}{\|\mathbf{u}\|}\mathbf{u}\cdot\frac{1}{\|\mathbf{u}\|}\mathbf{u}=\frac{1}{\|\mathbf{u}\|^2}\mathbf{u}\cdot\mathbf{u}=1.$$

同理得 $\|\mathbf{v}_0\|^2=1$, 故由 Lemma 4.1.4 得知

$$\begin{aligned}\|\mathbf{u}_0+\mathbf{v}_0\|^2&=2+2\mathbf{u}_0\cdot\mathbf{v}_0, \qquad (4.1)\\ \|\mathbf{u}_0-\mathbf{v}_0\|^2&=2-2\mathbf{u}_0\cdot\mathbf{v}_0.\end{aligned}$$

因為 $\|\mathbf{u}_0+\mathbf{v}_0\|^2\geq 0$ 且 $\|\mathbf{u}_0-\mathbf{v}_0\|^2\geq 0$, 故由式子 (4.1) 得 $-1\leq\mathbf{u}_0\cdot\mathbf{v}_0\leq 1$. 換回 $\mathbf{u},\mathbf{v}$ 得

$$-\|\mathbf{u}\|\,\|\mathbf{v}\|\leq\mathbf{u}\cdot\mathbf{v}\leq\|\mathbf{u}\|\,\|\mathbf{v}\|.$$

亦即 $|\mathbf{u}\cdot\mathbf{v}|\leq\|\mathbf{u}\|\,\|\mathbf{v}\|$.

從上可知當 $\mathbf{u},\mathbf{v}$ 皆不為零向量時, 此不等式之等式會成立等同於 $\mathbf{u}_0\cdot\mathbf{v}_0=1$ 或 $\mathbf{u}_0\cdot\mathbf{v}_0=-1$. 此時由式子 (4.1) 分別得 $\|\mathbf{u}_0-\mathbf{v}_0\|^2=0$ 或 $\|\mathbf{u}_0+\mathbf{v}_0\|^2=0$, 也就是說 $\mathbf{u}_0=\mathbf{v}_0$ 或 $\mathbf{u}_0=-\mathbf{v}_0$. 換回 $\mathbf{u},\mathbf{v}$ 我們得

$$\mathbf{v}=\frac{\|\mathbf{v}\|}{\|\mathbf{u}\|}\mathbf{u}\text{ 或}\mathbf{v}=-\frac{\|\mathbf{v}\|}{\|\mathbf{u}\|}\mathbf{u}.$$

故此時只要令 λ 分別為 $\|\mathbf{v}\|/\|\mathbf{u}\|$ 或 $-\|\mathbf{v}\|/\|\mathbf{u}\|$, 即可得 $\mathbf{v}=\lambda\mathbf{u}$.

反之若 $\mathbf{v}=\lambda\mathbf{u}$, 則由 Proposition 4.1.2 可得

$$|\mathbf{u}\cdot\mathbf{v}|=|\lambda|\,|\mathbf{u}\cdot\mathbf{u}|=|\lambda|\,\|\mathbf{u}\|^2=\|\mathbf{u}\|\,\|\lambda\mathbf{u}\|=\|\mathbf{u}\|\,\|\mathbf{v}\|.$$

□

在證明 Proposition 4.1.5 時, 我們用了一個很特殊的技巧, 就是將 $\mathbf{u}$ 化成長度為 1 的向量 $\mathbf{u}_0=\mathbf{u}/\|\mathbf{u}\|$. 一般來說一個長度為 1 的向量, 我們稱之為 *unit vector*. 任意的非零向量 $\mathbf{u}$ 都可以化成 unit vector, 即取 $\mathbf{u}_0=\mathbf{u}/\|\mathbf{u}\|$. 這種化成 unit vector 的方法不只讓我們確定向量的長度且又保有原向量的方向性, 是線性代數處理內積有關的問題很好用的技巧.

利用 Proposition 4.1.5, 我們可以得到所謂的三角不等式.

Corollary 4.1.6 (Triangle inequality)**.** 若 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, 則 $\|\mathbf{u}+\mathbf{v}\|\leq\|\mathbf{u}\|+\|\mathbf{v}\|$.

Proof. 由 Lemma 4.1.4 以及 Proposition 4.1.5, 我們有

$$\|\mathbf{u}+\mathbf{v}\|^2=\|\mathbf{u}\|^2+2\mathbf{u}\cdot\mathbf{v}+\|\mathbf{v}\|^2\leq\|\mathbf{u}\|^2+2\|\mathbf{u}\|\,\|\mathbf{v}\|+\|\mathbf{v}\|^2=(\|\mathbf{u}\|+\|\mathbf{v}\|)^2.$$

不等式兩邊開根號得證 $\|\mathbf{u}+\mathbf{v}\|\leq\|\mathbf{u}\|+\|\mathbf{v}\|$. □

Question 4.2. 試找到充分必要條件使得 $\|\mathbf{u}+\mathbf{v}\| = \|\mathbf{u}\|+\|\mathbf{v}\|$.

利用內積我們可以知道坐標平面或空間中向量之間的一些幾何關係. 例如若兩非零向量 $\mathbf{u},\mathbf{v}$ 的夾角為 θ, 因為 $\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta$, 所以我們可以利用內積得知此二非零向量所夾角度. 特別地當 $\mathbf{u}\cdot\mathbf{v} = 0$ 即表示 $\mathbf{u}$ 和 $\mathbf{v}$ 垂直. 我們也可將此幾何意義推廣到更一般的 $\mathbb{R}^n$. 雖然當 $n \geq 4$ 時, 我們無法 "看到" $\mathbb{R}^n$ 中的向量 (無法用幾何的方式來定義夾角), 此時我們可以沿習 $\mathbb{R}^2,\mathbb{R}^3$ 上的結果定義兩非零向量 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 的夾角為 θ, 其中 $0 \leq \theta \leq \pi$ 使得

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}.$$

當我們定義一個東西時要注意這個定義是否 "well-defined". 也就是說要確認這樣定義出來的夾角 θ 是否可以找得到, 這是所謂「存在性」的問題. 我們都知道當 $0 \leq \theta \leq \pi$ 時, $|\cos\theta| \leq 1$. 所以這裡夾角 θ 的存在性就關係到 $\mathbb{R}^n$ 中兩個非零向量 $\mathbf{u},\mathbf{v}$ 是否會滿足

$$\left|\frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}\right| \leq 1.$$

然而 Proposition 4.1.5 告訴我們這是一定對的, 所以這裡 θ 的存在性沒問題. 另一個要確認的問題是, 這樣定出來的夾角會不會有兩個或更多呢? 這是所謂「唯一性」的問題. 就是因為會有 $\theta' \neq \theta$ 但 $\cos\theta = \cos\theta'$ 的情形發生, 所以這裡我們要求 θ 要滿足 $0 \leq \theta \leq \pi$, 如此才能確保所得的夾角會是唯一的. 也就是說用這種方法定義兩非零向量的夾角是沒有問題的, 我們就稱這樣的定義是 well-defined.

Example 4.1.7. 在 $\mathbb{R}^4$ 中設 $\mathbf{u} = (1,1,1,1), \mathbf{v} = (1,0,-2,-2)$ 的夾角為 θ, 則由

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \frac{-3}{2\times 3} = -\frac{1}{2},$$

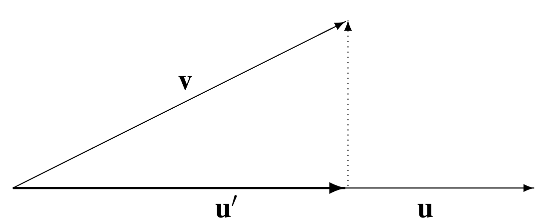
得知 $\theta = 120^\circ$.

利用夾角的定義我們進而定義出何謂「垂直」.

Definition 4.1.8. 令 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 為非零向量, 我們說 $\mathbf{u}$ 和 $\mathbf{v}$ 為 *orthogonal* 若且唯若 $\mathbf{u}\cdot\mathbf{v} = 0$.

注意這裡因在 $\mathbb{R}^n$ 空間, 習慣上垂直我們稱為 *orthogonal* 而較少用一般幾何上的 *perpendicular*. 有了垂直概念後, 我們也可以將 $\mathbb{R}^2$ 或 $\mathbb{R}^3$ 上的向量在另一向量上的投影 (projection) 之概念推廣至 $\mathbb{R}^n$.

我們先看 $\mathbb{R}^2$ 的情況, 給定一非零向量 $\mathbf{u}\in\mathbb{R}^2$, 對任意 $\mathbf{v}\in\mathbb{R}^2$, 若 $\mathbf{u}'$ 為 $\mathbf{v}$ 在 $\mathbf{u}$ 上的投影, 表示向量 $\mathbf{v}-\mathbf{u}'$ (參考下圖虛線表示的向量) 會和 $\mathbf{u}$ 垂直, 即 $(\mathbf{v}-\mathbf{u}')\cdot\mathbf{u} = 0$, 也就是說 $\mathbf{v}\cdot\mathbf{u} = \mathbf{u}'\cdot\mathbf{u}$.



因為 $\mathbf{u}'$ 會落在 $\mathrm{Span}(\mathbf{u})$, 也就是說要找到 $r\in\mathbb{R}$ 使得 $\mathbf{u}'=r\mathbf{u}$, 且符合 $(\mathbf{v}-\mathbf{u}')\cdot\mathbf{u}=0$. 將 $\mathbf{u}'=r\mathbf{u}$ 代入得 $\mathbf{v}\cdot\mathbf{u}=r\mathbf{u}\cdot\mathbf{u}=r\|\mathbf{u}\|^2$, 亦即 $r=(\mathbf{v}\cdot\mathbf{u})/\|\mathbf{u}\|^2$. 也就是說, 若 $\mathbf{v}$ 在 $\mathbf{u}$ 的投影是存在的, 那們它的投影一定就是 $\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u}$ (這說明了投影的唯一性). 然而若令 $r=(\mathbf{v}\cdot\mathbf{u})/\|\mathbf{u}\|^2$ (注意 $\mathbf{u}$ 為非零向量的假設), 則 $\mathbf{u}'=r\mathbf{u}$ 確實符合 $(\mathbf{v}-\mathbf{u}')\cdot\mathbf{u}=0$ (這說明了投影的存在性). 我們可以將以上的概念推廣到 $\mathbb{R}^n$ 的情形.

Proposition 4.1.9. 給定一非零向量 $\mathbf{u}\in\mathbb{R}^n$, 對任意 $\mathbf{v}\in\mathbb{R}^n$, 皆可寫成 $\mathbf{v}=\mathbf{u}'+\mathbf{v}'$, 其中 $\mathbf{u}',\mathbf{v}'\in\mathbb{R}^n$ 滿足 $\mathbf{v}'\cdot\mathbf{u}=0$ 且 $\mathbf{u}'=r\mathbf{u}$, $r\in\mathbb{R}$. 事實上這樣的寫法是唯一的, 即

$$r=\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}.$$

Proof. 前面的論述在 $\mathbb{R}^n$ 亦成立, 亦即 $r=(\mathbf{v}\cdot\mathbf{u})/\|\mathbf{u}\|^2$ 是唯一的實數會使得 $(\mathbf{v}-r\mathbf{u})\cdot\mathbf{u}=0$. 換言之,

$$\mathbf{u}'=\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u}.$$

是唯一的向量會滿足 $\mathbf{u}'=r\mathbf{u}$ 且 $(\mathbf{v}-\mathbf{u}')\cdot\mathbf{u}=0$. 既然 $\mathbf{u}'$ 是唯一的, 故而 $\mathbf{v}'$ 要滿足 $\mathbf{v}'+\mathbf{u}'=\mathbf{v}$, 即 $\mathbf{v}'=\mathbf{v}-\mathbf{u}'$, 自然也就唯一確定了. □

Question 4.3. 能否不用找到 r 的方法得到 *Proposition 4.1.9* 的唯一性?

Proposition 4.1.9, 大致上是說給定一 $\mathbb{R}^n$ 中的非零向量 $\mathbf{u}$ 後, 我們都可以將 $\mathbb{R}^n$ 中任一向量 $\mathbf{v}$ 分解成兩個向量之和, 其中一個向量會落在 $\mathrm{Span}(\mathbf{u})$ (即定理中的 $\mathbf{u}'$) 而另一個與 $\mathbf{u}$ 垂直 (即定理中的 $\mathbf{v}'$), 且這個表法是唯一的. 我們稱落在 $\mathrm{Span}(\mathbf{u})$ 的那個向量

$$\frac{\mathbf{v}\cdot\mathbf{u}}{\|\mathbf{u}\|^2}\mathbf{u}$$

為 $\mathbf{v}$ 在 $\mathbf{u}$ 的 *projection* (投影).

Example 4.1.10. 在 $\mathbb{R}^4$ 中考慮 $\mathbf{u}=(1,1,1,1),\mathbf{v}=(1,0,-2,-2)$. 因 $\|\mathbf{u}\|=2$ 且 $\mathbf{v}\cdot\mathbf{u}=-3$, 得 $\mathbf{v}$ 在 $\mathbf{u}$ 的 projection 為

$$-\frac{3}{4}\mathbf{u}=-\frac{3}{4}(1,1,1,1).$$

又我們有

$$\mathbf{v}=(1,0,-2,-2)=-\frac{3}{4}(1,1,1,1)+(\frac{7}{4},\frac{3}{4},-\frac{5}{4},-\frac{5}{4}),$$

其中

$$-\frac{3}{4}(1,1,1,1)\in\mathrm{Span}((1,1,1,1))\quad\text{and}\quad(\frac{7}{4},\frac{3}{4},-\frac{5}{4},-\frac{5}{4})\cdot(1,1,1,1)=0.$$

4.2. Inner Product

在 $\mathbb{R}^m$ 中有關於 dot product 的性質, 可以推廣到一般 over $\mathbb{R}$ 的 vector space. 在一般的 vector space, 我們就不再用 dot product 而用所謂 *inner product* 來稱之. 由於我們要談的是一班的 inner product, 為了和原本 $\mathbb{R}^n$ 的 dot product 區分, 對於 $\mathbf{v},\mathbf{w}\in V$, 原來我們用 $\langle\mathbf{v},\mathbf{w}\rangle$ 來表示 $\mathbf{v},\mathbf{w}$ 的 inner product.

考慮 over $\mathbb{R}$ 的 vector space V. 若對於任意 $\mathbf{v},\mathbf{w}\in V$ 皆有 $\langle\mathbf{v},\mathbf{w}\rangle\in\mathbb{R}$ 滿足 Proposition 4.1.2 有關內積的四個性質, 即

(1) $\langle \mathbf{v},\mathbf{w}\rangle = \langle \mathbf{w},\mathbf{v}\rangle$, $\forall \mathbf{v},\mathbf{w}\in V$.

(2) 對任意 $r\in\mathbb{R}$ 皆有 $\langle r\mathbf{v},\mathbf{v}\rangle = \langle \mathbf{v},r\mathbf{w}\rangle = r\langle \mathbf{v},\mathbf{w}\rangle$.

(3) $\langle \mathbf{u},\mathbf{v}+\mathbf{w}\rangle = \langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{u},\mathbf{w}\rangle$, $\forall \mathbf{u},\mathbf{v},\mathbf{w}\in V$.

(4) $\langle \mathbf{v},\mathbf{v}\rangle \geq 0$ 且 $\langle \mathbf{v},\mathbf{v}\rangle = 0$ 若且唯若 $\mathbf{v}=\mathbf{0}$.

