大學線性代數初步

李華介

國立台灣師範大學數學系

前言

本講義主要目的是針對數學系大一學生介紹有關線性代數基本的理論, 大致上僅談論實係數的向量空間.

一般經驗上大一學生會覺得線性代數學習起來會比微積分吃力, 主要是因素可能是在高中時期學習有關線性代數部分主要著重於操作而較少論證. 這一點在大學線性代數中就較不同了. 不過經驗上也告訴我們當學生到大三大四時 (因準備研究所考試) 再重新溫習線性代數時, 會覺得它不再那麼難以親近. 由此可知大學時期的線性代數其理論並不難懂, 較大的障礙是要開始學習數學的論證. 這個障礙或許到大三大四一些數學思維較成熟時稍可解除, 不過對一些同學來說可能為時已晚, 畢竟線性代數的理論或概念與其他課程都脫不了關係. 基於這個原因寫下這份講義, 希望藉由較平易近人的方式介紹線性代數也慢慢引導熟悉數學的論證方式. 本講義希望以淺顯易懂為主旨, 而不是生動有趣. 畢竟有些事情要說明清楚就會顯得囉唆, 當然就不有趣了.

研讀本講義的同學要有心理準備, 本講義是針對數學系學生而寫, 自然偏重於整個線性代數的理論架構. 對於線性代數在其他領域的應用幾乎沒有著墨. 我們依循一貫的原則就是理論清楚了, 接下來的應用或推廣就不難了. 所以對應用有興趣的同學應再參考其他的參考書籍. 另外本講義並未提供習題, 不過在某些概念講述之後有時會提供一些問題 (Question). 這些問題幾乎是檢視觀念是否正確或是對內容是否了解, 大部分問題若觀念已清楚應可以立即回答. 所以這些問題的份量仍不及一般習題, 對熟習線性代數所給予的訓練. 針對這一點, 請欲學習好線性代數的同學務必參閱一般線性代數書籍的習題, 自行磨練.

本講義雖然主要以中文撰寫, 不過當涉及定義或專有名詞時, 為免翻譯的困擾將以英文取代. 因此將以中英夾雜較不傳統的方式顯現, 若有不便請見諒.

本講義編寫費時, 編寫完後並未經過嚴謹的校對. 疏漏在所難免, 雖不至於有理論性上嚴重的錯誤, 但讀者仍應注意不宜概括全收. 若發現錯誤, 歡迎提出寶貴的意見.

Part I

Vectors, Systems of Linear Equations, Matrices

在高中階段大家接受到與線性代數相關的數學, 有向量 (vectors), 一次聯立方程式 (systems of linear equations) 以及矩陣 (matrices). 在本講義的第一部份, 我們想利用這三個大家較為熟悉的題材慢慢引入線性代數的概念. 在此要強調一下, 千萬不要以為這三種題材在高中沒學好, 線性代數就學不好. 我們僅是利用這些大家可能較熟悉的題材當成動機, 讓大家剛接受到抽象的線性代數理論時不會絕得太突兀摸不著頭緒.

Chapter 1

Vectors in $\mathbb{R}^n$

我們藉由大家熟悉的向量來介紹向量間運算的性質. 要注意我們不再去定義何謂向量, 而是著重於如何用這些已知的向量性質, 利用較抽象的方法來推導出一些幾何的性質.

1.1. 坐標平面中的向量

本節針對對抽象數學論述不熟悉的同學, 想利用大家熟悉坐標平面的向量慢慢引導進入狀況. 也因此本節在細節的說明特別繁瑣. 若對坐標平面的向量相當清楚的同學, 可放心略過此節.

在坐標平面中的向量, 我們都可用 (a,b) 來表示, 其中 $a,b \in \mathbb{R}$ (我們用 $\mathbb{R}$ 來表示所有實數所成的集合, 所以 $a,b \in \mathbb{R}$ 表示 a,b 屬於實數, 也就是說 a,b 皆為實數). 意思就是說如果你在坐標平面中任給一點 P, 然後從 P 點開始往水平方向走 a 單位 ($a > 0$ 時往右; $a < 0$ 時往左), 再沿鉛直方向走 b 單位 ($b > 0$ 時往上; $b < 0$ 時往下), 最後到達的點若記為 Q. 那麼從 P 點開始到 Q 點為止的這一個向量就可用 (a,b) 來表示, 記為 $\overrightarrow{PQ} = (a,b)$.

用坐標來表示一個向量 (即用 (a,b) 這種方法) 有許多好處, 例如大家很容易理解: 當兩個向量 (a,b) 和 (c,d) 相等時 (即 $(a,b) = (c,d)$), 這表示 $a = c$ 且 $b = d$; 從這觀點, 若用點來表示向量時就較麻煩, 因為如果 P,P',Q,Q' 為平面中四個點 $\overrightarrow{PQ} = \overrightarrow{P'Q'}$ 並不代表 P 和 P' 為同一點且 Q 和 Q' 為同一點. 不過若已知 $P = P'$, 則可得 $Q = Q'$. 反之若 $Q = Q'$ 則可得 $P = P'$. (我們說明一下這裡符號的使用: 因為 P,P' 皆為 “點” 而不是 “數”, 所以這裡 $P = P'$ 這個等號表示同一 “點” 而不是同一 “數”). 上面這個論述可用剛才定義 $\overrightarrow{PQ}$ 的方法驗證, 這裡就不再驗證. 基於符號的方便性, 當我們要用符號來表示一個向量時, 除非已知此向量為特定兩點所決定的向量, 通常會僅用 $\mathbf{u},\mathbf{v}$ 這類的粗體字符號來表示. 一般來說我們用 $\mathbb{R}^2$ 來表示坐標平面上的向量所成的集合, 所以若我們說 $\mathbf{v} \in \mathbb{R}^2$, 就表示 $\mathbf{v}$ 是坐標平面上的一個向量, 也就是說可以找到 $a,b \in \mathbb{R}$ 使得 $\mathbf{v} = (a,b)$.

坐標表示法的另一個好處是很容易幫助我們定義向量的加法 (*addition*) 以及係數積 (*scalar multiplication*).

Definition 1.1.1. 令 $\mathbf{u}=(a_1,a_2),\mathbf{v}=(b_1,b_2)\in\mathbb{R}^2$ 以及 $r\in\mathbb{R}$. 我們定義

$$\mathbf{u}+\mathbf{v}=(a_1+b_1,a_2+b_2)\quad\text{and}\quad r\mathbf{u}=(ra_1,ra_2).$$

要注意這裡所謂的定義 (即 definition) 指的是規定, 也就是說我們規定向量必須這樣相加及乘以常數. 當然了你也可以自行規定一套向量加法的規則, 不過一般在數學上的定義都有其必要性. 通常一個定義對理論推導或實用上都會有相當的幫助. 例如這一個定義與我們直觀上認為若 P,Q,R 為坐標平面上三點, 則 $\overrightarrow{PQ}+\overrightarrow{QR}=\overrightarrow{PR}$ 相吻合. 這是因為: 假設 P,Q,R 三點的坐標分別為 $P(x_1,y_1),Q(x_2,y_2),R(x_3,y_3)$ (注意: 為了區分點和向量的坐標表示法, 當我們提到點的坐標時都會加上該點的代號), 則

$$\overrightarrow{PQ}=(x_2-x_1,y_2-y_1),\overrightarrow{QR}=(x_3-x_2,y_3-y_2),\overrightarrow{PR}=(x_3-x_1,y_3-y_1).$$

然而依前面向量加法的定義確實

$$\overrightarrow{PQ}+\overrightarrow{QR}=((x_2-x_1)+(x_3-x_2),(y_2-y_1)+(y_3-y_2))=(x_3-x_1,y_3-y_1).$$

故得

$$\overrightarrow{PQ}+\overrightarrow{QR}=\overrightarrow{PR}. \tag{1.1}$$

一個定義一定要清楚明確, 例如若 $\mathbf{u}=(1,2),\mathbf{v}=(3,4)$ 且 $r=5$, 由 Definition 1.1.1 我們知 $\mathbf{u}+\mathbf{v}=(1+3,2+4)=(4,6)$ 且 $r\mathbf{u}=(5\times1,5\times2)=(5,10)$. 不過我們絕對不能用: 若 $\mathbf{u}=(1,2),\mathbf{v}=(3,4)$ 且 $r=5$, 則定義 $\mathbf{u}+\mathbf{v}=(1+3,2+4)=(4,6)$ 且 $r\mathbf{u}=(5\times1,5\times2)=(5,10)$ 這樣的說法來定義向量的加法及係數積. 因為這樣的說法只是定義出兩個特殊的向量的加法以及和一特殊實數的乘積, 不能依這特例要求別人 "依此類推". 這就是當初定義時用 a_1,a_2,b_1,b_2,r 這些符號來代替具體數字的用意. 另外要注意在 Definition 1.1.1 中我們沒有提 $r\mathbf{v}$ 的定義, 不過當初既然 $\mathbf{u}$ 已表成任意 $\mathbb{R}^2$ 上的向量, 定義 $r\mathbf{u}$ 已足夠, 所以不必再去定 $r\mathbf{v}$. 最後要強調的是: 這裡我們並不是定義兩向量的乘法, 而是定義實數和向量相乘, 所以我們稱為係數積 (scalar multiplication) 而不能說是向量的乘法.

有了定義之後, 我們就需依定義處理相關問題, 但通常直接依定義處理較繁複, 我們可依定義推導出一些性質, 利用這些性質簡化處理程序. 例如在微積分, 我們定義出一個函數在某一點的極限後, 若每次都得依定義處理極限問題論證起來很複雜; 但當我們利用定義推導出一些極限的性質後, 用這些性質處理極限問題就簡單方便多了. 所以在定義之後我們會有一些定理 (Proposition 或 Theorem) 來論證一些依定義可得的性質, 以方便我們處理更進一步的問題. 以下就是要談向量加法及係數積有關的性質.

Proposition 1.1.2. 對於 $\mathbb{R}^2$ 上的向量, 我們有以下的性質:

(1) 對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^2$, 皆有 $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

(2) 對任意 $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^2$, 皆有 $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

(3) 存在一向量 $\mathbf{O}\in\mathbb{R}^2$ 滿足對任意 $\mathbf{u}\in\mathbb{R}^2$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$.

(4) 對任意 $\mathbf{u}\in\mathbb{R}^2$ 皆可找到 $\mathbf{u}'\in\mathbb{R}^2$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$.

(5) 對任意 $r,s\in\mathbb{R}$ 以及 $\mathbf{u}\in\mathbb{R}^2$, 皆有 $r(s\mathbf{u})=(rs)\mathbf{u}$.

(6) 對任意 $r,s \in \mathbb{R}$ 以及 $\mathbf{u} \in \mathbb{R}^2$, 皆有 $(r+s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$.

(7) 對任意 $r \in \mathbb{R}$ 以及 $\mathbf{u},\mathbf{v} \in \mathbb{R}^2$ 皆有 $r(\mathbf{u}+\mathbf{v}) = r\mathbf{u} + r\mathbf{v}$.

(8) 對任意 $\mathbf{u} \in \mathbb{R}^2$, 皆有 $1\mathbf{u} = \mathbf{u}$.

通常一個定理敘述完就要證明, 不過在這裡我們建議先緩一緩. 對同學來說了解定理說些什麼比起證明來得重要. 在這裡我們就一一說明一下這個定理說些什麼.

(1) 敘述的是所謂向量加法的交換性. 它告訴我們在處理向量加法時可以依方便交換順序. 或許同學覺得這個很自然為何還要證明. 事實上只要是定義未提的事情都要證明, 不能因為覺得自然而不去處理. 這裡大家會覺得自然是因為大家對實數的加法運算很清楚, 不過數學上是存在許多"抽象"的數系它的加法是不能交換的. 所以經由證明不只讓我們確認事情是對的, 也能幫助我們釐清事情是對的其背後的主要因素.

(2) 說的就是所謂的結合律, 它依然是因為實數加法的性質而成立. 這裡 $(\mathbf{u}+\mathbf{v})+\mathbf{w}$ 是說先將 $\mathbf{u}$ 和 $\mathbf{v}$ 相加後所得的向量再和 $\mathbf{w}$ 相加. 這樣所得的向量和先將 $\mathbf{v}$ 和 $\mathbf{w}$ 相加後再和 $\mathbf{u}$ 相加會是同樣的向量. 這裡雖然也是談向量加法的順序問題, 不過和 (1) 所談的順序是兩回事, 大家應該要分清楚.

(3) 談的就是所謂的零向量, 零向量的特點就是加上任何向量都不動. 為什麼要特別談零向量的存在性? 這就好比在實數上若沒有零的概念就沒有減法一樣, 在向量的運算上是相當重要的. 尤其以後要用抽象的方式談向量系統時零向量的存在性更不容忽視.

(4) 談的就是所謂的反向量, 要注意需有零向量的存在才能談反向量. 而且要區分清楚這裡的敘述是給了 $\mathbf{u}$ 後可找到 $\mathbf{u}'$ 使得 $\mathbf{u}+\mathbf{u}' = \mathbf{O}$. 這裡 $\mathbf{u}'$ 是會隨著 $\mathbf{u}$ 而改變, 而不是一個固定的向量和所有的向量加起來會是零向量. 數學的敘述要弄清楚否則差之毫釐失之千里.

(5),(6),(7) 談的是係數積的性質, 例如 $r(s\mathbf{u})$ 表示是先將 $\mathbf{u}$ 乘上 s 倍後所得的向量再乘上 r. 這幾個性質也都和實數乘法性質息息相關, 雖然看起來不顯眼但在處理向量的運算時非常重要.

(8) 指的是所有向量乘上 1 後仍不動. 這裡特別提出來其實和零向量意義很像, 唯有 1 的引入以後才能談係數的除法. 例如已知 $2\mathbf{u} = \mathbf{v}$, 就可利用 (5) 的性質兩邊乘上 $1/2$, 得

$$\mathbf{u} = 1\mathbf{u} = \frac{1}{2}(2\mathbf{u}) = \frac{1}{2}\mathbf{v}.$$

最後要強調一下: 這裡將這些性質列出, 並不是要求大家將這幾個性質背下來. 一來我們希望大家知道有些性質不能覺得理所當然就不去證明, 另一方面也讓大家知道以後在處理向量運算時可以放心且自然的使用這幾個性質. 現在我們就來看看這些證明.

Proof. (of Proposition 1.1.2) 這幾個性質其實很簡單, 我們寫下證明是希望不熟悉寫抽象證明的同學利用這個簡單的證明學學看如何寫好證明. 若自覺對這些性質的證明清楚的同學可跳過此證明.

(1) 假設 $\mathbf{u} = (a_1,a_2), \mathbf{v} = (b_1,b_2)$, 則依定義知

$$\mathbf{u}+\mathbf{v} = (a_1+b_1, a_2+b_2), \mathbf{v}+\mathbf{u} = (b_1+a_1, b_2+a_2).$$

故由實數加法的交換性 (即 $a+b=b+a$) 可得 $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

(2) 假設 $\mathbf{u}=(a_1,a_2),\mathbf{v}=(b_1,b_2),\mathbf{w}=(c_1,c_2)$, 則依定義知 $\mathbf{u}+\mathbf{v}=(a_1+b_1,a_2+b_2)$ 故得

$$(\mathbf{u}+\mathbf{v})+\mathbf{w}=((a_1+b_1)+c_1,(a_2+b_2)+c_2).$$

同理由 $\mathbf{v}+\mathbf{w}=(b_1+c_1,b_2+c_2)$ 可得

$$\mathbf{u}+(\mathbf{v}+\mathbf{w})=(a_1+(b_1+c_1),a_2+(b_2+c_2)).$$

因此由實數加法的結合律 (即 $(a+b)+c=a+(b+c)$) 得知 $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{v}+(\mathbf{u}+\mathbf{w})$.

(3) 這是一個存在性的問題, 也就是說要找到一個向量 $\mathbf{O}$ 滿足所求. 這裡我們只要令 $\mathbf{O}=(0,0)$, 則對任意向量 $\mathbf{u}=(a_1,a_2)$ 皆有

$$\mathbf{O}+\mathbf{u}=(0+a_1,0+a_2)=(a_1,a_2)=\mathbf{u}.$$

故知確實存在這樣的向量.

(4) 上面我們已知可令 $\mathbf{O}=(0,0)$, 故此時對任意 $\mathbf{u}=(a_1,a_2)$ 我們只要考慮 $\mathbf{u}'=(-a_1,-a_2)$, 則可得

$$\mathbf{u}+\mathbf{u}'=(a_1+(-a_1),a_2+(-a_2))=(0,0)=\mathbf{O}.$$

(5) 假設 $\mathbf{u}=(a_1,a_2)$, 依定義知 $s\mathbf{u}=(sa_1,sa_2)$, 因此得

$$r(s\mathbf{u})=(r(sa_1),r(sa_2)).$$

另一方面

$$(rs)\mathbf{u}=((rs)a_1,(rs)a_2),$$

故由實數乘法結合律 (即 $r(sa)=(rs)a$) 可得 $r(s\mathbf{u})=(rs)\mathbf{u}$.

(6) 假設 $\mathbf{u}=(a_1,a_2)$, 依定義知

$$(r+s)\mathbf{u}=((r+s)a_1,(r+s)a_2).$$

另一方面 $r\mathbf{u}=(ra_1,ra_2),s\mathbf{u}=(sa_1,sa_2)$, 可得

$$r\mathbf{u}+s\mathbf{u}=(ra_1+sa_1,ra_2+sa_2),$$

故由實數加法與乘法的分配律 (即 $(r+s)a=ra+sa$) 可得 $(r+s)\mathbf{u}=r\mathbf{u}+s\mathbf{u}$.

(7) 假設 $\mathbf{u}=(a_1,a_2),\mathbf{v}=(b_1,b_2)$, 依定義知

$$r(\mathbf{u}+\mathbf{v})=(r(a_1+b_1),r(a_2+b_2)).$$

另一方面 $r\mathbf{u}=(ra_1,ra_2),r\mathbf{v}=(rb_1,rb_2)$, 可得

$$r\mathbf{u}+r\mathbf{v}=(ra_1+rb_1,ra_2+rb_2),$$

故由實數加法與乘法的分配律 (即 $r(a+b)=ra+rb$) 可得 $r(\mathbf{u}+\mathbf{v})=r\mathbf{u}+r\mathbf{v}$.

(8) 假設 $\mathbf{u}=(a_1,a_2)$, 由於對任意實數 a 皆有 $1a=a$, 故由 $1\mathbf{u}=(1a_1,1a_2)$ 可得 $1\mathbf{u}=\mathbf{u}$.

□

Question 1.1. 利用 $\mathbb{R}^2$ 向量加法的定義, 試證明以下性質:

(1) $\mathbf{O}=(0,0)$ 是 $\mathbb{R}^2$ 中唯一的向量滿足對任意 $\mathbf{u}\in\mathbb{R}^2$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$.

(2) 給定 $\mathbf{u}=(a,b)\in\mathbb{R}^2$, 試證明 $\mathbf{u}'=(-a,-b)$ 是 $\mathbb{R}^2$ 中唯一的向量滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$.

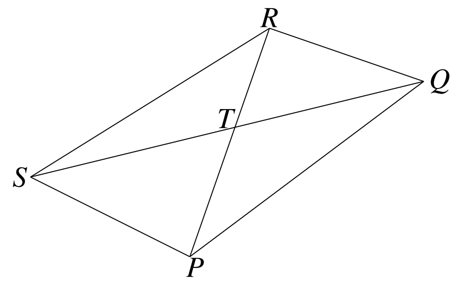
接下來我們將舉一個例子來說明我們可以用向量的性質 (Proposition 1.1.2) 來處理一些幾何的問題. 由於這些性質以後我們不會用到, 所以在此不用定理的形式呈現而是用例題的形式呈現.

Example 1.1.3. 假設 P,Q,R,S 是平面中四個點, 其中 $\overrightarrow{PQ}=\overrightarrow{SR}$, 則我們有以下性質:

(1) $\overrightarrow{QR}=\overrightarrow{PS}$.

(2) 假設 T 為線段 $\overline{PR}$ 和線段 $\overline{SQ}$ 的交點, 則 $\overrightarrow{PT}=\overrightarrow{TR}$ 且 $\overrightarrow{ST}=\overrightarrow{TQ}$.

我們利用下圖來解釋這兩個性質會成立的原因:



(1) 我們要利用 $\overrightarrow{PQ}=\overrightarrow{SR}$ 來得到 $\overrightarrow{QR}=\overrightarrow{PS}$. 由前面式子 (1.1) 我們知道 $\overrightarrow{PQ}+\overrightarrow{QR}=\overrightarrow{PR}$ 且 $\overrightarrow{PS}+\overrightarrow{SR}=\overrightarrow{PR}$. 也就是說

$$\overrightarrow{PQ}+\overrightarrow{QR}=\overrightarrow{PS}+\overrightarrow{SR}. \tag{1.2}$$

由 Proposition 1.1.2 (4), 我們知存在 $\mathbf{u}$ 使得 $\overrightarrow{PQ}+\mathbf{u}=\mathbf{O}$. 要注意由於 Proposition 1.1.2 (1) 以及 $\overrightarrow{PQ}=\overrightarrow{SR}$ 的假設, 這等於

$$\overrightarrow{PQ}+\mathbf{u}=\mathbf{u}+\overrightarrow{PQ}=\overrightarrow{SR}+\mathbf{u}=\mathbf{u}+\overrightarrow{SR}=\mathbf{O}$$

現將上面式子 (1.2) 兩邊加上 $\mathbf{u}$ 可得

$$\mathbf{u}+(\overrightarrow{PQ}+\overrightarrow{QR})=\mathbf{u}+(\overrightarrow{PS}+\overrightarrow{SR}) \tag{1.3}$$

然而利用 Proposition 1.1.2 (2), (3) 我們有

$$\mathbf{u}+(\overrightarrow{PQ}+\overrightarrow{QR})=(\mathbf{u}+\overrightarrow{PQ})+\overrightarrow{QR}=\mathbf{O}+\overrightarrow{QR}=\overrightarrow{QR}.$$

同理可得

$$\mathbf{u}+(\overrightarrow{PS}+\overrightarrow{SR})=(\overrightarrow{PS}+\overrightarrow{SR})+\mathbf{u}=\overrightarrow{PS}+(\overrightarrow{SR}+\mathbf{u})=\overrightarrow{PS}+\mathbf{O}=\overrightarrow{PS}.$$

故由式子 (1.3) 可得 $\overrightarrow{QR}=\overrightarrow{PS}$.

(2) 通常我們用坐標來處理問題時, 碰到交點的問題可以用解方程式的方法處理, 不過這裡我們要用抽象的向量處理, 遇到交點問題就比較麻煩, 需要用比較特殊的方法處理. 這裡原要證明線段 $\overline{PR}$ 和線段 $\overline{SQ}$ 的交點 T 會是 $\overline{PR}$ 的中點, 也會是 $\overline{SQ}$ 的中點. 我們用反過來的看法處理, 也就是要去說明 $\overline{PR}$ 的中點和 $\overline{SQ}$ 的中點會是同

一點. 如此一來這一點因同時在 $\overline{PR}$ 和 $\overline{SQ}$ 上, 所以自然就是 $\overline{PR}$ 和 $\overline{SQ}$ 的交點. 又因為 $\overline{PR}$ 和 $\overline{SQ}$ 不平行, 僅能有一個交點, 這就說明了線段 $\overline{PR}$ 和線段 $\overline{SQ}$ 的交點就是 $\overline{PR}$ 以及 $\overline{SQ}$ 的中點. 具體來說我們假設 T',T'' 分別為 $\overline{PR}$ 和 $\overline{SQ}$ 的中點. 然後說明 $T'=T''$, 如此一來便得證 T' 為 $\overline{PR}$ 和 $\overline{SQ}$ 的交點, 也就是說 $T=T'=T''$, 故得證所求.

依假設 T',T'' 分別為 $\overline{PR}$ 和 $\overline{SQ}$ 的中點, 故知
$$\overrightarrow{PT'}=\overrightarrow{T'R}=\frac{1}{2}\overrightarrow{PR} \tag{1.4}$$
且
$$\overrightarrow{ST''}=\overrightarrow{T''Q}=\frac{1}{2}\overrightarrow{SQ}. \tag{1.5}$$
若我們能證明 $\overrightarrow{PT'}=\overrightarrow{PT''}$ 便得證 $T'=T''$. 然而 $\overrightarrow{PT''}=\overrightarrow{PS}+\overrightarrow{ST''}$, 故由式子 (1.5) 可得
$$\overrightarrow{PT''}=\overrightarrow{PS}+\overrightarrow{ST''}=\overrightarrow{PS}+\frac{1}{2}\overrightarrow{SQ}.$$
又由於 (1) 告訴我們 $\overrightarrow{PS}=\overrightarrow{QR}$ 上式可改寫為
$$\overrightarrow{PT''}=\overrightarrow{QR}+\frac{1}{2}\overrightarrow{SQ}=(\frac{1}{2}\overrightarrow{QR}+\frac{1}{2}\overrightarrow{QR})+\frac{1}{2}\overrightarrow{SQ}=\frac{1}{2}\overrightarrow{QR}+\frac{1}{2}(\overrightarrow{QR}+\overrightarrow{SQ})=\frac{1}{2}(\overrightarrow{QR}+\overrightarrow{SR})$$
再利用已知 $\overrightarrow{PQ}=\overrightarrow{SR}$, 上式又可改寫為
$$\overrightarrow{PT''}=\frac{1}{2}(\overrightarrow{QR}+\overrightarrow{SR})=\frac{1}{2}(\overrightarrow{QR}+\overrightarrow{PQ})=\frac{1}{2}\overrightarrow{PR}$$
故由式子 (1.4) 得證 $\overrightarrow{PT'}=\overrightarrow{PT''}$, 也因此得知 $T'=T''$ 就是 $\overline{PR}$ 和 $\overline{SQ}$ 的交點, 也就是說 $T=T'=T''$. 故由式子 (1.4,1.5) 得證 $\overrightarrow{PT}=\overrightarrow{TR}$ 且 $\overrightarrow{ST}=\overrightarrow{TQ}$.

同學或許會想到用設定坐標的方式處理 Example 1.1.3 的問題. 這裡我們故意不去設定坐標系, 而僅用 Proposition 1.1.2 中列出的向量運算性質去處理, 主要就是要強調這幾個運算性質就足以處理向量有關的性質. 也就是說不一定需要架設坐標系, 只要符合 Proposition 1.1.2 中列出的運算性質, 都可享有 Example 1.1.3 的性質. 另外由 Example 1.1.3 的處理過程中我們了解到, Proposition 1.1.2 的運算性質可幫助我們在處理向量有關的等式運算時可如處理實數的等式運算一樣 (例如移項, 消去... 等). 這些等我們以後更進一步談向量空間時, 大家就更能體會了.

1.2. $\mathbb{R}^n$ 中的向量

在高中時我們也學坐標空間中的向量, 也就是 $\mathbb{R}^3$. 很容易理解 $\mathbb{R}^3$ 中的向量可以說是 $\mathbb{R}^2$ 中的向量的推廣. 同樣的我們也可將之推廣而定義 $\mathbb{R}^n$ 中的向量, 其中 $n\in\mathbb{N}$ 是任意的正整數. 本節中我們將探討 $\mathbb{R}^n$ 中的向量. 或許同學們會覺得 $\mathbb{R}^n$ 中的向量已看不到, 而疑惑為何要探討它. 事實上線性代數的應用有許多情況就是在這類抽象且看不到的狀況, 這甚至可以說是線性代數發展的主因 (如果僅為了 $\mathbb{R}^2$ 或 $\mathbb{R}^3$ 就沒必要發展這一套理論). 在這一節中我們會發現, 雖然它看不到, 但因為具有如 $\mathbb{R}^2$ 中向量的運算性質, 我們還是可以如處理 $\mathbb{R}^2$ 中的向量一般的方式處理它們相關的性質.

$\mathbb{R}^3$ 中的向量, 即坐標空間中的向量我們可以用 (a_1,a_2,a_3) 其中 $a_1,a_2,a_3 \in \mathbb{R}$ 來表示, 而且我們定義兩向量 $(a_1,a_2,a_3),(b_1,b_2,b_3)$ 相等表示 $a_1=b_1,a_2=b_2$ 且 $a_3=b_3$. 對於任意的 $n\in\mathbb{N}$ 我們也有如下的定義:

Definition 1.2.1. 給定任意 $n\in\mathbb{N}$, 我們定義 $\mathbb{R}^n$ 中的向量為 $(a_1,\ldots,a_n)$, 其中 $a_1,\ldots,a_n\in\mathbb{R}$. 我們說兩向量 $(a_1,\ldots,a_n)$ 以及 $(b_1,\ldots,b_n)$ 相等若且唯若 $a_1=b_1,\ldots,a_n=b_n$.

我們多說明一下符號的表示法, 這裡 $(a_1,\ldots,a_n)$ 表示有 n 個位置, 每個位置我們依次填入實數其中第 1 個位置的元素我們用 a_1 來表示, 第二個位置用 a_2 來表示, 這樣一直下去直到第 n 個位置用 a_n 來表示. 例如 $n=4$ 時, $\mathbb{R}^4$ 中的向量可以用 (a_1,a_2,a_3,a_4) 來表示. 因為 n 可以是任意的正整數不能如 $n=4$ 時將所有的 $a_1,a_2,\ldots$ 都列出來, 所以我們就用 $(a_1,\ldots,a_n)$ 來表示. 這裡我們也沿習這樣省略的方法說 $a_1,\ldots,a_n\in\mathbb{R}$, 表示 a_1 到 a_n 這 n 個元素都是實數. 有的書用較嚴謹的說法, 會用 $a_i\in\mathbb{R}, \forall 1\leq i\leq n$ 來表示, 此即對所有的 $1\leq i\leq n$ 皆有 $a_i\in\mathbb{R}$ 的意思. 這種說法的意義是: i 是任意 1 到 n 的整數, 而對於這個 i 所對應的 a_i 會是實數. 所以這等同於說 a_1 到 a_n 這 n 個數都是實數. 另一方面談論向量的相等是必要的, 這是因為在談論向量的運算時就如同實數的運算, 我們必須明確規定等式的意義. 實數的相等很明確, 所以我們就利用實數的相等來定義向量的相等. 這裡 $a_1=b_1,\ldots,a_n=b_n$ 也是一種省略的說法, 一般也可以用 $a_i=b_i, \forall 1\leq i\leq n$ 來表示.

接著我們沿用 $\mathbb{R}^2$ 中向量的加法及係數積來定義 $\mathbb{R}^n$ 中向量的加法 (*addition*) 以及係數積 (*scalar multiplication*).

Definition 1.2.2. 令 $\mathbf{u}=(a_1,\ldots,a_n),\mathbf{v}=(b_1,\ldots,b_n)\in\mathbb{R}^n$ 以及 $r\in\mathbb{R}$. 我們定義

$$\mathbf{u}+\mathbf{v}=(a_1+b_1,\ldots,a_n+b_n) \quad \text{and} \quad r\mathbf{u}=(ra_1,\ldots,ra_n).$$

依此定義, 給定 $\mathbb{R}^n$ 中的兩向量我們可以明確地計算它們之和, 例如若

$$\mathbf{u}=(1,1,2,2,3),\mathbf{v}=(5,4,3,2,1)\in\mathbb{R}^5$$

則

$$\mathbf{u}+\mathbf{v}=(1+5,1+4,2+3,2+2,3+1)=(6,5,5,4,4).$$

這裡特別要注意的是必須是 $\mathbf{u},\mathbf{v}$ 皆在相同的 $\mathbb{R}^n$ 中才能定義 $\mathbf{u}+\mathbf{v}$. 例如若 $\mathbf{u}$ 在 $\mathbb{R}^3$ 而 $\mathbf{v}$ 在 $\mathbb{R}^4$ 是不能談 $\mathbf{u}+\mathbf{v}$ 的了!

如同 $\mathbb{R}^2$ 的情形, $\mathbb{R}^n$ 中向量加法及係數積有以下的性質. 由於這些性質的證明和 $\mathbb{R}^2$ 的情形完全相同 (用到實數相對應的性質), 此處我們就不再證明了.

Proposition 1.2.3. 對於 $\mathbb{R}^n$ 上的向量, 我們有以下的性質*:*

(1) 對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, 皆有 $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

(2) 對任意 $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^n$, 皆有 $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

(3) 存在一向量 $\mathbf{O}\in\mathbb{R}^n$ 滿足對任意 $\mathbf{u}\in\mathbb{R}^n$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$.

(4) 對任意 $\mathbf{u}\in\mathbb{R}^n$ 皆可找到 $\mathbf{u}'\in\mathbb{R}^n$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$.

(5) 對任意 $r,s\in\mathbb{R}$ 以及 $\mathbf{u}\in\mathbb{R}^n$, 皆有 $r(s\mathbf{u})=(rs)\mathbf{u}$.

(6) 對任意 $r,s\in\mathbb{R}$ 以及 $\mathbf{u}\in\mathbb{R}^n$, 皆有 $(r+s)\mathbf{u}=r\mathbf{u}+s\mathbf{u}$.

(7) 對任意 $r\in\mathbb{R}$ 以及 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 皆有 $r(\mathbf{u}+\mathbf{v})=r\mathbf{u}+r\mathbf{v}$.

(8) 對任意 $\mathbf{u}\in\mathbb{R}^n$, 皆有 $1\mathbf{u}=\mathbf{u}$.

Proposition 1.2.3 (3) 所提的 $\mathbf{O}$ 就是所謂 *additive identity*, 這裡其實就是向量 $(a_1,\ldots,a_n)$, 其中 $a_i=0$, $\forall 1\leq i\leq n$, 我們也稱此 $\mathbf{O}$ 為零向量. 另外若 $\mathbf{u}=(a_1,\ldots,a_n)\in\mathbb{R}^n$, 則 Proposition 1.2.3 (4) 所提的 $\mathbf{u}'$ 就是所謂 *additive inverse*, 這裡其實就是 $(b_1,\ldots,b_n)$, 其中 $b_i=-a_i$, $\forall 1\leq i\leq n$. 如同 Question 1.1 所提, 利用 $\mathbb{R}^n$ 向量加法的定義, 我們知這裡 $\mathbf{O}$ 是 $\mathbb{R}^n$ 中唯一滿足對任意 $\mathbf{u}\in\mathbb{R}^n$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$ 的向量, 而給定 $\mathbf{u}\in\mathbb{R}^n$, $\mathbf{u}'=-1\mathbf{u}$ 是唯一滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$ 的向量. 事實上, 不必利用 $\mathbb{R}^n$ 向量加法的定義, 由 Proposition 1.2.3 (3) 的性質本身就可以確定 $\mathbf{O}$ 是唯一的. 這是因為若有另一個 $\mathbf{O}'\in\mathbb{R}^n$ 也是 additive identity (即滿足 Proposition 1.2.3 (3) 的性質), 則由 $\mathbf{O}$ 為 additive identity 可得 $\mathbf{O}+\mathbf{O}'=\mathbf{O}'$. 另一方面, 由 $\mathbf{O}'$ 為 additive identity 可得 $\mathbf{O}+\mathbf{O}'=\mathbf{O}$. 也就是說我們有 $\mathbf{O}=\mathbf{O}'$, 得到唯一性. 同樣的道理, 給定 $\mathbf{u}\in\mathbb{R}^n$, 若 $\mathbf{u}',\mathbf{u}''$ 皆滿足 Proposition 1.2.3 (3) 的性質, 即 $\mathbf{u}+\mathbf{u}'=\mathbf{u}+\mathbf{u}''=\mathbf{O}$, 則得

$$\mathbf{u}'=\mathbf{u}'+\mathbf{O}=\mathbf{u}'+(\mathbf{u}+\mathbf{u}'')=(\mathbf{u}'+\mathbf{u})+\mathbf{u}''=\mathbf{O}+\mathbf{u}''=\mathbf{u}''.$$

所以我們有以下之性質.

Corollary 1.2.4. *在 $\mathbb{R}^n$ 中存在唯一的向量 $\mathbf{O}$ 滿足對任意 $\mathbf{u}\in\mathbb{R}^n$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$. 另外, 給定 $\mathbf{u}\in\mathbb{R}^n$, 存在唯一的 $\mathbf{u}'\in\mathbb{R}^n$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$.*

Remark 1.2.5. 在數學中, 一個定理若是由前一個定理直接得到 (不需引用新的性質), 通常會以 Corollary 稱之. Proposition 1.2.3 (3),(4) 其實談的是 additive identity 和 additive inverse 的存在性而 Corollary 1.2.4 談的是它們的唯一性. 同學或許會奇怪為何要將它們分開敘述呢? 這也就是我們用 Corollary 稱之的用意, 就是要強調我們雖然可以用 $\mathbb{R}^n$ 的加法及係數積的定義得到 Corollary 1.2.4, 但這裡我們避開定義, 直接用 Proposition 1.2.3 的結果得到 Corollary 1.2.4 (以後我們在抽象的向量空間中就會這麼處理). 就如在前一節所述, 很多向量的運算性質, 不需回歸到定義, 利用 Proposition 1.2.3 中所述的這些性質就足以推導出一般向量有關的性質.

有了唯一性, 以後我們將一律用 $\mathbf{O}$ 表示 $\mathbb{R}^n$ 中的 additive identity. 而給定 $\mathbf{u}\in\mathbb{R}^n$, 我們用 $-\mathbf{u}$ 表示 $\mathbf{u}$ 的 additive inverse.

我們利用 Proposition 1.2.3 (3) 有關 additive identity 的存在性證得 Corollary 1.2.4 中有關 additive identity 的唯一性. 證明過程中用了一個很重要的關鍵, 就是 Proposition 1.2.3 (3) 中要求 $\mathbf{O}$ 必須滿足對所有 $\mathbf{u}\in\mathbb{R}^n$ 皆滿足 $\mathbf{u}+\mathbf{O}=\mathbf{u}$. 如果當初僅只有特定 $\mathbf{u}\in\mathbb{R}^n$ 會滿足 $\mathbf{u}+\mathbf{O}=\mathbf{u}$, 則就無法證出 Corollary 1.2.4 中有關 additive identity 的唯一性了. 不過因為我們知道向量加法有 Proposition 1.2.3 這們完善的性質, 當我們要確認一個向量 $\mathbf{v}$ 是否為零向量時, 不必檢查所有的向量, 只要找到一個向量 $\mathbf{u}$ 滿足 $\mathbf{u}+\mathbf{v}=\mathbf{u}$, 就可以確認 $\mathbf{v}$ 是零向量了. 事實上我們有以下之結果.

Corollary 1.2.6. 假設 $\mathbf{v} \in \mathbb{R}^n$ 且存在 $\mathbf{u} \in \mathbb{R}^n$ 使得 $\mathbf{u}+\mathbf{v}=\mathbf{u}$, 則 $\mathbf{v}=\mathbf{O}$.

Proof. 由 $\mathbf{u}=\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$, 將等式兩邊加上 $-\mathbf{u}$, 可得

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

□

這裡的邏輯順序一定要弄清楚. 我們是先知道 Proposition 1.2.3 是對的, 才能推得 Corollary 1.2.4 以及 Corollary 1.2.6. Corollary 1.2.6 告訴我們如果要確定一個 $\mathbb{R}^n$ 中的向量 $\mathbf{v}$ 是零向量, 我們只要在 $\mathbb{R}^n$ 中找到一個向量 $\mathbf{u}$, 使得 $\mathbf{v}+\mathbf{u}=\mathbf{u}$ 即可, 不必去試 $\mathbb{R}^n$ 上所有的向量. 所以我們馬上有以下的性質.

Corollary 1.2.7. 令 $\mathbf{v} \in \mathbb{R}^n$, 我們有以下之結果.

(1) $0\mathbf{v}=\mathbf{O}$.

(2) $(-1)\mathbf{v}=-\mathbf{v}$.

Proof. 若直接用 $\mathbb{R}^n$ 上向量加法及係數積的定義, 馬上可得到此結果. 不過這裏我們利用 Proposition 1.2.3 以及其衍伸的 Corollaries 來證明.

(1) 利用 Corollary 1.2.6, 我們只要找到一個向量 $\mathbf{u}$ 滿足 $0\mathbf{v}+\mathbf{u}=\mathbf{u}$ 即可. 若考慮 $\mathbf{u}=\mathbf{v}$, 此時

$$0\mathbf{v}+\mathbf{v}=0\mathbf{v}+1\mathbf{v}=(0+1)\mathbf{v}=1\mathbf{v}=\mathbf{v}.$$

得證 $0\mathbf{v}=\mathbf{O}$.

(2) 要證明 $(-1)\mathbf{v}=-\mathbf{v}$, 利用 Corollary 1.2.4 additive inverse 的唯一性我們只要證明 $(-1)\mathbf{v}+\mathbf{v}=\mathbf{O}$ 即可. 然而,

$$(-1)\mathbf{v}+\mathbf{v}=(-1)\mathbf{v}+1\mathbf{v}=(-1+1)\mathbf{v}=0\mathbf{v},$$

故由 (1) 的結果知 $(-1)\mathbf{v}+\mathbf{v}=\mathbf{O}$.

□

再次強調一次, Corollaries 1.2.4, 1.2.6, 1.2.7, 都可以很簡單的利用 $\mathbb{R}^n$ 上向量加法與係數積的定義得到. 它們很自然也很容易理解, 不必去死記. 這裡特別提到它們是要讓大家知道, 這些性質都可以直接由 Proposition 1.2.3 得到, 也讓大家更加確定 Proposition 1.2.3 中的性質足以使得 $\mathbb{R}^n$ 中向量的運算都可像實數一般處理. 同樣的, 我們可以如實數一樣引用 "減法" 的符號, 也就是說將 $\mathbf{w}+(-\mathbf{v})$ 寫成 $\mathbf{w}-\mathbf{v}$. 如此一來以後我們在一些等式的推演時就直接沿用大家習慣的「移項」的說法. 例如 $2\mathbf{u}+\mathbf{v}=\mathbf{w}$, 我們就直接移項且乘以 1/2 得 $\mathbf{u}=\frac{1}{2}(\mathbf{w}-\mathbf{v})$.

1.3. Span of Vectors

在 $\mathbb{R}^n$ 中有許多由 $\mathbb{R}^n$ 中的向量展成的“子空間”. 這些子空間是這一節要探討的課題. 所謂子空間以後我們後有更正式的定義, 這裡我們僅暫時借用這個名詞, 大略指的是 $\mathbb{R}^n$ 中一些向量所組成的特殊子集合 (subset).

我們先從熟悉的坐標平面開始. 在坐標平面中我們最常談的就是直線. 一個直線通常有很多方式來表示, 若用通俗一點的說法一個直線可以說是通過一特定點且沿著一特定方向前進 (或後退) 所得的點所成的集合. 如果這個特定點用 P 來表示, 而 Q 為此直線上另外一點, 則用向量的觀點來說就是此直線是通過 P 點且沿著 $\overrightarrow{PQ}$ 方向的直線. 這個向量 $\overrightarrow{PQ}$ 我們稱為此直線的“方向向量” (directional vector). 一個直線的方向向量並不唯一, 因為若 Q' 為此直線上 P,Q 以外的另一點, 則 $\overrightarrow{PQ'}$ 也會是此直線的一個方向向量. 不過因為 P,Q,Q' 皆在同一直線上, 依向量的定義我們知道會存在一實數 r 使得 $\overrightarrow{PQ'} = r\overrightarrow{PQ}$. 因此若已知一非零向量 $\mathbf{v}$ 為一直線的方向向量, 則此直線上任兩點 P,Q 所成的向量都會存在一實數 r 使得 $\overrightarrow{PQ} = r\mathbf{v}$. 反之若已知一直線 L 通過 P 點且其方向向量為 $\mathbf{v}$, 若 Q 點滿足 $\overrightarrow{PQ} = r\mathbf{v}$, 其中 $r \in \mathbb{R}$, 則知 Q 點會在此直線 L 上. 因此我們可以用以下的集合來表示通過 P 點且方向向量為 $\mathbf{v}$ 的直線 L 上的點, 即

$$L = \{Q \mid \overrightarrow{PQ} = r\mathbf{v}, r \in \mathbb{R}\}.$$

通常在集合的表示法中, 如果無法用列舉的方式一一列舉此集合的元素時, 我們會用上面的方法來表示. 也就是在 “$|$” 的左邊寫下此集合元素的形式, 右邊寫下這些元素所需符合的性質. 在這裡左邊的 Q 表示此集合所組成的元素是像 Q 這樣的點, 而右邊 $\overrightarrow{PQ} = r\mathbf{v}$, $r \in \mathbb{R}$ 表示 Q 點需滿足 $\overrightarrow{PQ} = r\mathbf{v}$, 其中 r 為實數.

從上面可知若 $\mathbf{v}$ 是直線 L 的方向向量那麼對任意非零實數 r, 若 $\mathbf{w} = r\mathbf{v}$, 則 $\mathbf{w}$ 也是 L 的方向向量. 為了方便起見我們就稱 $\mathbf{w}$ 和 $\mathbf{v}$ 為 *parallel* (平行), 用 $\mathbf{v} \parallel \mathbf{w}$ 來表示. 若兩直線它們的方向向量是平行的, 我們就稱此二直線相平行.

我們可以發現, 上面這種用向量來描述一直線上的點的方法, 不僅在坐標平面中適用, 在坐標空間甚至更高維度的空間皆適用. 所以在 $\mathbb{R}^n$ 中若 $\mathbf{v}$ 是一個非零向量,

$$\{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} = r\mathbf{v}, r \in \mathbb{R}\}$$

這一類由向量所組成的集合便顯得重要. 提醒一下, 這個集合表示法 $|$ 的左邊是“向量”, 表示是由向量所成的集合, 和前面直線 L 的集合表法 $|$ 左邊是“點” 表示是由點所成的集合有所不同, 大家應區分清楚.

在坐標空間中除了直線另一個大家常探討的便是平面. 同樣的一個平面也有許多表示法. 和向量有關的最常見的就是法向量的表示法, 不過關於法向量的看法因牽涉內積我們留待以後再說明, 這裡我們依然沿用剛才直線的看法. 也就是說每一個平面, 我們都可以在其上找到兩個不平行的直線 L_1, L_2, 而整個平面就是沿著其中一條直線 L_1 畫出與 L_2 平行的直線而得. 換言之, 若 L_1, L_2 分別為此平面上不平行的兩條直線, 其方向向量分別為 $\mathbf{v}_1, \mathbf{v}_2$ 且相交於 P 點, 則對 L_1 上任一點 Q 皆可找到 $r_1 \in \mathbb{R}$ 滿足 $\overrightarrow{PQ} = r_1\mathbf{v}_1$. 而通過 Q 點且與 L_2 平行的直線會在此平面上, 亦即若 R 點會在此與 L_2 平行的直線上, 則可找到 $r_2 \in \mathbb{R}$ 使得

$\overrightarrow{QR} = r_2\mathbf{v}_2$. 利用 $\overrightarrow{PR} = \overrightarrow{PQ} + \overrightarrow{QR}$ 我們可得

$$\overrightarrow{PR} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2,$$

也就是說對於平面上任一點 R 皆可找到 $r_1, r_2 \in \mathbb{R}$ 滿足 $\overrightarrow{PR} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2$. 反之若 R 滿足 $\overrightarrow{PR} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2$, 我們也可知 R 會落在一個過 L_1 且與 L_2 平行的直線上, 也就是說 R 會在此平面上. 所以若一平面 H 通過 P 點且 $\mathbf{v}_1, \mathbf{v}_2$ 為 H 上兩條不平行的直線的方向向量, 我們也可用以下的集合表示平面 H 上的點, 即

$$H = \{R \mid \overrightarrow{PR} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2, r_1, r_2 \in \mathbb{R}\}.$$

要注意表示 H 的這兩個向量 $\mathbf{v}_1, \mathbf{v}_2$ 並不唯一, 至於要怎樣的兩個向量可以同樣描述 H 這個平面, 這裡我們暫不探討, 留待以後我們學習更多線性代數理論時再探討. 這裡我們將專注於 $\{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} = r\mathbf{v}, r \in \mathbb{R}\}$ 以及 $\{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2, r_1, r_2 \in \mathbb{R}\}$ 這一類由向量所組成的集合.

上述有關平面的向量表示法, 不僅在坐標空間上適用, 我們可以將上述概念推廣至 $\mathbb{R}^n$ 上且考慮更一般的情形. 在 $\mathbb{R}^n$ 上我們不只可談直線和平面, 還有許多和直線平面有類似的特性的事物值得探討. 我們自然引進以下的定義.

Definition 1.3.1. 給定 $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$, 以及 $r_1, \ldots, r_m \in \mathbb{R}$ 令

$$\mathbf{w} = r_1\mathbf{v}_1 + \cdots + r_m\mathbf{v}_m,$$

則稱 $\mathbf{w}$ 為 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 的一個 *linear combination* (線性組合). 所有 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 的 linear combinations 所組成的集合稱為它們的 *span* (展成的空間), 我們用 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ 來表示這個集合, 也就是說

$$\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m) = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} = r_1\mathbf{v}_1 + \cdots + r_m\mathbf{v}_m, r_1, \ldots, r_m \in \mathbb{R}\}.$$

要注意因為 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 在 $\mathbb{R}^n$ 中, 所以 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 的一個線性組合仍在 $\mathbb{R}^n$ 中. 也就是說 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ 是 $\mathbb{R}^n$ 的一個子集合. 這個子集合具有一重要的性質, 我們稱之為 $\mathbb{R}^n$ 的「子空間」(以後會給正式定義). 這個子集合具有怎樣的特殊性質呢? 比方說利用定義我們可以知零向量 $\mathbf{O}$ 一定會在 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ 中 (即每個 r_i 皆取為 0), 又若 $\mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$, 則因 $\mathbf{w} = r_1\mathbf{v}_1 + \cdots + r_m\mathbf{v}_m$ 可知 $-\mathbf{w} = (-r_1)\mathbf{v}_1 + \cdots + (-r_m)\mathbf{v}_m$ 也就是說 $-\mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$. 這些性質我們可以推廣到更一般的情況而得到以下的定理.

Proposition 1.3.2. 給定 $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^n$, 則對任意 $\mathbf{u}, \mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ 以及 $s, t \in \mathbb{R}$ 皆有

$$s\mathbf{u} + t\mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m).$$

Proof. 因 $\mathbf{u}, \mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$ 由定義知存在 $r_1, \ldots, r_m \in \mathbb{R}$ 以及 $r'_1, \ldots, r'_m \in \mathbb{R}$ 使得 $\mathbf{u} = r_1\mathbf{v}_1 + \cdots + r_m\mathbf{v}_m$ 以及 $\mathbf{w} = r'_1\mathbf{v}_1 + \cdots + r'_m\mathbf{v}_m$. 因此得

$$s\mathbf{u} + t\mathbf{w} = (sr_1 + tr'_1)\mathbf{v}_1 + \cdots + (sr_m + tr'_m)\mathbf{v}_m.$$

也就是說 $s\mathbf{u} + t\mathbf{w}$ 仍為 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 的一個線性組合, 故得證 $s\mathbf{u} + t\mathbf{w} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$. □

雖然在 Proposition 1.3.2 中我們僅提及 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)$ 中的兩個向量的線性組合仍在 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)$ 中, 不過利用數學歸納法, 我們可以證得任意有限多個 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)$ 中的向量的線性組合也會在 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)$ 中.

Question 1.2. 設 $\mathbf{v}_1,\ldots,\mathbf{v}_m \in \mathbb{R}^n$, 試證明若 $\mathbf{w}_1,\ldots,\mathbf{w}_k \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m)$, 則

$$\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k) \subseteq \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_m).$$

對於 span of vectors, 其中有一個基本的問題便是判斷哪些 vectors 會在特定 vectors 的 span 裡. 例如我們可以問 $(1,2,3) \in \mathbb{R}^3$ 是否屬於 $\mathrm{Span}((1,-1,2),(2,1,-2))$? 這樣的問題就等同於是否找得到 $r,s \in \mathbb{R}$ 使得 $(1,2,3) = r(1,-1,2)+s(2,1,-2)$. 比較向量各坐標, 我們得到一次聯立方程組

$$\left\{\begin{array}{rrcl} 1r & +2s & = & 1 \\ -1r & +1s & = & 2 \\ 2r & -2s & = & 3 \end{array}\right. .$$

所以知道這一類的問題會和解聯立方程組有關. 另外大家可以發現, 若將向量寫成直行 (即 column vector), $r(1,-1,2)+s(2,1,-2) = (1,2,3)$ 可寫成

$$r\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} + s\begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

這樣看成聯立方程組就更清楚方便了. 所以有的書都會將 vector 寫成 column vector 的樣子. 不過寫成 column vector 的缺點便是太占版面了, 本講義採 column vector 和 row vector (橫列) 混用的方法. 會視方便有時用 row vector 有時用 column vector. 這一點等介紹矩陣運算時我們會再釐清.

還有一個解決 column vector 占版面的方法就是利用所謂的 *standard basis*. 在 $\mathbb{R}^2$ 我們令 $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, 每一個 $\mathbb{R}^2$ 的 vector 都可以唯一寫成 $\mathbf{i},\mathbf{j}$ 的線性組合, 例如 $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ 便可以寫成 $2\mathbf{i}+3\mathbf{j}$. 一般我們令 $\mathbb{R}^3$ 上的 standard basis vectors 為

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

而 $\mathbb{R}^n$ 上的 standard basis vectors 定為

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix},$$

其中對任意 $1 \leq i \leq n$, $\mathbf{e}_i$ 為第 i 個位置是 1, 其他位置為 0 的 column vector. 不難發現, 任意 $\mathbb{R}^n$ 的 vector 都可以唯一寫成 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 的 linear combination. 事實上 $\mathbb{R}^n$ 還有其他組和 standard basis vectors 有類似性質的 vectors, 我們留待以後再談.

1.4. Dot Product

在 $\mathbb{R}^2$ 和 $\mathbb{R}^3$ 中大家熟悉內積的定義也可以推廣到一般的 $\mathbb{R}^n$. 將來我們會知道內積可以幫助我們定義出 $\mathbb{R}^n$ 中許多重要的子空間, 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 1.4.1. 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 1.4.2. 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 $\mathbf{u}=\mathbf{O}$.

(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, 不同的是 n 可以是任意自然數我們無法完整地寫下 $\mathbb{R}^n$ 中的向量, 而需藉由符號的幫助.

假設 $\mathbf{u}=(a_1,\ldots,a_n),\mathbf{v}=(b_1,\ldots,b_n),\mathbf{w}=(c_1,\ldots,c_n)$. 我們想利用 $\sum$ (summation) 這個符號來處理內積, 讓大家習慣這個便利的符號.

(1) 依定義

$$\mathbf{u}\cdot\mathbf{v}=\sum_{i=1}^{n}a_ib_i$$

這表示 $\mathbf{u}\cdot\mathbf{v}$ 是這些 a_ib_i 的和其中 i 是跑遍 1 到 n 的所有正整數. 由於這 n 項的每一項 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

$$\mathbf{u} = (a_1,\ldots,a_n) = (0,\ldots,0) = \mathbf{O}.$$

反之若 $\mathbf{u} = (a_1,\ldots,a_n) = \mathbf{O}$ 表示對任意 $1 \leq i \leq n$ 皆有 $a_i = 0$, 故得

$$\mathbf{u}\cdot\mathbf{u} = \sum_{i=1}^n a_i a_i = 0.$$

(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 1.4.2 (2) 告訴我們除了零向量 $\mathbf{O}$ 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 1.4.3. 令 $\mathbf{v} = (a_1,\ldots,a_n) \in \mathbb{R}^n$, 我們定義 $\mathbf{v}$ 的長度 (*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 1.4.2 的處理一些有關於內積的性質, 而不必涉及內積的定義.

Lemma 1.4.4. 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 1.4.2 (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 1.4.2 (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 1.3. 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 1.4.4 僅用到內積的性質, 所以在一般的情形若我們不是利用 Definition 1.4.1 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 1.4.2 中的性質, 我們依然可得到 Lemma 1.4.4 中的性質. Lemma 1.4.4 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 1.4.5 (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 1.4.4 得知

$$\|\mathbf{u}_0+\mathbf{v}_0\|^2=2+2\mathbf{u}_0\cdot\mathbf{v}_0, \tag{1.6}$$

$$\|\mathbf{u}_0-\mathbf{v}_0\|^2=2-2\mathbf{u}_0\cdot\mathbf{v}_0.$$

因為對任意的 $\mathbf{w}\in\mathbb{R}^n$ 皆有 $\|\mathbf{w}\|^2\geq 0$, 故得 $-1\leq\mathbf{u}_0\cdot\mathbf{v}_0\leq 1$. 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 (1.6) 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 1.4.2 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 1.4.5 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 1.4.5, 我們可以得到所謂的三角不等式.

Corollary 1.4.6 (Triangle inequality)**.** 若 $\mathbf{u},\mathbf{v} \in \mathbb{R}^n$, 則 $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

Proof. 由 Lemma 1.4.4 以及 Proposition 1.4.5, 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 1.4. 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 1.4.5 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 1.4.7. 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

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

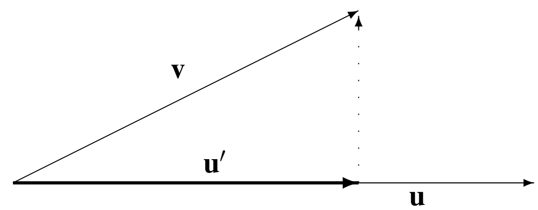
得知 $\theta = 120^\circ$.

利用夾角的定義我們進而定義出何謂「垂直」.

Definition 1.4.8. 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 1.4.9. 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 1.5. 能否不用找到 r 的方法得到 *Proposition 1.4.9* 的唯一性?

Proposition 1.4.9, 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 1.4.10. 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

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

1.5. 結論

在本章中我們將大家熟悉 $\mathbb{R}^2,\mathbb{R}^3$ 上的向量加法, 係數積, 內積等運算推廣至 $\mathbb{R}^n$ 上. 這些推廣而得的運算與 $\mathbb{R}^2,\mathbb{R}^3$ 上的運算有共同的性質. 直接利用這些性質而不必利用這些運算的定義, 我們就可以得到許多和 $\mathbb{R}^2,\mathbb{R}^3$ 相類似的結果. 換言之, 只要有這些運算性質, 即使更抽象的空間, 有些事情我們依然可以如平面或空間一樣的情形“看到”. 希望大家能體會到這些運算性質的重要性, 將來我們就是要利用這些性質進入較抽象的線性代數世界.

Chapter 2

Systems of Linear Equations

這一章要探討的是多元一次的聯立方程組. 我們依然利用大家熟悉的加減消去法 (或高斯消去法) 來處理這類方程組. 不過我們不再只關心如何解特定的聯立方程組, 而會更著重於有系統地探討一般聯立方程組解的情況的理論. 我們會用矩陣來表示一個聯立方程組, 不過這裡的矩陣僅是為了方便起見而使用, 不會涉及矩陣的性質. 至於真正矩陣的運算及性質, 我們留待下一章再詳述.

2.1. 解一次聯立方程組

所謂 n 元一次的方程式就是有 n 個未知數 (variable) 的一次方程式 (linear equation). 例如 $2x_1+5x_2-x_3+x_4=1$ 就是一個 4 元一次的聯立方程組 (當然也可看成是 5 元或更高元). n 元一次的方程式抽象的表示法就是

$$a_1x_2+\cdots+a_nx_n=b,$$

其中這些 $a_1,\ldots,a_n$ 和 b 都是實數, 而這些 x_i 表未知數. 當我們有多個 n 元一次的方程式要討論它們的共同解時, 就稱為解一次聯立方程組 (system of linear equations). 一般抽象的表示法

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

表示有 m 個 n 元一次方程式所成的方程組. 這裡 $a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1$ 表示第一個方程式, $a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2$ 表示第二個方程式, 而當 $1\leq i\leq m$ 時, 第 i 個方程式就是 $a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n=b_i$, 所以最後一個 (即第 m 個) 方程式就是 $a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n=b_m$. 這裡 a_{ij},b_i 皆為實數, 這些實數才是真正影響到聯立方程組

的因素, 所以我們也可特別把它們標明出來, 令

$$A=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

然後將上面的聯立方程組用 $A\mathbf{x}=\mathbf{b}$ 來表示. 矩陣 A 中的每一個 a_{ij} 稱為 A 的一個 *entry*. 因為 A 的每一個 entry 對應到聯立方程組中某個未知數的係數, 通常我們會稱矩陣 A 為此聯立方程式的係數矩陣. 一個矩陣的一個橫排稱為一個 row (列), 而一個豎排稱為一個 column (行). 我們算 row 時是從上而下來數的, 也就是說最上面的一個 row 稱為第一個 row, 下一個 row 稱為第二個 row, 依此類推. 而算 column 是由左而右來數的, 也就是說最左邊的一個 column 稱為第一個 column, 再往右一個 column 稱為第二個 column, 依此類推. 大家可以看出矩陣 A 的 row 對應的就是此聯立方程組的方程式, 第一個 row 對應到第一個方程式, 第二個 row 對應到第二個方程式, 依此類推. 而 column 對應到的是方程組的未知數, 第一個 column 對應到的是未知數 x_1 的係數, 第二個 column 對應到的是未知數 x_2 的係數, 依此類推. 因為這裡是由 m 個方程式而且每個方程式有 n 個未知數所組成的聯立方程組, 所以 A 共有 m 個 row 以及 n 個 column, 我們稱這樣的矩陣為 $m\times n$ matrix. 注意這裡 $\mathbf{x}$ 表示是一個未知的向量而且我們將向量 $\mathbf{x},\mathbf{b}$ 都寫成 column vector (行向量) 是為了配合將來矩陣乘法的寫法. 目前大家只要記住這也是聯立方程式的一種表示法即可.

例如解聯立方程組

$$\begin{array}{rcl} 3x_1-2x_2+9x_4 & = & 4 \\ 2x_1+2x_2-4x_4 & = & 6 \end{array} \tag{2.1}$$

我們就可以表成

$$\begin{bmatrix} 3 & -2 & 0 & 9 \\ 2 & 2 & 0 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}=\begin{bmatrix} 4 \\ 6 \end{bmatrix}$$

注意這裡係數矩陣多出 $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ 這個 column 因為 x_3 的係數為 0.

為何要探討解多元一次聯立方程組呢? 事實上解多元一次聯立方程式和線性代數的許多問題息息相關. 例如在 1.3 節中我們提到 span 的概念. 若 $\mathbf{u}=(1,-1,2,2),\mathbf{v}=(3,1,-1,2)$ 我們要問 $\mathbf{w}=(1,0,1,0)$ 是否在 $\mathrm{Span}(\mathbf{u},\mathbf{v})$ 中就等同於問是否存在 $c_1,c_2\in\mathbb{R}$ 使得 $\mathbf{w}=c_1\mathbf{u}+c_2\mathbf{v}$. 利用向量相等的定義, 這表示

$$(1,0,1,0)=c_1(1,-1,2,2)+c_2(3,1,-1,2).$$

亦即要解

$$\begin{array}{rcrcl} x_1 & + & 3x_2 & = & 1 \\ -x_1 & + & x_2 & = & 0 \\ 2x_1 & - & x_2 & = & 1 \\ 2x_1 & + & 2x_2 & = & 0 \end{array}$$

這一聯立方程組. 又如要問哪些向量 $\mathbf{x}=(x_1,x_2,x_3,x_4)$ 會同時滿足 $\mathbf{u}\cdot\mathbf{x}=0$ 且 $\mathbf{v}\cdot\mathbf{x}=0$, 就等同於要解聯立方程組

$$\begin{array}{rcrcrcrcl} x_1 & - & x_2 & + & 2x_3 & + & 2x_4 & = & 0 \\ 3x_1 & + & x_2 & - & x_3 & + & 2x_4 & = & 0. \end{array}$$

將來我們還會碰到許多和解聯立方程組有關的問題, 這裡我們就不再多談而將重點放在如何解一個多元一次聯立方程組.

過去學習解一次聯立方程組的方法不外加減消去法或高斯消去法, 它們的原理都是一樣的, 即利用以下三種基本方法:

(1) 變換式子的順序

(2) 將某一式乘上一非零實數

(3) 將某一式乘上一實數後加到另一式上

利用這三種基本方法將方程式的某些變數消去, 最後求出解來. 當然這裡有些問題是要探討的: 第一就是要消到甚麼地步才可確認可求出解來? 第二就是為什麼經過這些過程所得的解就會是原方程組的解? 在本節中我們將先介紹一個較有系統的方法解聯立方程組的步驟, 讓大家知道何時就可確認此方程組有解或無解, 且有解時如何求解. 下一節我們再說明為何每一個聯立方程組都可以利用這個方式找到其解集合.

當我們要解

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

這一個聯立方程組時, 先寫出如下的 augmented matrix (增廣矩陣)

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

例如式子 (2.1) 中的聯立方程組所對應的 augmented matrix 為

$$\left[\begin{array}{rrrr|r} 3 & -2 & 0 & 9 & 4 \\ 2 & 2 & 0 & -4 & 6 \end{array}\right]$$

換言之, 若我們要解 $A\mathbf{x}=\mathbf{b}$ 這一個聯立方程組, 就要寫下 $[A\,|\,\mathbf{b}]$ 這一個 matrix. 反之一個 augmented matrix $[A\,|\,\mathbf{b}]$ 就對應到一個聯立方程組 $A\mathbf{x}=\mathbf{b}$.

接下來我們將如加減消去法的三種步驟, 利用所謂的 elementary row operation (基本列運算) 處理這個 augmented matrix. 所謂 elementary row operation 即表示對矩陣進行如下三種的列運算:

(1) 將矩陣的某兩個 row 對調

(2) 將矩陣的某一個 row 乘上一非零實數

(3) 將矩陣的某一個 row 乘上一實數後加到另一個 row.

大家應很容易看出一個 augmented matrix 經過以上這三種列運算後所得的 augmented matrix 所對應的聯立方程組就是前面所提加減消去法的三種步驟所得的方程組. 我們的目的就是要將 augmented matrix $[A \mid \mathbf{b}]$ 中的係數矩陣 A 利用這三種 elementary row operation 化成所謂的 *echelon form*.

我們先解釋一下何謂 echelon form. 首先我們將矩陣每一個 row 從左到右來看第一個不為 0 的項稱為這個 row 的 *leading entry*. 因為係數矩陣中的每一個 entry 對應到聯立方程組中某個 variable 的係數, 所以 leading entry 若是 variable x_i 的係數, 我們就說這個 leading entry 發生在 x_i 的位置. 要注意, 這也等同於這個 leading entry 是位於從左到右算來第 i 個 column. 例如矩陣

$$\begin{bmatrix} 1 & 2 & 1 & 1 & 4 \\ 0 & 0 & 5 & 0 & 2 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$$

第一個 row 的 leading entry 為 1 不過因為第一個 row 還有其他位置 1, 所以我們特別要說明第一個 row 的 leading entry 發生在 x_1 的位置, 而第二個 row 和第三個 row 的 leading entry 分別為 5 和 1 且發生的位置皆在 x_3.

所謂一個矩陣是 echelon form 表示這個矩陣沒有 leading entry 的 row (即該 row 每一項皆為 0) 必需在最下方, 而有 leading entry 的 row 其 leading entry 所在位置從上到下來看是往右移的. 換言之, 若上一個 row 的 leading entry 所在的位置是 x_i, 而下一個 row 的 lading entry 是 x_j, 則必需 $i < j$. 例如上一個矩陣並非 echelon form, 因為第 3 個 row 和第 2 個 row 的 leading entry 的位置皆為 x_3, 並未右移. 另外矩陣

$$\begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 1 & 2 \\ 0 & 0 & 2 & -1 \\ 3 & 0 & 0 & 0 \end{bmatrix}$$

都不是 echelon form, 因為前一個矩陣全為 0 的 row 並未置於最下方, 而後一個矩陣第 3 個 row 的 leading entry 在第 2 個 row 的 leading entry 的左方. 至於矩陣

$$\begin{bmatrix} 0 & 2 & 1 & 1 & 4 \\ 0 & 0 & 3 & 0 & 2 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

就是 echelon form. 當一個矩陣是 echelon form 時, 我們稱每一個 row 的 leading entry 為 *pivot*, 而 pivot 所在的位置我們稱為 *pivot variable*.

當我們將 augmented matrix $[A \mid \mathbf{b}]$ 利用 elementary row operation 將之化成 $[A' \mid \mathbf{b}']$ 且 A' 為 echelon form 後. A' 有兩種情形. 一種情形為 A' 每一個 row 皆不全為 0; 另一種為 A' 有些 row 全為 0. 我們分別依這兩種情形來討論聯立方程組的解.

(1) A' 每一個 row 皆不全為 0: 此時聯立方程組為 *consistent*, 即一定有解. 我們又可細分成兩種情況.

(a) 第一種情況是每一個變數 (variable) x_i 皆為 pivot variable. 亦即 pivot 的個數等於方程組未知數的個數 (即係數矩陣 A 的 column 個數). 例如

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 4 \\ 0 & 3 & 1 & 2 \\ 0 & 0 & -1 & 1 \end{array}\right]$$

此時 echelon form 的 pivot variable 分別為 x_1,x_2,x_3 恰就是聯立方程組的未知數 x_1,x_2,x_3. 在這種情況之下此聯立方程組會有唯一解, 而且我們可利用從下往上 "代回" 的方式求得解. 例如前面的 augmented matrix 所對應的聯立方程組為

$$\begin{array}{rrrcl} 2x_1 & +x_2 & +x_3 & = & 4 \\ & 3x_2 & +x_3 & = & 2 \\ & & -x_3 & = & 1 \end{array}$$

所以我們從最下面的 $-x_3=1$ 可得 $x_3=-1$. 再將 $x_3=-1$ 代入其上一式 $3x_2+x_3=2$, 得 $3x_2-1=2$, 即 $x_2=1$. 最後將 $x_3=-1,x_2=1$ 代入 $2x_1+x_2+x_3=4$, 得 $x_1=2$. 故得其解為 $x_1=2,x_2=1,x_3=-1$.