我們便稱 $\langle\,,\,\rangle$ 為 V 上的一個 *inner product* 而稱 V 為 *inner product space.*

簡單的來說, 當我們稱 V 是一個 inner product space 時, 表示我們已給定 V 中的一個 inner product $\langle\,,\,\rangle$. 例如在 $\mathbb{R}^n$ 上, 若我們定義 $\langle \mathbf{v},\mathbf{w}\rangle = \mathbf{v}\cdot\mathbf{w}$, $\forall \mathbf{v},\mathbf{w}\in\mathbb{R}^n$ (原先的 dot product), 那麼我們就說 $\mathbb{R}^n$ 是一個以 $\langle\,,\,\rangle$ 為內積的 inner product space. 另外要注意的是, 依我們的定義, 我們考慮的 inner product 是實數, 所以只有在 V 是 vector space over $\mathbb{R}$ 時, 才會談論是否為 inner product space. 事實上還有定義在複數 $\mathbb{C}$ 上的 inner product, 不過由於本課程並不需要用到這種情況, 我們就略過不談.

接下來我們介紹一個定義在 $\mathsf{M}_{m\times n}(\mathbb{R})$ 這一個 vector space 的 inner product. 其實當 $M\in\mathsf{M}_{m\times n}(\mathbb{R})$, 我們可以將 M 視為一個在 $\mathbb{R}^{mn}$ 上的向量, 也就是說先寫第一個 column 再接著串接第二個 column, 這樣一直下去將 n 個 column 寫成一個長長的 column. 所以我們可以將兩個 $\mathsf{M}_{m\times n}(\mathbb{R})$ 上的矩陣看成是兩個 $\mathbb{R}^{mn}$ 的向量, 因此再利用 $\mathbb{R}^{mn}$ 上的 dot product 就可以定義 $\mathsf{M}_{m\times n}(\mathbb{R})$ 上的 inner product 了. 也就是說若 $A,B\in\mathsf{M}_{m\times n}$ 其中 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 依序為 A 的 column vectors 而 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 依序為 B 的 column vectors, 則可定義 $\langle A,B\rangle = \mathbf{v}_1\cdot\mathbf{w}_1+\cdots+\mathbf{v}_n\cdot\mathbf{w}_n$. 由於 dot product 符合 inner product 的性質, 所以我們知道這個方法確實給了 $\mathsf{M}_{m\times n}(\mathbb{R})$ 上的一個 inner product. 事實上若從矩陣的乘法來看, 我們可以將 $\mathbb{R}^n$ 上的 column vector 看成 $n\times 1$ matrix, 所以當 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$, 將 $\mathbf{v},\mathbf{w}$ 視為 $\mathsf{M}_{n\times 1}(\mathbb{R})$ 中的矩陣, 則 $\mathbf{v}\cdot\mathbf{w} = \mathbf{w}^{\mathrm{t}}\mathbf{v}$. 從這個角度來看若 $A,B\in\mathsf{M}_{m\times n}$ 其中 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 依序為 A 的 column vectors 而 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 依序為 B 的 column vectors, 則 $\mathbf{v}_i\cdot\mathbf{w}_i$ 就是 $B^{\mathrm{t}}A$ 這個矩陣對角線上 (i,i)-th entry. 我們曾定義過一個方陣 M 其對角線上的 entries 之和稱為 M 的 *trace*, 記為 $\mathrm{tr}(M)$. 因此上面定義的 $\langle A,B\rangle = \mathbf{v}_1\cdot\mathbf{w}_1+\cdots+\mathbf{v}_n\cdot\mathbf{w}_n$ 也等於 $\mathrm{tr}(B^tA)$. 也就是說對於 $A,B\in\mathsf{M}_{m\times n}(\mathbb{R})$, 定義 $\langle A,B\rangle = \mathrm{tr}(B^{\mathrm{t}}A)$ 就給了 $\mathsf{M}_{m\times n}(\mathbb{R})$ 上的一個 inner product, 在此 inner product 之下 $\mathsf{M}_{m\times n}(\mathbb{R})$ 就形成一個 inner product space. 有些同學或許會疑惑為何不定義 $\langle A,B\rangle = \mathrm{tr}(A^{\mathrm{t}}B)$ 呢? 事實上對於任意方陣 M, 皆有 $\mathrm{tr}(M)=\mathrm{tr}(M^{\mathrm{t}})$, 所以 $\mathrm{tr}(A^{\mathrm{t}}B)=\mathrm{tr}((A^{\mathrm{t}}B)^{\mathrm{t}})=\mathrm{tr}(B^{\mathrm{t}}A)$. 不過以後當我們介紹更多 inner product 和矩陣關係時, $B^{\mathrm{t}}A$ 有其方便性, 所以一般來說我們會用 $\langle A,B\rangle = \mathrm{tr}(B^{\mathrm{t}}A)$ 來表示這一個 inner product.

Question 4.4. 對任意 $A,B\in\mathsf{M}_{m\times n}(\mathbb{R})$, 證明 $\mathrm{tr}(AB)=\mathrm{tr}(BA)$. 並利用此證明 $\langle A,B\rangle = \mathrm{tr}(B^{\mathrm{t}}A),\forall A.B\in\mathsf{M}_{m\times n}(\mathbb{R})$ 是 $\mathsf{M}_{m\times n}(\mathbb{R})$ 上的 *inner product.*

Question 4.5. 試說明 $\langle A,B\rangle = \det(B^{\mathrm{t}}A)$, $\forall A,B\in\mathsf{M}_{m\times n}(\mathbb{R})$ 是否為 $\mathsf{M}_{m\times n}(\mathbb{R})$ 的 *inner product?* 又若對任意 $M,M'\in\mathsf{M}_{n\times n}(\mathbb{R})$, 定義 $\langle M,M'\rangle = \mathrm{tr}(MM')$, 是否為 $\mathsf{M}_{n\times n}(\mathbb{R})$ 上的 *inner product?*

對於次數小於等於 n 的實係數多項式所形成的向量空間 $\mathsf{P}_n(\mathbb{R})$, 也有一個有趣的 inner product, 我們看以下的例子.

Example 4.2.1. 選取 $n+1$ 個相異的實數 $c_1,\ldots,c_{n+1}$, 定義

$$\langle f,g\rangle = f(c_1)g(c_1)+\cdots+f(c_{n+1})g(c_{n+1}) = \sum_{i=1}^{n+1} f(c_i)g(c_i), \quad \forall f,g\in \mathsf{P}_n(\mathbb{R}).$$

我們說明在此定義之下 $\langle,\rangle$ 是 $\mathsf{P}_n(\mathbb{R})$ 的 inner product.

考慮 $f,g,h\in\mathsf{P}_n(\mathbb{R})$ 以及 $r\in\mathbb{R}$, 關於 $\langle f,g\rangle=\langle g,f\rangle$, $\langle f,g+h\rangle=\langle f,g\rangle+\langle f,h\rangle$ 以及 $\langle f,rg\rangle=r\langle f,g\rangle$ 等性質很容易由實數的乘法加法性質推得, 這裡就不再證明. 我們僅證明 $\langle f,f\rangle\geq 0$ 以及 $\langle f,f\rangle=0$ 若且唯若 $f=0$ 這個性質. 依定義 $\langle f,f\rangle=f(c_1)^2+\cdots+f(c_{n+1})^2$. 由於 $f(c_i)$ 是實數, 故知 $f(c_i)^2\geq 0$, $\forall i=1,\ldots,n+1$. 由此得證 $\langle f,f\rangle\geq 0$. 又若 $\langle f,f\rangle=0$, 表示 $f(c_i)^2=0$, $\forall i=1,\ldots,n+1$. 亦即 $c_1,\ldots,c_{n+1}$ 是 $f(x)=0$ 的 $n+1$ 個相異實根. 因此由 f 是次數小於等於 n 的多項式知, 唯一的可能就是 f 是零多項式, 即 $f=0$.

Question 4.6. *在 Example 4.2.1 中, 若僅選 n 個相異實數, 是否可定出 $\mathsf{P}_n(\mathbb{R})$ 的 inner product? 而若選 $n+2$ 個相異實數, 是否可定出 $\mathsf{P}_n(\mathbb{R})$ 的 inner product?*

依照 inner product 的定義可以推導出一些 inner product 性質, 以下我們列出幾個常用的性質以利以後操作. 要注意以下的性質是可由 inner product 定義推得的, 它們並不屬於 inner product 的定義. 也就是說以後當我們要說明一個定義出來的是否為 inner product, 我們僅要驗證它是否符合 inner product 的四個要求. 若符合, 自然地它便會符合以下的性質.

Proposition 4.2.2. *假設 V 是以 $\langle,\rangle$ 為 inner product 的 inner product space. 則有以下的性質.*

(1) $\langle \mathbf{v},\mathbf{0}\rangle=0$, $\forall \mathbf{v}\in V$.

(2) 若 $\mathbf{v},\mathbf{w}\in V$ 且 $\langle\mathbf{v},\mathbf{u}\rangle=\langle\mathbf{w},\mathbf{u}\rangle$, $\forall\mathbf{u}\in V$, 則 $\mathbf{v}=\mathbf{w}$.

Proof. (1) 任取 $\mathbf{v}\in V$, 由 $\langle\mathbf{v},\mathbf{0}\rangle=\langle\mathbf{v},\mathbf{0}+\mathbf{0}\rangle=\langle\mathbf{v},\mathbf{0}\rangle+\langle\mathbf{v},\mathbf{0}\rangle$, 得 $\langle\mathbf{v},\mathbf{0}\rangle=0$.

(2) 由 $\langle\mathbf{v},\mathbf{u}\rangle=\langle\mathbf{w},\mathbf{u}\rangle$ 得 $\langle\mathbf{v}-\mathbf{w},\mathbf{u}\rangle=0$. 由於此等式依假設是對所有的 $\mathbf{u}\in V$ 皆成立, 故令 $\mathbf{u}=\mathbf{v}-\mathbf{w}$, 可得 $\langle\mathbf{v}-\mathbf{w},\mathbf{v}-\mathbf{w}\rangle=0$. 故依 innner product 的定義知 $\mathbf{v}-\mathbf{w}=\mathbf{0}$, 得證 $\mathbf{v}=\mathbf{w}$. □

從 Proposition 4.2.2 (2) 的證明中, 我們了解到 inner product 的定義中 $\langle\mathbf{v},\mathbf{v}\rangle=0$ 若且唯若 $\mathbf{v}=\mathbf{0}$ 是相當重要的性質. 再加上 $\langle\mathbf{v},\mathbf{v}\rangle\geq 0$, 相當符合我們對長度的要求. 因此我們很自然地可以定義一個 inner product space 中的長度概念.

Definition 4.2.3. 假設 V 是以 $\langle,\rangle$ 為 inner product 的 inner product space. 對任意 $\mathbf{v}\in V$, 定義 $\mathbf{v}$ 的 *norm* (or *length*) 為

$$\|\mathbf{v}\|=\sqrt{\langle\mathbf{v},\mathbf{v}\rangle}.$$

稱之為 norm 自然會是符合長度定義的要求, 即三角不等式. 在介紹 dot product 時, 我們完整的推導出 norm 的性質, 當初證明這些性質因為僅用到 inner product 定義的基本性質, 這裡我們就不再證明, 僅列出其結果.

Proposition 4.2.4. 假設 V 是以 $\langle,\rangle$ 為 *inner product* 的 *inner product space* 且 $\|\,\|$ 為以 $\langle,\rangle$ 定義的 *norm,* 則對任意 $r\in\mathbb{R}$, $\mathbf{v},\mathbf{w}\in V$ 皆有以下的性質.

(1) $\|r\mathbf{v}\|=|r|\,\|\mathbf{v}\|$.

(2) $\|\mathbf{v}\|\geq 0$ 且 $\|\mathbf{v}\|=0$ 若且唯若 $\mathbf{v}=\mathbf{0}$.

(3) $\|\mathbf{v}+\mathbf{w}\|^2=\|\mathbf{v}\|^2+2\langle\mathbf{v},\mathbf{w}\rangle+\|\mathbf{w}\|^2$.

(4) $\|\mathbf{v}+\mathbf{w}\|^2+\|\mathbf{v}-\mathbf{w}\|^2=2\|\mathbf{v}\|^2+2\|\mathbf{w}\|^2$.

(5) $|\langle\mathbf{v},\mathbf{w}\rangle|\leq\|\mathbf{v}\|\,\|\mathbf{w}\|$. 特別地當 $\mathbf{v},\mathbf{w}$ 皆不為零向量時, $|\langle\mathbf{v},\mathbf{w}\rangle|=\|\mathbf{v}\|\,\|\mathbf{w}\|$ 若且唯若存在 $\lambda\in\mathbb{R}$ 使得 $\mathbf{w}=\lambda\mathbf{v}$.

(6) $\|\mathbf{v}+\mathbf{w}\|\leq\|\mathbf{v}\|+\|\mathbf{w}\|$.

將來用到 norm 的時候, 我們都會用到 Proposition 4.2.4 上的性質. 例如當 $\mathbf{v}\neq\mathbf{0}$ 時, 令 $\mathbf{u}=(1/\|\mathbf{v}\|)\mathbf{v}$, 則由 (1) 知

$$\|\mathbf{u}\|=\|\frac{1}{\|\mathbf{v}\|}\mathbf{v}\|=\left|\frac{1}{\|\mathbf{v}\|}\right|\|\mathbf{v}\|=1.$$

這種符合 $\|\mathbf{u}\|=1$ 的 $\mathbf{u}$, 就稱為在此 norm 之下的 *unit vector.* Proposition 4.2.4 的性質 (4) 稱為 parallelogram relation; (5) 就是 Cauchy-Schwarz inequality; (6) 就是 triangle inequality.

Question 4.7. 考慮 $\mathsf{P}_2(\mathbb{R})$ 上利用 $-1,0,1$ 三相異實數所訂出的 *inner product (*參見 *Example 4.2.1).* 試利用此 *inner product* 求 $\|x\|$ 並找到 $f\in\mathrm{Span}(x)$ 滿足 $\|f\|=1$.

4.3. Projection and Gram-Schmidt Process

我們曾經介紹 $\mathbb{R}^n$ 上的 projection. 簡單來說, 給定非零向量 $\mathbf{w}\in\mathbb{R}^n$, 我們定義 $\mathbf{v}\in\mathbb{R}^n$ 在 $\mathbf{w}$ 上的投影向量 $\mathbf{w}'$ 必須是在 $\mathbf{w}$ 所展成的空間中 (即 $\mathbf{w}'\in\mathrm{Span}(\mathbf{w})$), 而且 $\mathbf{v}-\mathbf{w}'$ 必須與 $\mathbf{w}$ 垂直, 也因此需與 $\mathrm{Span}(\mathbf{w})$ 上所有向量垂直. 依此, 我們將投影的定義推廣到一般對 inner product space V 的一個 subspace W 上的投影. 也就是說, 當 W 為 V 的 subspace, 對於 $\mathbf{v}\in V$, 我們定義 $\mathbf{v}$ 在 W 的 projection 為 W 上的一個向量 $\mathbf{w}'$ (即 $\mathbf{w}'\in W$), 滿足 $\mathbf{v}-\mathbf{w}'$ 和 W 上所有向量垂直. 首先我們給以下的定義.

Definition 4.3.1. 假設 V 為 inner product space. 給定 W 為 V 的 subspace. 令

$$W^\perp=\{\mathbf{v}\in V\mid\langle\mathbf{v},\mathbf{w}\rangle=0,\forall\mathbf{w}\in W\}.$$

也就是說 $W^\perp$ 為 V 中和所有 W 中的向量垂直的向量所成的集合. 一般稱 $W^\perp$ 為 *orthogonal complement* of W.

有了 orthogonal complement 的定義我們就可以對上述的 projection 給了以下的定義.

Definition 4.3.2. 假設 V 為 inner product space. 給定 W 為 V 的 subspace. 對於 $\mathbf{v}\in V$, 若 $\mathbf{w}\in W$ 滿足 $\mathbf{v}-\mathbf{w}\in W^\perp$, 則稱 $\mathbf{w}$ 為 the *orthogonal projection* of $\mathbf{v}$ on W.

等一下我們將說明對於任意 $\mathbf{v} \in V$, the projection of $\mathbf{v}$ on W 一定存在, 而且唯一. 因此為了方便起見我們將此向量用 $\mathrm{Proj}_W(\mathbf{v})$ 表示. 我們將先證明唯一性, 有了唯一性, 將來我們要說明 $\mathbf{w}$ 就是 $\mathrm{Proj}_W(\mathbf{v})$, 就僅要檢查 $\mathbf{w} \in W$ 且滿足 $\mathbf{v} - \mathbf{w} \in W^{\perp}$ 就可以了.