(b) 第二種情況是有些 variable x_i 不是 pivot variable. 也就是方程組未知數的個數多於 pivot 的個數. 例如

$$\left[\begin{array}{cccc|c} 2 & 1 & 3 & 1 & 4 \\ 0 & 3 & 3 & 1 & 2 \\ 0 & 0 & 0 & -1 & 1 \end{array}\right]$$

此時 echelon form 的 pivot variable 分別為 x_1,x_2,x_4 少於立方程組的未知數 x_1,x_2,x_3,x_4. 在此情形之下此聯立方程組會有無窮多解. 要得到這種方程組所有的解, 首先我們要找到 *free variables*. 所謂 free variable 指的是方程組不是 pivot variable 的 variable. 例如前面這個例子, x_3 就是 free variable. Free variable 意指它可以任意取值, 所以找到 free variables 後你可以給它們任意的參數, 然後再利用如上一情況中由下往上代回的方式找到聯立方程組所有的解. 例如上一個 augmented matrix 所對應的聯立方程組為

$$\begin{array}{rrrrcl} 2x_1 & +x_2 & +3x_3 & +x_4 & = & 4 \\ & 3x_2 & +3x_3 & +x_4 & = & 2 \\ & & & -x_4 & = & 1 \end{array}$$

首先令 free variable x_3 為一參數 t (表示它可以是任意實數 $t\in\mathbb{R}$). 接著我們從最下面的 $-x_4=1$ 可得 $x_4=-1$. 再將 $x_3=t,x_4=-1$ 代入其上一式 $3x_2+3x_3+x_4=2$, 得 $3x_2+3t-1=2$, 即 $x_2=1-t$. 最後將 $x_2=1-t,x_3=t,x_4=-1$ 代入 $2x_1+x_2+3x_3+x_4=4$, 得 $x_1=2-t$. 故得其解為 $x_1=2-t,x_2=1-t,x_3=t,x_4=-1$, 其中 t 為任意實數. 因為 t 可以是任意實數, 由此我們也知此方程組有無窮多解.

(2) A' 有些 row 全為 0: 此時聯立方程組可能無解, 我們分成兩種情況:

(a) A' 有一個 row 全為 0 但 $\mathbf{b}'$ 在該 row 不為 0. 例如

$$[A' \mid \mathbf{b}'] = \left[\begin{array}{ccc|c} 2 & 1 & 1 & 4 \\ 0 & 3 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

A' 最後一個 row 皆為 0, 但 $\mathbf{b}'$ 在該 row 的位置為 1. 在此情形之下聯立方程組為 *inconsistent*, 即無解. 例如上一個 augmented matrix 其最後一個 row 所對應的方程式為

$$0x_1 + 0x_2 + 0x_3 = 1$$

但不管 x_1, x_2, x_3 代任何的實數都無法滿足 $0x_1 + 0x_2 + 0x_3 = 1$, 所以此方程組無解.

(b) A' 全為 0 的 row, $\mathbf{b}'$ 在該 row 亦為 0. 例如

$$\left[\begin{array}{cc|c} 2 & 1 & 4 \\ 0 & 3 & 2 \\ 0 & 0 & 0 \end{array}\right], \quad \left[\begin{array}{cccc|c} 2 & 1 & 3 & 1 & 4 \\ 0 & 3 & 3 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

這兩個 augmented matrices 皆為這種情形. 在此情形之下聯立方程組一定是 consistent. 事實上在此情形我們可以忽略全為 0 的 row, 例如前兩個 augmented matrices 所對應的方程組和

$$\left[\begin{array}{cc|c} 2 & 1 & 4 \\ 0 & 3 & 2 \end{array}\right], \quad \left[\begin{array}{cccc|c} 2 & 1 & 3 & 1 & 4 \\ 0 & 3 & 3 & 1 & 2 \end{array}\right]$$

所對應的方程組一樣. 所以我們可依前面 (1) A' 每一個 row 皆不全為 0 的情況找出聯立方程組所有的解.

我們要強調, 絕不會有 pivot 的個數多於方程組 variables (未知數) 的個數的情形發生. 這是因為當係數矩陣 A 是 echelon form 時, 每一個 column 最多僅能有一個 pivot (因為不能有兩個 leading term 在同一個位置), 所以 pivot 的個數不能多於 column 的個數. 而 A 的 column 個數表示的就是此聯立方程組 variables 的個數, 因此 pivot 的個數不會多於 variables 的個數. 另一方面依定義每一個 row 最多僅能有一個 pivot, 所以 pivot 的個數也不會多於該方程組的方程式個數 (即係數矩陣 row 的個數).

Question 2.1. 考慮一個由 n 個 *variables* 的 m 個方成式所組成的聯立方程組. 試說明前面討論 (1)(a) 的情形只有在 $m = n$ 的時候才有可能發生; 而 (1)(b) 的情形只有在 $m < n$ 的情形才有可能發生;

在這一節中我們介紹解一次聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的步驟. 也就是先將 augmented matrix $[A \mid \mathbf{b}]$ 利用 elementary row operations 化成 $[A' \mid \mathbf{b}']$ 其中 A' 為 echelon form 的情形. 再利用上面談論的情況找出聯立方程組的解. 下一節中我們將說明為何可利用 elementary row operations 將 A 變成 echelon form A', 且要說明為何這樣所得的解便是原方程組的解.

2.2. Elementary Row Operations

前一節中我們知道要解一個聯立方程組 $A\mathbf{x}=\mathbf{b}$, 可以先將 augmented matrix $[A\,|\,\mathbf{b}]$ 經由一系列的 elementary row operations 化成 $[A'\,|\,\mathbf{b}']$ 其中 A' 為 echelon form 後再求解. 在本節中我們要說明為何經由 elementary row operations 我們可以將一個矩陣化為 echelon form 且解釋為何經由 elementary row operations 後所對應的聯立方程組與原方程組會有相同的解集合.

我們利用數學歸納法來說明為何一定可以將一個矩陣化為 echelon form. 或許有些人會對這裡數學歸納法處理的方式覺得奇怪, 不過若能仔細體會其真意, 會發現這是最好的處理方式. 我們是對矩陣的 row 的個數作數學歸納法. 先說明所有只有一個 row 的矩陣一定是 echelon form, 然後利用這件事實證明所有有兩個 row 的矩陣皆可利用 elementary row operations 化為 echelon form. 再利用兩個 row 的矩陣會成立的事實證明有 3 個 row 的矩陣也可利用 elementary row operations 化為 echelon form, 如此一直下去我們可證有 4,5,6,... 個 row 的矩陣會成立. 不過這樣的方法我們可以證得有特定個數的 row 的矩陣會成立 (例如 10 個 row), 但無法證得一般的情形 (即任意個數的 row). 此時數學歸納法是最好的論證工具了. 若我們能知道有 k 個 row 的矩陣一定能利用 elementary row operations 化為 echelon form 這個事實且利用這個事實證得有 $k+1$ 個 row 的矩陣一定能利用 elementary row operations 化為 echelon form, 這就表示當我們知道有一個 row 的矩陣能利用 elementary row operations 化為 echelon form 就能推得有兩個 row 的矩陣能利用 elementary row operations 化為 echelon form, 也進而推得有 3 個 row 的矩陣亦成立, 再進而推得有 4 個 row 的矩陣亦成立, 如此一直下去當然可知任意的矩陣皆能利用 elementary row operations 化為 echelon form.

由於這裡的論證不容易說明清楚, 我們先由一個例子來說明. 考慮一個有 3 個 row 的矩陣

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

要將之化為 echelon form 首先第一個 row 必需是 leading entry 的位置在最左邊, 所以我們利用將第一, 二兩個 row 交換的 elementary row operation 將此矩陣變換為

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

由於第三個 row 的 leading entry 位置與第一個 row 相同所以須將此位置的數消掉. 利用第一個 row 乘上 -2 加到第三個 row 的 elementary row operation 將此矩陣變換為

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -2 & -7 \end{bmatrix}$$

如此一來第一個 row 以下的各 row 的 leading entry 所在位置都在第一個 row 的 leading entry 所在位置的右方. 接下來我們可以不再管第一個 row 而處理第一個 row 以下的部份.

此部份是一個僅有兩個 row 的矩陣

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -2 & -7 \end{bmatrix}$$

若我們知道變換兩個 row 的矩陣成 echelon 的方法, 直接套用就可以完成了. 事實上我們就是將上面的方法再處理一遍, 將第一個 row 乘以 2 加到第二個 row 即可得

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -5 \end{bmatrix}$$

這一個 echelon form. 要注意我們故意忽略原來矩陣的第一個 row 的原因就是為了能套用數學歸納法的假設 (即將 row 的個數變少). 事實上若用原來的矩陣, 我們是將矩陣

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -2 & -7 \end{bmatrix}$$

的第二個 row 乘以 2 加到第三個 row 得

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 3 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & -5 \end{bmatrix}$$

這一個 echelon form.

接下來我們處理一般的情形. 首先我們來看只有一個 row 的矩陣. 此時由於沒有任何的 row 在其下方所以依定義自然是 echelon form. 接著看有兩個 row 的矩陣. 首先注意依定義一個 echelon form 的第一個 row 其 leading entry (若有的話) 必在所有其他 row 的 leading entry 所在位置的左方. 所以我們在此有兩個 row 的矩陣挑出 leading entry 在最左方的一個 row (若兩個 row 的 leading entry 所在位置相同就任取一個 row) 利用 row 交換的 row operation 將之置於第一個 row. 接下來注意依定義下一個 row 的 leading entry 所在位置需在第一個 row 的 leading entry 的右方. 現若第二個 row 的 leading entry 所在位置和第一個 row 不同, 則因已知第一個 row 的 leading entry 所在位置在最左方, 第二個 row 的 leading entry 所在位置一定在第一個 row 的 leading entry 的右方, 故依定義此時已為 echelon form. 而若第二個 row 的 leading entry 所在位置和第一個 row 相同, 我們可將第一個 row 乘以 $-b/a$, 其中 a 為第一個 row 的 leading entry 而 b 為第二個 row 的 leading entry, 再加到第二個 row 上. 如此一來第二個 row 原本的 leading entry 所在位置變為 0, 故其 leading entry 所在位置往右移了, 依定義此時為 echelon form.

我們可以如法泡製處理有 3 個 row 的矩陣, 但由於要使用數學歸納法, 此時我們可直接假設我們已處理到有 k 個 row 的矩陣了, 亦即有 k 個 row 的矩陣皆可利用 elementary row operation 化為 echelon form. 現在我們要處理有 $k+1$ 個 row 的矩陣. 如前面的方法, 首先我們將 leading entry 的位置在最左邊的那個 row 利用兩 row 互換的 row operation 將之置於第一個 row. 假設此時第一個 row 的 leading entry 為 a. 接下來我們將 leading entry 的位置與第一個 row 的 leading entry 位置一樣的 row 挑出, 若該 row 的 leading entry 為 b, 我們便將第一個 row 乘上 $-b/a$ 後加到該 row 上. 如此一來該 row 的 leading entry 所在位置便往右移了. 一直重複此步驟, 直到第一個 row 以外的 row 其 leading entry 所在位置皆與第一個 row 的 leading entry 所在位置相異. 注意, 此時第一個 row 以下的各 row 其

leading entry 所在位置皆在第一個 row 的 leading entry 所在位置的右方. 若我們不看第一個 row, 所剩下的是一個有 k 個 row 的矩陣, 所以利用前面已知有 k 個 row 的矩陣皆可利用 elementary row operations 化為 echelon form, 我們可以利用 elementary row operations 將此矩陣第一個 row 以下的部份化為 echelon form. 但此時因各個 row 的 leading entry 所在位置皆在第一個 row 的 leading entry 所在位置的右方, 所以整個矩陣亦為 echelon form. 故得證所有矩陣皆可利用 elementary row operations 化為 echelon form. 大家或許注意到我們在化成 echelon form 的過程皆沒有用到將某個 row 乘上一非 0 實數這一個 elementary row operation. 事實上在化成 echelon form 的過程確實不需要這一種 row operation, 不過它在以後要談的化為 "reduced" echelon form 的過程是需要的, 留待以後再談.

既然每一個矩陣都能用 elementary row operations 化為 echelon form, 接下來我們要說明的是利用 elementary row operation 處理後的聯立方程組其解集合不會改變. 要注意這裡指的是將 augmented matrix 用 elementary row operations 變換後的 augmented matrix 其對應的聯立方程組其解集合不會改變. 亦即聯立方程組 $A\mathbf{x}=\mathbf{b}$ 所對應的 augmented matrix $[A \mid \mathbf{b}]$ 若經一些 elementary row operations 後變換成 $[A' \mid \mathbf{b}']$, 那麼其對應的聯立方程組 $A'\mathbf{x}=\mathbf{b}'$ 和原方程組 $A\mathbf{x}=\mathbf{b}$ 有相同的解集合. 不要誤以為是將係數矩陣 A 利用 elementary row operations 變換成 A' 後聯立方程組 $A'\mathbf{x}=\mathbf{b}$ 的解集合和原方程組 $A\mathbf{x}=\mathbf{b}$ 的解集合相同.

首先觀察若將一聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的 augmented matrix $[A \mid \mathbf{b}]$ 利用三種 elementary row operation 的任一種變換成 $[A' \mid \mathbf{b}']$ 表示將原方程組利用加減消去法的三個基本方法之一將之變成方程組 $A'\mathbf{x}=\mathbf{b}'$. 然而方程組 $A\mathbf{x}=\mathbf{b}$ 若利用加減消去法的三種方法 (即將兩式子對調順序或將某一式乘上某個非 0 實數或將一個式子乘上某個實數加到另一個式子) 變換成方程組 $A'\mathbf{x}=\mathbf{b}'$, 原來滿足 $A\mathbf{x}=\mathbf{b}$ 的一組解仍會滿足 $A'\mathbf{x}=\mathbf{b}'$. 換句話說 $A\mathbf{x}=\mathbf{b}$ 的解集合會包含於 $A'\mathbf{x}=\mathbf{b}'$ 的解集合. 不過加減消去法的三個基本方法是可以還原回去的, 也就是說方程組 $A'\mathbf{x}=\mathbf{b}'$ 也可以用加減消去法的三種方法還原回原方程組 $A\mathbf{x}=\mathbf{b}$. 因此 $A'\mathbf{x}=\mathbf{b}'$ 的解集合也會包含於 $A\mathbf{x}=\mathbf{b}$ 的解集合. 因此得證 $A\mathbf{x}=\mathbf{b}$ 和 $A'\mathbf{x}=\mathbf{b}'$ 會有相同的解集合. 我們證得了 $[A \mid \mathbf{b}]$ 若經由一個 elementary row operation 後得 $[A' \mid \mathbf{b}']$, 則它們所對應的聯立方程組會有相同的解集合. 因此若 $[A \mid \mathbf{b}]$ 經由好幾次的 elementary row operation 變換成 $[A' \mid \mathbf{b}']$, 它們所對應的聯立方程組當然也會有相同的解集合.

Example 2.2.1. Solve the linear system

$$\begin{array}{rrrcl} & x_2 & -3x_3 & = & -5 \\ 2x_1 & +3x_2 & -1x_3 & = & 7 \\ 4x_1 & +5x_2 & -2x_3 & = & 10. \end{array}$$

此聯立方程組的 augmented matrix 為

$$\left[\begin{array}{ccc|c} 0 & 1 & -3 & -5 \\ 2 & 3 & -1 & 7 \\ 4 & 5 & -2 & 10 \end{array}\right].$$

由於第二, 三 row 的 leading entry 在最左端. 但第二 row 的 leading entry 的值較小, 為了計算方便, 我們將之置於第一個 row, 即將一, 二 row 交換得

$$\left[\begin{array}{ccc|c} 2 & 3 & -1 & 7 \\ 0 & 1 & -3 & -5 \\ 4 & 5 & -2 & 10 \end{array}\right].$$

接下來由於第三 row 的 leading entry 也在 x_1 的位置需要消去, 所以將第一 row 乘上 -2 加到第三 row 得

$$\left[\begin{array}{rrr|r} 2 & 3 & -1 & 7 \\ 0 & 1 & -3 & -5 \\ 0 & -1 & 0 & -4 \end{array}\right].$$

此時係數矩陣仍不是 echelon form, 需將第三 row 的 x_2 位置的 entry 消去. 故將第二 row 加至第三 row 得

$$\left[\begin{array}{ccc|c} 2 & 3 & -1 & 7 \\ 0 & 1 & -3 & -5 \\ 0 & 0 & -3 & -9 \end{array}\right].$$

這是 echelon form. 由於係數矩陣沒有全為 0 的 row, 我們知此 linear system 為 consistent. 而又 pivot 的個數等於 variable 的個數, 故知此 linear system 的解唯一. 事實上, 最下面第三 row 表示 $-3x_3 = -9$, 得 $x_3 = 3$. 代入第二 row 表示的 $x_2 - 3x_3 = -5$, 得 $x_2 = 4$. 最後代入第一 row 表示的 $2x_1 + 3x_2 - x_3 = 7$, 得 $x_1 = -1$. 故知此 linear system 的解為 $(x_1, x_2, x_3) = (-1, 4, 3)$.

Example 2.2.2. Solve the linear system

$$\begin{array}{rrrrcr} x_1 & -1x_2 & +2x_3 & +3x_4 & = & 2 \\ 2x_1 & +1x_2 & +1x_3 & & = & 1 \\ x_1 & +2x_2 & -1x_3 & -3x_4 & = & 7. \end{array}$$

此聯立方程組的 augmented matrix 為

$$\left[\begin{array}{rrrr|r} 1 & -1 & 2 & 3 & 2 \\ 2 & 1 & 1 & 0 & 1 \\ 1 & 2 & -1 & -3 & 7 \end{array}\right].$$

第二, 三 row 的 leading entry 需被消去. 故將第一 row 分別乘上 $-2, -1$ 加到第二, 三 row 得

$$\left[\begin{array}{rrrr|r} 1 & -1 & 2 & 3 & 2 \\ 0 & 3 & -3 & -6 & -3 \\ 0 & 3 & -3 & -6 & 5 \end{array}\right].$$

接下來由於第三 row 的 leading entry 需要消去, 所以將第二 row 乘上 -1 加到第三 row 得

$$\left[\begin{array}{rrrr|r} 1 & -1 & 2 & 3 & 2 \\ 0 & 3 & -3 & -6 & -3 \\ 0 & 0 & 0 & 0 & 8 \end{array}\right].$$

這是 echelon form. 由於第三 row 表示 $0x_1 + 0x_2 + 0x_3 = 8$, 知此 linear system 為 inconsistent.

Example 2.2.3. Solve the linear system

$$\begin{array}{rrrrcr} x_1 & -2x_2 & +1x_3 & -1x_4 & = & 4 \\ 2x_1 & -3x_2 & +4x_3 & -3x_4 & = & -1 \\ 3x_1 & -5x_2 & +5x_3 & -4x_4 & = & 3 \\ -x_1 & +1x_2 & -3x_3 & +2x_4 & = & 5. \end{array}$$

此聯立方程組的 augmented matrix 為

$$\left[\begin{array}{rrrr|r} 1 & -2 & 1 & -1 & 4 \\ 2 & -3 & 4 & -3 & -1 \\ 3 & -5 & 5 & -4 & 3 \\ -1 & 1 & -3 & 2 & 5 \end{array}\right].$$

第二, 三, 四 row 的 leading entry 需被消去. 故將第一 row 分別乘上 $-2,-3,1$ 加到第二, 三, 四 row 得

$$\left[\begin{array}{rrrr|r} 1 & -2 & 1 & -1 & 4 \\ 0 & 1 & 2 & -1 & -9 \\ 0 & 1 & 2 & -1 & -9 \\ 0 & -1 & -2 & 1 & 9 \end{array}\right].$$

接下來第三, 四 row 的 leading entry 需要消去, 所以將第二 row 分別乘上 $-1,1$ 加到第三, 四 row 得

$$\left[\begin{array}{rrrr|r} 1 & -2 & 1 & -1 & 4 \\ 0 & 1 & 2 & -1 & -9 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

這是 echelon form. 由於係數矩陣全為 0 的第三, 四 row 全為 0, 知此 linear system 為 consistent.

事實上此 linear system 的 pivot variables 為 x_1,x_2, 而 free variables 為 x_3,x_4. 我們可以令 $x_4=r$, $x_3=s$, 代入第二 row 表示的 $x_2+2x_3-x_4=-9$, 得 $x_2=-9+r-2s$. 再代入第一 row 表示的 $x_1-2x_2+x_3-x_4=4$, 得 $x_1=-14+3r-5s$. 故知此 linear system 的解為

$$(x_1,x_2,x_3,x_4)=(-14+3r-5s,-9+r-2s,s,r), r,s\in\mathbb{R}.$$

通常我們習慣寫成 row vector 且將 r,s 提出. 故將解寫成

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -14 \\ -9 \\ 0 \\ 0 \end{bmatrix} + r\begin{bmatrix} 3 \\ 1 \\ 0 \\ 1 \end{bmatrix} + s\begin{bmatrix} -5 \\ -2 \\ 1 \\ 0 \end{bmatrix}, r,s\in\mathbb{R}.$$

在解 linear system 的過程中還可以進一步將 echelon form 化為所謂的 *reduced echelon form*. Reduced echelon form 事實上仍為 echelon form, 不過再加上兩個限制. 第一個限制是每一個 pivot 需為 1. 另一個限制為 pivot 的位置上方全為 0. 要注意, 依定義 echelon form 的 pivot 位置下方已全為 0 所以 reduced echelon form 每一個 pivot 所在的 column, 除了自己需為 1 外其他部分皆為 0. 例如

$$A=\begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B=\begin{bmatrix} 1 & 1 & 3 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

都不是 reduced echelon form 但是

$$A' = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

就是 reduced echelon form. 每一個 echelon form 皆可利用 elementary row operations 換為 reduced echelon form. 這是因為, 若有一個 row 的 pivot 為 a (注意依定義 $a \neq 0$), 我們只要將該 row 乘上 $1/a$, 則該 row 的 pivot 便是 1 了. 例如上面 A 這一個 echelon form 若將第二個 row 乘上 $1/3$, 就可得 A' 這一個 reduced echelon form. 當我們將每個 pivot 都變為 1 後, 就可利用將該 row 乘上某一實數加到另一個 row 的方法將 pivot 所在的 column 的其他部分化為 0. 例如上面 B 這一個 echelon form 若將第三個 row 分別乘上 -3, -1 加到第一個 row 和第二個 row, 得 $\begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{bmatrix}$. 再將第二個 row 乘上 -1 加到第一個 row, 就可得 B' 這一個 reduced echelon form. 注意一般我們都是從上而下將矩陣換成 echelon form, 不過得到 echelon form 後是從下而上將 echelon form 換成 reduced echelon form 較為方便.

既然每一個矩陣都可以經由 elementary row operations 化為 reduced echelon form, 所以我們可以利用這個方法求出聯立方程組的解. 化成 reduced echelon form 後由於每一個 row 除了該 row 的 pivot 外, 只剩 free variables (其他的 pivot variable 所在的 entry 皆為 0), 所以可以很快地看出解的形式. 例如方程組 $B'\mathbf{x} = \mathbf{O}$ 為

$$\begin{array}{cccccc} x_1 & & & & = & 0 \\ & x_2 & & +3x_4 & = & 0 \\ & & x_3 & -x_4 & = & 0 \end{array}$$

因僅 x_4 為 free variable, 令 $x_4 = t$, 代入第三 row 得 $x_3 = t$. 代入第二 row 得 $x_2 = -3t$. 最後由第一 row 得 $x_1 = 0$. 故知解為 $(x_1, x_2, x_3, x_4) = (0, -3t, t, t) = t(0, -3, 1, 1), t \in \mathbb{R}$.

化成 reduced echelon form 雖然在最後可以很快地看出解的形式, 但一般來說化為 reduced echelon form 比僅化為 echelon form 所需的步驟多了許多, 所以利用 echelon form 來求解還是會比較快. 利用 echelon form 求解的方法一般稱為 *Gauss method*, 而用 reduced echelon form 求解一般稱為 *Gauss-Jordan method*.

Example 2.2.4. Example 2.2.1 的 linear system, 化成 echelon form 後為

$$\left[\begin{array}{ccc|c} 2 & 3 & -1 & 7 \\ 0 & 1 & -3 & -5 \\ 0 & 0 & -3 & -9 \end{array}\right].$$

將第三 row 乘以 $-1/3$ 得

$$\left[\begin{array}{ccc|c} 2 & 3 & -1 & 7 \\ 0 & 1 & -3 & -5 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

再將第三 row 分別乘以 $3, 1$ 加到第二, 第一 row 得

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & 10 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

接著將第二 row 乘以 -3 加到第一 row 得

$$\left[\begin{array}{ccc|c} 2 & 0 & 0 & -2 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right].$$

最後將第一 row 乘以 1/2 得 reduced echelon form

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right],$$

且馬上看出解為 $(x_1,x_2,x_3)=(-1,4,3)$.

Example 2.2.3 的 linear system, 化成 echelon form 後為

$$\left[\begin{array}{cccc|c} 1 & -2 & 1 & -1 & 4 \\ 0 & 1 & 2 & -1 & -9 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

將第二 row 乘以 2 加到第一 row 得 reduced echelon form

$$\left[\begin{array}{cccc|c} 1 & 0 & 5 & -3 & -14 \\ 0 & 1 & 2 & -1 & -9 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

因 x_4,x_3 為 free variables, 令 $x_4=r$, $x_3=s$, 代入第二 row 得 $x_2=-9+r-2s$. 再代入第一 row 得 $x_1=-14+3r-5s$.

2.3. The Rank of a Matrix

一個矩陣經由 elementary row operations 化為 echelon form 之後, 其中不全為 0 的 row 的個數 (即 pivot 的個數) 是固定的, 我們稱之為此矩陣的 rank. 其實一個一次聯立方程組其係數矩陣的 rank 可以提供我們有關此聯立方程組是否有解以及解是否唯一等訊息. 在本節中我們便是要探討這個課題.

雖然一個矩陣化為 echelon form 的方法不唯一, 其 echelon form 也不唯一, 不過同一個矩陣不管其化成的 echelon form 為何, 它的 pivot 的個數是固定的. 這個事實等到我們介紹好對的概念便會清楚了. 不過這裡我們先接受這個事實, 而給出以下的定義.

Definition 2.3.1. 假設 A 為一矩陣. 若 A 利用 elementary row operations 化為 echelon form 後其 pivot 的個數為 r, 我們稱 r 為 A 的 *rank*. 用 $\mathrm{rank}(A)=r$ 來表示.

考慮一次聯立方程組 $A\mathbf{x}=\mathbf{b}$, 其中 A 為 $m\times n$ matrix 且 $\mathrm{rank}(A)=r$. 現若 $r=m$, 表示將 A 利用 elementary row operations 化為 echelon form 後每一個 row 皆不全為 0, 此即 2.1 節 (1) 的情形. 因此此時不管 $\mathbf{b}$ 為何, $A\mathbf{x}=\mathbf{b}$ 為 consistent, 亦即聯立方程組必有解. 但若 $r<m$ (回顧一下 $r>m$ 不可能發生) 會怎樣呢? 此時 A 利用 elementary row operations 化為 echelon form A' 後, 有些 row 是全為 0 的. 因此有可能存在 $\mathbf{b}$ 使得增廣矩陣 $[A\,|\,\mathbf{b}]$ 利用 elementary row operations 化為 $[A'\,|\,\mathbf{b}']$ 後, $\mathbf{b}'$ 對應到 echelon form A' 中某個全為 0 的

row 的位置不是 0 (即 2.1 節 (2)(a) 的情形), 此時 $A\mathbf{x}=\mathbf{b}$ 為 inconsistent, 亦即聯立方程組 $A\mathbf{x}=\mathbf{b}$ 無解. 事實上只要任取一個 $\mathbf{b}'$ 使得 $\mathbf{b}'$ 對應到 echelon form A' 中某個全為 0 的 row 的位置不是 0, 然後利用 elementary row operations 逆推回去所得的 $\mathbf{b}$ 就會使得聯立方程組 $A\mathbf{x}=\mathbf{b}$ 無解. 也就是說當 $r<m$ 時一定存在 $\mathbf{b}$ 使得 $A\mathbf{x}=\mathbf{b}$ 無解. 所以我們有興趣於探討那些 $\mathbf{b}$ 可以使得 $A\mathbf{x}=\mathbf{b}$ 有解, 那些會使得 $A\mathbf{x}=\mathbf{b}$ 無解.

給定 $m\times n$ matrix A, 如何決定哪些 $\mathbf{b}$ 會使得 $A\mathbf{x}=\mathbf{b}$ 有解, 那些會使得 $A\mathbf{x}=\mathbf{b}$ 無解呢? 我們可以將 $\mathbf{b}=\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ 視為 $\mathbb{R}^m$ 中的未知向量, 然後考慮 $[A\mid\mathbf{b}]$ 這一個增廣矩陣. 此時利用 elementary row operations 將 A 化成 echelon form A', 同時便把 $[A\mid\mathbf{b}]$ 化成了 $[A'\mid\mathbf{b}']$. 注意由於我們是利用 elementary row operations, 此時 $\mathbf{b}'=\begin{bmatrix} b'_1 \\ \vdots \\ b'_m \end{bmatrix}$ 上每一個位置 b'_i 都是以 $b_1,\dots,b_m$ 為未知數的一次多項式. 由於假設 $\operatorname{rank}(A)=r$, echelon form A' 的前 r 個 row 皆不全為 0, 而 A' 最後 $m-r$ 個 row 皆全為 0. 因此要使得聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 由 2.1 節 (2) 的情形知, 等同於 $\mathbf{b}'$ 最後 $m-r$ 個位置 $b'_{r+1},\dots,b'_m$ 應為 0. 也就是說要使的 $A\mathbf{x}=\mathbf{b}$ 有解, 就等同於 $\mathbf{b}=\begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ 會滿足 $b'_{r+1},\dots,b'_m$ 所代表的 $m-r$ 個以 $b_1,\dots,b_m$ 為未知數的一次多項式皆為 0. 換言之, 使得聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解的 $\mathbf{b}$ 必須滿足一組有 $m-r$ 個以 $b_1,\dots,b_m$ 為未知數的 m 元一次方程組. 我們有以下的結論.

Proposition 2.3.2. *給定一 $m\times n$ matrix A 且假設 $\operatorname{rank}(A)=r$. 考慮以 A 為係數矩陣的聯立方程組.*

(1) *$r=m$ 若且唯若對任意 $\mathbf{b}\in\mathbb{R}^m$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 一定有解.*

(2) *若 $r<m$, 則存在一個 $(m-r)\times m$ 的 matrix B, 使得 $A\mathbf{x}=\mathbf{b}$ 有解若且唯若 $\mathbf{y}=\mathbf{b}$ 為 $B\mathbf{y}=\mathbf{O}$ 的一組解.*

由於在 Proposition 2.3.2 (2) 中的聯立方程組 $B\mathbf{y}=\mathbf{O}$ 的一組解 $\mathbf{y}=\mathbf{b}$ 會使得聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 反之當 $A\mathbf{x}=\mathbf{b}$ 有解時, $\mathbf{y}=\mathbf{b}$ 就會是聯立方程組 $B\mathbf{y}=\mathbf{O}$ 的一組解. 因此我們稱此聯立方程組 $B\mathbf{y}=\mathbf{O}$ 為以 A 為係數矩陣的聯立方程組的 *constrain equations*. 要注意給定 A 其所對應的聯立方程組的 constrain equations 並不唯一, 但是不管是哪一組 constrain equations, 它們的解集合都會相同, 就是那些可以使的 $A\mathbf{x}=\mathbf{b}$ 為 consistent 的 $\mathbf{b}$ 所成的集合.

Example 2.3.3. 考慮 $A=\begin{bmatrix} 1 & -1 & 1 \\ 3 & 2 & -1 \\ 1 & 4 & -3 \\ 3 & -3 & 3 \end{bmatrix}$ 我們要找到以 A 為係數矩陣的聯立方程組的 constrain equations. 利用 elementary row operation, 我們可以得到 $\operatorname{rank}(A)=2$, 所以 A 的 constrain equations 是一個 $(4-2)\times 2$ 的矩陣所形成的聯立方程組.

首先考慮 augmented matrix $\left[\begin{array}{ccc|c} 1 & -1 & 1 & b_1 \\ 3 & 2 & -1 & b_2 \\ 1 & 4 & -3 & b_3 \\ 3 & -3 & 3 & b_4 \end{array}\right]$. 將 1-st row 分別乘上 -3, -1, -3 加到 2-nd, 3-rd, 4-th rows 得 $\left[\begin{array}{ccc|c} 1 & -1 & 1 & b_1 \\ 0 & 5 & -4 & b_2-3b_1 \\ 0 & 5 & -4 & b_3-b_1 \\ 0 & 0 & 0 & b_4-3b_1 \end{array}\right]$. 最後將 2-nd row 乘上 -1 加到 3-rd row 得 $\left[\begin{array}{ccc|c} 1 & -1 & 1 & b_1 \\ 0 & 5 & -4 & b_2-3b_1 \\ 0 & 0 & 0 & b_3-b_2+2b_1 \\ 0 & 0 & 0 & b_4-3b_1 \end{array}\right]$. 由此知 $\mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$ 會使得 $A\mathbf{x}=\mathbf{b}$ 為 consistent 若且唯若 b_1,b_2,b_3,b_4 會滿足 constrain equations

$$\begin{array}{rrrrcl} 2b_1 & -b_2 & +b_3 & & = & 0 \\ -3b_1 & & & +b_4 & = & 0 \end{array}.$$

所以令 $B=\begin{bmatrix} 2 & -1 & 1 & 0 \\ -3 & 0 & 0 & 1 \end{bmatrix}$, 此 2×4 matrix 所形成的聯立方程式 $B\mathbf{y}=\mathbf{O}$ 就是以 A 為係數矩陣的聯立方程組的 constrain equations.

我們可以再解 $B\mathbf{y}=\mathbf{O}$ 得其解集合為 $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}=r\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}+s\begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}$, $r,s\in\mathbb{R}$. 也就是說會使得 $A\mathbf{x}=\mathbf{b}$ 為 consistent 的 $\mathbf{b}$ 所成的集合為

$$\mathrm{Span}(\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 3 \end{bmatrix}).$$

Question 2.2. *假設 A 為 $m\times n$ matrix. 試說明當 $n<m$ 時, 一定存在 $\mathbf{b}\in\mathbb{R}^m$ 使得聯立方程組 $A\mathbf{x}=\mathbf{b}$ 為 inconsistent. 也就是當未知數個數少於方程式個數時, 一定存在 $\mathbf{b}$ 使得 $A\mathbf{x}=\mathbf{b}$ 無解.*

接下來我們探討 rank 和解的個數的關係. 假設 A 為 $m\times n$ matrix 且 $\mathrm{rank}(A)=r$. 當 $r=n$ 時, 表示 $x_1\ldots,x_n$ 這 n 個 variable 皆為 pivot variable. 此時 $A\mathbf{x}=\mathbf{b}$ 是可能無解的 (即 2.1 節 (2)(a) 的情形); 不過若 $A\mathbf{x}=\mathbf{b}$ 有解, 則由於沒有 free variable, $x_1,\ldots,x_n$ 的取值都是固定的, 所以 $A\mathbf{x}=\mathbf{b}$ 的解便會是唯一. 反之當 $r<n$, 由於有 $n-r$ 個 free variables, 當 $A\mathbf{x}=\mathbf{b}$ 為 consistent 時, 這些 free variables 任意取值都能得到一組 $A\mathbf{x}=\mathbf{b}$ 的解, 故知有無窮多解. 由此我們可得以下的結果.

Proposition 2.3.4. *給定一 $m\times n$ matrix A. 假設 $\mathrm{rank}(A)=r$ 且假設聯立方程組 $A\mathbf{x}=\mathbf{b}$ 為 consistent.*

(1) *若 $r=n$, 則聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有唯一解.*

(2) *若 $r<n$, 則聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有無窮多解.*

Question 2.3. 假設 A 為 $m\times n$ *matrix.* 試說明當 $n>m$ 時, 對於 $\mathbf{b}\in\mathbb{R}^m$ 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 不可能僅有唯一解. 也就是當未知數個數多於方程式個數時, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 要不然無解; 要不然會有無窮多解.

有一種特殊形式的聯立方程組一定是 consistent, 就是 $A\mathbf{x}=\mathbf{O}$ 這樣的 linear system. 因為 $x_1=0,\ldots,x_n=0$ 就是它的一組解. 我們特別稱 $A\mathbf{x}=\mathbf{O}$ 這樣的 linear system 為 *homogeneous system.* 我們有以下的結果.

Corollary 2.3.5. 給定一 $m\times n$ *matrix* A. 則 $A\mathbf{x}=\mathbf{O}$ 有唯一解 $x_1=\cdots=x_n=0$ 若且唯若 $\operatorname{rank}(A)=n$.

一個 homogeneous system $A\mathbf{x}=\mathbf{O}$ 一定有 $x_1=0,\ldots,x_n=0$ 這一組解, 我們稱這組解為 $A\mathbf{x}=\mathbf{O}$ 的 *trivial solution.* 而除了 $x_1=0,\ldots,x_n=0$ 以外 $A\mathbf{x}=\mathbf{O}$ 若還有其他的解, 我們便稱那些解為 $A\mathbf{x}=\mathbf{O}$ 的 *nontrivial solution.* 所以 $A\mathbf{x}=\mathbf{O}$ 的解唯一就等同於它沒有 nontrivial solution; 而解不唯一就等同於它有 nontrivial solution.

Question 2.4. 假設 A 為 $m\times n$ *matrix.* 試說明 $A\mathbf{x}=\mathbf{O}$ 有 *nontrivial solution* 若且唯若 $\operatorname{rank}(A)\neq n$.

事實上當 $A\mathbf{x}=\mathbf{b}$ 是 consistent 時, $A\mathbf{x}=\mathbf{b}$ 的解和 $A\mathbf{x}=\mathbf{O}$ 的解是息息相關的. 具體來說, 若 $x_1=c_1,\ldots,x_n=c_n$ 和 $x_1=c_1',\ldots,x_n=c_n'$ 皆為聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的一組解. 則由

$$\left\{\begin{array}{ccccccc} a_{11}c_1 & + & \cdots & + & a_{1n}c_n & = & b_1\\ a_{21}c_1 & + & \cdots & + & a_{2n}c_n & = & b_2\\ & & \vdots & & & & \\ a_{m1}c_1 & + & \cdots & + & a_{mn}c_n & = & b_m \end{array}\right. \quad\text{以及}\quad \left\{\begin{array}{ccccccc} a_{11}c_1' & + & \cdots & + & a_{1n}c_n' & = & b_1\\ a_{21}c_1' & + & \cdots & + & a_{2n}c_n' & = & b_2\\ & & \vdots & & & & \\ a_{m1}c_1' & + & \cdots & + & a_{mn}c_n' & = & b_m \end{array}\right.$$

將相對應的式子相減得

$$\begin{array}{ccccccc} a_{11}(c_1-c_1') & + & \cdots & + & a_{1n}(c_n-c_n') & = & 0\\ a_{21}(c_1-c_1') & + & \cdots & + & a_{2n}(c_n-c_n') & = & 0\\ & & \vdots & & & & \\ a_{m1}(c_1-c_1') & + & \cdots & + & a_{mn}(c_n-c_n') & = & 0 \end{array}.$$

也就是說 $x_1=c_1-c_1',\ldots,x_n=c_n-c_n'$ 就會是 $A\mathbf{x}=\mathbf{O}$ 的一組解. 同樣的道理, 若 $x_1=c_1,\ldots,x_n=c_n$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解而 $x_1=u_1,\ldots,x_n=u_n$ 為 $A\mathbf{x}=\mathbf{O}$ 的一組解, 則 $x_1=c_1+u_1,\ldots,x_n=c_n+u_n$ 也會是 $A\mathbf{x}=\mathbf{b}$ 的一組解. 因此我們有以下的結論.

Proposition 2.3.6. 假設聯立方程組 $A\mathbf{x}=\mathbf{b}$ 為 *consistent* 且 $\mathbf{x}=\mathbf{c}$ 為其一組解. 則 $\mathbf{x}=\mathbf{c}'$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解若且唯若 $\mathbf{x}=\mathbf{c}-\mathbf{c}'$ 為 $A\mathbf{x}=\mathbf{O}$ 的一組解.

利用 Proposition 2.3.6, 我們馬上有以下之結果.

Corollary 2.3.7. 假設 A 為 $m\times n$ *matrix.* 下列敘述是等價的.

(1) 若聯立方程組 $A\mathbf{x}=\mathbf{b}$ 為 *consistent* 則 $A\mathbf{x}=\mathbf{b}$ 的解唯一.

(2) $A\mathbf{x}=\mathbf{O}$ 沒有 *nontrivial solution.*

(3) $\text{rank}(A) = n$.

Proof. 要證明一系列 (多於兩個) 的敘述是等價, 我們不必證明每兩個敘述皆為等價. 只要證明 $(1) \Rightarrow (2)$, $(2) \Rightarrow (3)$ 以及 $(3) \Rightarrow (1)$ 即可. 這並沒有少證到其他部分, 例如 $(2) \Rightarrow (1)$ 就可由 $(2) \Rightarrow (3)$ 以及 $(3) \Rightarrow (1)$ 推得. 這是一個相當簡便的方法來證明一系列的 statement 是等價的, 也是在數學證明中常用的技巧.

$(1) \Rightarrow (2)$: 利用反證法, 假設 $x_1 = u_1, \ldots, x_n = u_n$ 為 $A\mathbf{x} = \mathbf{O}$ 的一組 nontrivial solution 而 $x_1 = c_1, \ldots, x_n = c_n$ 為 $A\mathbf{x} = \mathbf{b}$ 的一組解, 則由 Proposition 2.3.6 知 $x_1 = c_1 + u_1, \ldots, x_n = c_n + u_n$ 會是 $A\mathbf{x} = \mathbf{b}$ 的另一組解. 此與 $A\mathbf{x} = \mathbf{b}$ 的解唯一相矛盾, 故知 $A\mathbf{x} = \mathbf{O}$ 沒有 nontrivial solution.

$(2) \Rightarrow (3)$: $A\mathbf{x} = \mathbf{O}$ 沒有 nontrivial solution, 表示 $A\mathbf{x} = \mathbf{O}$ 的解唯一, 故由 Corollary 2.3.5 知 $\text{rank}(A) = n$.

$(3) \Rightarrow (1)$: 此為 Proposition 2.3.4 (1). □

最後我們綜合上面的結果處理當 A 為 $n \times n$ 的方陣的情形.

Corollary 2.3.8. 假設 A 為 $n \times n$ *square matrix.* 下列敘述是等價的.

(1) 對於任意 $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ 一定有解.

(2) 對於任意 $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ 一定有解且解唯一.

(3) $A\mathbf{x} = \mathbf{O}$ 沒有 *nontrivial solution.*

(4) $\text{rank}(A) = n$.

Proof. 要證明一系列的敘述是等價, 不一定要按照定理敘述的順序證明. 也時可以挑選較易證明的順序來證明. 不過要注意證明的路徑一定要形成一個回圈這樣證明才完整. 這裡我們要證明的路徑為 $(1) \Rightarrow (4)$, $(4) \Leftrightarrow (3)$, $(4) \Rightarrow (2)$ 以及 $(2) \Rightarrow (1)$.

$(1) \Rightarrow (4)$: 由 Proposition 2.3.2 (1), 我們知對於任意 $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ 一定有解表示 $\text{rank}(A)$ 等於 A 的 row 的個數, 故得 $\text{rank}(A) = n$.

$(4) \Leftrightarrow (3)$: 此為 Corollary 2.3.7 $(3) \Leftrightarrow (2)$ 的情形.

$(4) \Rightarrow (2)$: 當 $\text{rank}(A) = n$ 表示 $\text{rank}(A)$ 等於 A 的 row 的個數, 故由 Proposition 2.3.2 (1) 知對於任意 $\mathbf{b} \in \mathbb{R}^n$, $A\mathbf{x} = \mathbf{b}$ 一定有解. 再利用 Corollary 2.3.7 $(3) \Rightarrow (1)$ 得證其解唯一.

$(2) \Rightarrow (1)$: (2) 的條件比 (1) 強, 此敘述當然成立. □

一般我們會將符合 Corollary 2.3.8 的 $n \times n$ matrix (即 rank 為 n) 稱為 *nonsingular matrix*, 而 rank 小於 n 的 $n \times n$ matrix 稱為 *singular matrix*.

2.4. 結論

在本章中我們學習解聯立方程組的技巧. 利用 elementary row operations 將 augmented matrix 中的係數矩陣化為 echelon form 後, 我們很快的可以知道此聯立方程組是否有解, 而有解時也可利用此 echelon form 完整的得到此聯立方程組所有的解. 由 echelon form 的解

法我們了解到 pivot 對聯立方程組是否有解以及解是否唯一有著重要的關連. 在下一章, 我們將介紹矩陣的運算, 這些運算都可以和解聯立方程組的問題相連結. 所以本章中有關聯立方程組的理論對後面理論的建立影響深遠, 千萬不要以為會解具體的聯立方程組就可以了, 而忽視這些理論.

Chapter 3

Matrix

在上一章我們僅利用矩陣來表示一個聯立方程組, 這種表示法不只有其方便性其實是有另一層的意義. 在這一章中我們將介紹有關矩陣的運算, 利用矩陣的運算我們對聯立方程組將有另一種看法. 利用這新的看法, 我們對聯立方程組的解可以有更進一步的了解.

3.1. 矩陣的運算

在本節中我們將簡單地回顧有關於矩陣的定義. 一般來說一個矩陣是由數個 (橫) 列 (row) 以及 (直) 行 (column) 的數組成. 若一矩陣由 m 個 row 和 n 個 column 的數所組成, 我們便稱該矩陣為一個 $m\times n$ matrix. 特別的, 一個 $n\times n$ matrix (即 row 的個數等於 column 的個數), 我們稱之為 *square matrix*. 在本講義中, 我們用 $\mathscr{M}_{m\times n}$ 來表示所有 $m\times n$ 的矩陣所成的集合. 通常我們會用大寫的英文字母來表示一個矩陣. 例如若

$$A=\begin{bmatrix} 1 & 0 & 2 & 3 \\ 0 & 1 & 5 & 8 \\ 2 & 1 & 1 & 0 \end{bmatrix}, \tag{3.1}$$

則 A 為一個 3×4 matrix, 即 $A\in\mathscr{M}_{3\times 4}$. 當我們要抽象地描述一個矩陣時, 我們也常用 $A=[a_{ij}]$ 這樣的方法來描述. 這種表示法意指 A 中在第 i 個 row 和 j 個 column 的位置我們用 a_{ij} 來表示, 並稱之為此矩陣的 (i,j)-th entry. 因此當我們說 $A=[a_{ij}]$ 為 $m\times n$ 矩陣, 這表示 $1\leq i\leq m$ 且 $1\leq j\leq n$. 例如對於式子 (3.1) 中的矩陣 A, 若 $A=[a_{ij}]$, 則

$$\begin{array}{llll} a_{11}=1, & a_{12}=0, & a_{13}=2, & a_{14}=3, \\ a_{21}=0, & a_{22}=1, & a_{23}=5, & a_{24}=8, \\ a_{31}=2, & a_{32}=1, & a_{33}=1, & a_{34}=0. \end{array}$$

另外為了方便起見, 我們也會將矩陣 A 的每一個 row 和 column 用向量的方法來表示, 這些稱為 A 的 row vectors 和 column vectors. 在本講義我們會將矩陣 $A=[a_{ij}]$ 第 i 個 row 所成的 row vector 用 ${}_i\mathbf{a}$ 來表示, 而第 j 個 column 所成的 column vector 用 $\mathbf{a}_j$ 來表示. 例如對於式子 (3.1) 中的矩陣 A, 我們有

$${}_1\mathbf{a}=\begin{bmatrix}1 & 0 & 2 & 3\end{bmatrix}, \quad {}_2\mathbf{a}\begin{bmatrix}0 & 1 & 5 & 8\end{bmatrix}, \quad {}_3\mathbf{a}=\begin{bmatrix}2 & 1 & 1 & 0\end{bmatrix}$$

以及
$$\mathbf{a}_1=\begin{bmatrix}1\\0\\2\end{bmatrix},\quad \mathbf{a}_2=\begin{bmatrix}0\\1\\1\end{bmatrix},\quad \mathbf{a}_3=\begin{bmatrix}2\\5\\1\end{bmatrix},\quad \mathbf{a}_4=\begin{bmatrix}3\\8\\0\end{bmatrix}.$$
注意由於我們也想將向量看成是一個矩陣, 這裡的 row vector 和 column vector 們都用矩陣的形式呈現.

我們想給矩陣一個運算, 既然要談運算就會牽涉相等的概念. 所以我們要先定義何謂矩陣的相等 (就如同我們曾定義 $\mathbb{R}^n$ 中向量的相等).

Definition 3.1.1. 假設 $A=[a_{ij}]$ 為一個 $m\times n$ matrix 且 $A'=[a'_{ij}]$ 為一個 $m'\times n'$ matrix. 我們定義 $A=A'$ 若且唯若 $m=m',n=n'$ 且對所有的 $1\le i\le m$ 以及 $1\le j\le n$ 皆有 $a_{ij}=a'_{ij}$.

很容易看出矩陣的相等的定義是向量相等的延伸. 在向量中只有同在 $\mathbb{R}^n$ 的向量我們才談是否相等, 且兩個 $\mathbb{R}^n$ 中的向量相等表示這兩個向量在每一個相同位置的數皆相等. 同樣的只有同在 $\mathscr{M}_{m\times n}$ 的矩陣才談是否相等, 且兩個 $\mathscr{M}_{m\times n}$ 中的矩陣相等表示這兩個矩陣在每一個相同位置的數皆相等.

我們也延伸向量加法與係數積的定義來定義矩陣的加法與係數積. 也就是說只有同為 $\mathscr{M}_{m\times n}$ 的矩陣我們才定義它們之間的加法, 且兩矩陣相加表示將這兩個矩陣在相同位置的數加起來. 而一個實數乘上一個矩陣即為將該矩陣每一個位置上的數乘上該實數. 具體來說我們有以下的定義.

Definition 3.1.2. 假設 $A=[a_{ij}]$, $B=[b_{ij}]$ 皆為 $m\times n$ matrix. 定義 $A+B=[c_{ij}]$, 其中對所有的 $1\le i\le m$ 以及 $1\le j\le n$ 皆有 $c_{ij}=a_{ij}+b_{ij}$. 對任意實數 r, 我們定義 $rA=[d_{ij}]$ 其中對所有的 $1\le i\le m$ 以及 $1\le j\le n$ 皆有 $d_{ij}=ra_{ij}$.

Definition 3.1.2 告訴我們若
$$A=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix},B=\begin{bmatrix}b_{11}&b_{12}&\cdots&b_{1n}\\b_{21}&b_{22}&\cdots&b_{2n}\\\vdots&\vdots&\ddots&\vdots\\b_{m1}&b_{m2}&\cdots&b_{mn}\end{bmatrix}$$
則
$$A+B=\begin{bmatrix}a_{11}+b_{11}&a_{12}+b_{12}&\cdots&a_{1n}+b_{1n}\\a_{21}+b_{21}&a_{22}+b_{22}&\cdots&a_{2n}+b_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{m1}+b_{m1}&a_{m2}+b_{m2}&\cdots&a_{mn}+b_{mn}\end{bmatrix}$$
且
$$rA=\begin{bmatrix}ra_{11}&ra_{12}&\cdots&ra_{1n}\\ra_{21}&ra_{22}&\cdots&ra_{2n}\\\vdots&\vdots&\ddots&\vdots\\ra_{m1}&ra_{m2}&\cdots&ra_{mn}\end{bmatrix}$$

既然矩陣的加法與係數積的定義是由向量的加法與係數積延伸而來, 我們可以預期矩陣的加法與係數積應有和向量的加法與係數積相同的性質. 事實上我們確有以下的性質, 這些性質的證明和向量的情形相同 (用到實數相對應的性質), 我們就不再重複了.

Proposition 3.1.3. 對於 $\mathscr{M}_{m\times n}$ 上的矩陣, 我們有以下的性質:

(1) 對任意 $A,B\in\mathscr{M}_{m\times n}$, 皆有 $A+B=B+A$.

(2) 對任意 $A,B,C\in\mathscr{M}_{m\times n}$, 皆有 $(A+B)+C=A+(B+C)$.

(3) 存在一矩陣 $O\in\mathscr{M}_{m\times n}$ 滿足對任意 $A\in\mathscr{M}_{m\times n}$ 皆有 $O+A=A$.

(4) 對任意 $A\in\mathscr{M}_{m\times n}$ 皆可找到 $A'\in\mathscr{M}_{m\times n}$ 滿足 $A+A'=O$.

(5) 對任意 $r,s\in\mathbb{R}$ 以及 $A\in\mathscr{M}_{m\times n}$, 皆有 $r(sA)=(rs)A$.

(6) 對任意 $r,s\in\mathbb{R}$ 以及 $A\in\mathscr{M}_{m\times n}$, 皆有 $(r+s)A=rA+sA$.

(7) 對任意 $r\in\mathbb{R}$ 以及 $A,B\in\mathscr{M}_{m\times n}$ 皆有 $r(A+B)=rA+rB$.

(8) 對任意 $A\in\mathscr{M}_{m\times n}$, 皆有 $1A=A$.

接著我們定義矩陣間的乘法. 首先回顧在 Chapter 1 我們曾說明 $(1,2,3)\in\mathbb{R}^3$ 是否屬於 $\mathrm{Span}((1,-1,2),(2,1,-2))$ 的問題, 等同於是否存在 $x_1,x_2\in\mathbb{R}$ 滿足 $(1,2,3)=x_1(1,-1,2)+x_2(2,1,-2)$. 若將它們寫成 column vector 的形式, 即

$$x_1\begin{bmatrix}1\\-1\\2\end{bmatrix}+x_2\begin{bmatrix}2\\1\\-2\end{bmatrix}=\begin{bmatrix}1\\2\\3\end{bmatrix},$$

可看出此又等同於聯立方程組

$$\begin{array}{rrcl} x_1 & +2x_2 & = & 1\\ -x_1 & +1x_2 & = & 2\\ 2x_1 & -2x_2 & = & 3\end{array}$$

是否有解的問題. 而這樣的聯立方程組, 我們又把它寫成

$$\begin{bmatrix}1&2\\-1&1\\2&-2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\2\\3\end{bmatrix}.$$

所以若我們定義矩陣 $\begin{bmatrix}1&2\\-1&1\\2&-2\end{bmatrix}$ 和 $\begin{bmatrix}x_1\\x_2\end{bmatrix}$ 的乘法為

$$\begin{bmatrix}1&2\\-1&1\\2&-2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=x_1\begin{bmatrix}1\\-1\\2\end{bmatrix}+x_2\begin{bmatrix}2\\1\\-2\end{bmatrix}$$

那麼聯立方程組與矩陣的關係就不只是為了列式方便而已, 聯立方程組和矩陣的運算產生了緊密的關係.

從這個角度出發, 我們有以下定義.

Definition 3.1.4. 設 $A=[a_{ij}]$ 為 $m\times n$ matrix 以及 $\mathbf{b}=[b_j]$ 為 $n\times 1$ matrix (即 $\mathbb{R}^n$ 中的 column vector). 若 $\mathbf{a}_i$ 表示 A 的 i-th column 則定義

$$A\mathbf{b}=\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&\ddots&\vdots\\a_{m1}&a_{m2}&\cdots&a_{mn}\end{bmatrix}\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}=b_1\,\mathbf{a}_1+b_2\,\mathbf{a}_2+\cdots+b_n\,\mathbf{a}_n.$$

注意依此定義, 必需 $A \in \mathscr{M}_{m\times n}$ 的 column 的個數 n 等於 $\mathbf{b} \in \mathscr{M}_{n\times 1}$ 的 row 的個數 n, 才能定義 $A\mathbf{b}$ 且此時 $A\mathbf{b}$ 會是 $m\times 1$ matrix (即 $\mathbb{R}^m$ 中的 column vector). 觀察此 column vector, 我們有

$$A\mathbf{b} = b_1\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + b_2\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + b_n\begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} b_1a_{11}+b_2a_{12}+\cdots+b_na_{1n} \\ b_1a_{21}+b_2a_{22}+\cdots+b_na_{2n} \\ \vdots \\ b_1a_{m1}+b_2a_{m2}+\cdots+b_na_{mn} \end{bmatrix}. \tag{3.2}$$

特別的, 當 $\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$ 為 $1\times n$ matrix 而 $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ 為 $n\times 1$ matrix, 依 Definition 3.1.4 的矩陣乘法定義

$$\mathbf{a}\mathbf{b} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = b_1a_1 + b_2a_2 + \cdots + b_na_n. \tag{3.3}$$

也就是說此時若將 $\mathbf{a},\mathbf{b}$ 看成 $\mathbb{R}^n$ 中的 vector, 則 $\mathbf{ab}$ 就是 $\mathbf{a},\mathbf{b}$ 的內積. 依此看法, 由式子 (3.2), 我們可將 $A\mathbf{b}$ 寫成

$$A\mathbf{b} = \begin{bmatrix} {}_1\mathbf{a}\mathbf{b} \\ {}_2\mathbf{a}\mathbf{b} \\ \vdots \\ {}_m\mathbf{a}\mathbf{b} \end{bmatrix}. \tag{3.4}$$

也就是說 $A\mathbf{b}$ 這一個 $m\times 1$ matrix 的 i-th entry 為 ${}_i\mathbf{a}\mathbf{b}$ 也就是 A 的 i-th row ${}_i\mathbf{a}$ 和 $\mathbf{b}$ 的內積.

Remark 3.1.5. 注意式子 (3.3) 是說當 $\mathbf{a},\mathbf{b}$ 為 $\mathbb{R}^n$ 上的向量, 若將 $\mathbf{a}$ 寫成 row vector 的形式, $\mathbf{b}$ 寫成 column vector 的形式, 那麼我們可以將 $\mathbf{a},\mathbf{b}$ 看成矩陣, 即 $\mathbf{a} \in \mathscr{M}_{1\times n}$, $\mathbf{b} \in \mathscr{M}_{n\times 1}$. 此時由矩陣相乘的定義 $\mathbf{ab}$ 就是一個 1×1 的矩陣 (即實數) 且其值就是 $\mathbf{a},\mathbf{b}$ 視為 $\mathbb{R}^n$ 上的向量後取內積. 也就是說, 當看成向量的內積時, 我們不管兩向量是寫成 column vector 或 row vector. 但若看成矩陣的乘法時, 我們就必需說清楚哪一個是 column vector 哪一個是 row vector 了. 絕不能說由於向量內積是可交換的, 便認為 $\mathbf{b} \in \mathscr{M}_{n\times 1}$, $\mathbf{a} \in \mathscr{M}_{1\times n}$ 看成矩陣相乘所得的 $\mathbf{ba}$ 也會等於 $\mathbf{a}$ 和 $\mathbf{b}$ 看成 $\mathbb{R}^n$ 上的向量後取內積. 下面我們定義更一般矩陣的乘法時, 便會知道看成矩陣的乘法時 $\mathbf{ab}$ 和 $\mathbf{ba}$ 是不同的 ($\mathbf{ba}$ 會是一個 $n\times n$ matrix).

現在我們將矩陣乘法推廣到更一般的情況, 當 $A = [a_{ij}]$ 是一個 $m\times n$ matrix, $B = [b_{jk}]$ 是一個 $n\times l$ matrix. 由於對 B 的每一個 column vector $\mathbf{b}_k \in \mathscr{M}_{n\times 1}$, $1 \le k \le l$, 我們已定義了 $A\mathbf{b}_k$ 為何, 現在我們定義 AB 為 $m\times l$ matrix, 其中 AB 的 k-th column vector 為 $A\mathbf{b}_k$. 我們大致上有以下的圖示.

$$A\begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_l \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_l \\ | & | & & | \end{bmatrix}$$

由於 $A\mathbf{b}_k$ 為 $m\times 1$ matrix, 依此定義確實 AB 為 $m\times l$ matrix. 現在我們來看正式的定義.

Definition 3.1.6. 設 $A=[a_{ij}]$ 為 $m\times n$ matrix 以及 $B=[b_{jk}]$ 為 $n\times l$ matrix, 則定義 $AB=C=[c_{ik}]$ 為 $m\times l$ matrix, 其中對於 $1\leq k\leq l$, C 的 k-th column $\mathbf{c}_k$ 為

$$\mathbf{c}_k=A\mathbf{b}_k=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}\begin{bmatrix} b_{1k}\\ b_{2k}\\ \vdots\\ b_{nk}\end{bmatrix}=b_{1k}\mathbf{a}_1+b_{2k}\mathbf{a}_2+\cdots+b_{nk}\mathbf{a}_n. \tag{3.5}$$

由此定義, 我們知對於 $1\leq i\leq m$, $1\leq k\leq l$, AB 的 (i,k)-th entry 應為其 k-th column (即 $A\mathbf{b}_k$) 從上往下算的第 i 個 entry. 由式子 (3.4) 我們知此即 A 的 i-th row ${}_i\mathbf{a}$ 和 B 的 k-th column $\mathbf{b}_k$ 看成 $\mathbb{R}^n$ 上的向量後取內積. 換言之, 若 $AB=[c_{ik}]$, 則 AB 的 (i,k)-th entry c_{ik} 為

$$c_{ik}={}_i\mathbf{a}\,\mathbf{b}_k=a_{i1}b_{1k}+a_{i2}b_{2k}+\cdots+a_{in}b_{nk}=\sum_{j=1}^{n}a_{ij}b_{jk}. \tag{3.6}$$

再次強調一次, 並不是任取兩個矩陣都可以定義乘法, 必須是左邊矩陣的 column 個數和右邊矩陣的 row 個數相同才能相乘.

Example 3.1.7. 令

$$A=\begin{bmatrix} -2 & 4\\ 3 & 6\\ 2 & 2\end{bmatrix}, B=\begin{bmatrix} 4 & -1 & 2 & 5\\ 3 & 0 & 1 & 1\end{bmatrix}$$

考慮矩陣乘法 AB. 依定義矩陣 AB 的 3-rd column 為

$$A\mathbf{b}_3=\begin{bmatrix} -2 & 4\\ 3 & 6\\ 2 & 2\end{bmatrix}\begin{bmatrix} 2\\ 1\end{bmatrix}=2\mathbf{a}_1+1\mathbf{a}_2=2\begin{bmatrix} -2\\ 3\\ 2\end{bmatrix}+1\begin{bmatrix} 4\\ 6\\ 2\end{bmatrix}=\begin{bmatrix} 0\\ 12\\ 6\end{bmatrix}.$$

所以 AB 的 $(2,3)$ entry 為 12 等於 A 的 2-nd row 和 B 的 3-rd column 看成 $\mathbb{R}^2$ 中的向量所得的內積, 即 $(3,6)\cdot(2,1)=12$. 事實上我們有

$$AB=\begin{bmatrix} -2 & 4\\ 3 & 6\\ 2 & 2\end{bmatrix}\begin{bmatrix} 4 & -1 & 2 & 5\\ 3 & 0 & 1 & 1\end{bmatrix}=\begin{bmatrix} 4 & 2 & 0 & -6\\ 30 & -3 & 12 & 21\\ 14 & -2 & 6 & 12\end{bmatrix}.$$

Question 3.1. *設* $\mathbf{a},\mathbf{b}$ *為* $\mathbb{R}^n$ *上的向量, 若將* $\mathbf{a}$ *寫成 row vector 的形式,* $\mathbf{b}$ *寫成 column vector 的形式, 且將* $\mathbf{a},\mathbf{b}$ *看成矩陣, 即* $\mathbf{a}\in\mathscr{M}_{1\times n}$*,* $\mathbf{b}\in\mathscr{M}_{n\times 1}$*. 試問依矩陣乘法定義* $\mathbf{ba}$ *應為何種矩陣? 它和* $\mathbf{a},\mathbf{b}$ *看成* $\mathbb{R}^n$ *上的向量後取內積有關嗎?*

大部分的書都會用式子 (3.6) 當成矩陣乘法的定義. 我們選用式子 (3.5) 的用意, 主要是它較能描繪當初矩陣乘法定義的用意. 另外它是由 column 來描繪矩陣的乘法, 在證明或推導有關矩陣乘法性質時, 有時比式子 (3.6) 利用 entry 來看方便多了. 例如我們有以下的性質.

Proposition 3.1.8. *假設* $A,A'\in\mathscr{M}_{m\times n}$*,* $B,B'\in\mathscr{M}_{n\times l}$*. 我們有以下的性質.*

(1) $A(B+B')=AB+AB'$.

(2) $(A+A')B=AB+A'B$.

Proof. 首先注意因 $A+A'$ 仍為 $m\times n$ 矩陣且 $B+B'$ 為 $n\times l$ 矩陣, 所以這些矩陣的階數是符合矩陣乘法的規定. 我們假設 $A=[a_{ij}]$, $A'=[a'_{ij}]$, $B=[b_{jk}]$, $B=[b'_{jk}]$, 其中 $1\leq i\leq m$, $1\leq j\leq n$ 以及 $1\leq k\leq l$.

(1) 我們證明當 $1\leq k\leq l$ 時, $A(B+B')$ 的 k-th column 會等於 $AB+AB'$ 的 k-th column. 依定義 $A(B+B')$ 的 k-th column 為 A 的右邊乘上 $B+B'$ 的 k-th column. 然而由矩陣加法定義, $B+B'$ 的 k-th column 為 $\mathbf{b}_k+\mathbf{b}'_k$, 即 B 的 k-th column 加上 B' 的 k-th column. 因此我們有 $A(B+B')$ 的 k-th column 為

$$A(\mathbf{b}_k+\mathbf{b}'_k)=A\begin{bmatrix}b_{1k}+b'_{1k}\\b_{2k}+b'_{2k}\\\vdots\\b_{nk}+b'_{nk}\end{bmatrix}=(b_{1k}+b'_{1k})\mathbf{a}_1+(b_{2k}+b'_{2k})\mathbf{a}_2+\cdots+(b_{nk}+b'_{nk})\mathbf{a}_n. \tag{3.7}$$

另一方面, $AB+AB'$ 的 k-th column 為 AB 的 k-th column 加上 AB' 的 k-th column, 因此 $AB+AB'$ 的 k-th column 為

$$A\begin{bmatrix}b_{1k}\\b_{2k}\\\vdots\\b_{nk}\end{bmatrix}+A\begin{bmatrix}b'_{1k}\\b'_{2k}\\\vdots\\b'_{nk}\end{bmatrix}=(b_{1k}\mathbf{a}_1+b_{2k}\mathbf{a}_2+\cdots+b_{nk}\mathbf{a}_n)+(b'_{1k}\mathbf{a}_1+b'_{2k}\mathbf{a}_2+\cdots+b'_{nk}\mathbf{a}_n). \tag{3.8}$$

由 Proposition 1.2.3 向量加法相關性質, 我們得證式子 (3.7) 和式子 (3.8) 相等.

(2) 我們證明當 $1\leq k\leq l$ 時, $(A+A')B$ 的 k-th column 會等於 $AB+A'B$ 的 k-th column. 依定義 $(A+A')B$ 的 k-th column 為 $A+A'$ 的右邊乘上 B 的 k-th column. 然而由矩陣加法定義, 當 $1\leq j\leq n$ 時, $A+A'$ 的 j-th column 為 $\mathbf{a}_j+\mathbf{a}'_j$, 即 A 的 j-th column 加上 A' 的 j-th column. 因此我們有 $(A+A')B$ 的 k-th column 為

$$(A+A')\mathbf{b}_k=(A+A')\begin{bmatrix}b_{1k}\\b_{2k}\\\vdots\\b_{nk}\end{bmatrix}=b_{1k}(\mathbf{a}_1+\mathbf{a}'_1)+b_{2k}(\mathbf{a}_2+\mathbf{a}'_2)+\cdots+b_{nk}(\mathbf{a}_n+\mathbf{a}'_n). \tag{3.9}$$

另一方面, $AB+A'B$ 的 k-th column 為 AB 的 k-th column 加上 $A'B$ 的 k-th column, 因此 $AB+A'B$ 的 k-th column 為

$$A\begin{bmatrix}b_{1k}\\b_{2k}\\\vdots\\b_{nk}\end{bmatrix}+A'\begin{bmatrix}b_{1k}\\b_{2k}\\\vdots\\b_{nk}\end{bmatrix}=(b_{1k}\mathbf{a}_1+b_{2k}\mathbf{a}_2+\cdots+b_{nk}\mathbf{a}_n)+(b_{1k}\mathbf{a}'_1+b_{2k}\mathbf{a}'_2+\cdots+b_{nk}\mathbf{a}'_n). \tag{3.10}$$

再次由 Proposition 1.2.3 向量加法相關性質, 我們得證式子 (3.9) 和式子 (3.10) 相等. □

矩陣乘法和 scalar multiplication (係數積) 也有以下關係

Proposition 3.1.9. 設 $r\in\mathbb{R}$, $A\in\mathscr{M}_{m\times n}, B\in\mathscr{M}_{n\times l}$. 則

$$r(AB)=(rA)B=A(rB).$$

Proof. 假設 $A=[a_{ij}]$, $B=[b_{jk}]$, 其中 $1\leq i\leq m$, $1\leq j\leq n$ 以及 $1\leq k\leq l$. $r(AB)$ 的 k-th column 為 r 乘上 AB 的 k-th column, 故為

$$r(b_{1k}\mathbf{a}_1+b_{2k}\mathbf{a}_2+\cdots+b_{nk}\mathbf{a}_n). \tag{3.11}$$

而當 $1\leq j\leq n$ 時, (rA) 的 j-th column 為 $r\mathbf{a}_j$. 故 $(rA)B$ 的 k-th column 為

$$b_{1k}(r\mathbf{a}_1)+b_{2k}(r\mathbf{a}_2)+\cdots+b_{nk}(r\mathbf{a}_n). \tag{3.12}$$

最後由於 rB 的 k-th column 為

$$\begin{bmatrix} rb_{1k}\\ rb_{2k}\\ \vdots\\ rb_{nk}\end{bmatrix},$$

故 $A(rB)$ 的 k-th column 為

$$(rb_{1k})\mathbf{a}_1+(rb_{2k})\mathbf{a}_2+\cdots+(rb_{nk})\mathbf{a}_n. \tag{3.13}$$

再次由 Proposition 1.2.3 向量相關性質, 我們得證 (3.11), (3.12), (3.13) 三個式子皆相等. □

由 Proposition 3.1.8 和 Proposition 3.1.9 的證明我們可以看出, 有些矩陣乘法性質的推導可以簡化成右邊的矩陣是一個 column 的情形處理. 其實利用 row 來看矩陣的乘法也很很有用, 不過這個留待下一節介紹矩陣的 transpose (轉置) 後會更清楚.