我們先了解以下一些有關於 orthogonal complement 的性質. 假設 W 為 V 的 subspace, 並令 $\mathbf{v}, \mathbf{v}' \in W^{\perp}$ 以及 $r, s \in \mathbb{R}$. 利用內積的性質, 對於任意 $\mathbf{w} \in W$, 我們有 $\langle r\mathbf{v} + s\mathbf{v}', \mathbf{w} \rangle = r\langle \mathbf{v}, \mathbf{w} \rangle + s\langle \mathbf{v}', \mathbf{w} \rangle$. 再利用 $\mathbf{v}, \mathbf{v}' \in W^{\perp}$, 即 $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}', \mathbf{w} \rangle = 0$, 得知 $\langle r\mathbf{v} + s\mathbf{v}', \mathbf{w} \rangle = 0$. 此即表示對於任意 $\mathbf{v}, \mathbf{v}' \in W^{\perp}$ 以及 $r, s \in \mathbb{R}$, 皆有 $r\mathbf{v} + s\mathbf{v}' \in W^{\perp}$, 得證 $W^{\perp}$ 為 V 的 subspace.

依定義要說明 $\mathbf{v} \in W^{\perp}$ 就得說明 $\langle \mathbf{v}, \mathbf{w} \rangle = 0, \forall \mathbf{w} \in W$. 這個過程看似複雜, 因為要對所有 W 檢查. 不過由於 W 是 vector space, 我們僅要找到 W 的一組 basis, 然後對這組 basis 檢查即可, 因為我們有以下之結果.

Lemma 4.3.3. 假設 V 為 *inner product space* 且 W 為 V 的 *subspace.* 假設 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 是 W 的一組 *basis,* 則 $\mathbf{v} \in W^{\perp}$ 若且唯若 $\langle \mathbf{v}, \mathbf{w}_i \rangle = 0, \forall i = 1, \ldots, n.$

Proof. 依定義若 $\mathbf{v} \in W^{\perp}$ 則對任意 $\mathbf{w} \in W$ 皆有 $\langle \mathbf{v}, \mathbf{w} \rangle = 0$, 所以當然 $\langle \mathbf{v}, \mathbf{w}_i \rangle = 0, \forall i = 1, \ldots, n$ 成立. 我們僅要證明若 $\langle \mathbf{v}, \mathbf{w}_i \rangle = 0$, $\forall i = 1, \ldots, n$ 則對任意 $\mathbf{w} \in W$ 皆有 $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. 對任意 $\mathbf{w} \in W$, 因為 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 是 W 的一組 basis, 故存在 $c_1, \ldots, c_n \in \mathbb{R}$ 使得 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$, 故得

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n \rangle = c_1\langle \mathbf{v}, \mathbf{w}_1 \rangle + \cdots + c_n\langle \mathbf{v}, \mathbf{w}_n \rangle = 0.$$

□

要注意 orthogonal complement 的 complement 不是指集合的補集. W 的 orthogonal complement $W^{\perp}$ 並不是 W 的補集. 甚至 $W \cap W^{\perp}$ 並不會是空集合. 這是因為 $W \cap W^{\perp}$ 也會是一個 subspace, 所以 $\mathbf{0}$ 一定在其中. 事實上我們有以下的結果.

Lemma 4.3.4. 假設 V 為 *inner product space* 且 W 為 V 的 *subspace.* 則 $W \cap W^{\perp} = \{\mathbf{0}\}.$

Proof. 因 $W \cap W^{\perp}$ 為 subspace, 故知 $\mathbf{0} \in W \cap W^{\perp}$. 現假設 $\mathbf{w} \in W \cap W^{\perp}$. 由於 $\mathbf{w} \in W^{\perp}$, 對任意 $\mathbf{w}' \in W$ 皆有 $\langle \mathbf{w}, \mathbf{w}' \rangle = 0$. 而又 $\mathbf{w} \in W$, 故得 $\langle \mathbf{w}, \mathbf{w} \rangle = 0$. 因此由內積的性質知 $\mathbf{w} = \mathbf{0}$, 得證 $W \cap W^{\perp} = \{\mathbf{0}\}$. □

現在我們可以證明 projection 的唯一性.

Proposition 4.3.5. 假設 V 為 *inner product space.* 給定 W 為 V 的 *subspace.* 對於 $\mathbf{v} \in V$, $\mathbf{v}$ 對於 W 的 *projection* 是唯一的.

Proof. 假設 $\mathbf{w}, \mathbf{w}' \in W$ 皆為 $\mathbf{v}$ 在 W 的 projection. 也就是說 $\mathbf{w}, \mathbf{w}'$ 皆滿足 $\mathbf{v} - \mathbf{w} \in W^{\perp}$ 以及 $\mathbf{v} - \mathbf{w}' \in W^{\perp}$. 由於 $W^{\perp}$ 為 subspace, 我們有 $(\mathbf{v} - \mathbf{w}) - (\mathbf{v} - \mathbf{w}') \in W^{\perp}$, 亦即 $\mathbf{w} - \mathbf{w}' \in W^{\perp}$. 又因 W 為 subspace, 我們也知 $\mathbf{w} - \mathbf{w}' \in W$. 故由 Lemma 4.3.4 知 $\mathbf{w} - \mathbf{w}' = \mathbf{0}$, 即 $\mathbf{w} = \mathbf{w}'$, 證得唯一性. □

有了 Proposition 4.3.5 的唯一性, 以後我們要說明 $\mathbf{w}$ 是 $\mathbf{v}$ 在 W 的 projection, 就僅要檢查 $\mathbf{w} \in W$ 且滿足 $\mathbf{v}-\mathbf{w} \in W^{\perp}$ 就可以了.

接下來我們探討 projection 的存在性. 要考慮 $\mathbf{v}$ 對 W 的 projection, 我們必須找到 W 的一組 basis, 然後再看看這組 basis 所有可能的線性組合, 哪一個可以符合 projection 的要求. 這裡較麻煩的地方就是, 我們不知道要選取哪一個線性組合. 是否有可能找到一組特殊的 basis, 可以讓上述找到線性組合變簡單呢? 我們便是要回答這一個問題, 並探討如何找到這種特殊的 basis.

假設 W 為 V 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為其一組 basis 滿足 $\langle \mathbf{w}_i,\mathbf{w}_j\rangle = 0, \forall i \neq j$. 對於任意 $\mathbf{w} \in W$, 因為 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis, 存在 $c_1,\ldots,c_n \in \mathbb{R}$ 使得 $\mathbf{w} = c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$. 一般來說我們都是利用解聯立方程組的方法找到 $c_1,\ldots,c_n$, 不過這裡由於這些 $\mathbf{w}_i$ 之間兩兩互相垂直, 我們可以利用內積求出 c_i. 事實上對於任意 $i=1,\ldots,n$, 考慮 $\langle \mathbf{w},\mathbf{w}_i\rangle$. 我們有

$$\langle \mathbf{w},\mathbf{w}_i\rangle = \langle c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n,\mathbf{w}_i\rangle = c_1\langle \mathbf{w}_1,\mathbf{w}_i\rangle+\cdots+c_n\langle \mathbf{w}_n,\mathbf{w}_i\rangle = c_i\langle \mathbf{w}_i,\mathbf{w}_i\rangle.$$

因為 $\mathbf{w}_i \neq \mathbf{0}$, 我們有 $\langle \mathbf{w}_i,\mathbf{w}_i\rangle = \|\mathbf{w}_i\|^2 \neq 0$, 故得 $c_i = \langle \mathbf{w},\mathbf{w}_i\rangle/\langle \mathbf{w}_i,\mathbf{w}_i\rangle = \langle \mathbf{w},\mathbf{w}_i\rangle/\|\mathbf{w}_i\|^2$. 特別的, 若 $\|\mathbf{w}_i\| = 1$, 則 $c_i = \langle \mathbf{w},\mathbf{w}_i\rangle$. 我們有以下的定理.

Proposition 4.3.6. 假設 V 為 *inner product space*, W 為 V 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *basis* 滿足 $\langle \mathbf{w}_i,\mathbf{w}_j\rangle = 0, \forall i \neq j$. 則對於任意 $\mathbf{w} \in W$, 我們有

$$\mathbf{w} = \frac{\langle \mathbf{w},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\cdots+\frac{\langle \mathbf{w},\mathbf{w}_i\rangle}{\|\mathbf{w}_i\|^2}\mathbf{w}_i+\cdots+\frac{\langle \mathbf{w},\mathbf{w}_n\rangle}{\|\mathbf{w}_n\|^2}\mathbf{w}_n.$$

由於這種兩兩互相垂直的 basis, 對於寫下一個向量的 linear combination 相當的方便, 我們有以下的定義.

Definition 4.3.7. 假設 V 為 inner product space 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis 滿足對於任意 $i \neq j$ 皆有 $\langle \mathbf{v}_i,\mathbf{v}_j\rangle = 0$. 則稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 *orthogonal basis*. 若又要求 $\langle \mathbf{v}_i,\mathbf{v}_i\rangle = 1$ (即 $\|\mathbf{v}_i\| = 1$), $\forall i = 1,\ldots,n$, 則稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 *orthonormal basis*.

要注意, 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的一組 orthogonal basis, 對於所有 $i=1,\ldots,n$, 令 $\mathbf{u}_i = \mathbf{v}_i/\|\mathbf{v}_i\|$, 則 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 就會是 V 的一組 orthonormal basis. 這是因為

$$\langle \mathbf{u}_i,\mathbf{u}_j\rangle = \langle \frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i,\frac{1}{\|\mathbf{v}_j\|}\mathbf{v}_j\rangle = \frac{1}{\|\mathbf{v}_i\|}\frac{1}{\|\mathbf{v}_j\|}\langle \mathbf{v}_i,\mathbf{v}_j\rangle.$$

因此當 $i \neq j$, 我們有 $\langle \mathbf{u}_i,\mathbf{u}_j\rangle = \langle \mathbf{v}_i,\mathbf{v}_j\rangle = 0$; 而當 $i = j$, 我們有 $\langle \mathbf{u}_i,\mathbf{u}_i\rangle = \langle \mathbf{v}_i,\mathbf{v}_i\rangle/\|\mathbf{v}_i\|^2 = 1$.

由前面已知, 若能找到 W 的一組 orthogonal basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$, 則我們可以很容易的將任意 W 中的向量寫成 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 的線性組合. 這似乎克服了前面所述較複雜的部份. 事實上確實如此, 若能找到 W 的一組 orthogonal basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$, 則我們便可以輕易地得到 $\mathbf{v}$ 在 W 的 projection 了. 這是因為若 $\mathbf{w} = \mathrm{Proj}_W(\mathbf{v})$ 且 $\mathbf{w} = c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$. 此時由於 $\mathbf{v}-\mathbf{w} \in W^{\perp}$, 對於所有 $i=1,\ldots,n$, 我們有 $\langle \mathbf{v},\mathbf{w}_i\rangle - \langle \mathbf{w},\mathbf{w}_i\rangle = \langle \mathbf{v}-\mathbf{w},\mathbf{w}_i\rangle = 0$. 因此得 $\langle \mathbf{v},\mathbf{w}_i\rangle = \langle \mathbf{w},\mathbf{w}_i\rangle = c_i\langle \mathbf{w}_i,\mathbf{w}_i\rangle$, 亦即 $c_i = \langle \mathbf{v},\mathbf{w}_i\rangle/\langle \mathbf{w}_i,\mathbf{w}_i\rangle$. 所以我們得到一個很簡捷求 projection 的方法.

Theorem 4.3.8. 假設 V 為 *inner product space*, W 為 V 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthogonal basis*. 若 $\mathbf{v}\in V$, 則 $\mathbf{v}$ 在 W 的 *projection* 為

$$\mathrm{Proj}_W(\mathbf{v})=\frac{\langle\mathbf{v},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1+\cdots+\frac{\langle\mathbf{v},\mathbf{w}_n\rangle}{\|\mathbf{w}_n\|^2}\mathbf{w}_n.$$

特別的當 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthonormal basis*, 則

$$\mathrm{Proj}_W(\mathbf{v})=\langle\mathbf{v},\mathbf{w}_1\rangle\mathbf{w}_1+\cdots+\langle\mathbf{v},\mathbf{w}_n\rangle\mathbf{w}_n.$$

Proof. 令 $\mathbf{w}=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$, 其中 $c_i=\langle\mathbf{v},\mathbf{w}_i\rangle/\langle\mathbf{w}_i,\mathbf{w}_i\rangle$, $\forall i=1,\ldots,n$. 此時 $\mathbf{w}\in W$, 且對任意 $i=1,\ldots,n$ 皆有

$$\langle\mathbf{v}-\mathbf{w},\mathbf{w}_i\rangle=\langle\mathbf{v},\mathbf{w}_i\rangle-\langle\mathbf{w},\mathbf{w}_i\rangle=\langle\mathbf{v},\mathbf{w}_i\rangle-\sum_{j=1}^{n}c_j\langle\mathbf{w}_j,\mathbf{w}_i\rangle=\langle\mathbf{v},\mathbf{w}_i\rangle-c_i\langle\mathbf{w}_i,\mathbf{w}_i\rangle=0.$$

故由 Lemma 4.3.3 知 $\mathbf{v}-\mathbf{w}\in W^{\perp}$. 因此由 projection 的唯一性 (Proposition 4.3.5) 知 $\mathbf{w}=\mathrm{Proj}_W(\mathbf{v})$. □

我們已經知道只要找到 W 的一組 orthogonal basis, 就可以輕易求得 $\mathbf{v}$ 在 W 的 projection. 對一般 inner product space, 我們將介紹一個方法找到它的一組 orthogonal basis, 也因此證明了對於一般的 inner product space 一定存在 orthogonal basis (以及 orthonormal basis). 這個方法就是所謂的 *Gram-Schmidt process*.

給定 V 的一個 nonzero subspace W, 且假設 $\dim(W)=n$. 首先我們說明若 $\mathbf{w}_1,\ldots,\mathbf{w}_k\in W$ 為非零向量且滿足 $\langle\mathbf{w}_i,\mathbf{w}_j\rangle=0$, $\forall i\neq j$, 則 $\mathbf{w}_1,\ldots,\mathbf{w}_k$ 為 linearly independent. 這是因為若不是 linearly independent 表示存在 $c_1,\ldots,c_k\in\mathbb{R}$ 不全為 0 使得 $c_1\mathbf{w}_1+\cdots+c_k\mathbf{w}_k=\mathbf{0}$. 然而對任意 $i=1,\ldots,k$ 由於 $0=\langle c_1\mathbf{w}_1+\cdots+c_k\mathbf{w}_k,\mathbf{w}_i\rangle=c_i\|\mathbf{w}_i\|^2$. 也因此由 $\|\mathbf{w}_i\|\neq 0$, 得證 $c_i=0$. 此和 $c_1,\ldots,c_k$ 不全為 0 的假設相矛盾, 故知 $\mathbf{w}_1,\ldots,\mathbf{w}_k$ 為 linearly independent. 因此要找到 W 的一組 orthogonal basis, 我們只要在 W 中找到 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 滿足 $\langle\mathbf{w}_i,\mathbf{w}_j\rangle=0$, $\forall i\neq j$ 即可, 因為它們是 linearly independent 且 $\dim(W)=n$, 故知它們是 W 的一組 basis. 接下來我們要說明在 W 中如何找到這樣的一組 nonzero vectors.