利用矩陣乘法定義, 也可直接推得乘法具有結合律的性質 (即 $(AB)C=A(BC)$). 不過由於直接推導有點複雜, 這裡請大家自行驗證. 不過下一節我們學習了用 row 來處理矩陣乘法時, 我們會給一個較簡明的證明. 最後我們要強調的是矩陣乘法雖具有許多和實數乘法類似的性質, 但它卻沒有交換律. 事實上有可能 A 乘以 B 有定義, 但 B 卻不能乘以 A, 例如 $A\in\mathscr{M}_{2\times3}$, $B\in\mathscr{M}_{3\times4}$ 的情形. 也有可能即使 A 乘以 B 和 B 乘以 A 都有定義, 但由於乘了以後階數不同, 仍會使得 $AB\neq BA$, 例如 $A\in\mathscr{M}_{2\times3}, B\in\mathscr{M}_{3\times2}$ 的情形. 僅有在 A,B 為同階方陣時, 才有可能使得 AB 和 BA 的階數相同. 但此時仍有可能 $AB\neq BA$, 例如

$$A=\begin{bmatrix} a & b\\ c & d\end{bmatrix}, B=\begin{bmatrix} 1 & 0\\ 0 & -1\end{bmatrix}, AB=\begin{bmatrix} a & -b\\ c & -d\end{bmatrix}, BA=\begin{bmatrix} a & b\\ -c & -d\end{bmatrix},$$

這種情形只有在 $b=c=0$ 時, 才會使得 $AB=BA$. 所以在處理矩陣乘法時要特別小心. 例如當 A,B 為同階方陣時由 Proposition 3.1.8 和 Proposition 3.1.9 可推得 $(A-B)(A+B)=A^2-AB+BA-B^2$, 但由於可能 $AB\neq BA$, 我們不見得會有 $(A-B)(A+B)=A^2-B^2$.

當然了, 仍然有許多方陣會和所有的同階方陣相乘是可交換的. 一個常見的就是 *zero matrix* (零矩陣) O (即 $O=[a_{i,j}]$ 滿足每一個 entry $a_{i,j}=0$). 很容易驗證若 O 是一個 $n\times n$ square matrix, 則對任意 $A\in\mathscr{M}_{n\times n}$, 皆有 $OA=AO=O$. 另一個常見的便是所謂的 *identity matrix*. 通常 $n\times n$ 階的 identity matrix, 我們會用 I_n 來表示. I_n 的 i-th column 為 $\mathbf{e}_i$, 其中

$\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ 就是我門曾提過 $\mathbb{R}^n$ 的 standard basis (寫成 column vector). 例如

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

利用矩陣乘法的定義, 很容易知道對任意 $A \in \mathscr{M}_{m\times n}, B \in \mathscr{M}_{n\times l}$ 我們皆有 $AI_n = A, I_nB = B$. 特別的, 當 A 為 $n\times n$ matrix, 我們有 $AI_n = I_nA = A$.

Question 3.2. 假設 $A \in \mathscr{M}_{n\times n}$, 是否 $(A-2I_n)^2 = A^2 - 4A + 4I_n$ 為對?

Question 3.3. 試證明 I_n 是唯一的 $n\times n$ 滿足對任意 $A \in \mathscr{M}_{m\times n}$ 皆滿足 $AI_n = A$.

一個 $n\times n$ 的 square matrix 其 (i,i)-th entry 稱為 diagonal entry. 若除了 diagonal entries 以外, 其他的 entry 皆為 0, 我們便稱之為 *diagonal matrix*. Identity matrix 就是一個 diagonal matrix. 因為它的 diagonal entry 皆為 1, 其他的 entry 皆為 0. 另外, 對於任意 $r \in \mathbb{R}$, rI_n 亦為 diagonal matrix. 因為它 diagonal entry 皆為 r, 其他 entry 皆為 0. 對於任意 $A \in \mathscr{M}_{m\times n}, B \in \mathscr{M}_{n\times l}$ 我們很容易驗證 $rA = A(rI_n), rB = (rI_n)B$.

Question 3.4. 試利用 *Proposition 3.1.9* 驗證對任意 $n\times n$ *square matrix* A, 皆有 $(rI_n)A = A(rI_n)$.

要注意, 並不是所有 $n\times n$ 的 diagonal matrix 都會和 $n\times n$ 的 square matrix 相乘可交換. 前面曾給過例子 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 就不能和所有的 2×2 相乘可交換.

3.2. Transpose Operation

這一節中我們將介紹矩陣取 transpose (即轉置矩陣) 的概念, 並利用它來探討如何從 row 的角度來看矩陣相乘. 最後結合 column 和 row 的關係來處理矩陣相乘的結合律.

Definition 3.2.1. 對任意 $m\times n$ matrix $A = [a_{ij}]$. 我們定義 A 的 *transpose* 為一個 $n\times m$, 記為 A^{T}, 滿足對任意 $1 \le i \le n$, $1 \le j \le m$, A^{T} 的 (i,j)-th entry 為 A 的 (j,i)-th entry.

首先我們看看此定義是否為 well-defined, 也就是說給定一個 $m\times n$ matrix A, 我們是否真的可以得到一個 $n\times m$ matrix A^{T}. 依定義對任意 $1 \le i \le n$, $1 \le j \le m$, 在 A^{T} 的 i-th row 和 j-th column 位置的數應為 A 的 j-th row 和 i-th column 位置的數. 因為 A 的 row 和 column 的個數分別為 m 和 n, 而 j,i 又分別滿足 $1 \le j \le m$, $1 \le i \le n$, 所以我們真的可以在 A 中找到此 entry 放到 A^{T} 的 i-th row 和 j-th column 的位置. 可知此定義是沒問題的. 我們看以下的例子.

Example 3.2.2. 令

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

依定義 A^{T} 應為 3×2 matrix. 設其為

$$A^{\mathrm{T}}=\begin{bmatrix} x_{11} & x_{12}\\ x_{21} & x_{22}\\ x_{31} & x_{32}\end{bmatrix}.$$

依定義 A^{T} 的 $(1,1)$-th entry x_{11} 應為 A 的 $(1,1)$-th entry, 即 1. A^{T} 的 $(2,1)$-th entry x_{21} 應為 A 的 $(1,2)$-th entry, 即 2, 再來 A^{T} 的 $(3,1)$-th entry x_{31} 應為 A 的 $(1,3)$-th entry, 即 3. 同理可得 $x_{12}=-1$, $x_{22}=-2$, $x_{32}=-3$, 故知

$$A^{\mathrm{T}}=\begin{bmatrix} 1 & -1\\ 2 & -2\\ 3 & -3\end{bmatrix}.$$

可看出 A^{T} 的 1-st column 就是將 A 的 1-st row 寫成 column 而得, 且 A^{T} 的 2-nd column 就是將 A 的 2-nd row 寫成 column 而得. 而 A^{T} 的 1-st, 2-nd 和 3-rd row 就是將 A 的 1-st, 2-nd 和 3-rd column 寫成 row 而得.

由於我們希望整體的看一個矩陣的 row 和 column, 所以我們需要探討 A 和 A^{T} 它們的 row 與 column 之間的關係.

Lemma 3.2.3. 假設 A 為 $m\times n$ *matrix.* 對於 $1\le i\le n$, A^{T} 的 i*-th row* 就是將 A 的 i*-th column* 寫成 *row vector.* 對於 $1\le j\le m$, A^{T} 的 j*-th column* 就是將 A 的 j*-th row* 寫成 *column vector.*

Proof. 依定義 A^{T} 為 $n\times m$ matrix. 設 $A=[a_{kl}], A^{\mathrm{T}}=[a'_{ij}]$. 對於 $1\le i\le n$, A^{T} 的 i-th row 為

$$\begin{bmatrix} a'_{i1} & a'_{i2} & \cdots & a'_{im}\end{bmatrix}.$$

依定義 $a'_{i1}=a_{1i}, a'_{i2}=a_{2i},\dots,a'_{im}=a_{mi}$, 所以 A^{T} 的 i-th row 為

$$\begin{bmatrix} a_{1i} & a_{2i} & \cdots & a_{mi}\end{bmatrix},$$

就是 A 的 i-th column

$$\begin{bmatrix} a_{1i}\\ a_{2i}\\ \vdots\\ a_{mi}\end{bmatrix}$$

寫成 row vector 的形式. 同理可得, 對於 $1\le j\le m$, A^{T} 的 j-th column 就是將 A 的 j-th row 寫成 column vector. □

現在我們來看矩陣取 transpose 的基本性質.

Proposition 3.2.4. 假設 A,B 為 $m\times n$ *matrix,* C 為 $n\times l$ *matrix.* 我們有以下之性質.

(1) $(A^{\mathrm{T}})^{\mathrm{T}}=A$.

(2) $(A+B)^{\mathrm{T}}=A^{\mathrm{T}}+B^{\mathrm{T}}$.

(3) $(AC)^{\mathrm{T}}=C^{\mathrm{T}}A^{\mathrm{T}}$.

Proof. 首先觀察 A^{T} 為 $n\times m$ matrix, 故 $(A^{\mathrm{T}})^{\mathrm{T}}$ 為 $m\times n$ matrix, 與 A 階數相同. 同樣的, $A^{\mathrm{T}}+B^{\mathrm{T}}$ 為 $n\times m$ matrix 與 $(A+B)^{\mathrm{T}}$ 的階數相同. 另一方面 C^{T} 為 $l\times n$ matrix, 故 $C^{\mathrm{T}}A^{\mathrm{T}}$ 為 $l\times m$ matrix. 而 AC 為 $m\times l$ matrix, 所以 $(AC)^{\mathrm{T}}$ 為 $l\times m$ matrix 與 $C^{\mathrm{T}}A^{\mathrm{T}}$ 階數相同.

(1) 因 $(A^{\mathrm{T}})^{\mathrm{T}}$ 與 A 皆為 $m\times n$ matrix, 對於 $1\leq i\leq n$, 我們只要檢查 $(A^{\mathrm{T}})^{\mathrm{T}}$ 的 i-th column 就是 A 的 i-th column. 然而 $(A^{\mathrm{T}})^{\mathrm{T}}$ 的 i-th column 依 Lemma 3.2.3 知就是 A^{T} 的 i-th row 寫成 column vector, 而 A^{T} 的 i-th row 就是 A 的 i-th column. 故得證 $(A^{\mathrm{T}})^{\mathrm{T}}=A$.

(2) 因 $A^{\mathrm{T}}+B^{\mathrm{T}}$ 與 $(A+B)^{\mathrm{T}}$ 皆為 $n\times m$ matrix, 對於 $1\leq i\leq m$, 我們只要檢查 $A^{\mathrm{T}}+B^{\mathrm{T}}$ 的 i-th column 就是 $(A+B)^{\mathrm{T}}$ 的 i-th column. 依定義 $A^{\mathrm{T}}+B^{\mathrm{T}}$ 的 i-th column 就是 A^{T} 和 B^{T} 的 i-th column 之和. 依 Lemma 3.2.3 知它就是 A 和 B 的 i-th row 之和. 另一方面, $(A+B)^{\mathrm{T}}$ 的 i-th column 就是 $A+B$ 的 i-th row, 也就是 A 和 B 的 i-th row 之和. 得證 $(A+B)^{\mathrm{T}}=A^{\mathrm{T}}+B^{\mathrm{T}}$.

(3) 由於 $(AC)^{\mathrm{T}}$ 的 column 是由 AC 的 row 所決定, 而我們尚未討論 A 和 C 相乘 row 之間的關係, 所以這裡我們利用 entry 相同來證明相等. 對於 $1\leq i\leq l$, $1\leq j\leq m$, $(AC)^{\mathrm{T}}$ 的 (i,j)-th entry 為 AC 的 (j,i)-th entry, 即 A 的 j-th row 和 C 的 i-th column (看成 $\mathbb{R}^n$ 的向量) 的內積. 另一方面, $C^{\mathrm{T}}A^{\mathrm{T}}$ 的 (i,j)-th entry 為 C^{T} 的 i-th row 和 A^{T} 的 j-th column 的內積. 這也是 C 的 i-th column 和 A 的 j-th row 的內積, 故得證 $(AC)^{\mathrm{T}}=C^{\mathrm{T}}A^{\mathrm{T}}$. □

Question 3.5. 假設 A 為 $m\times n$ *matrix,* $r\in\mathbb{R}$. 試證明 $(rA)^{\mathrm{T}}=rA^{\mathrm{T}}$.

一個 $n\times n$ square matrix, 若滿足 $A^{\mathrm{T}}=A$, 我們稱 A 為 *symmetric matrix*. 上一節介紹過的 diagonal matrix 就是 symmetric matrix. 以後我們會學到 symmetric matrix 的重要性質, 現在我們先看和 symmetric matrix 有關的幾個簡單情形.

Corollary 3.2.5. 假設 A 為 $n\times n$ *square matrix,* B 為 $m\times n$ *matrix.* 以下皆為 *symmetric matrix.*

$$A+A^{\mathrm{T}}, BB^{\mathrm{T}}, B^{\mathrm{T}}B.$$

Proof. 由 Proposition 3.2.4, 我們有 $(A+A^{\mathrm{T}})^{\mathrm{T}}=A^{\mathrm{T}}+(A^{\mathrm{T}})^{\mathrm{T}}=A^{\mathrm{T}}+A$, 故知 $A+A^{\mathrm{T}}$ 為 symmetric matrix. 另一方面, $(BB^{\mathrm{T}})^{\mathrm{T}}=(B^{\mathrm{T}})^{\mathrm{T}}B^{\mathrm{T}}=BB^{\mathrm{T}}$, 故得 BB^{T} 為 symmetric matrix. 同理可得 $B^{\mathrm{T}}B$ 亦為 symmetric matrix. □

利用 Proposition 3.2.4, 我們可以從 row 的角度處理矩陣的乘法. 首先我們看一個 $1\times m$ matrix 乘上一個 $m\times n$ matrix 的情形. 假設 $A\in\mathscr{M}_{1\times m}, B\in\mathscr{M}_{m\times n}$, 令

$$A=\begin{bmatrix}a_1 & a_2 & \cdots & a_m\end{bmatrix}, B=\begin{bmatrix}b_{11} & b_{12} & \cdots & b_{1n}\\ b_{21} & b_{22} & \cdots & b_{2n}\\ \vdots & \vdots & \cdots & \vdots\\ b_{m1} & b_{m2} & \cdots & b_{mn}\end{bmatrix}.$$

則由 $(AB)^{\mathrm{T}}=B^{\mathrm{T}}A^{\mathrm{T}}$, 以及矩陣右邊乘 column vector 的定義得

$$(AB)^{\mathrm{T}}=\begin{bmatrix}b_{11} & b_{21} & \cdots & b_{m1}\\ b_{12} & b_{22} & \cdots & b_{m2}\\ \vdots & \vdots & \cdots & \vdots\\ b_{1n} & b_{2n} & \cdots & b_{mn}\end{bmatrix}\begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_m\end{bmatrix}=a_1\begin{bmatrix}b_{11}\\ b_{12}\\ \vdots\\ b_{1n}\end{bmatrix}+a_2\begin{bmatrix}b_{21}\\ b_{22}\\ \vdots\\ b_{2n}\end{bmatrix}+\cdots+a_m\begin{bmatrix}b_{m1}\\ b_{m2}\\ \vdots\\ b_{mn}\end{bmatrix}.$$

亦即 $(AB)^{\mathrm{T}} = a_1\,({}_1\mathbf{b})^{\mathrm{T}} + a_2\,({}_2\mathbf{b})^{\mathrm{T}} + \cdots + a_m\,({}_m\mathbf{b})^{\mathrm{T}}$, 這裡 $({}_i\mathbf{b})^{\mathrm{T}}$ 指的是將 B 的 i-th row 取轉置 (寫成 column 的形式). 故利用 Proposition 3.2.4 將 $(AB)^{\mathrm{T}}$ 再取轉置還原得

$$AB = a_1\,({}_1\mathbf{b}) + a_2\,({}_2\mathbf{b}) + \cdots + a_m\,({}_m\mathbf{b}).$$

也就是說

$$\begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ldots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = a_1 \begin{bmatrix} b_{11} & \cdots & b_{1n} \end{bmatrix} + \cdots + a_m \begin{bmatrix} b_{m1} & \cdots & b_{mn} \end{bmatrix} \tag{3.14}$$

現在我們來看一般的情形, 設 $A = [a_{ij}]$ 為 $m \times n$ matrix 以及 $B = [b_{jk}]$ 為 $n \times l$ matrix. 考慮 $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$. 依定義 $B^{\mathrm{T}}A^{\mathrm{T}}$ 的 i-th column, 為 B^{T} 右邊乘上 A^{T} 的 i-th column. 然而 A^{T} 的 i-th column, 為 A 的 i-th row 取轉置, 即 $({}_i\mathbf{a})^{\mathrm{T}}$. 也就是說 $(AB)^{\mathrm{T}}$ 的 i-th column 為 $B^{\mathrm{T}}({}_i\mathbf{a})^{\mathrm{T}}$. 利用 Proposition 3.2.4 再取轉置還原得, AB 的 i-th row 為

$$(B^{\mathrm{T}}({}_i\mathbf{a})^{\mathrm{T}})^{\mathrm{T}} = (({}_i\mathbf{a})^{\mathrm{T}})^{\mathrm{T}}(B^{\mathrm{T}})^{\mathrm{T}} = {}_i\,\mathbf{a}B.$$

換言之, 我們有以下的圖示

$$\begin{bmatrix} - & {}_1\mathbf{a} & - \\ - & {}_2\mathbf{a} & - \\ & \vdots & \\ - & {}_m\mathbf{a} & - \end{bmatrix} B = \begin{bmatrix} - & {}_1\mathbf{a}B & - \\ - & {}_2\mathbf{a}B & - \\ & \vdots & \\ - & {}_m\mathbf{a}B & - \end{bmatrix}. \tag{3.15}$$

結合式子 (3.14), 我們有以下之結果.

Proposition 3.2.6. *設 $A = [a_{ij}]$ 為 $m \times n$ matrix 以及 $B = [b_{jk}]$ 為 $n \times l$ matrix, 則對於 $1 \leq i \leq m$, AB 的 i-th row 為*

$${}_i\mathbf{a}B = \begin{bmatrix} a_{i1} & \cdots & a_{in} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1l} \\ \vdots & \ldots & \vdots \\ b_{n1} & \cdots & b_{nl} \end{bmatrix} = a_{i1} \begin{bmatrix} b_{11} & \cdots & b_{1l} \end{bmatrix} + \cdots + a_{in} \begin{bmatrix} b_{n1} & \cdots & b_{nl} \end{bmatrix}.$$

最後我們介紹矩陣乘法的結合律.

Proposition 3.2.7. *假設 $A \in \mathscr{M}_{m\times n}, B \in \mathscr{M}_{n\times l}, C \in \mathscr{M}_{l\times k}$, 則 $(AB)C = A(BC)$.*

Proof. 依定義 $AB \in \mathscr{M}_{m\times l}$, 故 $(AB)C \in \mathscr{M}_{m\times k}$. 而 $BC \in \mathscr{M}_{n\times k}$, 故 $A(BC) \in \mathscr{M}_{m\times k}$ 與 $(AB)C$ 同階. 對於 $1 \leq i \leq m$, $1 \leq j \leq k$, 我們要證明 $(AB)C$ 和 $A(BC)$ 的 (i,j)-th entry 相等. 令 ${}_i\mathbf{a}$ 為 A 的 i-th row, 且令 $\mathbf{c}_j$ 為 C 的 j-th column. 依定義 $(AB)C$ 的 (i,j)-th entry 為 AB 的 i-th row 乘上 C 的 j-th column. 由 Proposition 3.2.6 知 AB 的 i-th row 為 ${}_i\mathbf{a}B$, 故 $(AB)C$ 的 (i,j)-th entry 為 $({}_i\mathbf{a}B)\mathbf{c}_j$. 另一方面, $A(BC)$ 的 (i,j)-th entry 為 A 的 i-th row 乘上 BC 的 j-th column. 由 Definition 3.1.6 知 BC 的 j-th row 為 $B\mathbf{c}_j$, 故 $A(BC)$ 的 (i,j)-th entry 為 ${}_i\mathbf{a}(B\mathbf{c}_j)$. 也就是說, 要證明 $(AB)C$ 和 $A(BC)$ 的 (i,j)-th entry 相等, 就是要證明 $({}_i\mathbf{a}B)\mathbf{c}_j = {}_i\mathbf{a}(B\mathbf{c}_j)$. 由

於 ${}_i\mathbf{a}$ 和 $\mathbf{c}_j$ 分別只有一個 row 和 column, 為了方便考量, 我們將 ${}_i\mathbf{a}$ 和 $\mathbf{c}_j$ 用單一足碼表達. 也就是說假設

$$ {}_i\mathbf{a}=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}, B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix}, \mathbf{c}_j=\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_l \end{bmatrix} $$

依矩陣乘法定義

$$ {}_i\mathbf{a}B=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix}=\begin{bmatrix} {}_i\mathbf{a}\mathbf{b}_1 & {}_i\mathbf{a}\mathbf{b}_2 & \cdots & {}_i\mathbf{a}\mathbf{b}_l \end{bmatrix} $$

因此得

$$ ({}_i\mathbf{a}B)\mathbf{c}_j=\begin{bmatrix} {}_i\mathbf{a}\mathbf{b}_1 & {}_i\mathbf{a}\mathbf{b}_2 & \cdots & {}_i\mathbf{a}\mathbf{b}_l \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_l \end{bmatrix}=c_1\,({}_i\mathbf{a}\mathbf{b}_1)+c_2\,({}_i\mathbf{a}\mathbf{b}_2)+\cdots+c_l\,({}_i\mathbf{a}\mathbf{b}_l). $$

另一方面, 由矩陣乘法的定義 (Definition 3.1.4) 知

$$ B\mathbf{c}_j=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_l \end{bmatrix}=c_1\mathbf{b}_1+c_2\mathbf{b}_2+\cdots+c_l\mathbf{b}_l, $$

故由內積性質 (Proposition 1.4.2), 得

$$ {}_i\mathbf{a}\,(B\mathbf{c}_j)={}_i\mathbf{a}(c_1\mathbf{b}_1+c_2\mathbf{b}_2+\cdots+c_l\mathbf{b}_l)=c_1\,({}_i\mathbf{a}\mathbf{b}_1)+c_2\,({}_i\mathbf{a}\mathbf{b}_2)+\cdots+c_l\,({}_i\mathbf{a}\mathbf{b}_l). $$

得證 $({}_i\mathbf{a}B)\mathbf{c}_j={}_i\mathbf{a}\,(B\mathbf{c}_j)$, 故知 $(AB)C=A(BC)$. □

Question 3.6. 假設 $A\in\mathscr{M}_{m\times n}$, $r\in\mathbb{R}$, 試證明 $(rI_m)A=rA=A(rI_n)$ 並利用這個等式以及 *Proposition 3.2.7* 證明 *Proposition 3.1.9.*

有了矩陣乘法的結合律 (Proposition 3.2.7), 以後我們談多個矩陣相乘時, 為了方便起見, 我們會捨去括號例如直接用 ABC 表示. 特別的, 當 A 為方陣時, 既然 $(AA)A=A(AA)$, 我們就用 A^3 來表示. 同理, 當 n 個 A 相乘時, 我們就用 A^n 來表示.

3.3. Elementary Matrix

我們知道矩陣 A 左邊乘上另一矩陣 E, 可以視為 E 的 row 對矩陣 A 的作用. 事實上, 若將 A 做一個 elementary row operation, 會是將 A 的左邊乘上一個矩陣. 這樣的矩陣我們稱之為 elementary matrix.

設 $A=[a_{ij}]$ 為 $m\times n$ matrix. 首先觀察 identity matrix I_m 對 A 的作用. 由於 I_m 的 i-th row 為

$$ \begin{bmatrix} 0 & \cdots & \underset{\hat{i}}{1} & \cdots & 0 \end{bmatrix} $$

即 i-th entry 為 1, 其他 entry 皆為 0. 所以依 Proposition 3.2.6, I_mA 的 i-th row 為

$$\begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} A = 0\,{}_1\mathbf{a} + \cdots + 1\,{}_i\mathbf{a} + \cdots + 0\,{}_m\mathbf{a} = {}_i\mathbf{a},$$

(就是將 1 乘上 A 的 i-th row, 而將 0 乘上 A 的其他 row 再加起來, 故為 A 的 i-th row.) 換言之, 將 I_m 乘在 A 的左邊, 會將 A 的每一個 row 都固定不變, 所以知 $I_mA = A$. 現若 $j \neq i$ 且 E 為將 I_m 的 i-th row 改為 j-th entry 為 1 其他 entry 為 0, 而 i-th row 以外的其他 row 不變. 從上面的看法知 EA 的 i-th row 會是 A 的 j-th row, 也就是說 EA 會是將 A 的 i-th row 換成 A 的 j-th row, 而其他的 row 不動的矩陣.

現若用 i-th row 和 j-th row 交換的 elementary row operation 將 I_m 轉換成矩陣 E, 即

$$E = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \tag{3.16}$$

則利用前述的說法, EA 的 i-th row 是 A 的 j-th row, 而 EA 的 j-th row 是 A 的 i-th row, 而其他的 row 都不變. 換言之, EA 就是將 A 利用 i-th row 和 j-th row 交換這樣的 elementary row operation 變換所得的矩陣.

同樣的若將 I_m 的 i-th row 乘上實數 r 加到 I_m 的 j-th row 所得的矩陣為 E, 即

$$E = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & r & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \tag{3.17}$$

則因 E 的 j-th row 的 i-th entry 為 r, j-th entry 為 1. 故由 Proposition 3.2.6, EA 的 j-th row 就是將 r 乘上 A 的 i-th row 後再加上 A 的 j-th row, 而其他的 row 都不變. 換言之, EA 就是將 A 的 i-th row 乘上實數 r 加到 A 的 j-th row 這樣的 elementary row operation 變換所得的矩陣.

最後若將 I_m 的 i-th row 乘上非零實數 r 所得的矩陣為 E, 即

$$E = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & r & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \tag{3.18}$$

則很容易看出 EA 的 i-th row 就是將 A 的 i-th row 乘上 r, 而其餘的 row 不變. 也就是說, EA 就是將 A 的 i-th row 乘上非零實數 r 這樣的 elementary row operation 變換所得的矩陣.

從上面的說明我們知道, 對一個 $m\times n$ matrix 做一個 elementary row operation, 事實上就是將此矩陣左邊乘上一個 $m\times m$ matrix. 這個 $m\times m$ matrix 就是將 $m\times m$ identity matrix I_m 做同樣的 elementary row operation 所得的矩陣. 這樣的矩陣我們稱之為 *elementary matrix.* (3.16), (3.17), (3.18) 就是 elementary matrix 的三種形式.

當我們對一個 $m\times n$ matrix A, 進行多次的 elementary row operations, 就是將 A 左邊逐次的乘上相對應的 elementary matrix. 比方說做兩次 elementary row operations, 就是將 A 的左邊乘上第一次 elementary row operation 所對應的 elementary matrix E_1. 做第二次時就是將 E_1A 左邊再乘上第二次 elementary row operation 所對應的 elementary matrix E_2. 故所得的矩陣 $E_2(E_1A)$ 就是將 A 做這兩次 elementary row operations 所得的矩陣. 又由於矩陣乘法的結合律, 我們又可以將 $E_2(E_1A)$ 寫成 $(E_2E_1)A$. 同理, 對一個矩陣 A 進行一連串的 elementary row operations, 就是將 A 左邊乘上一個矩陣, 而這個矩陣就是這一連串 elementary row operations 所對應的 elementary matrices 的乘積. 不過要注意, 這些 elementary matrices 乘在一起的順序很重要, 因為 elementary matrices 之間的乘法不一定可以交換.

Question 3.7. 試找出那些同階的 *elementary matrices* 其相乘是不可以交換的.

有時我們需知道一個矩陣經由一連串的 elementary row operations, 其左邊到底是乘上哪一個矩陣. 當然我們可以如前述將所對應的 elementary matrices 乘在一起即可, 但這樣做其實很麻煩費時. 接下來我們來看一個很 “clever” 的方法, 可以在做 elementary row operation 時便幫我們將這個矩陣記錄下來. 這個方法就是, 若要對一個 $m\times n$ matrix A 做 elementary row operations, 我們先寫下一個 augmented matrix $[A|I_m]$. 也就是一個 $m\times(n+m)$ 的增廣矩陣, 其左邊 n 個 columns (即前 n 個 columns) 為矩陣 A, 而右邊 m 個 columns (即後 m 個 column) 為 I_m. 現將 A 做第一次的 elementary row operation, 假設其對應的 elementary matrix 為 E_1, 則對 $[A|I_m]$ 做相同的 elementary row operation 的話, 所得的結果為 $E_1[A|I_m]$. 然而依矩陣乘法的定義我們知

$$E_1[A|I_m]=[E_1A|E_1I_m]=[E_1A|E_1].$$

也就是說, 當我們對 $[A|I_m]$ 做同樣的 elementary row operation, 所得的增廣矩陣其左邊就是將 A 做此 elementary row operation 所得的矩陣, 而右邊就是此 elementary row operation 所對應的 elementary matrix. 接著當我們做下一個 elementary row operation, 假設此 elementary row operation 所對應的 elementary matrix 為 E_2, 則此 elementary row operation 對 $[E_1A|E_1]$ 作用後所得的矩陣便是 $E_2[E_1A|E_1]=[E_2(E_1A)|E_2E_1]$. 這樣繼續下去, 當我們對增廣矩陣 $[A|I_m]$ 進行一連串的 elementary row operations 後, 所得的矩陣 $[A'|E]$, 其左邊 A' 就是 A 經由這一連串的 elementary row operations 作用後所得的矩陣, 而右邊的

E 就是這些 elementary row operations 所對應的 elementary matrices 依序從右到左相乘所得的結果, 因此 $EA = A'$. 我們有以下的結論.

Lemma 3.3.1. 假設 A 為 $m \times n$ *matrix.* 若將 A 經由一連串的 *elementary row operations* 轉換成 A', 則存在一個 $m \times m$ *matrix* E 使得 $EA = A'$, 其中 E 為這一連串 *elementary row operations* 所對應的 *elementary matrix* 由右而左依序相乘的乘積. 事實上若將 *augmented matrix* $[A|I_m]$ 經由同樣的 *elementary row operations* 作用後, 所得的 *augmented matrix* 就是 $[A'|E]$.

Example 3.3.2. 將矩陣

$$A = \begin{bmatrix} 2 & -4 & 4 & -6 \\ 1 & -2 & 1 & -1 \\ 4 & -8 & 4 & -4 \end{bmatrix}$$

化為 reduced echelon form, 並找到 elementary matrices 的乘積 E 使得 EA 為此 reduced echelon form.

首先寫下 augmented matrix

$$[A|I_3] = \left[\begin{array}{cccc|ccc} 2 & -4 & 4 & -6 & 1 & 0 & 0 \\ 1 & -2 & 1 & -1 & 0 & 1 & 0 \\ 4 & -8 & 4 & -4 & 0 & 0 & 1 \end{array}\right]$$

並將此 augmented matrix 的 1-st 和 2-nd row 交換, 得

$$\left[\begin{array}{cccc|ccc} 1 & -2 & 1 & -1 & 0 & 1 & 0 \\ 2 & -4 & 4 & -6 & 1 & 0 & 0 \\ 4 & -8 & 4 & -4 & 0 & 0 & 1 \end{array}\right]$$

接著我們將 augmented matrix 的 1-st row 乘上 -2 加到 2-nd row 上, 得

$$\left[\begin{array}{cccc|ccc} 1 & -2 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 2 & -4 & 1 & -2 & 0 \\ 4 & -8 & 4 & -4 & 0 & 0 & 1 \end{array}\right].$$

然後將 augmented matrix 的 1-st row 乘上 -4 加到 3-rd row 得

$$\left[\begin{array}{cccc|ccc} 1 & -2 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 2 & -4 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & 1 \end{array}\right].$$

繼續將 augmented matrix 的 2-nd row 乘上 $1/2$ 得

$$\left[\begin{array}{cccc|ccc} 1 & -2 & 1 & -1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & 1 \end{array}\right].$$

最後將 augmented matrix 的 2-nd row 乘上 -1 加到 1-st row 得

$$\left[\begin{array}{cccc|ccc} 1 & -2 & 0 & 1 & -\frac{1}{2} & 2 & 0 \\ 0 & 0 & 1 & -2 & \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -4 & 1 \end{array}\right].$$

令最後所得的 augmented matrix 為 $[A'|E]$, 我們檢查是否 A 的 reduced echelon form A' 就是 EA. 事實上, 我們確有

$$EA=\begin{bmatrix} -\frac{1}{2} & 2 & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & -4 & 1 \end{bmatrix}\begin{bmatrix} 2 & -4 & 4 & -6 \\ 1 & -2 & 1 & -1 \\ 4 & -8 & 4 & -4 \end{bmatrix}=\begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

另外我們想確認 E 確為這五個 elementary row operations 所對應的 elementary matrices 的乘積. 因為 A 為 3×4 matrix, 所以第一個 elementary row operation 所對應的 elementary matrix E_1 就是將 3×3 的 identity matrix I_3 的 1-st 和 2-nd row 交換, 而第二個 elementary matrix E_2 為將 I_3 的 1-st row 乘上 -2 加到 2-nd row 上. 第三個 elementary matrix E_3 為將 I_3 的 1-st row 乘上 -4 加到 3-rd row 上. 接下來的 elementary matrix E_4 為將 I_3 的 2-nd row 乘上 $1/2$, 而最後一個 elementary matrix E_5 為將 I_3 的 2-nd row 乘上 -1 加到 1-st row 上. 也就是說, 我們有

$$E_1=\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_2=\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_3=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix},$$

$$E_4=\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_5=\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

將這五個 elementary matrices 由右而左依序相乘, 確實得

$$E_5E_4E_3E_2E_1=\begin{bmatrix} -\frac{1}{2} & 2 & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & -4 & 1 \end{bmatrix}=E.$$

最後我們要再次強調 elementary matrices 之間相乘是沒有交換性的. 事實上一個矩陣若右邊乘上一個 elementary matrix, 是對此矩陣做所謂的 elementary column operation. 這用類似前面的看法, 改用 column 的觀點來處理矩陣乘法 (Definition 3.1.6) 便可看出. 也就是說若 E 為將 identity I_n 的 i-th row 和 j-th row 交換的 elementary matrix, 其實也可將 E 視為將 I_n 的 i-th column 和 j-th column 交換. 所以若將 E 乘在一個 $m\times n$ matrix A 的右邊, 所得的矩陣 AE, 事實上就是將 A 的 i-th column 和 j-th column 交換. 同樣的, 若 E 為將 identity I_n 的 i-th row 乘上一非零實數 r 的 elementary matrix, 我們也可將 E 視為將 I_n 的 i-th column 乘上 r. 所以若將 E 乘在一個 $m\times n$ matrix A 的右邊, 所得的矩陣 AE, 事實上就是將 A 的 i-th column 乘上 r. 不過要特別注意的是將 I_n 的 i-th row 乘上 r 加到 j-th row 的 elementary matrix E. 這個 elementary matrix 若視為 column operation, 是將 I_n 的 j-th column 乘上 r 加在 i-th column 上 (不是 i-th column 乘上 r 加在 j-th column 上). 所以若將 E 乘在一個 $m\times n$ matrix A 的右邊, 所得的矩陣 AE, 事實上就是將 A 的 j-th column 乘上 r 加在 i-th column 上.