首先因 $W\neq\{\mathbf{0}\}$, 故可在 W 中取一 nonzero vector $\mathbf{v}_1$. 為了方便起見我們令 $\mathbf{w}_1=\mathbf{v}_1$ 且 $W_1=\mathrm{Span}(\mathbf{v}_1)=\mathrm{Span}(\mathbf{w}_1)$. 若 $\dim(W)=1$, 則 $W=W_1$ 故 $\mathbf{w}_1$ 就是 W 的一個 orthogonal basis. 而若 $\dim(W)>1$, 則因 $W_1\subsetneq W$, 我們可以找到 nonzero vector $\mathbf{v}_2\in W$ 且 $\mathbf{v}_2\not\in W_1$. 現在我們要利用 $\mathbf{v}_2$, 找到 $\mathbf{w}_2\in W$ 滿足 $\mathbf{w}_2\neq\mathbf{0}$ 且 $\langle\mathbf{w}_1,\mathbf{w}_2\rangle=0$. 很自然的, 我們會考慮 $\mathbf{w}_2=\mathbf{v}_2-\mathrm{Proj}_{W_1}(\mathbf{v}_2)$, 因為此時 $\mathbf{w}_2\in W_1^{\perp}$, 而 $W_1=\mathrm{Span}(\mathbf{w}_1)$, 故當然有 $\langle\mathbf{w}_1,\mathbf{w}_2\rangle=0$. 我們也要說明 $\mathbf{w}_2\neq\mathbf{0}$. 這是因為若 $\mathbf{w}_2=\mathbf{0}$, 會得到 $\mathbf{v}_2=\mathrm{Proj}_{W_1}(\mathbf{v}_2)\in W_1$, 此與當初 $\mathbf{v}_2\not\in W_1$ 的假設相矛盾. 另一方面因為 $W=\mathrm{Span}(\mathbf{w}_1)$, 利用 Proposition 4.1.9 我們知 $\mathrm{Proj}_{W_1}(\mathbf{v}_2)=(\langle\mathbf{v}_2,\mathbf{w}_1\rangle/\|\mathbf{w}_1\|^2)\mathbf{w}_1$, 所以我們知

$$\mathbf{w}_2=\mathbf{v}_2-\frac{\langle\mathbf{v}_2,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1.$$

另外要注意的是 $\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$, 這是因為依 $\mathbf{w}_1,\mathbf{w}_2$ 的選取, 我們有 $\mathbf{w}_1,\mathbf{w}_2\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$, 因此 $\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)\subseteq\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$. 然而因為 $\mathbf{v}_1,\mathbf{v}_2$ 為 linearly independent 且 $\mathbf{w}_1,\mathbf{w}_2$ 為 linearly independent, 故由 $\dim(\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2))=\dim(\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2))=2$ 得證

$\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$. 為了方便起見, 我們令 $W_2=\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$. 現若 $\dim(W)=2$, 則 $W=W_2$, 故 $\mathbf{w}_1,\mathbf{w}_2$ 為 W 的一組 orthogonal basis. 而若 $\dim(W)>2$, 則因 $W_2\subsetneq W$, 我們可以找到 nonzero vector $\mathbf{v}_3\in W$ 且 $\mathbf{v}_3\not\in W_2$. 現在我們要利用 $\mathbf{v}_3$, 找到 $\mathbf{w}_3\in W$ 滿足 $\mathbf{w}_3\neq\mathbf{0}$ 且 $\langle\mathbf{w}_1,\mathbf{w}_3\rangle=\langle\mathbf{w}_2,\mathbf{w}_3\rangle=0$. 同前, 我們考慮 $\mathbf{w}_3=\mathbf{v}_3-\mathrm{Proj}_{W_2}(\mathbf{v}_3)$, 因為此時 $\mathbf{w}_3\in W_2^{\perp}$, 而 $W_2=\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)$, 故當然有 $\langle\mathbf{w}_1,\mathbf{w}_3\rangle=\langle\mathbf{w}_2,\mathbf{w}_3\rangle=\mathbf{0}$. 另外因 $\mathbf{v}_3\not\in W_2$, 同前面的理由我們有 $\mathbf{w}_3\neq\mathbf{0}$. 另一方面因為 $\mathbf{w}_1,\mathbf{w}_2$ 為 W_2 的 orthogonal basis, 利用 Theorem 4.3.8 我們得

$$\mathbf{w}_3=\mathbf{v}_3-\mathrm{Proj}_{W_2}(\mathbf{v}_3)=\mathbf{v}_3-\frac{\langle\mathbf{v}_3,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1-\frac{\langle\mathbf{v}_3,\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2.$$

最後和前面同樣的理由, 我們有 $\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3)=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3)$. 這樣一直下去, 我們可以得到 $W_k=\mathrm{Span}(\mathbf{w}_1,\dots,\mathbf{w}_k)=\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_k)$ 且 $\mathbf{w}_1,\dots,\mathbf{w}_k$ 是 W_k 的一組 orthogonal basis. 現若 $k=\dim(W)=n$, 則得 $W_k=W$, 所以 $\mathbf{w}_1,\dots,\mathbf{w}_k$ 是 W 的一組 orthogonal basis. 而若 $k<n$, 則存在 $\mathbf{v}_{k+1}\in W$ 且 $\mathbf{v}_{k+1}\not\in W_k$. 故令

$$\mathbf{w}_{k+1}=\mathbf{v}_{k+1}-\mathrm{Proj}_{W_k}(\mathbf{v}_{k+1})=\mathbf{v}_{k+1}-\frac{\langle\mathbf{v}_{k+1},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1-\frac{\langle\mathbf{v}_{k+1},\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2-\cdots-\frac{\langle\mathbf{v}_{k+1},\mathbf{w}_k\rangle}{\|\mathbf{w}_k\|^2}\mathbf{w}_k,$$

則得 $\mathbf{w}_{k+1}\neq\mathbf{0}$ 且 $\langle\mathbf{w}_{k+1},\mathbf{w}_i\rangle=0,\forall i=1,\dots,k$. 另外因 $\mathbf{w}_1,\dots,\mathbf{w}_k,\mathbf{w}_{k+1}\in\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_k,\mathbf{v}_{k+1})$, 同上可得 $\mathrm{Span}(\mathbf{w}_1,\dots,\mathbf{w}_k,\mathbf{w}_{k+1})=\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_k,\mathbf{v}_{k+1})$, 故令 $W_{k+1}=\mathrm{Span}(\mathbf{w}_1,\dots,\mathbf{w}_k,\mathbf{w}_{k+1})=\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_k,\mathbf{v}_{k+1})$, 我們有 $\mathbf{w}_1,\dots,\mathbf{w}_k,\mathbf{w}_{k+1}$ 為 W_{k+1} 的一組 orthogonal basis. 這樣一直下去直到得到 $W_n=W$, 這樣 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 就是 W 的一組 orthogonal basis.

上述 Gram-Schmidt process 中 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 的選取事實上和我們過去找 vector space 的 basis 方法是一樣的. 差別就是我們要將 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 這組 basis 修改成 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 這一組 orthogonal basis. 因此如果一開始已給定 W 的一組 basis $\mathbf{v}_1\dots,\mathbf{v}_n$, 我們可以將之直接套用, 因此有以下的結果.

Theorem 4.3.9 (Gram-Schmidt Process)**.** *假設 V 為 inner product space, W 為 V 的 subspace 且 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 W 的一組 basis. 令*

$$\mathbf{w}_1=\mathbf{v}_1,\quad \mathbf{w}_2=\mathbf{v}_2-\frac{\langle\mathbf{v}_2,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1,\dots$$

這樣一直下去, 即對於 $i=1,\dots,n-1$ 令

$$\mathbf{w}_{i+1}=\mathbf{v}_{i+1}-\frac{\langle\mathbf{v}_{i+1},\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1-\frac{\langle\mathbf{v}_{i+1},\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2-\cdots-\frac{\langle\mathbf{v}_{i+1},\mathbf{w}_i\rangle}{\|\mathbf{w}_i\|^2}\mathbf{w}_i,$$

則 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 為 W 的一組 orthogonal basis. 而且

$$\mathrm{Span}(\mathbf{w}_1,\dots,\mathbf{w}_i)=\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_i),\quad \forall i=1,\dots,n.$$

Gram-Schmidt process 確保了 orthogonal basis 的存在性, 而當 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為其 orthogonal basis 時, 我們可以除去其長度得到 $(1/\|\mathbf{v}_1\|)\mathbf{v}_1,\dots,(1/\|\mathbf{v}_n\|)\mathbf{v}_n$ 這一組 orthonormal basis. 利用 orthogonal basis 的存在性, 我們也就得到了 orthogonal projection 的存在性了. 也就是說當 V 為 inner product space 且 W 為其 subspace, 我們就可以利用 Gram-Schmidt process 找到 W 的一組 orthogonal basis, 然後利用 Theorem 4.3.8, 得到任意 V 中的向量 $\mathbf{v}$ 在 W 上的 orthogonal projection 了.

Example 4.3.10. 我們要用 orthogonal basis 來處理 orthogonal projection 的問題. 我們在 $\mathbb{R}^4$ 使用 dot product, 考慮 $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 4 \end{bmatrix}$ 在 $W = \mathrm{Span}(\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 2 \\ 3 \end{bmatrix})$ 的 orthogonal projection. 首先找 W 的一組 orthogonal basis. 令

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 3 \\ 4 \\ 2 \\ 3 \end{bmatrix} - \frac{18}{10}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} = \frac{2}{5}\begin{bmatrix} 3 \\ 1 \\ -4 \\ 3 \end{bmatrix}.$$

此時 $\mathbf{w}_1, \mathbf{w}_2$ 為 W 的一組 orthogonal basis, 故利用 Theorem 4.3.8 得

$$\mathrm{Proj}_W(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \frac{\langle \mathbf{v}, \mathbf{w}_2 \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 = \frac{8}{10}\begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} + \frac{28/5}{28/5} \times \frac{2}{5}\begin{bmatrix} 3 \\ 1 \\ -4 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ 2 \end{bmatrix}.$$

雖然 Theorem 4.3.9 的敘述是找到 V 的 subspace W 的 orthogonal basis, 不過因 W 是 V 中任意的 subspace, 所以當 W 為 V 時, 我們也就找到 V 的 orthogonal basis 了. 所以對任意 finite dimensional inner product space, Gram-Schmidt process 都能幫我們找到 orthogonal basis. 注意這裡需要有限維的假設, 因為整個過程我們是一個一個置換這些向量, 所以有限多個向量才可全部置換完成. Theorem 4.3.9 的敘述牽涉到 V 的 subspace W 主要用意是, 我們不只可找到 W 的 orthogonal basis, 也可繼續這個 process, 而將 W 的 orthogonal basis 擴大成 V 的 orthogonal basis. 這是因為若 $W \neq V$, 當找到 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 W 的 orthogonal basis 後我們可以繼續考慮 $\mathbf{v}_{n+1} \in V$ 但 $\mathbf{v}_{n+1} \notin W$, 然後利用 Gram-Schmidt process 得到 $\mathbf{w}_{n+1} = \mathbf{v}_{n+1} - \mathrm{Proj}_W(\mathbf{v}_{n+1}) \in W^{\perp}$. 這樣一直下去直到得到 V 的一組 orthogonal basis 為止. 也因此我們得到以下之結果.

Corollary 4.3.11. 假設 V 為 *inner product space,* W 為 V 的 *subspace.* 若 $\dim(V) = m$ 且 $\dim(W) = n$, 則存在 V 的一組 *orthogonal basis* $\mathbf{v}_1, \ldots, \mathbf{v}_n, \ldots, \mathbf{v}_m$, 其中 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 W 的 *orthogonal basis.*

Example 4.3.12. 考慮 $\mathbb{R}^4$ 中以 dot product 所形成的 inner product space. 令

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 3 \\ 3 \end{bmatrix},$$

我們要求 $W = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ 的一組 orthogonal basis, 並將之擴大成 $\mathbb{R}^4$ 的一組 orthogonal basis. 首先令 $\mathbf{w}_1 = \mathbf{v}_1$, 得

$$\mathbf{w}_2 = \mathbf{v}_2 - \frac{\langle \mathbf{v}_2, \mathbf{w}_1 \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 1 \end{bmatrix} - \frac{4}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 0 \end{bmatrix}.$$

最後得

$$\mathbf{w}_3=\mathbf{v}_3-\frac{\langle \mathbf{v}_3,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1-\frac{\langle \mathbf{v}_3,\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2=\begin{bmatrix}1\\1\\3\\3\end{bmatrix}-\frac{8}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}-\frac{-4}{8}\begin{bmatrix}2\\0\\-2\\0\end{bmatrix}=\begin{bmatrix}0\\-1\\0\\1\end{bmatrix}.$$

由於 $\dim(V)=3$ 且 $\dim(\mathbb{R}^4)=4$, 我們需要再找到一個向量以形成 $\mathbb{R}^4$ 的 basis. 考慮 $\mathbf{v}_4=\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$, 我們可以檢查 $\mathbf{v}_4\not\in W$ (或直接套用 Gram-Schmidt process, 若 $\mathbf{v}_4\in W$, 會得到 $\mathbf{v}_4-\mathrm{Proj}_W(\mathbf{v}_4)=\mathbf{0}$. 若真如此就再換一個向量). 得

$$\mathbf{w}_4=\mathbf{v}_4-\frac{\langle \mathbf{v}_4,\mathbf{w}_1\rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1-\frac{\langle \mathbf{v}_4,\mathbf{w}_2\rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2-\frac{\langle \mathbf{v}_4,\mathbf{w}_3\rangle}{\|\mathbf{w}_3\|^2}\mathbf{w}_3=\begin{bmatrix}0\\0\\0\\1\end{bmatrix}-\frac{1}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}-\frac{1}{2}\begin{bmatrix}0\\-1\\0\\1\end{bmatrix}=\frac{1}{4}\begin{bmatrix}-1\\1\\-1\\1\end{bmatrix}.$$

此時 $\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,\mathbf{w}_4$ 就是 $\mathbb{R}^4$ 的一組 orthogonal basis, 其中 $\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3$ 是 W 的 orthogonal basis.

對於每一個 $\mathbf{w}_i$, 我們可以除以其長度 $\|\mathbf{w}_i\|$, 得到一組 orthonormal basis

$$\mathbf{u}_1=\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\mathbf{u}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\\0\end{bmatrix},\mathbf{u}_3=\frac{1}{\sqrt{2}}\begin{bmatrix}0\\-1\\0\\1\end{bmatrix},\mathbf{u}_4=\frac{1}{2}\begin{bmatrix}-1\\1\\-1\\1\end{bmatrix}.$$

接下來我們看幾個有關 orthogonal basis 的應用. 由於利用 orthonormal basis 會比較方便 (省去除掉長度的麻煩), 所以以下都用 orthonormal basis 處理. 首先當 W 是 inner product space V 的 subspace, 我們知道 $W^\perp$ 也是 V 的 subspace. 很自然的我們會想要知道 $\dim(W^\perp)$ 和 $\dim(W)$ 的關係. 當我們利用 Corollary 4.3.11 找到 $\mathbf{u}_1,\dots,\mathbf{u}_n,\dots,\mathbf{u}_m$ 是 V 的一組 orthonormal basis, 其中 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 W 的 orthonormal basis, 我們就可以將任意 $\mathbf{v}\in W^\perp$ 寫成 $\mathbf{v}=c_1\mathbf{u}_1+\cdots+c_n\mathbf{u}_n+c_{n+1}\mathbf{u}_{n+1}+\cdots+c_m\mathbf{u}_m$. 其中 $c_i=\langle\mathbf{v},\mathbf{u}_i\rangle$ (Proposition 4.3.6 套用 $W=V$ 的情形). 由於 $\mathbf{v}\in W^\perp$, 我們有 $\langle\mathbf{v},\mathbf{u}_i\rangle=0$, $\forall i=1,\dots,n$ (因為這些 $\mathbf{u}_i\in W$). 因此得 $\mathbf{v}=c_{n+1}\mathbf{u}_{n+1}+\cdots+c_m\mathbf{u}_m\in\mathrm{Span}(\mathbf{u}_{n+1},\dots,\mathbf{u}_m)$. 反之若 $\mathbf{v}=c_{n+1}\mathbf{u}_{n+1}+\cdots+c_m\mathbf{u}_m\in\mathrm{Span}(\mathbf{u}_{n+1},\dots,\mathbf{u}_m)$, 由於當 $i\neq j$ 時 $\langle\mathbf{u}_i,\mathbf{u}_j\rangle=0$, 故當 $i=1,\dots,n$ 時

$$\langle\mathbf{v},\mathbf{u}_i\rangle=\sum_{j=n+1}^{m}c_j\langle\mathbf{u}_j,\mathbf{u}_i\rangle=0.$$

因此由 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 W 的 basis 以及 Lemma 4.3.3 得 $\mathbf{v}\in W^\perp$, 也因此我們證明了 $W^\perp=\mathrm{Span}(\mathbf{u}_{n+1},\dots,\mathbf{u}_m)$. 又因為 $\mathbf{u}_{n+1},\dots,\mathbf{u}_m$ 為 linearly independent, 故知 $\mathbf{u}_{n+1},\dots,\mathbf{u}_m$ 為 $W^\perp$ 的一組 basis (事實上也是 orthonormal basis).