Example 3.3.3. 考慮

$$E_1=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, E_2=\begin{bmatrix} 10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E_3=\begin{bmatrix} 1 & 0 & 10 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A=\begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix}$$

E_1 可視為將 I_3 的 2-nd row 和 3-rd row 交換, 也可視為將 I_3 的 2-nd column 和 3-rd column 交換. 事實上我們有

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 11 & 22 & 33 \\ -1 & -2 & -3 \end{bmatrix},$$

$$AE_1 = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ -1 & -3 & -2 \\ 11 & 33 & 22 \end{bmatrix}.$$

E_2 可視為將 I_3 的 1-st row 乘以 10, 也可視為將 I_3 的 1-st column 乘以 10. 事實上我們有

$$E_2A = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix} = \begin{bmatrix} 10 & 20 & 30 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix},$$

$$AE_2 = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix} \begin{bmatrix} 10 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 2 & 3 \\ -10 & -2 & -3 \\ 110 & 22 & 33 \end{bmatrix}.$$

E_3 可視為將 I_3 的 3-rd row 乘以 10 加到 1-st row, 也可視為將 I_3 的 1-st column 乘以 10 加到 3-rd column. 事實上我們有

$$E_3A = \begin{bmatrix} 1 & 0 & 10 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix} = \begin{bmatrix} 111 & 222 & 333 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix},$$

$$AE_1 = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 11 & 22 & 33 \end{bmatrix} \begin{bmatrix} 1 & 0 & 10 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 13 \\ -1 & -2 & -13 \\ 11 & 22 & 143 \end{bmatrix}.$$

3.4. Matrix 和 System of Linear Equations 的連結

在上一章, 我們曾經利用矩陣的 rank 來探討其所對應的聯立方程組何時有解以及解是否唯一的問題. 現在我們又知道解一次聯立方程組的問題可以看成矩陣乘法的問題, 這一節中我們就是要用這個觀點進一步探討聯立方程組何時有解以及解是否唯一.

首先由於我們都要用矩陣的乘法來探討, 為了方便起見對於 $\mathbb{R}^n$ 中的向量, 除非特別聲明為 row vector, 我們將一律用 column vector 來表示. 也就是說將它視為一個 $n\times 1$ matrix. 所以若 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, 我們用 $\mathbf{u}^{\mathrm{T}}\mathbf{v}$ 表示 $\mathbf{u},\mathbf{v}$ 的內積 (而不用 $\mathbf{u}\cdot\mathbf{v}$). 另外回顧, 給定一次聯立方程組

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

我們令

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},$$

然後將上面的聯立方程組用 $A\mathbf{x}=\mathbf{b}$ 來表示. 現若 $x_1=c_1,x_2=c_2,\dots,x_n=c_n$, 為此聯立方程組的一組解, 我們便會用

$$\mathbf{c}=\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix},$$

來表示這一組解, 而說 $\mathbf{x}=\mathbf{c}\in\mathbb{R}^n$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解. 依矩陣乘法定義這等同於說 A 這一個 $m\times n$ matrix 乘以 $\mathbf{c}$ 這一個 $n\times 1$ matrix 會等於 $\mathbf{b}$ 這一個 $m\times 1$ matrix, 即 $A\mathbf{c}=\mathbf{b}$.

3.4.1. 解的存在性. 我們再一次探討怎樣的 $m\times n$ matrix A 會滿足對任意的 $\mathbf{b}\in\mathbb{R}^m$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 皆有解.

首先假設 $\mathbf{b}\in\mathbb{R}^m$ 且 $A\mathbf{x}=\mathbf{b}$ 有解. 令 $\mathbf{c}=\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 為一解, 此即表示 $A\mathbf{c}=\mathbf{b}$. 利用矩陣乘法定義得

$$c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{b},$$

其中 $\mathbf{a}_1,\dots,\mathbf{a}_n$ 為 A 的 column vectors. 換句話說, $\mathbf{b}$ 可以寫成 A 的 column vectors 的 linear combination. 用符號來表示就是 $\mathbf{b}\in\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)$. 反之, 若 $\mathbf{b}\in\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)$, 表示存在 $c_1,\dots,c_n\in\mathbb{R}$ 使得 $\mathbf{b}=c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$. 故得 $x_1=c_1,\dots,x_n=c_n$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解. 我們證得了以下的性質.

Lemma 3.4.1. 假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathbf{b}\in\mathbb{R}^m$. 則 $A\mathbf{x}=\mathbf{b}$ 有解若且唯若 $\mathbf{b}\in\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)$, 其中 $\mathbf{a}_1,\dots,\mathbf{a}_n$ 為 A 的 *column vectors.*

我們有興趣於知道怎樣的 $m\times n$ matrix A 會使得對任意的 $\mathbf{b}\in\mathbb{R}^m$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 皆有解. 我們利用過去學過的幾種不同觀點, 發現有許多和它等價的條件. 首先觀察由於 A 的 column vectors 皆在 $\mathbb{R}^m$ 中, 所以自然有 $\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)\subseteq\mathbb{R}^m$. 然而由 Lemma 3.4.1 知, 若對於任意 $\mathbf{b}\in\mathbb{R}^m$ 皆會使得 $A\mathbf{x}=\mathbf{b}$ 有解, 表示對任意 $\mathbf{b}\in\mathbb{R}^m$ 皆有 $\mathbf{b}\in\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)$. 故知此時 $\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)=\mathbb{R}^m$. 反之, 若 $\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)=\mathbb{R}^m$, 表示對任意 $\mathbf{b}\in\mathbb{R}^m$ 皆有 $\mathbf{b}\in\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)$. 同樣由 Lemma 3.4.1 知此即對於任意 $\mathbf{b}\in\mathbb{R}^m$ 皆會使得 $A\mathbf{x}=\mathbf{b}$ 有解. 因此從這觀點來看, 對任意的 $\mathbf{b}\in\mathbb{R}^m$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 皆有解和 $\mathrm{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)=\mathbb{R}^m$ 是等價的.

另外我們可以考慮聯立方程組 $A\mathbf{x}=\mathbf{e}_i$, 其中 $\mathbf{e}_1,\dots,\mathbf{e}_m\in\mathbb{R}^m$ 為 $\mathbb{R}^m$ 的 standard basis. 若已知對任意的 $\mathbf{b}\in\mathbb{R}^m$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 皆有解, 則對所有的 $i=1,\dots,m$, 我們都可找到 $\mathbf{c}_i\in\mathbb{R}^n$ 使得 $\mathbf{x}=\mathbf{c}_i$ 為聯立方程組 $A\mathbf{x}=\mathbf{e}_i$ 的一組解. 也就是說對所有的 $i=1,\dots,m$ 皆有 $A\mathbf{c}_i=\mathbf{e}_i$. 現考慮 $n\times m$ matrix C, 其 i-th column 就是 $\mathbf{c}_i$. 此時依矩陣乘法的定義我們有

$$AC=A\begin{bmatrix} | & | & & | \\ \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_m \\ | & | & & | \end{bmatrix}=\begin{bmatrix} | & | & & | \\ A\mathbf{c}_1 & A\mathbf{c}_2 & \cdots & A\mathbf{c}_m \\ | & | & & | \end{bmatrix}=\begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_m \\ | & | & & | \end{bmatrix}=I_m.$$

也就是說, 此時必存在 $n \times m$ matrix C 使得 $AC = I_m$. 反之, 若 C 為 $n \times m$ matrix 滿足 $AC = I_m$, 則對任意 $\mathbf{b} \in \mathbb{R}^m$, 我們考慮 $\mathbf{c} = C\mathbf{b} \in \mathbb{R}^n$, 皆會有

$$A\mathbf{c} = A(C\mathbf{b}) = (AC)\mathbf{b} = I_m\mathbf{b} = \mathbf{b}.$$

也就是說此時對任意 $\mathbf{b} \in \mathbb{R}^m$, 我們都可以找到 $\mathbf{c} = C\mathbf{b} \in \mathbb{R}^n$, 使得 $\mathbf{x} = \mathbf{c}$ 是聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解. 因此從這觀點來看, 對任意的 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解和存在 $n \times m$ matrix C 使得 $AC = I_m$ 是等價的.

我們也曾經在 2.3 節中利用 rank 探討過這個問題 (參見 Proposition 2.3.2). 也就是說, 若 A 經由 elementary row operations 化為 echelon form 後, 其 pivot 的個數恰等於 A 的 row 的個數 m, 表示 A 的 echelon form 沒有一個 row 全為 0, 故由 2.1 節的討論 (即 Case (1)) 知此時任意的 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解. 反之, 如果 pivot 的個數不等於 m, 表示 A 的 echelon form A' 中最後一個 row 必全為 0. 此時我們一定可以找到 $\mathbf{b} \in \mathbb{R}^m$ 使得增廣矩陣 $[A|\mathbf{b}]$ 化為 echelon form $[A'|\mathbf{b}']$ 後, $\mathbf{b}'$ 最後一個 entry 不為 0 (即 Case 2(a)). 此時 $A\mathbf{x} = \mathbf{b}$ 會無解. 因此從這觀點來看, 對任意的 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解和 A 的 echelon form 的 pivot 的個數為 m (即 $\operatorname{rank}(A) = m$) 是等價的.

綜合上面這幾種看法, 我們證得了以下這個非常重要的定理.

Theorem 3.4.2. 假設 A 為 $m \times n$ *matrix,* 令 $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ 為 A 的 *column vectors.* 以下各敘述是等價的.

(1) 對任意的 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解.

(2) $\operatorname{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \mathbb{R}^m$.

(3) $\operatorname{rank}(A) = m$.

(4) 存在 $n \times m$ *matrix* C 使得 $AC = I_m$.

特別提醒一下, Theorem 3.4.2, 指的是對所有 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解的情況. 所以若僅知單一的 $\mathbf{b}$ 使得聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解, Theorem 3.4.2 並不適用 (不過 Lemma 3.4.1 是適用的).

我們曾提及, 當 A 為 $m \times n$ matrix, 將 A 化為 echelon form 後其 pivot 的個數不能多於 row 和 column 的個數. 也就是說 pivot 的個數應小於等於 $\min\{m,n\}$ (此指的是 m,n 中最小的那一個). 所以若 pivot 的個數為 m, 則表示 $n \geq m$. 換言之, 若 $n < m$, 我們便知 pivot 的個數不可能等於 m, 所以 Theorem 3.4.2 中的情況不可能發生. 我們有以下的結論.

Corollary 3.4.3. 假設 A 為 $m \times n$ *matrix,* 其中 $n < m$, 則必存在 $\mathbf{b} \in \mathbb{R}^m$ 使得聯立方程組 $A\mathbf{x} = \mathbf{b}$ 無解. 而且此時, 不會存在 $n \times m$ *matrix* C 使得 $AC = I_m$.

Proof. 由前所述, 當 $n < m$ 時 A 化為 echelon form 後, 其 pivot 的個數不可能為 m, 亦即 $\operatorname{rank}(A) < m$. 故由 Theorem 3.4.2 知不可能對任意的 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解. 亦即存在 $\mathbf{b} \in \mathbb{R}^m$ 使得聯立方程組 $A\mathbf{x} = \mathbf{b}$ 無解. 同理, 由 Theorem 3.4.2 知不會存在 $n \times m$ matrix C 使得 $AC = I_m$. □

Question 3.8. 假設 A 為 $m \times n$ *matrix*, 其中 $m < n$. 是否存在 $n \times m$ *matrix* C 使得 $CA = I_n$?

前面提過 Theorem 3.4.2 是個很重要的定理, 它可以告訴我們一些解聯立方程組的訊息. 例如 Corollary 3.4.3 就是告訴我們當方程式的個數多於未知數的個數時, 會存在 $\mathbf{b} \in \mathbb{R}^m$ 使得聯立方程組 $A\mathbf{x} = \mathbf{b}$ 無解. 我們也可已將它運用於有關於 spanning set 的問題.

Corollary 3.4.4. 設 m, n 為正整數且 $n < m$. 任取 $\mathbb{R}^m$ 中 n 個 *vectors* $\mathbf{a}_1, \ldots, \mathbf{a}_n$, 則 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \neq \mathbb{R}^m$.

Proof. 我們用反證法, 假設存在 $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ 使得 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \mathbb{R}^m$. 令 A 為由 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 為 column 所組成的 $m \times n$ matrix, 即對於 $i = 1, \ldots, n$, A 的 i-th column 為 $\mathbf{a}_i$. 則由 Theorem 3.4.2 知 $\mathrm{rank}(A) = m$. 但矩陣 A 的 column 的個數小於 row 的個數 (即 $n < m$), 得 $\mathrm{rank}(A) \leq n < m$ 之矛盾. 由此矛盾得知 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \neq \mathbb{R}^m$. □

Corollary 3.4.4 告訴我們, 在 $\mathbb{R}^m$ 中任取少於 m 個向量, 它們的線性組合是不能展成整個 $\mathbb{R}^m$ 的.

3.4.2. 解的唯一性. 所謂聯立方程組解的唯一性, 指的是假設聯立方程組有解時, 探討其解是否唯一. 所以唯一性並不涉及解是否存在的問題.

給定 $A \in \mathscr{M}_{m \times n}$ 以及 $\mathbf{b} \in \mathbb{R}^m$. 我們曾說明過如果 $A\mathbf{x} = \mathbf{b}$ 有解, 則 $A\mathbf{x} = \mathbf{b}$ 的解和 $A\mathbf{x} = \mathbf{O}$ 的解 (這裡 $\mathbf{O}$ 是 $\mathbb{R}^m$ 的零向量) 息息相關 (參見 Proposition 2.3.6). 這個事實現在由於可以利用矩陣的運算給予更簡明的證明, 所以我們再敘述及證明一次.

Lemma 3.4.5. 給定 $A \in \mathscr{M}_{m \times n}$ 以及 $\mathbf{b} \in \mathbb{R}^m$ 且假設 $\mathbf{x} = \mathbf{c} \in \mathbb{R}^n$ 是聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解. 則

(1) 若 $\mathbf{x} = \mathbf{c}' \in \mathbb{R}^n$ 是 $A\mathbf{x} = \mathbf{b}$ 的一組解, 則 $\mathbf{x} = \mathbf{c}' - \mathbf{c}$ 為 $A\mathbf{x} = \mathbf{O}$ 的一組解.

(2) 若 $\mathbf{x} = \mathbf{u} \in \mathbb{R}^n$ 為 $A\mathbf{x} = \mathbf{O}$ 的一組解, 則 $\mathbf{x} = \mathbf{c} + \mathbf{u}$ 是 $A\mathbf{x} = \mathbf{b}$ 的一組解.

Proof. (1) 假設 $\mathbf{x} = \mathbf{c}' \in \mathbb{R}^n$ 是 $A\mathbf{x} = \mathbf{b}$ 的一組解, 意即 $A\mathbf{c}' = \mathbf{b}$. 由已知 $A\mathbf{c} = \mathbf{b}$ 得

$$A(\mathbf{c}' - \mathbf{c}) = A\mathbf{c}' - A\mathbf{c} = \mathbf{b} - \mathbf{b} = \mathbf{O}.$$

因此 $\mathbf{x} = \mathbf{c}' - \mathbf{c}$ 會是 $A\mathbf{x} = \mathbf{O}$ 的一組解.

(2) 若 $\mathbf{x} = \mathbf{u} \in \mathbb{R}^n$ 為 $A\mathbf{x} = \mathbf{O}$ 的一組解, 則

$$A(\mathbf{c} + \mathbf{u}) = A\mathbf{c} + A\mathbf{u} = \mathbf{b} + \mathbf{O} = \mathbf{b}.$$

得證 $\mathbf{x} = \mathbf{c} + \mathbf{u}$ 為聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解. □

Lemma 3.4.5 告訴我們若已知 $\mathbf{x} = \mathbf{c}$ 為 $A\mathbf{x} = \mathbf{b}$ 的一組解, 且知道 $A\mathbf{x} = \mathbf{O}$ 所有的解, 就能利用 $\mathbf{c}$ 以及 $A\mathbf{x} = \mathbf{O}$ 所有的解得到 $A\mathbf{x} = \mathbf{b}$ 所有的解. 所以了解 $A\mathbf{x} = \mathbf{O}$ 所有的解是很重要的課題 (以後我們會深入探討). 回顧一下 $A\mathbf{x} = \mathbf{O}$ 這樣的 linear system, 我們稱之為 homogeneous linear system. Homogeneous linear system 一定有解, 事實上當 $A \in \mathscr{M}_{m \times n}$ 時, $x_1 = 0, \ldots, x_n = 0$ 就是 $A\mathbf{x} = \mathbf{O}$ 的一組解. 這組解 $\mathbf{x} = \mathbf{O} \in \mathbb{R}^n$ 因為不需任何計算就能得到,

我們稱之為 $A\mathbf{x}=\mathbf{O}$ 的 trivial solution. 注意 trivial solution $\mathbf{x}=\mathbf{O}$ 這裡的 $\mathbf{O}$ 是 $\mathbb{R}^n$ 的零向量, 而 $A\mathbf{x}=\mathbf{O}$ 這裡的 $\mathbf{O}$ 是 $\mathbb{R}^m$ 的零向量, 所以雖然我們用同樣的符號表示, 但當 $n\neq m$ 時它們是不同的, 大家需區分清楚. 當一個 homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 除了 trivial solution 外還有其他的 solution (即解不唯一), 我們稱這些不為 $\mathbf{O}$ 的 solution 為 nontrivial solution.

從 Lemma 3.4.5 我們知, 若 $A\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution (即解唯一), 則對於 $\mathbf{b}\in\mathbb{R}^m$, 若 $A\mathbf{x}=\mathbf{b}$ 有解, 其解必唯一. 由這觀點, 我們可以得到以下關於聯立方程組解的唯一性的重要定理.

Theorem 3.4.6. *假設 A 為 $m\times n$ matrix. 以下各敘述是等價的.*

(1) *若 $\mathbf{b}\in\mathbb{R}^m$ 且聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 則解唯一.*

(2) *Homogeneous system $A\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution.*

(3) $\operatorname{rank}(A)=n$.

(4) *存在 $n\times m$ matrix B 使得 $BA=I_n$.*

Proof. 在 Corollary 2.3.7 中我們已證得 (1),(2),(3) 是等價的. 故僅需證明 $(3)\Rightarrow(4)$ 以及 $(4)\Rightarrow(1)$.

$(3)\Rightarrow(4)$: 假設 $\operatorname{rank}(A)=n$, 即 A 化為 echelon form 後, 其 pivot 的個數為 n. 考慮將 A 化為 reduced echelon form A'. 此時 A' 由於有 n 個 pivot, 所以每一個 pivot 必分別在 A' 前面 n 個 row 上. 而又 A' 為 $m\times n$ matrix, 有 n 個 column. 所以 A' 每一個 pivot 必落在 (i,i)-th entry, 其中 $1\leq i\leq n$. 又因為 A' 為 reduced echelon form, 此 n 個 pivots 的值皆為 1. 然而 reduced echelon form 每一個 pivot 所在的 column, 除了 pivot 所在位置外, 其他位置應為 0, 所以我們知 A' 必為以下的 matrix $A'=\begin{bmatrix} I_n \\ \mathbf{O}\end{bmatrix}$, 即 A' 的前 n 個 row 就是 I_n. 由 Lemma 3.3.1, 我們知存在 $m\times m$ matrix E 使得 $EA=A'$. 現若令 E 的 i-th row 為 ${}_i\varepsilon$, 由 row 的觀點看矩陣乘法 (參見圖示 (3.15)), 我們有

$$EA=\begin{bmatrix} — & {}_1\varepsilon & — \\ & \vdots & \\ — & {}_n\varepsilon & — \\ & \vdots & \\ — & {}_m\varepsilon & — \end{bmatrix} A=\begin{bmatrix} — & {}_1\varepsilon A & — \\ & \vdots & \\ — & {}_n\varepsilon A & — \\ & \vdots & \\ — & {}_m\varepsilon A & — \end{bmatrix}=\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \\ & \mathbf{O} & \end{bmatrix}.$$

現若令 B 為 $n\times m$ matrix, 對於 $i=1,\ldots,n$, 其 i-th row 為 ${}_i\varepsilon$ (即 B 為截取 E 的前 n 個 row 的 $n\times m$ matrix), 則由前述的矩陣乘法性質知

$$BA=\begin{bmatrix} — & {}_1\varepsilon & — \\ & \vdots & \\ — & {}_n\varepsilon & — \end{bmatrix} A=\begin{bmatrix} — & {}_1\varepsilon A & — \\ & \vdots & \\ — & {}_n\varepsilon A & — \end{bmatrix}=\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}=I_n.$$

$(4)\Rightarrow(1)$: 我們利用反證法假設 $\mathbf{c}\neq\mathbf{c}'\in\mathbb{R}^n$ 且 $\mathbf{x}=\mathbf{c},\mathbf{x}=\mathbf{c}'$ 皆為 $A\mathbf{x}=\mathbf{b}$ 的一組解. 亦即, $A\mathbf{c}=\mathbf{b}$ 且 $A\mathbf{c}'=\mathbf{b}$. 現已知存在 $B\in\mathscr{M}_{n\times m}$ 使得 $BA=I_n$, 故得 $B\mathbf{b}=B(A\mathbf{c})=(BA)\mathbf{c}=\mathbf{c}$ 且

$B\mathbf{b} = B(A\mathbf{c}') = (BA)\mathbf{c}' = \mathbf{c}'$. 此結果 $\mathbf{c} = B\mathbf{b} = \mathbf{c}'$ 與當初假設 $\mathbf{c} \neq \mathbf{c}'$ 相矛盾, 故得證若 $A\mathbf{x} = \mathbf{b}$ 有解, 則解必唯一. □

再次提醒, Theorem 3.4.6, 並不能知道聯立方程組 $A\mathbf{x} = \mathbf{b}$ 是否有解. 它告訴我們若已知 $A\mathbf{x} = \mathbf{O}$ 有 nontrivial solution, 則 $A\mathbf{x} = \mathbf{b}$ 要不然無解, 要不然會有無窮多解.

Question 3.9. *假設 A 為 $m \times n$ matrix, $\mathbf{b}, \mathbf{b}' \in \mathbb{R}^m$. 若已知 $A\mathbf{x} = \mathbf{b}$, $A\mathbf{x} = \mathbf{b}'$ 皆有解, 且 $A\mathbf{x} = \mathbf{b}$ 的解唯一. 是否 $A\mathbf{x} = \mathbf{b}'$ 的也會唯一?*

前面已提過, 當 A 為 $m \times n$ matrix, 將 A 化為 echelon form 後其 pivot 的個數應小於等於 $\min\{m,n\}$. 所以若 pivot 的個數為 n, 則表示 $m \geq n$. 換言之, 若 $m < n$, 我們便知 pivot 的個數不可能等於 n, 所以 Theorem 3.4.6 中的情況不可能發生. 我們有以下的結論.

Corollary 3.4.7. *假設 A 為 $m \times n$ matrix, 其中 $m < n$. 若 $\mathbf{b} \in \mathbb{R}^m$ 且聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解, 則必有無窮多解. 而且此時, 不會存在 $n \times m$ matrix B 使得 $BA = I_n$.*

Proof. 由前所述, 當 $m < n$ 時 A 化為 echelon form 後, 其 pivot 的個數不可能為 n. 故由 Theorem 3.4.6 知 homogeneous linear system $A\mathbf{x} = \mathbf{O}$ 有 nontrivial solution. 亦即在 $\mathbb{R}^n$ 存在非零向量 $\mathbf{c}$ 使得 $\mathbf{x} = \mathbf{c}$ 為 $A\mathbf{x} = \mathbf{O}$ 之一組解. 此時考慮任何時數 r, 由矩陣乘法性質 (Proposition 3.1.9) 知,

$$A(r\mathbf{c}) = r(A\mathbf{c}) = r\mathbf{O} = \mathbf{O}.$$

亦即 $x = r\mathbf{c}$ 也是 $A\mathbf{x} = \mathbf{O}$ 的一組解. 故知 $A\mathbf{x} = \mathbf{O}$ 有無窮多解, 所以 Lemma 3.4.5 告訴我們, 若 $A\mathbf{x} = \mathbf{b}$ 有解, 則必有無窮多解.

另一方面 Theorem 3.4.6 也告訴我們若 pivot 的個數不是 n, 則不會存在 $n \times m$ matrix B 使得 $BA = I_n$. □

Theorem 3.4.6 也和 Theorem 3.4.2 一樣是很重要的定理, 它可以告訴我們一些解聯立方程組的訊息. 例如 Corollary 3.4.7 就是告訴我們當方程式的個數少於未知數的個數時, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 不可能解唯一. 我們也可已將它運用於有關於“線性相關”的問題.

Corollary 3.4.8. *設 m,n 為正整數且 $m < n$. 任取 $\mathbb{R}^m$ 中 n 個 vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, 則存在 $c_1, \ldots, c_n \in \mathbb{R}$ 不全為 0 使得 $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{O}$.*

Proof. 令 A 為由 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 為 column 所組成的 $m \times n$ matrix. 由於矩陣 A 的 row 的個數小於 column 的個數 (即 $m < n$), 由 Corollary 3.4.7 知聯立方程組 $A\mathbf{x} = \mathbf{O}$ 有 nontrivial solution $x_1 = c_1, \ldots, x_n = c_n$. 由矩陣乘法定義, 此即表示 $c_1, \ldots, c_n$ 不全為 0 且 $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{O}$. □

假設 $c_1, \ldots, c_n$ 不全為 0 且 $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n = \mathbf{O}$. 不失一般性我們可假設 $c_1 \neq 0$, 此時

$$\mathbf{a}_1 = \frac{-c_2}{c_1}\mathbf{a}_2 + \cdots + \frac{-c_n}{c_1}\mathbf{a}_n.$$

這表示 $\mathbf{a}_1$ 可以用其他的向量 $\mathbf{a}_2, \ldots, \mathbf{a}_n$ 的線性組合來表達. 因此我們便會稱 $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ 這一組向量為 *linearly dependent* (線性相關). 一組向量之間是否為 linearly dependent 也是

線性代數中很重要的課題, 以後我們會再正式的給定義並深入的探討. 這裡, Corollary 3.4.8 告訴我們, 在 $\mathbb{R}^m$ 中任取多於 m 個向量, 它們必為 linearly dependent.

3.5. Invertible Matrix

所謂 invertible matrix 就是 "可逆矩陣". 我們會發現只有 square matrix 才有可能是 invertible matrix, 但並不是所有的 square matrix 都是 invertible matrix. 這一節中我們會探討有關 invertible matrix 的相關性質, 並介紹判斷一個方陣是否為 invertible 且找出其反矩陣的方法.

當初我們將聯立方程組用矩陣乘法的方式 $A\mathbf{x} = \mathbf{b}$ 來表示, 其中有一個很大的目的就是希望將解聯立方程式的問題此簡化成類似實數上解 $ax = b$ 的情形. 在實數情況, 當 $a \neq 0$ 時, $ax = b$ 的解就是很簡單的 $x = ba^{-1}$. 但在矩陣的情形, 我們沒有除法, 所以只能借助乘法來幫忙. 由於實數中 a^{-1} 有 $a^{-1}a = aa^{-1} = 1$ 的性質, 所以推廣這個概念至矩陣, 我們便希望找到矩陣滿足 BA 以及 AB 為 identity. 不過當 A 為 $m \times n$ matrix 且 $m \neq n$ 時, 由 Corollary 3.4.3 以及 Corollary 3.4.7, 我們知這是不可能存在的. 所以我們僅對 $m = n$, 即 A 為 square matrix 時有以下的定義.

Definition 3.5.1. 假設 A 為 $n \times n$ square matrix, 若存在 $n \times n$ matrix B 使得 $AB = BA = I_n$, 則稱 A 為 *invertible*. 反之, 我們稱 A 為 *non-invertible*

再一次強調當 A 不是方陣時, 我們知 A 絕對不是 invertible. 因此當我們不知矩陣 A 的階數時, 絕對不能用存在 B 滿足 BA 為 identity 來說 A 為 invertible, 必須檢查另一邊 AB 亦為 identity 才可. 不過當 A 為 $n \times n$ square matrix, 確實檢查單邊就可以確定 A 為 invertible. 我們有以下的性質.

Theorem 3.5.2. 假設 A 為 $n \times n$ *square matrix.* 則下列是等價的.

(1) A 為 *invertible matrix.*

(2) 存在 $n \times n$ *matrix* B 使得 $BA = I_n$.

(3) $\mathrm{rank}(A) = n$.

(4) 存在 $n \times n$ *matrix* C 使得 $AC = I_n$.

Proof. 依 A 為 invertible 的定義, 我們知若 A 為 invertible, 則存在 $B \in \mathscr{M}_{n\times n}$ 使得 $BA = I_n$. 故 $(1) \Rightarrow (2)$.

由 A 為 $n \times n$ matrix 以及 Theorem 3.4.6 知存在 $B \in \mathscr{M}_{n\times n}$ 使得 $BA = I_n$ 若且唯若 A 化為 echelon form 後 pivot 的個數為 n. 故 $(2) \Leftrightarrow (3)$.

同理, 由 A 為 $n \times n$ matrix 以及 Theorem 3.4.2 知存在 $C \in \mathscr{M}_{n\times n}$ 使得 $AC = I_n$ 若且唯若 A 化為 echelon form 後 pivot 的個數為 n. 故 $(3) \Leftrightarrow (4)$.

最後, 由 A 化為 echelon form 後 pivot 的個數為 n 知存在 $B, C \in \mathscr{M}_{n\times n}$ 使得 $BA = I_n$ 以及 $AC = I_n$. 若能證得 $C = B$, 則由 $BA = AB = I_n$ 得證 A 為 invertible. 然而由 $BA = I_n$, 得 $(BA)C = I_nC = C$. 又由 $(BA)C = B(AC) = BI_n = B$, 得證 $B = C$. 故 $(3) \Rightarrow (1)$. 得證本定理. □

我們在前一章曾提過當一個 $n \times n$ matrix 的 rank 為 n 時, 我們稱之為 nonsingular matrix. 所以由 Theorem 3.5.2 我們知 invertible matrix 就是 nonsingular matrix. 反之, non-invertible matrix 就是 singular matrix. 不過為了讓大家不被這麼多名詞弄混. 以後我們僅用 invertible 和 non-invertible 而不用 nonsingular 和 singular.

由 Theorem 3.5.2 的證明我們知 A 為 $n \times n$ square matrix, 若存在 $B, C \in \mathscr{M}_{n\times n}$ 使得 $BA = I_n$ 且 $AC = I_n$, 則 $B = C$. 我們自然會問有沒有可能存在不同的 $B, B' \in \mathscr{M}_{n\times n}$ 皆滿足 $BA = I_n$ 以及 $B'A = I_n$. 下一個定理告訴我們這樣的方陣其實是唯一的.

Corollary 3.5.3. 假設 A 為 $n \times n$ *square matrix* 且 $B, B' \in \mathscr{M}_{n\times n}$ 滿足 $BA = I_n$ 以及 $B'A = I_n$. 則 $B = B'$.

Proof. 由 Theorem 3.5.2 我們知 A 為 invertible 且由其證明知 $BA = AB = I_n$ 以及 $B'A = AB' = I_n$. 故

$$B = I_nB = (B'A)B = B'(AB) = B'I_n = B'.$$

$\square$

由 Corollary 3.5.3, 我們知道若 A 為 $n \times n$ invertible matrix, 則僅會存在唯一的一個 $n \times n$ matrix B 滿足 $BA = AB = I_n$. 它和 A 的關係如同在實數上非零實數的乘法的 inverse (乘法反元素), 所以我們給以下的定義.

Definition 3.5.4. 假設 A 為 $n \times n$ invertible matrix. 我們稱唯一滿足 $BA = AB = I_n$ 的 $n \times n$ matrix B 為 A 的 *inverse* (反矩陣), 且用 A^{-1} 表示.

給定一 $n \times n$ invertible matrix A 由於其反矩陣是唯一的, 所以若要確定 $B = A^{-1}$ 我們僅要檢查是否 $BA = I_n$ 或 $AB = I_n$ 即可. 我們有以下之性質

Proposition 3.5.5. 假設 $A, B \in \mathscr{M}_{n\times n}$. 我們有以下之性質

(1) 若 A 為 *invertible,* 則 A^{-1} 亦為 *invertible* 且

$$(A^{-1})^{-1} = A.$$

(2) A 為 *invertible* 若且唯若 A^{T} 為 *invertible* 且此時

$$(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}.$$

(3) A, B 皆為 *invertible* 若且唯若 AB 為 *invertible.* 且此時

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. 由 Theorem 3.5.2, 我們要說一個 $n \times n$ matrix 為 invertible, 只要找到 $B \in \mathscr{M}_{n\times n}$ 使得 $BA = I_n$ 或 $AB = I_n$ 且此時由唯一性 (Corollary 3.5.3) 知 $B = A^{-1}$.

(1) 依定義 A^{-1} 亦為 $n \times n$ matrix 故 Theorem 3.5.2 適用. 利用 $A^{-1}A = I_n$, 得知 A^{-1} 亦為 invertible 且 $(A^{-1})^{-1} = A$.

(2) 依定義 A^{T} 亦為 $n\times n$ matrix 故 Theorem 3.5.2 適用. 由 $A^{-1}A=I_n$ 利用 Proposition 3.2.4 得

$$I_n=(A^{-1}A)^{\mathrm{T}}=A^{\mathrm{T}}(A^{-1})^{\mathrm{T}}$$

故知 A^{T} 為 invertible 且 $(A^{\mathrm{T}})^{-1}=(A^{-1})^{\mathrm{T}}$. 反之若 A^{T} 為 invertible, 由前知 $(A^{\mathrm{T}})^{\mathrm{T}}$ 為 invertible, 故由利用 Proposition 3.2.4 $(A^{\mathrm{T}})^{\mathrm{T}}=A$ 得證 A 為 invertible.

(3) 依定義 AB 為 $n\times n$ matrix 故 Theorem 3.5.2 適用. 現若 A,B 皆為 invertible, 則由

$$(AB)B^{-1}A^{-1}=A(BB^{-1})A^{-1}=AI_nA^{-1}=AA^{-1}=I_n,$$

得證 AB 為 invertible 且 $(AB)^{-1}=B^{-1}A^{-1}$. 反之, 若 AB 為 invertible, 且令 $C=(AB)^{-1}$. 此時由 $(AB)C=I_n$ 得 $A(BC)=I_n$, 故由假設 A 為 $n\times n$ matrix 以及 Theorem 3.5.2 得證 A 為 invertible. 同理, 由 $C(AB)=I_n$, 得 $(CA)B=I_n$, 得證 B 為 invertible. □

要注意 Proposition 3.5.5 (3) 中由 AB invertible 推得 A,B 皆為 invertible 是需要用到 A,B 皆為 $n\times n$ matrix. 否則當 $m\neq n$ 時, 在 Theorem 3.4.2 中我們知道有可能 $A\in\mathscr{M}_{m\times n},C\in\mathscr{M}_{n\times m}$ 滿足 $AC=I_m$. 此時 I_m 為 invertible, 但 A,C 皆為 non-invertible. 同樣的, 當 A,B 為方陣時, 因為由 AB 為 invertible 可推得 A,B 皆為 invertible, 故知 BA 亦為 invertible. 也就是說當 A,B 為方陣時 AB 為 invertible 和 BA 為 invertible 是等價的. 但在 A,B 不為方陣時, 若 AB 為 invertible 會導致 BA 不為 invertible.

Question 3.10. 試舉例 A,B 不為 *invertible* 但 AB 為 *invertible.* 同時也驗證此時 BA 為 *non-invertible.*

接下來我們探討如何判別一個具體的 $n\times n$ matrix 是否為 invertible, 且若為 invertible 如何找出其 inverse. 這個問題可藉由將方陣利用 elementary row operations 化為 reduced echelon form 來處理. 事實上, 當 A 為 $n\times n$ matrix, 由 Theorem 3.5.2 我們知道 A 為 invertible 若且唯若 A 化為 echelon form 後其 pivot 的個數等於 n. 因此我們只要將 A 化為 echelon form 後計算其 pivot 的個數, 便可以知道 A 是否為 invertible. 若 A 為 invertible, 即 pivot 的個數為 n, 此時由於 A 的 reduced echelon form 為 $n\times n$ matrix, 故得 A 的 reduced echelon form 為 I_n. 也就是說我們可以用 elementary row operations 將 A 化為 I_n. 故由 Lemma 3.3.1 我們知存在 $E\in\mathscr{M}_{n\times n}$ 為一些 elementary matrix 的乘積使得 $EA=I_n$. 事實上若將 augmented matrix $[A|I_n]$ 利用 elementary row operations 化為 $[I_n|E]$, 則 $EA=I_n$, 故此時 E 就是 A^{-1}. 我們看以下的例子.

Example 3.5.6. 考慮矩陣

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

我們要決定是否 A 是否為 invertible. 若為 invertible, 要找出 A^{-1}.

我們直接考慮 augmented matrix $[A|I_4]$, 利用 elementary row operation 將 A 的部分轉換成 echelon form. 首先將 1-st row 分別乘上 -1, 3 加至 3-rd, 4-th row, 即

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -3 & 4 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 2 & 0 & 0 & 1 & 0 \\ -3 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{cccc|cccc} 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -3 & 4 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 3 & -1 & 0 & 1 & 0 \\ 0 & 0 & 3 & -4 & 3 & 0 & 0 & 1 \end{array}\right].$$

接著將 3-rd row 乘上 $3/2$ 加至 4-th row 得

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -3 & 4 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 3 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{3}{2} & 0 & \frac{3}{2} & 1 \end{array}\right].$$

此時 augmented matrix 左半部為 echelon form, 其 pivot 的個數為 4, 故知 A 為 invertible. 我們繼續將左半部化為 reduced echelon form 便可得到 A^{-1}.

先將 4-th row 乘以 2, 然後將所得的 augmented matrix 的 4-th row 分別乘上 -3, -4, 1 加至 3-rd, 2-nd 和 1-st row, 即

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -3 & 4 & 0 & 1 & 0 & 0 \\ 0 & 0 & -2 & 3 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 & 3 & 2 \end{array}\right] \sim \left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 4 & 0 & 3 & 2 \\ 0 & -1 & -3 & 0 & -12 & 1 & -12 & -8 \\ 0 & 0 & -2 & 0 & -10 & 0 & -8 & -6 \\ 0 & 0 & 0 & 1 & 3 & 0 & 3 & 2 \end{array}\right].$$

接著將 3-rd row 乘以 $-1/2$, 然後將所得的 augmented matrix 的 3-rd row 分別乘上 3, -1 加至 2-nd 和 1-st row, 即

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 1 & 0 & 4 & 0 & 3 & 2 \\ 0 & -1 & -3 & 0 & -12 & 1 & -12 & -8 \\ 0 & 0 & 1 & 0 & 5 & 0 & 4 & 3 \\ 0 & 0 & 0 & 1 & 3 & 0 & 3 & 2 \end{array}\right] \sim \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -1 & 0 & -1 & -1 \\ 0 & -1 & 0 & 0 & 3 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 5 & 0 & 4 & 3 \\ 0 & 0 & 0 & 1 & 3 & 0 & 3 & 2 \end{array}\right].$$

最後將 augmented matrix 的 2-nd row 乘上 -1. 此時所得 augmented matrix 左半部為 reduced echelon form (即 I_4), 故其右半部為 A^{-1}, 即

$$A^{-1} = \begin{bmatrix} -1 & 0 & -1 & -1 \\ -3 & -1 & 0 & -1 \\ 5 & 0 & 4 & 3 \\ 3 & 0 & 3 & 2 \end{bmatrix}.$$

由前面討論我們知當 $A \in \mathscr{M}_{n\times n}$ 為 invertible, 則存在 elementary matrices $E_1,\ldots,E_k$ 使得 $(E_k\cdots E_1)A = I_n$. 亦即 $A^{-1} = E_k\cdots E_1$, 由 Proposition 3.5.5 (3), 我們知 $E_1,\ldots,E_k$ 皆為 invertible, 且由 $(A^{-1})^{-1} = A$, 得 $A = E_1^{-1}\cdots E_k^{-1}$. 事實上這些 elementary matrix E_i 的 inverse 就是將 E_i 還原成 I_n 的 elementary row operation 所對應的 elementary matrix. 也就是說 E_i^{-1} 亦為 elementary matrix. 因此我們有以下的定理.

Proposition 3.5.7. *A 為 invertible matrix 若且唯若 A 為一些 elementary matrices 的乘積.*

Example 3.5.8. 考慮矩陣
$$A=\begin{bmatrix}0&2&0\\1&-1&0\\0&2&1\end{bmatrix}.$$
在求 A 的 inverse 的過程中, 首先我們將 1-st row 和 2-nd row 交換. 令此 elementary row operation 所對應的 elementary matrix 為 E_1. 因用相同的 elementary row operation 可將 E_1 還原成 I_3, 故 $E_1=E_1^{-1}$, 即
$$\left[\begin{array}{ccc|ccc}0&2&0&1&0&0\\1&-1&0&0&1&0\\0&2&1&0&0&1\end{array}\right]\sim\left[\begin{array}{ccc|ccc}1&-1&0&0&1&0\\0&2&0&1&0&0\\0&2&1&0&0&1\end{array}\right], E_1=E_1^{-1}=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}.$$
接著將 augmented matrix 的 2-nd row 乘上 -1 加至 3-rd row. 令此 elementary row operation 所對應的 elementary matrix 為 E_2. 因將 2-nd row 乘上 1 加至 3-rd row 的 elementary row operation 可將 E_2 還原成 I_3, 故所得的 augmented matrix 及 E_2, E_2^{-1} 分別為
$$\left[\begin{array}{ccc|ccc}1&-1&0&0&1&0\\0&2&0&1&0&0\\0&0&1&-1&0&1\end{array}\right], E_2=\begin{bmatrix}1&0&0\\0&1&0\\0&-1&1\end{bmatrix}, E_2^{-1}=\begin{bmatrix}1&0&0\\0&1&0\\0&1&1\end{bmatrix}.$$
然後將 augmented matrix 的 2-nd row 乘上 $1/2$. 令此 elementary row operation 所對應的 elementary matrix 為 E_3. 因將 2-nd row 乘上 2 的 elementary row operation 可將 E_3 還原成 I_3, 故所得的 augmented matrix 及 E_3, E_3^{-1} 分別為
$$\left[\begin{array}{ccc|ccc}1&-1&0&0&1&0\\0&1&0&\frac{1}{2}&0&0\\0&0&1&-1&0&1\end{array}\right], E_3=\begin{bmatrix}1&0&0\\0&\frac{1}{2}&0\\0&0&1\end{bmatrix}, E_3^{-1}=\begin{bmatrix}1&0&0\\0&2&0\\0&0&1\end{bmatrix}.$$
最後將 augmented matrix 的 2-nd row 乘上 1 加至 1-st row. 令此 elementary row operation 所對應的 elementary matrix 為 E_4. 因將 2-nd row 乘上 -1 加至 3-rd row 的 elementary row operation 可將 E_4 還原成 I_3, 故所得的 augmented matrix 及 E_4, E_4^{-1} 分別為
$$\left[\begin{array}{ccc|ccc}1&0&0&\frac{1}{2}&1&0\\0&1&0&\frac{1}{2}&0&0\\0&0&1&-1&0&1\end{array}\right], E_4=\begin{bmatrix}1&1&0\\0&1&0\\0&0&1\end{bmatrix}, E_4^{-1}=\begin{bmatrix}1&-1&0\\0&1&0\\0&0&1\end{bmatrix}.$$
我們檢查可得
$$A^{-1}=E_4E_3E_2E_1=\begin{bmatrix}\frac{1}{2}&1&0\\\frac{1}{2}&0&0\\-1&0&1\end{bmatrix}, A=E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}=\begin{bmatrix}0&2&0\\1&-1&0\\0&2&1\end{bmatrix}.$$

當 A 為 $n\times n$ invertible matrix, Theorem 3.5.2 告訴我們 A 化為 echelon 後其 pivot 的個數為 n. 由於 A 為方陣, 其 row 的個數和 column 的個數皆為 n, 此時很自然地可以將 Theorem 3.4.2 和 Theorem 3.4.6 相連結得到 invertible matrix 與其所對應的聯立方程組相關的性質. 由於以下性質直接套用上述兩個定理就可推得, 我們就不再證明了.

Theorem 3.5.9. 假設 A 為 $n\times n$ *square matrix,* 令 $\mathbf{a}_1,\ldots,\mathbf{a}_n\in\mathbb{R}^n$ 為 A 的 *column vectors.* 則下列是等價的.

(1) A 為 *invertible matrix.*

(2) $\mathrm{Span}(\mathbf{a}_1,\ldots,\mathbf{a}_n)=\mathbb{R}^n$.

(3) 對於任意 $\mathbf{b}\in\mathbb{R}^n$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 皆有解.

(4) 聯立方程組 $A\mathbf{x}=\mathbf{O}$ 沒有 *nontrivial solution.*

(5) 對於任意 $\mathbf{b}\in\mathbb{R}^n$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 皆有解且解唯一.

設 A 為 $n\times n$ invertible matrix, 則對任意 $\mathbf{b}\in\mathbb{R}^n$, $\mathbf{x}=A^{-1}\mathbf{b}$ 為聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的一組解. 這是因為此時 $A\mathbf{x}=A(A^{-1}\mathbf{b})=(AA^{-1})\mathbf{b}=\mathbf{b}$. 又由 Theorem 3.5.9 知此時 $A\mathbf{x}=\mathbf{b}$ 的解唯一. 故 $\mathbf{x}=A^{-1}\mathbf{b}$ 為聯立方程組 $A\mathbf{x}=\mathbf{b}$ 唯一的一組解.

Example 3.5.10. 考慮聯立方程組

$$\begin{array}{rrrrcl} x_1 & & +x_3 & -x_4 & = & b_1 \\ & -x_2 & -3x_3 & +4x_4 & = & b_2 \\ x_1 & & -x_3 & +2x_4 & = & b_3 \\ -3x_1 & & & -x_4 & = & b_4 \end{array}$$

其中 b_1,b_2,b_3,b_4 為任意實數. 由於此時聯立方程組為 $A\mathbf{x}=\mathbf{b}$, 其中 A 為 Example 3.5.6 中的 4×4 matrix A 且 $\mathbf{b}=[b_1\ b_2\ b_3\ b_4]^{\mathrm{T}}$. 因 A 為 invertible, 故由 Theorem 3.5.9 知, 對任意實數 b_1,b_2,b_3,b_4, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 必有解且其解唯一. 事實上此唯一解為

$$\begin{bmatrix} x_1\\x_2\\x_3\\x_4\end{bmatrix}=A^{-1}\mathbf{b}=\begin{bmatrix} -1 & 0 & -1 & -1\\ -3 & -1 & 0 & -1\\ 5 & 0 & 4 & 3\\ 3 & 0 & 3 & 2\end{bmatrix}\begin{bmatrix} b_1\\b_2\\b_3\\b_4\end{bmatrix}=\begin{bmatrix} -b_1-b_3-b_4\\ -3b_1-b_2-b_4\\ 5b_1+4b_3+3b_4\\ 3b_1+3b_3+2b_4\end{bmatrix}.$$

3.6. 結論

在本章中我們定義了矩陣之間的運算且與解聯立方程組相連結. 我們進而發現 elementary row operations 也都可以看成矩陣的乘法. 因而得到聯立方程組解的存在與唯一性與矩陣乘法的特殊關係. 特別當一個矩陣為方陣時, 存在性和唯一性便合而為一. 我們推得一個係數矩陣為 invertible matrix 的聯立方程組一定有解且解唯一. 我們也了解可以利用 elementary row operations 來判斷一個方陣是否為 invertible, 並求出 invertible matrix 的 inverse.

Part II

Vector Space in $\mathbb{R}^m$

這一部分我們將探討線性代數最基本的核心, 就是有關於 vectors space. 我們先利用大家熟悉的 $\mathbb{R}^m$ 中的 subspace, 引進 vector space 的相關概念, 如 basis 及 dimension 等. 接著我們介紹 vector space 上最重要的函數 linear transformation, 不過這裡僅談論 $\mathbb{R}^m$ 上的 linear transformation. 由於 $\mathbb{R}^m$ 有所謂的坐標, 所以我們很自然會將 linear transformation 與矩陣相連結, 而得到所謂的 standard matrix representation. 我們希望讓大家先了解較簡單的 standard matrix representation, 以便將來更進一步了解一般的 matrix representation.

Chapter 4

Subspaces in $\mathbb{R}^m$

通常像 $\mathbb{R}^m$ 這樣, 裡面任意有限多個向量的線性組合仍在 $\mathbb{R}^m$ 中 (此稱為封閉性), 而向量的運算又符合 Proposition 1.2.3 的 8 項規則, 我們便稱之為 vector space. 在 $\mathbb{R}^m$ 中還有其他的 vector space, 我們稱之為 $\mathbb{R}^m$ 的 subspace. 在這一章中我們要探討 $\mathbb{R}^m$ 中 subspaces 的相關性質.

4.1. Subspace and basis

在這一節中, 我們將正式定義 subspace 並引進 basis 的概念.

$\mathbb{R}^m$ 中的非空子集 V, 由於它們裡面的向量仍在 $\mathbb{R}^m$ 中, Proposition 1.2.3 的 8 項規則中除了 (3), (4) 兩項外 (即 $\mathbf{O}\in V$ 以及若 $\mathbf{v}\in V$, 則 $-\mathbf{v}\in V$) 其他的性質皆會符合, 不過若再加上線性組合的封閉性 (即 V 中任意有限多個向量的線性組合仍在 V 中), 則對於 $\mathbf{v}\in V$, 我們便會有 $\mathbf{O}=0\mathbf{v}\in V$ 以及 $-\mathbf{v}=(-1)\mathbf{v}\in V$. 也就是說此時 V 不只有線性組合的封閉性而且向量的運算又符合 Proposition 1.2.3 的 8 項規則, 所以它就會像 $\mathbb{R}^m$ 一樣是一個 vector space 了. 我們有以下的定義.

Definition 4.1.1. 假設 V 為 $\mathbb{R}^m$ 的非空子集. 若任意有限多個 V 中的向量之線性組合仍在 V 中 (即對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 以及 $c_1,\ldots,c_n\in\mathbb{R}$ 皆有 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in V$), 則稱 V 為 $\mathbb{R}^m$ 的一個 *subspace* (子空間).

注意依此定義空集合 $\emptyset$ 不是 subspace. 不過 $\mathbb{R}^m$ 中的零向量所成的集合 $\{\mathbf{O}\}$ 是 $\mathbb{R}^m$ 的 subspace, 而 $\mathbb{R}^m$ 本身也是 $\mathbb{R}^m$ 的 subspace.

回顧一下, 我們用 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 來表示 $\mathbb{R}^m$ 中 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的所有線性組合所成的集合 (參見 Definition 1.3.1), 亦即

$$\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=\{\mathbf{v}\in\mathbb{R}^m \mid r_1\mathbf{v}_1+\cdots+r_n\mathbf{v}_n,\ \text{for some } r_1,\ldots,r_n\in\mathbb{R}\}.$$

所以 V 是 $\mathbb{R}^m$ 的 subspace 等同於對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 皆有

$$\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)\subseteq V. \tag{4.1}$$

依定義要檢查 V 是否為 subspace, 我們必須考慮 V 中任意有限多個向量的線性組合是否仍在 V 中, 感覺起來很麻煩. 事實上, 下一個定理告訴我們, 只要檢查任兩個向量的線性組合即可.

Proposition 4.1.2. 假設 V 為 $\mathbb{R}^m$ 的子集合, 則 V 為 $\mathbb{R}^m$ 的 *subspace* 若且唯若 $\mathbf{O}\in V$ 且對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{R}$ 皆有 $\mathbf{u}+r\mathbf{v}\in V$.

Proof. $(\Rightarrow)$: 依定義 V 為非空集合, 故必存在一向量 $\mathbf{v}\in V$. 現考慮 $0\mathbf{v}=\mathbf{O}$, 依 subspace 的定義 $0\mathbf{v}\in V$, 故得證 $\mathbf{O}\in V$. 同樣的依 subspace 的定義, 對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{R}$ 皆有 $\mathbf{u}+r\mathbf{v}\in V$.

$(\Leftarrow)$: 假設 $\mathbf{O}\in V$ 且對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{R}$ 皆有 $\mathbf{u}+r\mathbf{v}\in V$. 由 $\mathbf{O}\in V$, 我們知 V 為 $\mathbb{R}^m$ 的非空子集. 接著利用式子 (4.1), 我們要證明對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$, 皆有 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)\subseteq V$. 我們對向量的個數 n 作數學歸納法. 首先考慮只有一個向量的情形 (即 $n=1$), 此時我們要證明若 $\mathbf{v}_1\in V$, 則 $\mathrm{Span}(\mathbf{v}_1)\subseteq V$. 因為 $\mathrm{Span}(\mathbf{v}_1)$ 中的元素皆為 $r\mathbf{v}_1$ 的形式, 我們要證明對任意 $r\in\mathbb{R}$ 皆有 $r\mathbf{v}_1\in V$. 此時考慮 $\mathbf{u}=\mathbf{O}$, $\mathbf{v}=\mathbf{v}_1$, 我們有 $r\mathbf{v}_1=\mathbf{O}+r\mathbf{v}_1=\mathbf{u}+r\mathbf{v}$. 依假設知 $\mathbf{u}+r\mathbf{v}\in V$, 得證 $n=1$ 的情形成立. 現假設 $n=k$ 時成立, 亦即對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in V$ 皆有 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)\subseteq V$. 我們要證明對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1}\in V$ 皆有 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1})\subseteq V$. 然而對任意 $\mathbf{w}\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{k+1})$, 我們知存在 $c_1,\ldots,c_k,c_{k+1}\in\mathbb{R}$ 使得 $\mathbf{w}=c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k+c_{k+1}\mathbf{v}_{k+1}$. 此時令 $\mathbf{u}=c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k$, 依歸納假設我們知 $\mathbf{u}\in V$. 故考慮 $r=c_{k+1}\in\mathbb{R}$ 以及 $\mathbf{v}=\mathbf{v}_{k+1}\in V$, 我們有 $\mathbf{w}=\mathbf{u}+r\mathbf{v}\in V$, 得證 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{k+1})\subseteq V$. 故由數學歸納法知, 對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 皆有 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)\subseteq V$, 亦即 V 為 $\mathbb{R}^m$ 的 subspace. □

由 Proposition 4.1.2, 我們知道要檢查 V 是否為 subspace, 我們僅要檢查

(1) $\mathbf{O}\in V$

(2) $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{R}\Rightarrow\mathbf{u}+r\mathbf{v}\in V$.

是否成立即可.

Example 4.1.3. 設 $a,b,c,d\in\mathbb{R}$, 考慮 $S=\{(x,y,z)\in\mathbb{R}^3\mid ax+by+cz=d\}$. 我們想找到 a,b,c,d 使得 S 為 $\mathbb{R}^3$ 的 subspace. 現若 S 為 subspace, 由於已知 $(x,y,z)=(0,0,0)=\mathbf{O}\in S$, 故知 $d=0$. 反之, 若 $d=0$ 且 $\mathbf{u}=(x,y,z)\in S$, $\mathbf{v}=(x',y',z')\in S$ 以及對任意 $r\in\mathbb{R}$, 則因 $ax+by+cz=0, ax'+by'+cz'=0$ 可得

$$a(x+rx')+b(y+ry')+c(z+rz')=(ax+by+cz)+r(ax'+by'+cz')=0,$$

得證 $\mathbf{u}+r\mathbf{v}=(x+rx',y+ry',z+rz')\in S$. 由此知當 $d=0$ 且只有當 $d=0$ 時, S 為 $\mathbb{R}^3$ 的 subspace.

注意在 Example 4.1.3 中, 當 a,b,c,d 皆為 0 時 $S=\mathbb{R}^3$, 仍為 $\mathbb{R}^3$ 的 subspace.

$\mathbb{R}^m$ 中有許多 subspaces, 例如利用 Proposition 1.3.2 我們知道若 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbb{R}^m$, 則 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 是 $\mathbb{R}^m$ 的 subspace (事實上 $\mathbb{R}^m$ 中的 subspace 都可以寫成這種形式). 給定

一個 $m\times n$ matrix A. 我們可以考慮 A 的 column vectors 所展成的集合, 稱之為 A 的 *column space*. 即若 $\mathbf{a}_1,\ldots,\mathbf{a}_n\in\mathbb{R}^m$ 為 A 的 column vectors, 則 $\mathrm{Span}(\mathbf{a}_1,\ldots,\mathbf{a}_n)$ 為 A 的 column space. 同樣的 A 的 row vectors 所展的的集合, 稱為 A 的 *row space*. 注意因 A 為 $m\times n$ matrix, 故其 column space 會是 $\mathbb{R}^m$ 的 subspace, 而 row space 會是 $\mathbb{R}^n$ 的 subspace. 另外還有一個與矩陣 A 相關且重要的 subspace, 它就是 homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 所有解所成的集合, 即 $\{\mathbf{x}\in\mathbb{R}^n\mid A\mathbf{x}=\mathbf{O}\}$. 由於 $\mathbf{x}=\mathbf{O}\in\mathbb{R}^n$ 為 $A\mathbf{x}=\mathbf{O}$ 的一組解. 又若 $\mathbf{x}=\mathbf{u}\in\mathbb{R}^n$, $\mathbf{x}=\mathbf{v}\in\mathbb{R}^n$ 皆為 $A\mathbf{x}=\mathbf{O}$ 的解, 則對於任意 $r\in\mathbb{R}$, 我們有

$$A(\mathbf{u}+r\mathbf{v})=A\mathbf{u}+rA\mathbf{v}=\mathbf{O},$$

故知 $\mathbf{x}=\mathbf{u}+r\mathbf{v}$ 亦為 $A\mathbf{x}=\mathbf{O}$ 的解. 由 proposition 4.1.2 知 $\{\mathbf{x}\in\mathbb{R}^n\mid A\mathbf{x}=\mathbf{O}\}$ 會是 $\mathbb{R}^n$ 的 subspace, 我們稱之為 A 的 *nullspace*.

在 $\mathbb{R}^m$ 中我們曾介紹過所謂 standard basis $\mathbf{e}_1,\ldots,\mathbf{e}_m$, 這一組向量它的特點就是 $\mathbb{R}^m$ 的相量都可以寫成 $\mathbf{e}_1,\ldots,\mathbf{e}_m$ 的線性組合, 而且寫法唯一. 因為這組 standard basis 的存在, 我們可以將 $\mathbb{R}^m$ 的向量坐標化, 而推得 $\mathbb{R}^m$ 上很多性質. 我們自然有興趣知道是否 $\mathbb{R}^m$ 的 subspace 中可找到類似的一組向量有同樣的性質, 因此我們有以下的定義.

Definition 4.1.4. 假設 V 為 $\mathbb{R}^m$ 的 subspace. 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$, 且任意 $\mathbf{v}\in V$, 皆存在唯一的 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{v}$, 則稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 *basis* (基底).

Definition 4.1.4 很容易讓我們了解定義 basis 的目的, 不過要直接利用這個定義找出 basis 會有困難, 所以我們必需找到和此定義等價卻又較容易處理的條件. 首先若考慮以 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbb{R}^m$ 為 column vectors 的 $m\times n$ matrix A, 即

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

由於 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 等同於 $x_1=c_1,\ldots,x_n=c_n$ 是聯立方程組 $A\mathbf{x}=\mathbf{v}$ 的一組解, 所以對任意 $\mathbf{v}\in V$, 存在唯一的 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{v}$, 就等同於對任意 $\mathbf{v}\in V$ 聯立方程組 $A\mathbf{x}=\mathbf{v}$ 皆有解且其解唯一. 所以我們馬上有以下的性質.

Proposition 4.1.5. *假設 V 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$. 令 A 為以 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 column vectors 的 $m\times n$ matrix. 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的一組 basis 若且唯若對任意 $\mathbf{v}\in V$, 聯立方程組 $A\mathbf{x}=\mathbf{v}$ 皆有解且其解唯一.*

注意, 並不是只要一個 $m\times n$ matrix A 會滿足對任意 $\mathbf{v}\in V$, 聯立方程組 $A\mathbf{x}=\mathbf{v}$ 皆有解且解唯一, 則 A 的 column vectors 就會是 V 的一組 basis. 我們還需要 A 的 column vectors 皆在 V 內才行. 也就是說形成 V 的一組 basis 的 vectors 也都必須在 V 中才行. 否則若無此要求, V 中的所有向量都可以用 $\mathbb{R}^m$ 的 standard basis 唯一表示, 那麼 standard basis 會是 $\mathbb{R}^m$ 中所有 subspace 的 basis, 那麼談 basis 就沒甚麼意思了!

當我們實際上有一組向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 我們就可利用 Proposition 4.1.5 以及之前學過的聯立方程組有解且解唯一的等價條件來判斷 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是否為 V 的一組 basis. 不過這個方法

在處理抽象的一般情況就不好處理了, 所以我們還要另一個比較好處理的等價條件來判斷一組向量是否為 basis.

首先當 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$, 由於 V 為 $\mathbb{R}^m$ 的 subspace, 我們知 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n) \subseteq V$. 又當 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis 時, 由於對任意 $\mathbf{v} \in V$ 皆存在 $c_1,\ldots,c_n \in \mathbb{R}$ 使得 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 亦即對任意 $\mathbf{v} \in V$ 皆滿足 $\mathbf{v} \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$. 依此得 $V \subseteq \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 故此時我們有 $V = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$. 對於有這樣性質的 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 我們給它一個特殊名稱.

Definition 4.1.6. 假設 V 為 $\mathbb{R}^m$ 的 subspace. 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \mathbb{R}^m$ 滿足 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n) = V$, 我們稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 *spanning vectors*.

注意若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 spanning vectors, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$. 這是因為 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 故由 $V = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 得 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$.

另一方面當 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis 時, 由於任意 $\mathbf{v} \in V$ 皆存在唯一的 $c_1,\ldots,c_n \in \mathbb{R}$ 使得 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{v}$, 而又因 V 為 subspace, 我們知 $\mathbf{O} \in V$, 故當 $\mathbf{v} = \mathbf{O}$ 時, 存在唯一的 $c_1,\ldots,c_n \in \mathbb{R}$ 使得 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{O}$. 然而已知當 $c_1 = \cdots = c_n = 0$ 時 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{O}$, 故由唯一性知, 不可能會有其他的 $c_1,\ldots,c_n \in \mathbb{R}$ 會使得 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{O}$. 換言之, 只有當 $c_1 = \cdots = c_n = 0$ 才有可能滿足 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{O}$. 對於有這樣性質的 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 我們給它一個特殊名稱.

Definition 4.1.7. 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \mathbb{R}^m$. 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 滿足只有當 $c_1 = \cdots = c_n = 0$ 才有可能使得 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{O}$, 我們稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly independent*.

注意若令 A 為以 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 column vectors 的 $m \times n$ matrix, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent 就表示 homogeneous linear system $A\mathbf{x} = \mathbf{O}$ 沒有 nontrivial solution. 這是因為若存在 $x_1 = c_1,\ldots,x_n = c_n$ 是 $A\mathbf{x} = \mathbf{O}$ 的 nontrivial solution, 表示 $c_1,\ldots,c_n$ 不全為 0 使得 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{O}$, 此與 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent 相矛盾.

關於 spanning vectors 以及 linear independence 的性質, 我們會在下一節更深入討論. 由上面的討論與定義, 我們知若 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$ 為 V 的一組 basis, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的 spanning vectors 且為 linearly independent. 事實上這兩個條件和 basis 是等價的, 我們有以下的定理.

Proposition 4.1.8. *假設 V 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$. 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的一組 basis 若且唯若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的 spanning vectors 且為 linearly independent.*

Proof. 我們僅剩下要證明若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的 spanning vectors 且為 linearly independent 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 V 的一組 basis. 我們想利用 Proposition 4.1.5, 也就是令 A 為以 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 column vectors 的 $m \times n$ matrix, 我們要證明對任意 $\mathbf{v} \in V$, 聯立方程組 $A\mathbf{x} = \mathbf{v}$ 皆有解且其解唯一. 首先若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 spanning vectors, 我們知 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$, 亦即 A 的 column vectors 皆在 V 中, 且由 Lemma 3.4.1 知 $V = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 表示對所有 $\mathbf{v} \in V$, 聯立方程組 $A\mathbf{x} = \mathbf{v}$ 皆有解. 再加上若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent, 表示 homogeneous linear system $A\mathbf{x} = \mathbf{O}$ 沒有 nontrivial solution. 因此由 Theorem 3.4.6 可得聯立方程組 $A\mathbf{x} = \mathbf{v}$ 的解唯一. 故由 Proposition 4.1.5 知此時 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的一組 basis. □

Question 4.1. 假設 V 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. 令 A 為以 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 *column vectors* 的 $m \times n$ *matrix.* 若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的一組 *basis,* 則 A 的 *column space* 為何? A 的 *nullspace* 為何?

4.2. Spanning Vectors and Linear Dependence

我們介紹了 spanning vectors 以及 linear independence 的定義. 在這一節中, 我們將進一步探討它們的性質.

4.2.1. Spanning Vectors. 給定 $\mathbb{R}^m$ 中的一個 subspace V, 我們知道 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$, 要成為 V 的 spanning vectors 的先決條件就是 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. 所以要找的 V 的 spanning vectors, 我們需要在 V 中找到合適的向量再一一加入.

一般來說, 當向量的個數不夠多時, 是無法形成 spanning vectors 的. 例如由 Corollary 3.4.4 我們知, 當 $n < m$ 時, $\mathbb{R}^m$ 中任意 n 個向量就無法形成 $\mathbb{R}^m$ 的 spanning vectors. 當原來的向量 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 個數不夠, 我們加入新的向量 $\mathbf{v}_{n+1}$ 後所形成的 subspace $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$ 當然會包含原來的 subspace $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, 但並不一定會比它大, 下一個定理就是告訴我們如何可以將原來的 subspace 擴大.

Lemma 4.2.1. 假設 $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1} \in \mathbb{R}^m$. 我們有 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) \subseteq \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$, 且 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) \neq \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$ 若且唯若 $\mathbf{v}_{n+1} \notin \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

Proof. 首先 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$ 是 $\mathbb{R}^m$ 的 subspace, 又 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$, 故由式子 (4.1) 知

$$\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) \subseteq \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}). \tag{4.2}$$

同理若 $\mathbf{v}_{n+1} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, 則由 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 是 $\mathbb{R}^m$ 的 subspace, 以及 $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 得知

$$\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}) \subseteq \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n). \tag{4.3}$$

此時由式子 (4.2), (4.3) 知

$$\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}).$$

故得證當 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) \neq \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$ 時必有 $\mathbf{v}_{n+1} \notin \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$.

反之, 若 $\mathbf{v}_{n+1} \notin \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$, 則因 $\mathbf{v}_{n+1} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$, 得證 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) \neq \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$. □

假設 V 為 $\mathbb{R}^n$ 的 subspace. 當 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 不是 V 的 spanning vectors 時, 由於 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) \neq V$, 我們可以在 V 中選取 $\mathbf{v}_{n+1}$ 滿足 $\mathbf{v}_{n+1} \notin \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. 此時 Lemma 4.2.1 告訴我們, $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1})$ 會比 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 大, 所以我們有機會這樣一直下去, 而找到 V 的一組 spanning vectors.

4.2.2. Linear Dependence. 我們介紹過一組向量為 linearly independent 的定義, 而一組向量為 linearly dependent 的意思與 linear independent 的意思相反, 我們一般稱為 "線性相關".