Proposition 4.3.13. 假設 V 為 *inner product space*, W 為 V 的 *subspace* 且設 $\dim(V)=m$, $\dim(W)=n$. 若 $\mathbf{v}_1,\dots,\mathbf{v}_n,\dots,\mathbf{v}_m$ 為 V 的一組 *orthogonal basis* 且其中 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是 W 的 *orthogonal basis*, 則 $\mathbf{v}_{n+1},\dots,\mathbf{v}_m$ 為 $W^\perp$ 的 *orthogonal basis*. 特別的, 我們有

$$\dim(W^\perp)=\dim(V)-\dim(W).$$

Question 4.8. 在 *Example 4.3.12* 中, 試找到 $W^\perp$ 的一組 *basis.*

在補集的概念中, 我們知道一個集合的補集再取補集, 會是該集合本身. orthogonal complement 會不會也有同樣的情形呢? 也就是說當 W 是 inner product space V 的 subspace, 會不會有 $(W^\perp)^\perp = W$ 的情形發生? 一般要說明 $(W^\perp)^\perp = W$, 我們需證明 $W \subseteq (W^\perp)^\perp$ 以及 $(W^\perp)^\perp \subseteq W$. 證明 $W \subseteq (W^\perp)^\perp$ 這部分是簡單的, 因為 $(W^\perp)^\perp$ 依定義是所有和 $W^\perp$ 垂直的向量所成的集合, 所以當 $\mathbf{w} \in W$, 我們要說明 $\mathbf{w} \in (W^\perp)^\perp$ 僅要說明 $\langle \mathbf{w}, \mathbf{v} \rangle = 0, \forall \mathbf{v} \in W^\perp$ 即可. 然而任意 $\mathbf{v} \in W^\perp$ 依定義皆會和所有 W 中的向量垂直, 故由 $\mathbf{w} \in W$, 我們自然有 $\langle \mathbf{w}, \mathbf{v} \rangle = 0, \forall \mathbf{v} \in W^\perp$. 至於 $(W^\perp)^\perp \subseteq W$, 很不幸的它並不一定會成立. 事實上在 V 為無限維時可以找到反例. 由於本課程並不涉及無限維的向量空間, 這裡就略去不談. 不過在 V 為 finite dimensional inner product space, $(W^\perp)^\perp \subseteq W$ 就會成立. 只是它的證明是無法像前面 $W \subseteq (W^\perp)^\perp$ 的情況用集合元素方式推導 (否則就不會有在無限維時不成立的情況發生). 既然是有限維, 我們可以用為維度處理. 回顧一下當我們有 W' 為 W 的 subspace, 且知 $\dim(W') = \dim(W)$, 則可得 $W' = W$. 所以既然我們已知 $W \subseteq (W^\perp)^\perp$, 只要說明 $\dim(W) = \dim((W^\perp)^\perp)$, 就可得證 $(W^\perp)^\perp = W$.

Corollary 4.3.14. 假設 V 為 *finite dimensional inner product space* 且 W 為 V 的 *subspace.* 則 $(W^\perp)^\perp = W$.

Proof. 前面已證得 $W \subseteq (W^\perp)^\perp$, 所以現在僅要說明 $\dim(W) = \dim((W^\perp)^\perp)$, 就可得證 $(W^\perp)^\perp = W$. 然而由 Proposition 4.3.13, 我們知 $\dim((W^\perp)^\perp) = \dim(V) - \dim(W^\perp)$, 再由 $\dim(W^\perp) = \dim(V) - \dim(W)$, 得

$$\dim((W^\perp)^\perp) = \dim(V) - (\dim(V) - \dim(W)) = \dim(W).$$

□

在 $\mathbb{R}^2, \mathbb{R}^3$ 中的投影概念中, 還有一個重要的觀點就是一個點在直線 (或平面上) 的投影點就是這個線上 (或平面上) 距離該點最近的點. 這個概念對我們推廣到 inner product space 後的 orthogonal projection 也是對的. 我們有以下的性質.

Proposition 4.3.15. 假設 V 為 *inner product space* 且 W 為 V 的 *subspace.* 若 $\mathbf{w} = \mathrm{Proj}_W(\mathbf{v})$ 為 $\mathbf{v}$ 在 W 的 *orthogonal projection,* 則對於任意 $\mathbf{w}' \in W$ 且 $\mathbf{w}' \neq \mathbf{w}$, 皆有 $\|\mathbf{v} - \mathbf{w}'\| > \|\mathbf{v} - \mathbf{w}\|$.

Proof. 考慮 $\mathbf{v} - \mathbf{w}' = \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{w}'$. 因 $\mathbf{w} = \mathrm{Proj}_W(\mathbf{v})$, 故知 $\mathbf{v} - \mathbf{w} \in W^\perp$. 又因 W 為 vector space, 我們有 $\mathbf{w} - \mathbf{w}' \in W$. 故得

$$\|\mathbf{v} - \mathbf{w}'\|^2 = \langle \mathbf{v} - \mathbf{w}', \mathbf{v} - \mathbf{w}' \rangle = \langle \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{w}', \mathbf{v} - \mathbf{w} + \mathbf{w} - \mathbf{w}' \rangle = \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle + \langle \mathbf{w} - \mathbf{w}', \mathbf{w} - \mathbf{w}' \rangle.$$

亦即 $\|\mathbf{v} - \mathbf{w}'\|^2 = \|\mathbf{v} - \mathbf{w}\|^2 + \|\mathbf{w} - \mathbf{w}'\|^2$. 又因 $\mathbf{w} \neq \mathbf{w}'$ 我們有 $\|\mathbf{w} - \mathbf{w}'\| > 0$. 得證 $\|\mathbf{v} - \mathbf{w}'\|^2 > \|\mathbf{v} - \mathbf{w}\|^2$, 即 $\|\mathbf{v} - \mathbf{w}'\| > \|\mathbf{v} - \mathbf{w}\|$. □

4.4. 矩陣運算和內積的連結

前面提過矩陣的乘法, 和 $\mathbb{R}^n$ 的 dot product 有密切的關係. 當 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$, 我們用 $\langle\mathbf{v},\mathbf{w}\rangle$ 表示 $\mathbf{v},\mathbf{w}$ 的 dot product. 而當我們將 $\mathbf{v},\mathbf{w}$ 視為 $n\times 1$ 的矩陣則 $\mathbf{v}$, $\mathbf{w}$ 的 dot product 可視為矩陣乘法 $\langle\mathbf{v},\mathbf{w}\rangle=\mathbf{w}^{\mathrm{t}}\mathbf{v}$. 這一節我們將利用這個觀點以及上一節所談的 inner product 的性質, 來探討矩陣運算的性質.

當 $\mathbf{v}\in\mathbb{R}^m$, $\mathbf{w}\in\mathbb{R}^n$ 以及 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$, 由於 $A\mathbf{w}\in\mathbb{R}^m$, 我們可以討論 $\mathbf{v}$ 和 $A\mathbf{w}$ 在 $\mathbb{R}^m$ 的 dot product, 即 $\langle\mathbf{v},A\mathbf{w}\rangle$. 當我們將 $\mathbf{v},A\mathbf{w}$ 視為 $m\times 1$ 矩陣, 利用矩陣乘法可得

$$\langle\mathbf{v},A\mathbf{w}\rangle=(A\mathbf{w})^{\mathrm{t}}\mathbf{v}=(\mathbf{w}^{\mathrm{t}}A^{\mathrm{t}})\mathbf{v}=\mathbf{w}^{\mathrm{t}}(A^{\mathrm{t}}\mathbf{v}). \tag{4.2}$$

另一方面將 $A^{\mathrm{t}}\mathbf{v},\mathbf{w}$ 視為 $\mathbb{R}^n$ 上的向量, 考慮它們在 $\mathbb{R}^m$ 的 dot product, 我們有

$$\langle A^{\mathrm{t}}\mathbf{v},\mathbf{w}\rangle=\mathbf{w}^{\mathrm{t}}(A^{\mathrm{t}}\mathbf{v}). \tag{4.3}$$

結合式子 (4.2), (4.3) 我們可得以下性質.

Lemma 4.4.1. 假設 $\mathbf{v}\in\mathbb{R}^m$, $\mathbf{w}\in\mathbb{R}^n$ 以及 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$. 若考慮 $\langle\mathbf{v},A\mathbf{w}\rangle$ 為在 $\mathbb{R}^m$ 的 *dot product* 以及 $\langle A^{\mathrm{t}}\mathbf{v},\mathbf{w}\rangle$ 為在 $\mathbb{R}^n$ 的 *dot product,* 則

$$\langle\mathbf{v},A\mathbf{w}\rangle=\langle A^{\mathrm{t}}\mathbf{v},\mathbf{w}\rangle.$$

回顧當 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 時我們令 $\mathrm{Col}(A)$ 表示 A 的 column space, 亦即若 $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 分別為 A 的 column vectors, 則 $\mathrm{Col}(A)=\mathrm{Span}(\mathbf{a}_1,\ldots,\mathbf{a}_n)$. 另外我們也定義 A 的 null space 為 $N(A)$, 即 $N(A)=\{\mathbf{w}\in\mathbb{R}^n:A\mathbf{w}=\mathbf{0}\}$. 注意 $\mathrm{Col}(A)$ 為 $\mathbb{R}^m$ 的 subspace, 而 $N(A)$ 為 $\mathbb{R}^n$ 的 subspace.

我們有興趣知道在 $\mathbb{R}^m$ 中 $\mathrm{Col}(A)^\perp$ 為何? 若 $\mathbf{v}\in\mathrm{Col}(A)^\perp$, 表示 $\mathbf{v}$ 與 $\mathrm{Col}(A)$ 中所有的向量皆垂直. 利用 Lemma 4.3.3 相同的證明方法, 我們知道這等價於 $\langle\mathbf{v},\mathbf{a}_i\rangle=0$, $\forall i=1,\ldots,n$. 然而依矩陣的乘法, A 的 i-th column 等於 $A\mathbf{e}_i$, 其中 $\mathbf{e}_i$ 是 i-th entry 為 1 其他 entry 為 0 的 $\mathbb{R}^n$ 中的向量, 也就是說 $\mathbf{a}_i=A\mathbf{e}_i$. 因此由 Lemma 4.4.1 知 $0=\langle\mathbf{v},\mathbf{a}_i\rangle=\langle\mathbf{v},A\mathbf{e}_i\rangle=\langle A^{\mathrm{t}}\mathbf{v},\mathbf{e}_i\rangle$, $\forall i=1,\ldots,n$. 換言之 $A^{\mathrm{t}}\mathbf{v}$ 這個 $\mathbb{R}^n$ 的向量和每個 $\mathbf{e}_i$ 的 dot product 皆為 0, 得證 $A^{\mathrm{t}}\mathbf{v}=\mathbf{0}$, 也就是說 $\mathbf{v}\in N(A^{\mathrm{t}})$. 反之, 若 $\mathbf{v}\in N(A^{\mathrm{t}})$ 表示 $A^{\mathrm{t}}\mathbf{v}=\mathbf{0}$, 故得 $\langle A^{\mathrm{t}}\mathbf{v},\mathbf{e}_i\rangle=0$, $\forall i=1,\ldots,n$. 因此得 $\langle\mathbf{v},\mathbf{a}_i\rangle=0$, $\forall i=1,\ldots,n$, 即 $\mathbf{v}\in\mathrm{Col}(A)^\perp$. 我們有以下定理.

Theorem 4.4.2. 假設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$. 考慮 $\mathrm{Col}(A)$ 為 $\mathbb{R}^m$ 的 *subspace,* 使用 *dot product,* 我們有 $\mathrm{Col}(A)^\perp=N(A^{\mathrm{t}})$.

Theorem 4.4.2 可以用其他的形式表達. 例如利用 Corollary 4.3.14, 我們可以得到

$$\mathrm{Col}(A)=(\mathrm{Col}(A)^\perp)^\perp=N(A^{\mathrm{t}})^\perp.$$

另一方面, 利用 $(A^{\mathrm{t}})^{\mathrm{t}}=A$, 我們可得

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

回顧一下, 我們稱 $\dim(\mathrm{Col}(A))$ 為 A 的 rank, 用 $\mathrm{rank}(A)$ 表示, 而 $\dim(N(A))$ 稱為 A 的 nullity, 用 $\mathrm{nullity}(A)$ 表示. 當計算維度牽涉到 orthogonal complement 時要小心, 因為

$W^{\perp}$ 其實是和將 W 視為哪個 vector space 的 subspace 有關, 因此 Proposition 4.3.13 計算 $W^{\perp}$ 的維度其實是和 W 所在的向量空間 V (即將 W 視為 V 的 subspace) 的維度有關. 當 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$, 我們是將 $\mathrm{Col}(A)$ 視為 $\mathbb{R}^m$ 的 subspace (因 A 的每個 column vector 是在 $\mathbb{R}^m$ 中), 而將 $N(A)$ 視為 $\mathbb{R}^n$ 的 subspace (因為只有 $\mathbb{R}^n$ 的向量可以乘在 A 的右邊). 所以利用 $\mathrm{Col}(A^{\mathrm{t}}) = N(A)^{\perp}$, 我們可以得

$$\mathrm{rank}(A) = \dim(\mathrm{Col}(A^{\mathrm{t}})) = \dim(N(A)^{\perp}) = n - \dim(N(A)) = n - \mathrm{nullity}(A).$$

這與 Dimension Theorem 相吻合.

Lemma 4.4.1 還有許多的應用. 我們有以下的定理.

Proposition 4.4.3. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$. 則 $N(A^{\mathrm{t}}A) = N(A)$.

Proof. 假設 $\mathbf{v} \in N(A)$, 則由 $A\mathbf{v} = \mathbf{0}$ 得 $(A^{\mathrm{t}}A)\mathbf{v} = A^{\mathrm{t}}(A\mathbf{v}) = A^{\mathrm{t}}(\mathbf{0}) = \mathbf{0}$. 即 $\mathbf{v} \in N(A^{\mathrm{t}}A)$, 故得證 $N(A) \subseteq N(A^{\mathrm{t}}A)$. 反之, 若 $\mathbf{v} \in N(A^{\mathrm{t}}A)$, 則由 $(A^{\mathrm{t}}A)\mathbf{v} = \mathbf{0}$ 以及 Lemma 4.4.1, 可得

$$\langle A\mathbf{v}, A\mathbf{v}\rangle = \langle A^{\mathrm{t}}(A\mathbf{v}), \mathbf{v}\rangle = \langle (A^{\mathrm{t}}A)\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{0}, \mathbf{v}\rangle = 0.$$

故由內積性質得證 $A\mathbf{v} = \mathbf{0}$, 即 $\mathbf{v} \in N(A)$. 因此得 $N(A^{\mathrm{t}}A) \subseteq N(A)$. □

當 A 為 $m\times n$ matrix, 則 $A^{\mathrm{t}}A$ 就會是 $n\times n$ 方陣. 因此由 Dimension Theorem 以及 Proposition 4.4.3, 我們有

$$n - \mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{nullity}(A^{\mathrm{t}}A) = \dim(N(A^{\mathrm{t}}A)) = \dim(N(A)) = \mathrm{nullity}(A) = n - \mathrm{rank}(A)$$

因此得知 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$. 我們有以下的結果.

Corollary 4.4.4. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$. 則 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$. 特別的我們有 $\mathrm{rank}(A) = n$ 若且唯若 $A^{\mathrm{t}}A$ 為 *invertible matrix.*

Proof. 我們已經證得 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$. 由於 $A^{\mathrm{t}}A$ 為 $n\times n$ matrix, $\mathrm{rank}(A^{\mathrm{t}}A) = n$ 等價於 $A^{\mathrm{t}}A$ 為 invertible matrix (Theorem 2.5.2), 也因此由 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$ 知 $\mathrm{rank}(A) = n$ 等價於 $A^{\mathrm{t}}A$ 為 invertible matrix. □

要注意，雖然 $\dim(\mathrm{Col}(A)) = \mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{t}}A) = \dim(\mathrm{Col}(A^{\mathrm{t}}A))$，但這不代表 $\mathrm{Col}(A) = \mathrm{Col}(A^{\mathrm{t}}A)$. 這是因為當 A 為 $m\times n$ matrix 時 $\mathrm{Col}(A) \subseteq \mathbb{R}^m$, 而此時 $A^{\mathrm{t}}A$ 為 $n\times n$ matrix, 故 $\mathrm{Col}(A^{\mathrm{t}}A) \subseteq \mathbb{R}^n$. 也就是說，當 $n \neq m$ 時，$\mathrm{Col}(A)$ 和 $\mathrm{Col}(A^{\mathrm{t}}A)$ 根本就在不同空間中，當然不可能相等。即使當 $n = m$, 時 $\mathrm{Col}(A^{\mathrm{t}}A)$ 和 $\mathrm{Col}(A)$ 一般也沒有關係。事實上 $\mathrm{Col}(A^{\mathrm{t}}A)$ 會是 A 的 row space, $\mathrm{Col}(A^{\mathrm{t}})$.

Corollary 4.4.5. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$. 則 $\mathrm{Col}(A^{\mathrm{t}}A) = \mathrm{Col}(A^{\mathrm{t}})$.

Proof. 首先我們證明 $\mathrm{Col}(A^{\mathrm{t}}A) \subseteq \mathrm{Col}(A^{\mathrm{t}})$. 這是因為對任意 $\mathbf{v} \in \mathrm{Col}(A^{\mathrm{t}}A)$, 表示存在 $\mathbf{w} \in \mathbb{R}^n$ 使得 $\mathbf{v} = (A^{\mathrm{t}}A)\mathbf{w}$. 然而 $(A^{\mathrm{t}}A)\mathbf{w} = A^{\mathrm{t}}(A\mathbf{w})$, 故得證 $\mathbf{v} = A^{\mathrm{t}}(A\mathbf{w}) \in \mathrm{Col}(A^{\mathrm{t}})$.

接下來，我們只要說明 $\dim(\mathrm{Col}(A^{\mathrm{t}}A)) = \dim(\mathrm{Col}(A^{\mathrm{t}}))$, 即可得證 $\mathrm{Col}(A^{\mathrm{t}}A) = \mathrm{Col}(A^{\mathrm{t}})$ (Proposition 3.6.10). 然而 Corollary 4.4.4 告訴我們 $\mathrm{rank}(A^{\mathrm{t}}A) = \mathrm{rank}(A)$, 而 $\mathrm{rank}(A^{\mathrm{t}}A) = \dim(\mathrm{Col}(A^{\mathrm{t}}A))$ 且 $\mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{t}}) = \dim(\mathrm{Col}(A^{\mathrm{t}}))$ (Proposition 3.7.15), 故得證本定理。 □

當 W 為 $\mathbb{R}^m$ 的 subspace 時, Theorem 4.4.2 也可以幫助我們求 $\mathbf{v}\in\mathbb{R}^n$ 在 W 上的 orthogonal projection. 要注意在 Theorem 4.3.8 中找 W 的 orthogonal basis 的方法求 orthogonal projection 是適用於一般的 inner product space, 而這裡我們介紹的方法僅適用於 $\mathbb{R}^m$ 且使用 dot product.

假設 $W=\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)\subseteq\mathbb{R}^m$ 首先令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vectors 的 $m\times n$ matrix, 則 W 為 A 的 column space $\mathrm{Col}(A)$. 依 $\mathrm{Proj}_W(\mathbf{v})$ 的定義, 我們需要找到 $\mathbf{w}\in W$ 滿足 $\mathbf{v}-\mathbf{w}\in W^{\perp}$, 這樣 $\mathbf{w}$ 就會是 $\mathrm{Proj}_W(\mathbf{v})$ 了. 如何找到這樣的 $\mathbf{w}\in W$ 呢? 由 column space 的定義知, $\mathbf{w}\in W$ 表示存在 $\mathbf{x}\in\mathbb{R}^n$ 使得 $A\mathbf{x}=\mathbf{w}$. 至於 $\mathbf{v}-\mathbf{w}\in W^{\perp}=(\mathrm{Col}(A))^{\perp}$ 的要求, Theorem 4.4.2 知此即表示 $\mathbf{v}-\mathbf{w}=\mathbf{v}-A\mathbf{x}\in N(A^{\mathrm{t}})$. 也就是說我們必須找到 $\mathbf{x}\in\mathbb{R}^n$ 使得 $A^{\mathrm{t}}(\mathbf{v}-A\mathbf{x})=\mathbf{0}$. 利用矩陣乘法性質, 此即表示 $\mathbf{x}\in\mathbb{R}^n$ 須滿足聯立方程組 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$. 要注意，由於 $A^{\mathrm{t}}\mathbf{v}\in\mathrm{Col}(A^{\mathrm{t}})=\mathrm{Col}(A^{\mathrm{t}}A)$ (Corollary 4.4.5), 聯立方程組 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$ 一定有解。若能解 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$, 則所得的解 $\mathbf{x}$ 就會使得 $A\mathbf{x}=\mathrm{Proj}_W(\mathbf{v})$ 了. 我們有以下的定理.

Proposition 4.4.6. 假設 $W=\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)\subseteq\mathbb{R}^m$ 且考慮 $\mathbb{R}^m$ 的 *dot product.* 對於任意 $\mathbf{v}\in\mathbb{R}^m$, 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 *column vectors* 的 $m\times n$ *matrix* 且考慮聯立方程組 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$. 若 $\mathbf{x}_0$ 為此聯立方程組的一個解, 則 $\mathrm{Proj}_W(\mathbf{v})=A\mathbf{x}_0$.

特別的, 如果 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis, 則 $\mathrm{rank}(A)=\dim(W)=n$. 故利用 Corollary 4.4.4 可得 $A^{\mathrm{t}}A$ 為 invertible. 此時只要將聯立方程組 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$ 的兩邊乘上 $A^{\mathrm{t}}A$ 的 inverse, 即可得解為 $\mathbf{x}=(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}\mathbf{v}$. 注意此時我們僅解得 $(A^{\mathrm{t}}A)\mathbf{x}=A^{\mathrm{t}}\mathbf{v}$ 之解, 要將此解的左邊乘上 A 才得 $\mathrm{Proj}_W(\mathbf{v})$. 我們有以下的結論.

Corollary 4.4.7. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且考慮 $\mathbb{R}^m$ 的 *dot product.* 假設 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *basis,* 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 *column vector* 的 $m\times n$ *matrix.* 則對於任意 $\mathbf{v}\in\mathbb{R}^m$, $\mathbf{v}$ 在 W 的 *projection* 為

$$\mathrm{Proj}_W(\mathbf{v})=A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}\mathbf{v}.$$

Example 4.4.8. 我們要利用 Corollary 4.4.7 的結果求 $\mathbf{v}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$ 在 $W=\mathrm{Span}(\begin{bmatrix}1\\2\\2\\1\end{bmatrix},\begin{bmatrix}3\\4\\2\\3\end{bmatrix})$ 的投影. 首先考慮矩陣 $A=\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}$, 此時 $A^{\mathrm{t}}=\begin{bmatrix}1&2&2&1\\3&4&2&3\end{bmatrix}$, 故得 $A^{\mathrm{t}}A=\begin{bmatrix}10&18\\18&38\end{bmatrix}$ 以及其 inverse $(A^{\mathrm{t}}A)^{-1}=(1/28)\begin{bmatrix}19&-9\\-9&5\end{bmatrix}$. 因此由 Corollary 4.4.7 得

$$\mathrm{Proj}_W(\mathbf{v})=\frac{1}{28}\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}\begin{bmatrix}19&-9\\-9&5\end{bmatrix}\begin{bmatrix}1&2&2&1\\3&4&2&3\end{bmatrix}\begin{bmatrix}2\\0\\1\\4\end{bmatrix}=\begin{bmatrix}2\\2\\0\\2\end{bmatrix}.$$

這個結果和我們在 Example 4.3.10 利用 orthogonal basis 處理投影的結果一致.

對於 Corollary 4.4.7 要注意的是因為 W 為 $\mathbb{R}^m$ 的 subspace, 除非 $W = \mathbb{R}^m$, 否則 $\dim(W) = n$ 會小於 m. 然而當 $W = \mathbb{R}^m$ 時, 談論對 W 的 projection 是沒有意思的, 因為此時 $W^{\perp} = \{\mathbf{0}\}$, 所以任何 $\mathbb{R}^m$ 的向量對 $W = \mathbb{R}^m$ 的投影就是自己. 因此一般在談論投影時僅考慮 $\dim(W) = n < m$ 的情形. 也因此, 利用 W 的一組 basis 為 column vector 所成的矩陣 A, 是 $m \times n$ matrix 不會是一個方陣. 所以此時 A 和 A^{t} 皆不會是 invertible. 也因此我們不能將 $(A^{\mathrm{t}}A)^{-1}$ 寫成 $A^{-1}(A^{\mathrm{t}})^{-1}$. 因為這原因 Corollary 4.4.7 中 $A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 絕不能寫成 $A(A^{-1}(A^{\mathrm{t}})^{-1})A^{\mathrm{t}}$, 否則會變成 identity matrix.

Theorem 4.4.7 簡化了求 projection 的程序. 我們只要求出 W 的一組 basis 即可, 不必先求 W 的 orthogonal basis. 由於將矩陣 $A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 乘上任何 $\mathbb{R}^m$ 的向量 $\mathbf{v}$, 就可得 $\mathrm{Proj}_W(\mathbf{v})$. 因此我們將 $A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 稱之為對於 W 的 *projection matrix.*

Question 4.9. 在 *Example 4.4.8* 中對於 W 的 *projection matrix* 為何? 用 *Example 4.3.10* 中所得的 W 的 *orthogonal basis* 所得的 *projection matrix* 又是為何?

假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathrm{rank}(A) = n$, 則 A 的 column vectors 形成 $\mathrm{Col}(A)$ 的一組 basis. 假設 A 的 column 分別為 $\mathbf{v}_1, \ldots, \mathbf{v}_n$, 我們利用 Gram-Schmidt process 得到 $\mathrm{Col}(A)$ 的一組 orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. 由於 $\mathbf{u}_1, \ldots, \mathbf{u}_n$ 是 orthonormal basis, 將 $\mathbf{v}_j$ 寫成 $\mathbf{u}_1, \ldots, \mathbf{u}_n$ 的線性組合可得

$$\mathbf{v}_j = \langle \mathbf{v}_j, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{v}_j, \mathbf{u}_j \rangle \mathbf{u}_j + \cdots + \langle \mathbf{v}_j, \mathbf{u}_n \rangle \mathbf{u}_n.$$

因此若令 Q 為以 $\mathbf{u}_1, \ldots, \mathbf{u}_n$ 為 column vectors 的 $m \times n$ matrix, 依矩陣乘法定義我們可以將 A 寫成 QR, 即

$$\begin{bmatrix} | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_j & \cdots & \mathbf{v}_n \\ | & & | & & | \end{bmatrix} = \begin{bmatrix} | & & | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_j & \cdots & \mathbf{u}_n \\ | & & | & & | \end{bmatrix} \begin{bmatrix} \langle \mathbf{v}_1, \mathbf{u}_1 \rangle & & \langle \mathbf{v}_j, \mathbf{u}_1 \rangle & & \langle \mathbf{v}_n, \mathbf{u}_1 \rangle \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ \langle \mathbf{v}_1, \mathbf{u}_n \rangle & & \langle \mathbf{v}_j, \mathbf{u}_n \rangle & & \langle \mathbf{v}_n, \mathbf{u}_n \rangle \end{bmatrix},$$

其中 R 是一個 $n \times n$ matrix 且其 j-th column 為 $\begin{bmatrix} \langle \mathbf{v}_j, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{v}_j, \mathbf{u}_n \rangle \end{bmatrix}$, 也就是說 R 的 (i,j)-th entry 為 $\langle \mathbf{v}_j, \mathbf{u}_i \rangle = \langle \mathbf{u}_i, \mathbf{v}_j \rangle$. 在 Gram-Schmidt process, 對於 $j = 1, \ldots, n$, 我們都有 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_j) = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_j)$ 而且 $\mathbf{u}_{j+1}, \ldots, \mathbf{u}_n \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_j)^{\perp}$, 故知

$$\langle \mathbf{u}_{j+1}, \mathbf{v}_j \rangle = \langle \mathbf{u}_{j+2}, \mathbf{v}_j \rangle = \cdots = \langle \mathbf{u}_n, \mathbf{v}_j \rangle = 0.$$

另外由 $\mathbf{v}_j \not\in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}) = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{j-1})$, 我們知 $\langle \mathbf{u}_j, \mathbf{v}_j \rangle \neq 0$, 否則會造成 $\mathbf{v}_j = \langle \mathbf{u}_1, \mathbf{v}_j \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{u}_{j-1}, \mathbf{v}_j \rangle \mathbf{u}_{j-1} \in \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_{j-1})$ 之矛盾. 故由前面所述, 當 $i > j$ 時 $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = 0$, 我們知 R 為 $n \times n$ upper triangular matrix, 而且對角線的位置 (j,j)-th entry 為 $\langle \mathbf{u}_j, \mathbf{v}_j \rangle \neq 0$, 我們得 $\mathrm{rank}(R) = n$, 故 R 為 invertible. 這就是所謂 A 的 *QR decomposition.* 我們用一個例子來說明.

Example 4.4.9. 考慮 $A=\begin{bmatrix}1&3&1\\1&1&1\\1&-1&3\\1&1&3\end{bmatrix}$ 因 A 的 column vectors 就是 Example 4.3.12 的 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$, 我們直接套用其結果得

$$Q=\begin{bmatrix}\frac{1}{2}&\frac{1}{\sqrt{2}}&0\\\frac{1}{2}&0&-\frac{1}{\sqrt{2}}\\\frac{1}{2}&-\frac{1}{\sqrt{2}}&0\\\frac{1}{2}&0&\frac{1}{\sqrt{2}}\end{bmatrix},\quad R=\begin{bmatrix}\langle\mathbf{u}_1,\mathbf{v}_1\rangle&\langle\mathbf{u}_1,\mathbf{v}_2\rangle&\langle\mathbf{u}_1,\mathbf{v}_3\rangle\\\langle\mathbf{u}_2,\mathbf{v}_1\rangle&\langle\mathbf{u}_2,\mathbf{v}_2\rangle&\langle\mathbf{u}_2,\mathbf{v}_3\rangle\\\langle\mathbf{u}_3,\mathbf{v}_1\rangle&\langle\mathbf{u}_3,\mathbf{v}_2\rangle&\langle\mathbf{u}_3,\mathbf{v}_3\rangle\end{bmatrix}=\begin{bmatrix}2&2&4\\0&2\sqrt{2}&-\sqrt{2}\\0&0&\sqrt{2}\end{bmatrix}.$$

很容易檢查, 我們確有 $A=QR$.

將矩陣 A 寫成 QR decomposition 在許多應用上有其方便性. 特別是探討與 $A^{\mathrm{t}}A$ 有關的問題. 此時我們有 $A^{\mathrm{t}}A=(QR)^{\mathrm{t}}(QR)=(R^{\mathrm{t}}Q^{\mathrm{t}})(QR)=R^{\mathrm{t}}(Q^{\mathrm{t}}Q)R$. 然而 Q 的 column vectors 是 $\mathrm{Col}(A)$ 的一組 orthonormal basis, 很容易驗證 $Q^{\mathrm{t}}Q$ 會是 $n\times n$ diagonal matrix, 且其 (i,i)-th entry 為 $\langle\mathbf{u}_i,\mathbf{u}_i\rangle=\|\mathbf{u}_i\|^2=1$. 也就是說 $Q^{\mathrm{t}}Q$ 是 identity matrix I_n. 因此我們可得 $A^{\mathrm{t}}A=R^{\mathrm{t}}R$. 將 $A^{\mathrm{t}}A$ 寫成 $R^{\mathrm{t}}R$ 的好處是 R 是一個 upper triangular matrix 且為 invertible (注意 A 未必 invertible). 例如 Corollary 4.4.7 對 $W=\mathrm{Col}(A)$ 的 projection matrix $A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}$ 就可以寫成

$$A(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}=(QR)(R^{\mathrm{t}}R)^{-1}(QR)^{\mathrm{t}}=(QR)(R^{-1}(R^{\mathrm{t}})^{-1})(R^{\mathrm{t}}Q^{\mathrm{t}})=Q(RR^{-1})((R^{\mathrm{t}})^{-1}R^{\mathrm{t}})Q^{\mathrm{t}}=QQ^{\mathrm{t}}$$

這種簡單的形式了. 再次提醒我們知 $Q^{\mathrm{t}}Q$ 是 identity matrix, 但 QQ^{t} 就未必是 identity 了. 下一節中探討解聯立方程組問題時還會看到 QR decomposition 的應用.

4.5. 聯立方程組和內積的連結

在這一節中我們將利用內積的概念處理聯立方程組無解的情況. 我們要探討一個 linear system 無解時或是有解但解不唯一時, 如何可找到最佳的可能解. 我們也將利用這樣的概念處理大家高中所學有關統計二維資料的最適合直線.

當 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$, $\mathbf{b}\in\mathbb{R}^m$ 我們知道 $\mathrm{Col}(A)$ 可以幫助我們判斷聯立方程組 $A\mathbf{x}=\mathbf{b}$ 是否有解, 而 $N(A)$ 可幫助我們判斷 $A\mathbf{x}=\mathbf{b}$ 若有解其解是否唯一. 又 Theorem 4.4.2 告訴我們 $\mathrm{Col}(A)^{\perp}=N(A^{\mathrm{t}})$, 我們將利用這個關係來探討聯立方程組無解或解不唯一時如何處理問題.

首先我們探討無解的情形. 當 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$, 我們都知道聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 若且唯若 $\mathbf{b}\in\mathrm{Col}(A)$. 因此, 一般在日常應用中, 當我們要處理的聯立方程組 $A\mathbf{x}=\mathbf{b}$ 無解時, 我們認為很有可能是 $\mathbf{b}$ 發生誤差所導致, 所以會去找 $\mathbf{b}_0$ 為所有使得聯立方程組 $A\mathbf{x}=\mathbf{b}'$ 有解的 $\mathbf{b}'$ 中與 $\mathbf{b}$ 的距離最近的一個. 這樣的 $\mathbf{b}_0$ 所得的聯立方程組 $A\mathbf{x}=\mathbf{b}_0$ 的解, 我們認為是最佳的可能解. 然而符合 $A\mathbf{x}=\mathbf{b}'$ 有解的 $\mathbf{b}'$ 所成的集合就是 $\mathrm{Col}(A)$ 這一個 subspace, 所以依 Proposition 4.3.15 我們知道所有的 $\mathbf{b}'$ 中距離 $\mathbf{b}$ 最近的 $\mathbf{b}_0$ 應該就是 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的 orthogonal projection. 也因此由 orthogonal projection 的定義以及 Theorem 4.4.2 知, $\mathbf{b}-\mathbf{b}_0\in\mathrm{Col}(A)^{\perp}=N(A^{\mathrm{t}})$. 此即表示 $A^{\mathrm{t}}(\mathbf{b}-\mathbf{b}_0)=\mathbf{0}$, 也就是

$$A^{\mathrm{t}}\mathbf{b}_0=A^{\mathrm{t}}\mathbf{b}. \tag{4.4}$$

由於我們是要解 $A\mathbf{x}=\mathbf{b}_0$ 所以代入上式推得要解

$$A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}. \tag{4.5}$$

注意式子 (4.5) 和式子 (4.4) 的不同點在於, 我們不必求出 $\mathbf{b}_0$ 再解 $A\mathbf{x}=\mathbf{b}_0$, 而是直接解 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$. 不過我們必須說明式子 (4.5) 的解確實是我們希望得到的最佳的可能解.

假設 $\mathbf{x}=\mathbf{x}_0$ 是 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 的一解, 令 $\mathbf{b}_0=A\mathbf{x}_0$ 此即表示 $\mathbf{b}_0\in\mathrm{Col}(A)$ 且 $\mathbf{b}-\mathbf{b}_0\in N(A^{\mathrm{t}})=\mathrm{Col}(A)^{\perp}$. 因此得證 $\mathbf{b}_0$ 就是 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的 orthogonal projection, 故由 Proposition 4.3.15 知 $\mathbf{b}_0$ 確實是所有使得聯立方程組 $A\mathbf{x}=\mathbf{b}'$ 有解的 $\mathbf{b}'$ 中距離 $\mathbf{b}$ 最近的一個. 因此若 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 有解, 則此解的確會是我們所期望的一個最佳的可能解. 現在我們面臨的問題是式子 (4.5) 一定有解嗎?

Proposition 4.5.1. *假設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b}\in\mathbb{R}^m$. 聯立方程組 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 一定有解. 特別地, 若 $\mathrm{rank}(A)=n$, 則聯立方程組有解且解唯一.*

Proof. 由 Proposition 3.7.2, 我們知聯立方程組 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 有解，等同於 $A^{\mathrm{t}}\mathbf{b}\in\mathrm{Col}(A^{\mathrm{t}}A)$. 然而 Corollary 4.4.5 告訴我們 $\mathrm{Col}(A^{\mathrm{t}}A)=\mathrm{Col}(A^{\mathrm{t}})$, 故由 $A^{\mathrm{t}}\mathbf{b}\in\mathrm{Col}(A^{\mathrm{t}})$ 知 $A^{\mathrm{t}}\mathbf{b}\in\mathrm{Col}(A^{\mathrm{t}}A)$, 因此聯立方程組 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 一定有解.

當 $\mathrm{rank}(A)=n$, 由 Corollary 4.4.4 我們知 $A^{\mathrm{t}}A$ 為 invertible, 所以聯立方程組 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 有解且解唯一. □

由 Proposition 4.5.1, 我們特別有以下的定義.

Definition 4.5.2. 假設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b}\in\mathbb{R}^m$. 考慮聯立方程組 $A\mathbf{x}=\mathbf{b}$. 我們稱聯立方程組 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 為 $A\mathbf{x}=\mathbf{b}$ 的 *normal equation*, 而 normal equation 的解稱為原方程組的 *least squares solution*。

注意當 $\mathrm{rank}(A)<n$ 時, normal equation 雖然有解不過其解因等同於聯立方程 $A\mathbf{x}=\mathrm{Proj}_{\mathrm{Col}(A)}(\mathbf{b})$ 的解, 故由 A 為 $m\times n$ matrix 以及 Theorem 2.4.5 知其解不唯一 (會有無窮多解). 有關於解不唯一的情形, 通常我們又會去找長度最小的解（稱為 *minimal least squares solution*）我們以後再談.

另外要強調的是, 當求 $A\mathbf{x}=\mathbf{b}$ 無解時, 我們是求 normal equations 的解, 而不是要求 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的 projection. 所以在解出 normal equation 的後, 是不必再在解的左邊乘上 A 的. 另外即使聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 即 $\mathbf{b}\in\mathrm{Col}(A)$, 此時 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的 projection 就是 $\mathbf{b}$ 本身. 所以此時 $A\mathbf{x}=\mathbf{b}$ 的解和其 normal equations $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 的解是一致的. 因此我們可以不必擔心原方程組 $A\mathbf{x}=\mathbf{b}$ 是否有解, 直接求其 normal equations $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 的解即可.

Example 4.5.3. 考慮聯立方程組 $A\mathbf{x}=\mathbf{b}$ 其中 $A=\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}$ 且 $\mathbf{b}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$, 此時 $A^{\mathrm{t}}A=\begin{bmatrix}10&18\\18&38\end{bmatrix}$ 以及 $A^{\mathrm{t}}\mathbf{b}=\begin{bmatrix}8\\20\end{bmatrix}$, 故得 $A\mathbf{x}=\mathbf{b}$ 的 normal equations 為

$$\begin{array}{rcl}10x_1+18x_2&=&8\\18x_1+38x_2&=&20\end{array}.$$

解得 $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}$ 為其解. 我們也可由 $\mathrm{rank}(A)=2$, 得 $A^{\mathrm{t}}A$ 為 invertible 且其 inverse 為 $(A^{\mathrm{t}}A)^{-1}=(1/28)\begin{bmatrix}19&-9\\-9&5\end{bmatrix}$. 解得

$$\begin{bmatrix}x_1\\x_2\end{bmatrix}=(A^{\mathrm{t}}A)^{-1}A^{\mathrm{t}}\mathbf{b}=\frac{1}{28}\begin{bmatrix}19&-9\\-9&5\end{bmatrix}\begin{bmatrix}8\\20\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}.$$

我們可以檢查 $\begin{bmatrix}1&3\\2&4\\2&2\\1&3\end{bmatrix}\begin{bmatrix}-1\\1\end{bmatrix}=\begin{bmatrix}2\\2\\0\\2\end{bmatrix}$ 確為 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的 projection (參見 Example 4.4.8).

有關 normal equation 的應用, 最常見的就是二維資料的最適合直線. 也就是說當我們有一組二維的資料 $(x_1,y_1),\ldots,(x_n,y_n)$, 我們希望找到一條直線 $y=mx+c$, 希望這些點 (x_i,y_i) 盡可能地靠近這條直線. 注意這裡 $x_1,\ldots,x_n$ 以及 $y_1,\ldots,y_n$ 都是給定的數, 而我們要解的不是 x,y 而是這個直線的斜率 m 以及 y 截距 c. 依聯立方程組的觀點來看, 我們希望找到 m,c 使之符合

$$\begin{array}{rcl}mx_1+c&=&y_1\\mx_2+c&=&y_2\\&\vdots&\\mx_n+c&=&y_n.\end{array}$$

換句話說我們要解聯立方程組 $A\mathbf{x}=\mathbf{b}$, 其中 $A=\begin{bmatrix}x_1&1\\x_2&1\\\vdots&\vdots\\x_n&1\end{bmatrix}$, $\mathbf{x}=\begin{bmatrix}m\\c\end{bmatrix}$, $\mathbf{b}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}$. 當然了, 這些給定的資料 $(x_1,y_1),\ldots,(x_n,y_n)$ 不一定會使得 $A\mathbf{x}=\mathbf{b}$ 有解, 所以我們要求的是 $A\mathbf{x}=\mathbf{b}$ 的 normal equation 的解, 也就是解 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$. 注意, 因為一般要分析的二維資料是要探討 x_i,y_i 之間的關係, 所以這些資料中 $x_1,\ldots,x_n$ 是不會全相同的 (否則 $(x_1,y_1),\ldots,(x_n,y_n)$ 會同時落在一鉛直線上, 也沒甚麼好探討的了), 所以這裡 $\mathrm{rank}(A)=2$, 因此由 Proposition 4.5.1 知 normal equations $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 一定有解且解唯一. 接下來我們要說明的是, 這樣所得的直線其意義為何?

給定 $m,c\in\mathbb{R}$, 對於 $i=1,\ldots,n$, 令 $y_i'=mx_i+c$. 也就是說 (x_i,y_i') 會在直線 $y=mx+c$ 上. 為方便起見令 $\mathbf{b}'=\begin{bmatrix}y_1'\\\vdots\\y_n'\end{bmatrix}$, 此時 $\mathbf{b}'=\begin{bmatrix}x_1&1\\\vdots&\vdots\\x_n&1\end{bmatrix}\begin{bmatrix}m\\c\end{bmatrix}\in\mathrm{Col}(A)$. 又令 $\varepsilon_i=y_i'-y_i$, 此即將直線

$y=mx+c$ 代 x_i 所得的 y'_i 與實際資料中的 y_i 之誤差. 我們知道代這些 x_i 後所得的誤差越小越好, 不過由於 ε_i 會有正有負, 怕正負抵銷了影響誤差的判定, 所以我們取平方, 也就是說希望 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 的值越小越好. 然而依定義 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 就是 $\mathbf{b}'$ 和 $\mathbf{b}$ 距離的平方, 即 $\|\mathbf{b}-\mathbf{b}'\|^2$. 另一方面, 我們利用 $A\mathbf{x}=\mathbf{b}$ 求出其 normal equation $A^tA\mathbf{x}=A^t\mathbf{b}$ 的解 $\mathbf{x}=\begin{bmatrix} m \\ c \end{bmatrix}$, 這裡的 m,c 就是讓直線 $y=mx+c$ 所估計得的 $\mathbf{b}'$ 會是 $\mathbf{b}$ 在 $\mathrm{Col}(A)$ 的投影, 也就是說會讓 $\|\mathbf{b}-\mathbf{b}'\|$ 的值最小. 所以符合我們希望誤差 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 最小的要求. 也因此這樣所得的直線, 我們稱之為 *least squares line* (最適合直線, 或最小平方直線), 在統計學中稱之為 *line of regression* (迴歸直線).

Example 4.5.4. 考慮二維資料 $(-1,0),(1,1),(2,3)$ 我們要找出此資料的 least square line $y=mx+c$.

原先要解的方程組是

$$\begin{array}{rcl} -1m+c & = & 0 \\ 1m+c & = & 1 \\ 2m+c & = & 3, \end{array}$$

其矩陣表示法為 $\begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$. 故其 normal equations 為

$$\begin{bmatrix} -1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix}=\begin{bmatrix} -1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix},$$

即

$$\begin{array}{rcl} 6m+2c & = & 7 \\ 2m+3c & = & 4, \end{array}$$

解得 least squares solution 為 $m=13/14, c=5/7$ 故此組資料的 least squares line 為

$$y=\frac{13}{14}x+\frac{5}{7}.$$

接著我們探討幾個和 least squares line 有關的性質.

Proposition 4.5.5. 考慮二維資料 $(x_1,y_1),\ldots,(x_n,y_n)$. 令 $y=mx+c$ 為其 *least squares line* 且對於 $i=1,\ldots,n$ 令 $y'_i=mx_i+c$. 則有以下知性質:

(1) $y_1+\cdots+y_n=y'_1+\cdots+y'_n$.

(2) $x_1y_1+\cdots+x_ny_n=x_1y'_1+\cdots+x_ny'_n$.

(3) 令 $\overline{x}$ 和 $\overline{y}$ 分別為 $x_1,\ldots,x_n$ 以及 $y_1,\ldots,y_n$ 的平均數, 即 $\overline{x}=(x_1+\cdots+x_n)/n$, $\overline{y}=(y_1+\cdots+y_n)/n$. 則點 $(\overline{x},\overline{y})$ 在直線 $y=mx+c$ 上, 即 $\overline{y}=m\overline{x}+c$.

Proof. 為方便起見令 $\mathbf{b}=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix},\mathbf{b}'=\begin{bmatrix} y'_1\\ y'_2\\ \vdots\\ y'_n\end{bmatrix}$, 則依 $\mathbf{b}-\mathbf{b}'\in\mathrm{Col}(A)^{\perp}=N(A^{\mathrm{t}})$, 得

$$A^{\mathrm{t}}(\mathbf{b}-\mathbf{b}')=\begin{bmatrix} x_1 & \cdots & x_n\\ 1 & \cdots & 1\end{bmatrix}\begin{bmatrix} y_1-y'_1\\ y_2-y'_2\\ \vdots\\ y_n-y'_n\end{bmatrix}=\begin{bmatrix} x_1(y_1-y'_1)+\cdots+x_n(y_n-y'_n)\\ (y_1-y'_1)+\cdots+(y_n-y'_n)\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}.$$

得證 (1), (2). 又令 $\overline{y'}=(y'_1+\cdots+y'_n)/n$, 由 (1) 知 $\overline{y}=\overline{y'}$. 又由於對所有 $i=1,\ldots,n$, 皆有 $y'_i=mx_i+c$, 故知 $\overline{y'}=m\overline{x}+c$. 得證 (3), 即 $\overline{y}=m\overline{x}+c$. □

最後我們說明一下, 對於二維資料我們不只能談論這資料的最適合直線, 也能談論其他的最適合曲線. 例如有些二維資料我們覺得和拋物線有關, 我們就可以試圖找一條最適合的拋物線 $y=ax^2+bx+c$, 然後代入這些二維資料得到以 a,b,c 為變數的聯立方程組, 再找出此方程組的 least squares solution. 這樣所得的拋物線就會是誤差平方和最小的最適合拋物線了. 至於更高次的多項式也可如法炮製, 我們就不詳述了.

上一節提過 QR decomposition 可以幫忙處理這類的問題. 假設 $A\in\mathrm{M}_{m\times n}(\mathbb{R})$ 且 $\mathrm{rank}(A)=n$, 對於聯立方程組 $A\mathbf{x}=\mathbf{b}$ 要其解 normal equations $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$. 我們可以將 A 寫成 QR decomposition $A=QR$ 來處理. 此時我們要解 $(QR)^{\mathrm{t}}QR\mathbf{x}=(QR)^{\mathrm{t}}\mathbf{b}$, 利用 transpose 和矩陣乘法性質得 $R^{\mathrm{t}}Q^{\mathrm{t}}QR\mathbf{x}=R^{\mathrm{t}}Q^{\mathrm{t}}\mathbf{b}$. 然而 Q 的 column vectors 是 $C(Q)$ 的 orthonormal basis, 前面提過 $Q^{\mathrm{t}}Q$ 會是 identity matrix I_n. 因此我們可以將 normal equations 化簡成 $R^{\mathrm{t}}R\mathbf{x}=R^{\mathrm{t}}Q^{\mathrm{t}}\mathbf{b}$. 再利用 R 為 invertible, 所以 R^{t} 也是 invertible, 兩邊乘上 R^{t} 的 inverse, 我們再將 normal equations 化簡為 $R\mathbf{x}=Q^{\mathrm{t}}\mathbf{b}$. 這個聯立方程組就很好解了, 這是因為 R 為 upper triangular matrix 且對角線位置皆不為 0, 所以 R 本身就是 echelon form, 因此我們可以很快求出解.

接下來我們探討 linear system $A\mathbf{x}=\mathbf{b}$ 有解, 但解不唯一的情形. 在很多情況當 $A\mathbf{x}=\mathbf{b}$ 解不唯一時, 我們希望能找到長度最小的解. 我們有以下的定義.

Definition 4.5.6. 假設 $A\in\mathrm{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b}\in\mathbb{R}^m$. 考慮聯立方程組 $A\mathbf{x}=\mathbf{b}$. 若 $\mathbf{v}\in\mathbb{R}^n$ 滿足 $A\mathbf{v}=\mathbf{b}$ 且對任意 $A\mathbf{x}=\mathbf{b}$ 的一組解 $\mathbf{w}$ 皆滿足 $\|\mathbf{v}\|\leq\|\mathbf{w}\|$, 則稱 $\mathbf{v}$ 為 linear system $A\mathbf{x}=\mathbf{b}$ 的 *minimal solution.*

如何找到 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution 呢? 假設 $\mathbf{v},\mathbf{w}$ 皆為 $A\mathbf{x}=\mathbf{b}$ 的一組解, 則由 $A\mathbf{v}=\mathbf{b}$ 以及 $A\mathbf{w}=\mathbf{b}$ 可得 $A(\mathbf{v}-\mathbf{w})=\mathbf{0}$. 因此若令 $\mathbf{u}=\mathbf{v}-\mathbf{w}$, 則得 $\mathbf{u}\in N(A)$ 且 $\mathbf{w}=\mathbf{v}-\mathbf{u}$. 反之, 當我們已找到 $\mathbf{v}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解時, 對任意 $\mathbf{u}\in N(A)$, 皆有 $A(\mathbf{v}-\mathbf{u})=A\mathbf{v}-A\mathbf{u}=\mathbf{b}-\mathbf{0}=\mathbf{b}$, 亦即 $\mathbf{v}-\mathbf{u}$ 也是 $A\mathbf{x}=\mathbf{b}$ 的一組解. 因此聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解集合可以寫成 $\{\mathbf{v}-\mathbf{u}\mid\mathbf{u}\in N(A)\}$, 其中 $\mathbf{v}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解. 也就是說當我們已找到 $A\mathbf{x}=\mathbf{b}$ 的一組解 $\mathbf{v}$, 我們要找到 $\mathbf{u}\in N(A)$ 使得 $\|\mathbf{v}-\mathbf{u}\|\leq\|\mathbf{v}-\mathbf{u}'\|$, $\forall\mathbf{u}'\in N(A)$, 此時 $\mathbf{v}-\mathbf{u}$ 就會是 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution. 如何在 $N(A)$ 中找到 $\mathbf{u}$ 呢? 由於 $N(A)$ 是 $\mathbb{R}^n$ 的 subspace, Proposition 4.3.15 告訴我們 $\mathbf{u}$ 就是 $\mathbf{v}$ 在 $N(A)$ 的 orthogonal projection. 也因此可得 $\mathbf{v}-\mathbf{u}\in N(A)^{\perp}$. 再由 Theorem 4.4.2, 知

$N(A)^{\perp} = \mathrm{Col}(A^{\mathrm{t}})$, 故知此 minimal solution $\mathbf{v}-\mathbf{u} \in \mathrm{Col}(A^{\mathrm{t}})$. 這表示 $\mathbf{v}-\mathbf{u}$ 會在 $n \times m$ 矩陣 A^{t} 的 column space, 因此存在 $\mathbf{x}' \in \mathbb{R}^m$ 使得 $\mathbf{v}-\mathbf{u} = A^{\mathrm{t}}\mathbf{x}'$. 此時由 $A(\mathbf{v}-\mathbf{u}) = A\mathbf{v}-A\mathbf{u} = \mathbf{b}$, 得 $(AA^{\mathrm{t}})\mathbf{x}' = A(A^{\mathrm{t}}\mathbf{x}') = A(\mathbf{v}-\mathbf{u}) = \mathbf{b}$. 所以我們只要解聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$, 找出的解 $\mathbf{x}' \in \mathbb{R}^m$ 由於會使得 $A^{\mathrm{t}}\mathbf{x}' \in \mathrm{Col}(A^{\mathrm{t}})$ 且符合方程組 $A(A^{\mathrm{t}}\mathbf{x}') = \mathbf{b}$, 因此令 $\mathbf{x} = A^{\mathrm{t}}\mathbf{x}'$ 就會是 linear system $A\mathbf{x} = \mathbf{b}$ 的 minimal solution. 現在問題是 linear system $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$ 會有解嗎?

Proposition 4.5.7. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b} \in \mathbb{R}^m$. 假設聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解, 則聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$ 必有解, 且若 $\mathbf{x}' = \mathbf{w} \in \mathbb{R}^m$ 為其一組解, 則 $\mathbf{x} = A^{\mathrm{t}}\mathbf{w}$ 會是聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的 *minimal solution.*

Proof. 我們知道聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解表示 $\mathbf{b} \in \mathrm{Col}(A)$, 現要證明聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$ 有解, 我們只要證明 $\mathbf{b} \in \mathrm{Col}(AA^{\mathrm{t}})$ 即可. 回顧在 Proposition 3.7.16 (1) 當 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$, $B \in \mathsf{M}_{n\times l}(\mathbb{R})$, 則 $\mathrm{Col}(AB) \subseteq \mathrm{Col}(A)$. 因此由 $\mathrm{Col}(AA^{\mathrm{t}}) \subseteq \mathrm{Col}(A)$ 以及 $\mathrm{rank}(AA^{\mathrm{t}}) = \mathrm{rank}(A^{\mathrm{t}}) = \mathrm{rank}(A)$ (Corollary 4.4.4) 知 $\mathrm{Col}(AA^{\mathrm{t}}) = \mathrm{Col}(A)$. 因此 $\mathbf{v} \in \mathrm{Col}(AA^{\mathrm{t}})$, 得證聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$ 有解. 現若 $\mathbf{w} \in \mathbb{R}^m$ 是 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$ 的一解, 即 $(AA^{\mathrm{t}})\mathbf{w} = \mathbf{b}$, 則由 $(AA^{\mathrm{t}})\mathbf{w} = A(A^{\mathrm{t}}\mathbf{w})$ 知 $A^{\mathrm{t}}\mathbf{w}$ 為 $A\mathbf{x} = \mathbf{b}$ 的一解. 由前面的討論我們知道既然 $A^{\mathrm{t}}\mathbf{w}$ 為 $A\mathbf{x} = \mathbf{b}$ 的一解, 則聯立方程組 $A\mathbf{x} = \mathbf{b}$ 所有的解應為 $A^{\mathrm{t}}\mathbf{w}+\mathbf{u}$, 其中 $\mathbf{u} \in N(A)$ 這樣的形式. 此時

$$\|A^{\mathrm{t}}\mathbf{w}+\mathbf{u}\|^2 = \langle A^{\mathrm{t}}\mathbf{w}+\mathbf{u}, A^{\mathrm{t}}\mathbf{w}+\mathbf{u}\rangle = \langle A^{\mathrm{t}}\mathbf{w}, A^{\mathrm{t}}\mathbf{w}\rangle + 2\langle A^{\mathrm{t}}\mathbf{w}, \mathbf{u}\rangle + \langle \mathbf{u}, \mathbf{u}\rangle = \|A^{\mathrm{t}}\mathbf{w}\|^2 + \|\mathbf{u}\|^2.$$

這裡 $\langle A^{\mathrm{t}}\mathbf{w}, \mathbf{u}\rangle = 0$, 因為 $A^{\mathrm{t}}\mathbf{w} \in \mathrm{Col}(A^{\mathrm{t}}) = N(A)^{\perp}$. 由此可知當 $\mathbf{u} \in N(A)$ 且 $\mathbf{u} \neq \mathbf{0}$ 時 $\|A^{\mathrm{t}}\mathbf{w}+\mathbf{u}\| > \|A^{\mathrm{t}}\mathbf{w}\|$, 故 $\mathbf{x} = A^{\mathrm{t}}\mathbf{w}$ 會是聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的 minimal solution. □

在 Proposition 4.5.7 中因為 A 為 $m \times n$ matrix, 所以聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的解是在 $\mathbb{R}^n$ 中, 而要求其 minimal solution 我們要解的聯立方程組為 $AA^{\mathrm{t}}\mathbf{x}' = \mathbf{b}$ 其解是在 $\mathbb{R}^m$ 中, 而要得到 minimal solution 必須將 $AA^{\mathrm{t}}\mathbf{x}' = \mathbf{b}$ 的解左邊再乘上 A^{t}, 因此知 minimal solution 會在 A^{t} 的 column space, 也就是 A 的 row space.

Corollary 4.5.8. 假設 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$, $\mathbf{b} \in \mathbb{R}^m$ 且聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解. 則在 A 的 *row space* 中存在唯一的向量 $\mathbf{v}$ 滿足 $A\mathbf{v} = \mathbf{b}$, 且 $\mathbf{v}$ 為 $A\mathbf{x} = \mathbf{b}$ 的 *minimal solution.*

Proof. 我們已知 $A\mathbf{x} = \mathbf{b}$ 的 minimal solution 在 $\mathrm{Col}(A^{\mathrm{t}})$ 中, 亦即在 A 的 row space 中. 因此僅剩下要證明唯一性, 亦即假設 $\mathbf{v}, \mathbf{v}' \in \mathrm{Col}(A^{\mathrm{t}})$ 且 $A\mathbf{v} = \mathbf{b}$, $A\mathbf{v}' = \mathbf{b}$, 我們要證明 $\mathbf{v} = \mathbf{v}'$. 由 $A\mathbf{v} = A\mathbf{v}' = \mathbf{b}$ 可得 $A(\mathbf{v}-\mathbf{v}') = \mathbf{0}$, 亦即 $\mathbf{v}-\mathbf{v}' \in N(A)$. 又 $\mathrm{Col}(A^{\mathrm{t}})$ 為 $\mathbb{R}^n$ 的 subspace, 故 $\mathbf{v}-\mathbf{v}' \in \mathrm{Col}(A^{\mathrm{t}})$. 因此得 $\mathbf{v}-\mathbf{v}' \in N(A) \cap \mathrm{Col}(A^{\mathrm{t}})$. 然而 $\mathrm{Col}(A^{\mathrm{t}}) = N(A)^{\perp}$ (Theorem 4.4.2) 故得 $N(A) \cap \mathrm{Col}(A^{\mathrm{t}}) = N(A) \cap N(A)^{\perp} = \{\mathbf{0}\}$ (Lemma 4.3.4), 因此得證 $\mathbf{v} = \mathbf{v}'$. □

Example 4.5.9. 我們用一個簡單的例子說明 Proposition 4.5.7 以及 Corollary 4.5.8. 考慮 $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 0 & 1 & 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}$. 很容易看出此 linear system $A\mathbf{x} = \mathbf{b}$ 有一組解 $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$ 由於 $\mathrm{rank}(A) = 2$, 我們知到此方程組有無窮多解. 事實上此 linear system $A\mathbf{x} = \mathbf{b}$ 有另一組解 $\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$. 我們有 $\|\mathbf{v}_2\| = \sqrt{3} < \|\mathbf{v}_1\| = 2$, 還有沒有長度更小的解呢? 我們要找到其

minimal solution. 由於 $AA^{\mathrm{t}}=\begin{bmatrix}2&3&1\\3&6&3\\1&3&2\end{bmatrix}$, 我們解 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$, 即 $\begin{bmatrix}2&3&1\\3&6&3\\1&3&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}2\\2\\0\end{bmatrix}$ 得 $\mathbf{x}'=\begin{bmatrix}2\\-2/3\\0\end{bmatrix}$ 為其一組解, 故知 $\mathbf{x}=A^{\mathrm{t}}\mathbf{x}'=\begin{bmatrix}1&1&0\\0&1&1\\1&2&1\end{bmatrix}\begin{bmatrix}2\\-2/3\\0\end{bmatrix}=\begin{bmatrix}4/3\\-2/3\\2/3\end{bmatrix}$ 為原方程組 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution. 事實上 $N(A)=\mathrm{Span}(\begin{bmatrix}1\\1\\-1\end{bmatrix})$. 所以 $A\mathbf{x}=\mathbf{b}$ 所有的解為 $\mathbf{v}+r\mathbf{u}$ 的形式, 其中 $\mathbf{v}=\begin{bmatrix}4/3\\-2/3\\2/3\end{bmatrix},\mathbf{u}=\begin{bmatrix}1\\1\\-1\end{bmatrix}$, $r\in\mathbb{R}$. 由於 $\langle\mathbf{v},\mathbf{u}\rangle=0$, 我們有 $\|\mathbf{v}+r\mathbf{u}\|^2=\langle\mathbf{v}+r\mathbf{u},\mathbf{v}+r\mathbf{u}\rangle=\|\mathbf{v}\|^2+r^2\|\mathbf{u}\|^2=(24/9)+3r^2$. 知 $\mathbf{v}$ 確為 linear system $A\mathbf{x}=\mathbf{b}$ 之 minimal solution.