一組向量為 linearly dependent, 指的是這一組向量之間有關係, 也就是說其中有一個向量是其他向量的線性組合. 例如假設 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly dependent, 就表示其中有一個 $\mathbf{v}_i$ 可以寫成 $\mathbf{v}_1,\dots,\mathbf{v}_{i-1},\mathbf{v}_{i+1},\dots,\mathbf{v}_n$ 的線性組合. 每次要提有一個 $\mathbf{v}_i$ 是其他向量的線性組合有點麻煩. 不過若我們更進一步觀察, 此時 $\mathbf{v}_i = r_1\mathbf{v}_1+\cdots+r_{n-1}\mathbf{v}_{n-1}+r_{n+1}\mathbf{v}_{n+1}+\cdots+r_n\mathbf{v}_n$, 其中這些 r_j 皆為實數. 所以我們得

$$r_1\mathbf{v}_1+\cdots+r_{n-1}\mathbf{v}_{n-1}-\mathbf{v}_i+r_{n+1}\mathbf{v}_{n+1}+\cdots+r_n\mathbf{v}_n=\mathbf{O}.$$

也就是說我們找到一組不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$. 反之, 若存在一組不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$. 我們假設 $c_i\neq 0$, 此時得

$$\mathbf{v}_i=\frac{-c_1}{c_i}\mathbf{v}_1+\cdots+\frac{-c_{i-1}}{c_i}\mathbf{v}_{i-1}+\frac{-c_{i+1}}{c_i}\mathbf{v}_{i+1}+\cdots+\frac{-c_n}{c_i}\mathbf{v}_n,$$

也就是說 $\mathbf{v}_i$ 可以寫成 $\mathbf{v}_1,\dots,\mathbf{v}_{i-1},\mathbf{v}_{i+1},\dots,\mathbf{v}_n$ 的線性組合. 由此可知, 存在一組不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$ 就等同於 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 這一組向量之間有關係. 由於這個方式來表達線性相關不必敘述其中哪一個向量是其他向量的線性組合, 較為簡潔. 一般就以這個方式來定義線性相關.

Definition 4.2.2. 假設 $\mathbf{v}_1,\dots,\mathbf{v}_n\in\mathbb{R}^m$, 若存在一組不全為 0 的實數 $c_1,\dots,c_n$ 使得

$$c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O},$$

則稱 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 *linearly dependent.*

當 $\mathbf{v}_1,\dots,\mathbf{v}_n\in\mathbb{R}^m$ 不符合 Definition 4.2.2, 即無法找到一組不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$, 此時表示 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 之間沒有任何的線性關係. 這也就是為何在 Definition 4.1.7 中將這個情形稱為 linearly independent 的原因. 可以看出 linearly independent 和 linear dependent 是完全相反的概念. 若令 A 為以 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 column vectors 的 $m\times n$ matrix. 我們知道 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly independent 就等同於 homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution, 所以 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly dependent 就等同於 homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 有 nontrivial solution.

通常當我們要證明一組向量 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly independent, 我們有以下兩個方法: 第一個方法就是先設 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$, 再證明此時 $c_1,\dots,c_n$ 必全為 0. 第二種方法, 就是所謂的反證法, 亦即先假設 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly dependent (也就是說假設存在不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$), 再推得矛盾. 第一個方法通常在有具體的向量時使用, 而處理抽象的情形大多使用第二種方法, 如下面的例子.

Example 4.2.3. 假設 $\mathbf{v}_1,\dots,\mathbf{v}_n\in\mathbb{R}^m$ 為 linearly independent. 這表示 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 之間沒有線性關係. 因此可以理解若我們在 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 中移除 $\mathbf{v}_n$, 則 $\mathbf{v}_1,\dots,\mathbf{v}_{n-1}$ 這一組向量應仍為 linearly independent. 要證明這一個事實, 若我們用第一個方法, 很難由 $c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}=\mathbf{O}$ 推得 $c_1,\dots,c_{n-1}$ 必全為 0. 然而若利用第二個方法, 即假設存在不全為 0 的實

數 $c_1,\ldots,c_{n-1}$ 使得 $c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}=\mathbf{O}$. 此時令 $c_n=0$, 我們得到一組不全為 0 的實數 $c_1,\ldots,c_n$ 使得

$$c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}+c_n\mathbf{v}_n=c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}=\mathbf{O}.$$

此與 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbb{R}^m$ 為 linearly independent 的假設相矛盾, 故得證 $\mathbf{v}_1,\ldots,\mathbf{v}_{n-1}$ 為 linearly independent. 大家應可以看出, 我們其實是證明了當 $\mathbf{v}_1,\ldots,\mathbf{v}_{n-1}\in\mathbb{R}^m$ 為 linearly dependent 時, 加入任意的 $\mathbf{v}_n\in\mathbb{R}^n$ 後, $\mathbf{v}_1,\ldots,\mathbf{v}_{n-1},\mathbf{v}_n$ 也是 linearly dependent.

在 Example 4.2.3 中, 我們知道當 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbb{R}^m$ 這一組向量為 linearly independent 時, 在這一組向量中移除一些向量, 仍不會改變其 linearly independent 的性質. 但若加入新的向量情況可能改變. 下一個定理就是告訴我們何時加入新的向量仍會保持 linearly independent.

Lemma 4.2.4. 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}\in\mathbb{R}^m$. 若已知 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly independent,* 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 為 *linearly independent* 若且唯若 $\mathbf{v}_{n+1}\notin\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$.

Proof. 如果 $\mathbf{v}_{n+1}\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 表示 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 之間有線性關係, 即為 linearly dependent. 故知若 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 為 linearly independent, 不可能會有 $\mathbf{v}_{n+1}\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 的情形發生. 得證 $\mathbf{v}_{n+1}\notin\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$.

反之, 假設 $\mathbf{v}_{n+1}\notin\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 我們要證明 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 為 linearly independent. 利用反證法, 即設 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 為 linearly dependent, 也就是說存在一組不全為 0 的實數 $c_1,\ldots,c_n,c_{n+1}$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n+c_{n+1}\mathbf{v}_{n+1}=\mathbf{O}$. 我們知此時 c_{n+1} 必為 0, 否則由

$$\mathbf{v}_{n+1}=\frac{-c_1}{c_{n+1}}\mathbf{v}_1+\cdots+\frac{-c_n}{c_{n+1}}\mathbf{v}_n,$$

會得到 $\mathbf{v}_{n+1}\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 之矛盾. 故此時因 $c_{n+1}=0$, 得 $c_1,\ldots,c_n$ 是一組不全為 0 的實數使得

$$c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n+c_{n+1}\mathbf{v}_{n+1}=\mathbf{O},$$

亦即 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly dependent. 這和已知的假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent 相矛盾, 故得證 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 為 linearly independent. □

假設 V 為 $\mathbb{R}^n$ 的 subspace, 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 為 linearly independent. 如果 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 不是 V 的 spanning vectors, 則我們可以在 V 中選取 $\mathbf{v}_{n+1}$ 滿足 $\mathbf{v}_{n+1}\notin\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$. 此時 Lemma 4.2.1 告訴我們 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1}$ 仍保持 linearly independent.

一般來說, 當向量的個數太多時, 就不會是 linearly independent 了. 例如由 Corollary 3.4.8 我們知, 當 $n>m$ 時, $\mathbb{R}^m$ 中任意 n 個向量就一定會 linearly dependent. 我們可以將這個事實推廣到更一般的情況.

Lemma 4.2.5. 假設 $\mathbf{w}_1,\ldots,\mathbf{w}_k\in\mathbb{R}^m$. 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)$ 且 $n>k$, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly dependent.*

Proof. $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)$ 表示存在 a_{ij} 其中 $1 \le i \le n$, $1 \le j \le k$ 使得

$$\begin{array}{rcccccccc} \mathbf{v}_1 & = & a_{11}\mathbf{w}_1 & + & a_{12}\mathbf{w}_2 & + & \cdots & + & a_{1k}\mathbf{w}_k \\ & \vdots & & & & & & & \\ \mathbf{v}_n & = & a_{n1}\mathbf{w}_1 & + & a_{n2}\mathbf{w}_2 & + & \cdots & + & a_{nk}\mathbf{w}_k. \end{array}$$

因此對於 $c_1,\ldots,c_n \in \mathbb{R}$ 我們有

$$\begin{array}{rcccccccc} c_1\mathbf{v}_1 & = & c_1a_{11}\mathbf{w}_1 & + & c_1a_{12}\mathbf{w}_2 & + & \cdots & + & c_1a_{1k}\mathbf{w}_k \\ & \vdots & & & & & & & \\ c_n\mathbf{v}_n & = & c_na_{n1}\mathbf{w}_1 & + & c_na_{n2}\mathbf{w}_2 & + & \cdots & + & c_na_{nk}\mathbf{w}_k. \end{array}$$

因此, 若考慮 $n \times k$ matrix $A = [a_{ij}]$ 且令 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = b_1\mathbf{w}_1 + \cdots + b_k\mathbf{w}_k$ 則

$$\begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix} A = \begin{bmatrix} b_1 & \cdots & b_k \end{bmatrix}, \text{ 亦即 } A^{\mathrm{T}} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}.$$

由於 A^{T} 為 $k \times n$ matrix 且 $n > k$, 故由 Corollary 3.4.7 知 homogeneous linear system $A^{\mathrm{T}}\mathbf{x} = \mathbf{O}$ 有 nontrivial solution. 也就是說存在 $c_1,\ldots,c_n \in \mathbb{R}$ 不全為 0, 使得 $A^{\mathrm{T}} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$, 亦即 $\begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix} A = \begin{bmatrix} 0 & \cdots & 0 \end{bmatrix}$. 也因此這些不全為 0 的 $c_1,\ldots,c_k$ 會滿足

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = 0\mathbf{w}_1 + \cdots + 0\mathbf{w}_k = \mathbf{O}.$$

得證 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly dependent. □

Question 4.2. 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly independent.* 若 $\mathbf{w}_1,\ldots,\mathbf{w}_k \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 且 $k < n$, 試證明 $\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k) \neq \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$.

4.3. Dimension of subspace

我們定義過什麼是一個 $\mathbb{R}^m$ 的 subspace 的 basis, 但還不知道它是否存在. 在這一節中我們將證明只要是 $\mathbb{R}^m$ 中 nonzero subspace 都會有 basis. 並證明形成一個 subspace 的 basis 之向量個數是固定的, 因而定義此個數為這個 subspace 的 dimension. 接著我們會探討一些與 dimension 相關的性質.

假設 V 為 $\mathbb{R}^m$ 的 subspace. 若 $V = \{\mathbf{O}\}$, 此時 $\mathbf{O}$ 為 V 中唯一的元素, 所以 $\mathbf{O}$ 會是 V 的 spanning vector. 但是 $\mathbf{O}$ 本身不是 linearly independent, 這是因為我們可以找到 $c \neq 0$ 使得 $c\mathbf{O} = \mathbf{O}$. 所以此時 V 是沒有 basis 的. 而當 $V \neq \{\mathbf{O}\}$ 時, 表示存在 $\mathbf{v}_1 \in V$ 且 $\mathbf{v}_1 \neq \mathbf{O}$. 此時考慮 $V_1 = \mathrm{Span}(\mathbf{v}_1)$. 若 $V_1 = V$, 則知 $\mathbf{v}_1$ 為 V 的 spanning vector, 又 $\mathbf{v}_1$ 本身是 linearly independent, 所以知 $\mathbf{v}_1$ 是 V 的 basis. 而若 $V_1 \neq V$, 表示存在 $\mathbf{v}_2 \in V$ 且 $\mathbf{v}_2 \notin V_1 = \mathrm{Span}(\mathbf{v}_1)$. 此時考慮 $V_2 = \mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$. 注意由 Lemma 4.2.1, 我們知 $V_1 \subseteq V_2$ 且 $V_1 \neq V_2$, 也就是說 V_2 比 V_1 大. 另一方面 Lemma 4.2.4 也告訴我們 $\mathbf{v}_1,\mathbf{v}_2$ 為 linearly independent. 現若 $V_2 = V$, 則知 $\mathbf{v}_1,\mathbf{v}_2$ 為 V 的 spanning vectors, 再加上它們為 linearly independent, 我們得 $\mathbf{v}_1,\mathbf{v}_2$ 就是 V 的一組 basis. 而若 $V_2 \neq V$, 則繼續上面的步驟, 因此我們有以下的定理.

Theorem 4.3.1. 假設 V 為 $\mathbb{R}^m$ 的 *subspace* 且 $V \neq \{\mathbf{O}\}$, 則存在 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$, 其中 $n \leq m$, 為 V 的一組 *basis.*

Proof. 我們延續前面的符號, 當 $V_2 \neq V$ 時, 繼續處理下去. 當然了, 因為 m 不知有多大, 我們不可能如前這樣一步一步講下去, 這裡就可以用數學歸納法了. 我們必須說明按照前面的方法繼續下去, 在第 i 次時找出的 vectors $\mathbf{v}_1, \ldots, \mathbf{v}_i$ 皆為 linearly independent, 而且這個步驟一定會停止, 也就是說一直下去一定會有一個 n 滿足 $n \leq m$ 且 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$.

利用數學歸納法, 現假設第 k 次時所得的 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 為 linearly independent (別忘了我們知 $k = 1$ 成立). 令 $V_k = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. 若 $V_k = V$, 表示 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 為 V 的 spanning vectors, 再加上它們為 linearly independent, 得證 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 為 V 的一組 basis. 而若 $V_k \neq V$, 則由 $V_k \subseteq V$ 知必存在 $\mathbf{v}_{k+1} \in V$ 且 $\mathbf{v}_{k+1} \not\in V_k$. 此時考慮 $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$, 由 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 為 linearly independent 且 $\mathbf{v}_{K+1} \not\in V_k$ 利用 Lemma 4.2.4 得知 $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}$ 為 linearly independent. 我們用數學歸納法證得了用這個步驟所得的 vectors 為 linearly independent.

接著我們要證明, 依這個步驟一定會有 $n \leq m$, 使得 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$. 現假設到第 m 次時, 仍無法得 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m) = V$, 依前面數學歸納法我們知 $\mathbf{v}_1, \ldots, \mathbf{v}_m$ 為 linearly independent. 此時存在 $\mathbf{v}_{m+1} \in V$ 且 $\mathbf{v}_{m+1} \not\in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m)$, 故再次由 Lemma 4.2.4 知 $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}$ 為 linearly independent. 然而 $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1} \in V$ 所以當然 $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1} \in \mathbb{R}^m$, 因此由 Corollary 3.4.8 知 $\mathbf{v}_1, \ldots, \mathbf{v}_{m+1}$ 這 $m+1$ 個向量必為 linearly dependent. 此與 $\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{v}_{m+1}$ 為 linearly independent 相矛盾. 矛盾發生的原因在於我們假設到第 m 次時仍無法得 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_m) = V$, 因此知一定存在 $n \leq m$ 使得 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的 spanning vectors. 再加上已證得 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly independent, 得證 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的一組 basis. $\square$

從 Theorem 4.3.1 的證明我們也知道, 若 $\mathbf{v}_1 \ldots \mathbf{v}_k \in V$ 且為 linearly independent, 則我們可以將 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 擴大成 V 的一組 basis. 也就是說若 $\mathbf{v}_1, \ldots, \mathbf{v}_k$ 不是 V 的 spanning vectors, 我們便可利用前述的方法, 找到 $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n \in V$ 使得 $\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ 為 V 的一組 basis.

Theorem 4.3.1 告訴我們的是 basis 的存在性, 接下來我們要探討的是和 basis 有關的唯一性. 當然了我們很容易看出一個 $\mathbb{R}^m$ 的 subspace 其 basis 是不唯一的. 例如

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{和} \quad \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

都是 $\mathbb{R}^2$ 的 basis.

不過若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 $\mathbb{R}^m$ 的一組 basis, 由 Corollary 3.4.4 知如果 $n < m$, 則 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 不可能是 $\mathbb{R}^m$ 的 spanning vectors, 故得 $n \geq m$. 然而若 $n > m$, 由 Corollary 3.4.8 知 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 不可能為 linearly independent, 故知 $n = m$. 也就是說所有 $\mathbb{R}^m$ 的 basis 皆由 m 個向量所組成. 我們可以將此結果推廣成以下之定理.

Theorem 4.3.2. 假設 V 為 $\mathbb{R}^m$ 的 *subspace.* 若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 和 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 皆為 V 的 *basis,* 則 $n = k$.

Proof. 我們利用反證法, 假設 $n \neq k$. 不失一般性, 我們假設 $n > k$. 此時由於 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 V 的 spanning vectors, 故 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V = \mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$. 因為 $n > k$ 利用 Lemma 4.2.5 得 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly dependent. 此與 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的 basis 相矛盾. 矛盾發生於我們假設 $n \neq k$, 故得證 $n = k$. □

Theorem 4.3.2 告訴我們組成 V 的一組 basis 的向量個數是固定的. 也就是說若找到 n 個向量形成 V 的 basis, 則 V 其他的 basis 一定也會是由 n 個向量所組成. 由於這個重要的結果我們有以下的定義.

Definition 4.3.3. 假設 V 是 $\mathbb{R}^m$ 的 subspace. 組成 V 的一組 basis 的向量個數稱為 V 的 *dimension* (維度), 用 $\dim(V)$ 來表示.

例如 $\mathbb{R}^m$ 的 basis 皆由 m 個向量所組成, 所以我們有 $\dim(\mathbb{R}^m) = m$.

我們曾提過由 Theorem 4.3.1 的證明, 我們知道 V 中任一組 linearly independent 的向量都可以依保持 linearly independent 的方式一個一個加入向量直至形成 spanning vectors 為止. 另一方面, V 中的任一組 spanning vectors, 我們也可以依保持 spanning vectors 的方式一個一個將不需要的向量移除直至它們為 linearly independent 為止. 利用這樣的概念, 我們有以下的定理.

Proposition 4.3.4. 假設 V 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$.

(1) 若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的 *spanning vectors,* 則 $\dim(V) \leq n$. 特別的, 若此時 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 *linearly dependent,* 則 $\dim(V) < n$.

(2) 若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 *linearly independent,* 則 $\dim(V) \geq n$. 特別的, 若此時 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 不是 V 的 *spanning vectors,* 則 $\dim(V) > n$.

Proof. 假設 $\dim(V) = k$, 且令 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 V 的一組 basis.

(1) 若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的 spanning vectors, 表示 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$. 由於 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 V 的一組 basis, 故有 $\mathbf{w}_1, \ldots, \mathbf{w}_k \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 且 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 linearly independent. 現若 $k > n$, 則 Lemma 4.2.5 告訴我們 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 linearly dependent. 造成矛盾, 故得證 $\dim(V) = k \leq n$. 現若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly dependent, 表示 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 中有一個向量是其他向量的線性組合. 不失一般性我們假設此向量為 $\mathbf{v}_n$, 亦即 $\mathbf{v}_n \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$. 此時 Lemma 4.2.1 告訴我們 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) = V$, 亦即 $\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}$ 為 V 的 spanning vectors. 套用剛才已證的結果知 $\dim(V) \leq n-1$, 故得證 $\dim(V) < n$.

(2) 假設 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 為 linearly independent. 因 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 V 的一組 basis, 故有 $\mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_k) = V$, 亦即 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_k)$. 現若 $n > k$, 由 Lemma 4.2.5 會得到 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 為 linearly dependent 之矛盾. 故得證 $n \leq k = \dim(V)$. 現若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 不是 V 的 spanning vectors, 表示存在 $\mathbf{v}_{n+1} \in V$ 但 $\mathbf{v}_{n+1} \notin \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 此時 Lemma 4.2.4 告訴我們 $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}$ 仍為 linearly independent. 套用剛才已證的結果知 $\dim(V) \geq n+1$, 故得證 $\dim(V) > n$. □

當我們要說明 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis, 依定義我們必須檢查 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的 spanning vectors 以及它們是 linearly independent 這兩個條件. 不過若我們知道 $\dim(V)$ 恰好是 n, 下一個定理告訴我們僅要檢查 spanning vectors 或 linearly independent 其中一項就可.

Corollary 4.3.5. *假設 V 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$. 下列的敘述為等價.*

(1) $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 *basis.*

(2) $\dim(V)=n$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的 *spanning vectors.*

(3) $\dim(V)=n$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly independent.*

Proof. $(1)\Rightarrow(2)$: 由 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis, 我們知 V 的 basis 需由 n 個向量所組成, 故得 $\dim(V)=n$. 又組成 basis 的向量需為 spanning vectors, 故 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的 spanning vectors. 得證 $(1)\Rightarrow(2)$.

$(2)\Rightarrow(3)$: 因 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的 spanning vectors, 所以由 Proposition 4.3.4 (1) 知 $\dim(V)\leq n$. 現若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 不是 linearly independent, 則再由 Proposition 4.3.4 (1) 會得到 $\dim(V)<n$. 此與假設 $\dim(V)=n$ 矛盾, 故知 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent. 得證 $(2)\Rightarrow(3)$.

$(3)\Rightarrow(1)$: 因 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent, 所以由 Proposition 4.3.4 (2) 知 $\dim(V)\geq n$. 現若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 不是 V 的 spanning vectors, 則再由 Proposition 4.3.4 (2) 會得到 $\dim(V)>n$. 此與假設 $\dim(V)=n$ 矛盾, 故知 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的 spanning vectors. 得證 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis, 故知 $(3)\Rightarrow(1)$. □

回顧 Theorem 4.3.1 告訴我們若 V 為 $\mathbb{R}^m$ 的 subspace, 則組成 V 的 basis 的向量個數會小於等於 m, 所以我們有 $\dim(V)\leq m$, 也就是說 $\dim(V)\leq\dim(\mathbb{R}^m)$. 這個結果可以推廣到以下更一般的狀況.

Corollary 4.3.6. *假設 V,W 皆為 $\mathbb{R}^m$ 的 subspaces 且 $V\subseteq W$, 則 $\dim(V)\leq\dim(W)$. 此時 $\dim(V)=\dim(W)$ 若且唯若 $V=W$.*

Proof. 假設 $\dim(V)=n$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的 basis. 由於 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 且 $V\subseteq W$, 我們得 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in W$. 現因 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in W$ 為 linearly independent 故由 Proposition 4.3.4 (2) 得 $\dim(V)=n\leq\dim(W)$.

如果 $V=W$, 自然由 dimension 的唯一性知 $\dim(V)=\dim(W)$. 反之, 若 $\dim(V)=\dim(W)$ 但 $V\neq W$. 此時表示 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in W$ 為 linearly independent 但不是 W 的 spanning vectors, 故由 Proposition 4.3.4 (2) 得 $\dim(V)=n<\dim(W)$. 此與假設 $\dim(V)=\dim(W)$ 不合, 得證 $V=W$. □

注意一般來說 $\dim(V)=\dim(W)$ 並不代表 $V=W$. 例如在 $\mathbb{R}^2$ 中若 $\mathbf{v}\neq\mathbf{O}$, 則 $\mathrm{Span}(\mathbf{v})$ 就是 dimension 為 1 的 subspace. 然而只要 $\mathbf{w}\neq\mathbf{O}$ 且和 $\mathbf{v}$ 不平行, 就會有 $\mathrm{Span}(\mathbf{v})\neq\mathrm{Span}(\mathbf{w})$,

但它們皆為 dimension 1 的 subspace. 所以 Corollary 4.3.6 中我們需要 $V \subseteq W$ 這個前提才能由 $\dim(V) = \dim(W)$ 得到 $V = W$.

4.4. Column Space and Nullspace

我們將介紹如何找到一個矩陣的 column space, row space 以及 nullspace 的 basis. 我們會發現 column space 和 row space 的 dimension 皆相同且等於矩陣的 rank. 最後我們再探討如何得到一般 subspace 的 basis.

給定一個矩陣, 它的 column space 和 nullspace 和以該矩陣為係數矩陣所形成的聯立方程組是否有解以及解是否唯一息息相關. 由於 column space 和 nullspace 的重要性, 我們將之正式定義如下:

Definition 4.4.1. 假設 $A = \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{bmatrix}$ 為以 $\mathbb{R}^m$ 中的向量 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 為 column vectors 的 $m \times n$ matrix.

(1) 我們稱 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ 為 A 的 *column space*, 且用 $C(A)$ 來表示 A 的 column space.

(2) 我們稱 homogeneous linear system $A\mathbf{x} = \mathbf{O}$ 所有解所成的集合為 A 的 *nullspace* 且用 $N(A)$ 表示 A 的 nullspace. 即 $N(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{O}\}$.

利用 Lemma 3.4.1 以及 Theorem 3.4.6 我們馬上有以下的結果.

Proposition 4.4.2. *假設 A 為 $m \times n$ matrix 且 $\mathbf{b} \in \mathbb{R}^m$, 考慮聯立方程組 $A\mathbf{x} = \mathbf{b}$.*

(1) *$A\mathbf{x} = \mathbf{b}$ 有解若且唯若 $\mathbf{b} \in C(A)$.*

(2) *假設 $A\mathbf{x} = \mathbf{b}$ 有解則其解唯一若且唯若 $N(A) = \{\mathbf{O}\}$.*

接下來我們就是要找到一個矩陣的 column space 以及 nullspace 這兩個重要的 subspaces 的 basis. 一般來說要找到 $\mathbb{R}^m$ 的 subspace V 的一組 basis, 我們會先找 V 的一組 spanning vectors. 然後在其中再挑出仍保持為 spanning vectors 且為 linearly independent 的一組向量. 當只有兩個向量時, 我們可以馬上由它們是否為平行來判斷是否為 linearly independent. 不過通常有三個以上的向量時, 並不容易直接看出哪些向量會 linearly independent, 除非如以下的例子.

Example 4.4.3. 考慮

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ -5 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 7 \\ -1 \\ 0 \end{bmatrix}.$$

要說明 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 linearly independent, 我們必須說明只有當 $c_1=c_2=c_3=0$ 時, 才會使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{O}$. 然而

$$c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=c_1\begin{bmatrix}0\\3\\0\\1\end{bmatrix}+c_2\begin{bmatrix}2\\-5\\0\\0\end{bmatrix}+c_3\begin{bmatrix}0\\7\\-1\\0\end{bmatrix}=\begin{bmatrix}2c_2\\3c_1-5c_2+7c_3\\-c_3\\c_1\end{bmatrix}.$$

所以要使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{O}$, 就必須讓 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3$ 的 1-st entry $2c_2$, 3-rd entry $-c_3$ 以及 4-th entry c_1 皆為 0, 即 $c_1=c_2=c_3=0$. 得證只有當 $c_1=c_2=c_3=0$ 時, 才會使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{O}$, 故知 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 linearly independent.

從 Example 4.4.3 我們可以看出來, 當 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 中每一個向量 $\mathbf{v}_i$ 都可以找到一個 entry 不為 0, 而其他向量在該 entry 皆為 0, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent. (例如 Example 4.4.3 中 $\mathbf{v}_1$ 的 4-th entry 為 1, 而 $\mathbf{v}_2,\mathbf{v}_3$ 的 4-th entry 為 0; $\mathbf{v}_2$ 的 1-st entry 為 2, 而 $\mathbf{v}_1,\mathbf{v}_3$ 的 1-st entry 為 0; $\mathbf{v}_3$ 的 3-rd entry 為 -1, 而 $\mathbf{v}_1,\mathbf{v}_2$ 的 3-th entry 為 0, 就符合這個條件). 此時假設每個 $\mathbf{v}_i$ 的那個非 0 的特殊 entry 為 a_i, 由於 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 在該位置的 entry 為 c_ia_i, 所以若 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$, 則必 $c_ia_i=0$, 得每一個 c_i 皆為 0. 因此 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent.

當 A 為 $m\times n$ matrix, A 的 nullspace $N(A)$ 就是 homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 的所有解所成的集合. 由於我們已經知道如何找到 $A\mathbf{x}=\mathbf{O}$ 的解, 所以我們就從如何找 nullspace 的 basis 開始.

回顧我們找 $A\mathbf{x}=\mathbf{O}$ 的解集合的方法為, 利用 elementary row operations 將 A 化為 echelon form (或 reduced echelon form) A'. 此時 $A'\mathbf{x}=\mathbf{O}$ 的解集合就是 $A\mathbf{x}=\mathbf{O}$ 的解集合, 也就是說 A 和 A' 有相同的 nullspace. 接著我們找出 free variable, 再將每個 free variable 代入任意的實數, 從下往上推得出一組解. 注意在這個過程中, pivot variable 的值會由 free variables 的值所決定, 所以只要定出 free variable 的值, 就可以得到一組解. 現假設 free variables 為 $x_{i_1},\ldots,x_{i_k}$. 對每一個 $j=1,\ldots,k$, 我們考慮 $x_{i_j}=1$, 其他 free variable 為 0 的情形, 令這樣推得出來的解為 $\mathbf{v}_j$. 由於 $\mathbf{v}_j$ 的 i_j-th entry 為 1, 而其他 $\mathbf{v}_{i_{j'}}$ 的 i_j-th entry 為 0, 由前討論知 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 linearly independent. 而對於任意 $r_1,\ldots,r_k\in\mathbb{R}$, $r_1\mathbf{v}_1+\cdots+r_k\mathbf{v}_k$ 就等同於是將每個 free variables $x_{i_1},\ldots,x_{i_k}$ 分別代 $x_{i_1}=r_1,\ldots,x_{i_k}=r_k$ 所得的解. 換言之每個解都可以寫成 $r_1\mathbf{v}_1+\cdots+r_k\mathbf{v}_k$ 的形式, 也就是說 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 是 A 的 nullspace 的一組 spanning vectors. 我們證明了 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 就是 A 的 nullspace 的一組 basis, 也因此得知 A 的 nullspace 的 dimension 為 free variables 的個數, 亦即 A 的 column 的個數減去 pivot 的個數, 因此有以下之結果.

Proposition 4.4.4. *假設 A 為 $m\times n$ matrix. 若利用 row operations 將 A 化為 echelon form A' 後, A' 的 pivot 個數為 r, 則 A 的 nullspace 的 dimension 為 $n-r$. 假設 $A'\mathbf{x}=\mathbf{O}$ 的 free variables 為 $x_{i_1},\ldots,x_{i_k}$. 對每一個 $j=1,\ldots,k$, 我們取 $x_{i_j}=1$, 其他 free variable 為 0, 令這樣推得出來的解為 $\mathbf{v}_j$. 則 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 A 的 nullspace 的一組 basis.*

由於一個矩陣的 nullspace 不會因為其化為 echelon form 的不同而改變, 而且 nullspace 的 dimension 是固定的, 所以 Proposition 4.4.4 也證明了過去我們一直沒有證明的"不管一個矩陣利用 elementary row operations 所化得的 echelon form 為何, 其 pivot 的個數必相同". 也因此 Definition 2.3.1 有關於 rank 的定義是 well-defined.

Example 4.4.5. 考慮 A 的 nullspace, 其中

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

將 A 的 2-nd row 分別乘上 -2, -1, -1 加至 1-st, 3-rd 和 4-th row, 然後再將 1-st, 2-nd rows 交換得

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

接著將 2-nd row 分別乘上 $-1,-2$ 加至 3-rd 和 4-th row 得

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 1 & -2 & 0 & 0\\ 0 & 0 & 0 & 1 & 1 & 2\\ 0 & 0 & 0 & 1 & 1 & 2 \end{bmatrix}.$$

最後將 3-rd row 乘上 -1 加至 4-th row, 得 echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 1 & -2 & 0 & 0\\ 0 & 0 & 0 & 1 & 1 & 2\\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

我們就是要找到 homogeneous linear system

$$\begin{array}{cccccccc} x_1 & & & +x_4 & & & = & 0\\ & x_2 & +x_3 & -2x_4 & & & = & 0\\ & & & +x_4 & +x_5 & +2x_6 & = & 0 \end{array}$$

所有的解. 由 echelon form 看出 x_1,x_2,x_4 為 pivot variable, x_3,x_5,x_6 為 free variable. 現令 $x_6=1,x_5=0,x_3=0$, 解出 $x_4=-2$, $x_2=-4$, $x_1=2$, 而令 $x_6=0,x_5=1,x_3=0$ 解出 $x_4=-1$, $x_2=-2$, $x_1=1$, 最後令 $x_6=0,x_5=0,x_3=1$ 解出 $x_4=0$, $x_2=-1$, $x_1=0$. 故得

$$\mathbf{v}_1=\begin{bmatrix} 2\\ -4\\ 0\\ -2\\ 0\\ 1 \end{bmatrix}, \mathbf{v}_2=\begin{bmatrix} 1\\ -2\\ 0\\ -1\\ 1\\ 0 \end{bmatrix}, \mathbf{v}_3=\begin{bmatrix} 0\\ -1\\ 1\\ 0\\ 0\\ 0 \end{bmatrix}$$

為 A 的 nullspace 的一組 basis. 事實上, 若令 x_6,x_5,x_3 分別為任意的實數 r,s,t, 則可得 $x_4=-2r-s$, $x_2=-4r-2s-t$, $x_1=2r+s$. 也就是說 A 的 nullspace 中的向量都可以寫成

$$\begin{bmatrix} 2r+s \\ -4r-2s-t \\ t \\ -2r-s \\ s \\ r \end{bmatrix} = r\begin{bmatrix} 2 \\ -4 \\ 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} + s\begin{bmatrix} 1 \\ -2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = r\mathbf{v}_1 + s\mathbf{v}_2 + t\mathbf{v}_3.$$

故知 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 A 的 nullspace 的 spanning vectors, 又很容易看出 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 linearly independent, 得證 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 $N(A)$ 的一組 basis.

Question 4.3. *試將 Example 4.4.5 中的 A 化為 reduced echelon form. 是否更容易看出 $N(A)$ 的一組 basis 呢?*

接下來我們來看如何找 matrix A 的 column space $C(A)$ 的 basis. 首先一個直接的想法就是 A 的 column space, 就是使得聯立方程組 $A\mathbf{x}=\mathbf{v}$ 有解的 $\mathbf{v}$ 所成的集合. 所以我們只要找出這些 $\mathbf{v}$, 就可以得到 A 的 column space. 過去我們曾介紹過以 A 為係數矩陣的聯立方程組的 constrain equations 就可以找到這些 $\mathbf{v}$. 不過這個方法有點麻煩, 等一下我們會給一個更好的方法, 這裡就僅用例子回顧一下這個方法.

Example 4.4.6. 考慮 Example 4.4.5 中的 4×6 matrix A. 我們要找出 A 的 column vectors 的一組 basis. 假設 $\mathbf{b}$ 為 A 的 column space 裡的一個向量, 我們知道此時 $A\mathbf{x}=\mathbf{b}$ 有解, 因此令

$$\mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix},$$

我們要找到 b_1,b_2,b_3,b_4 的條件使得以下聯立方程組有解.

$$\begin{array}{rrrrrrcl} 2x_1 & +x_2 & +x_3 & & & & = & b_1 \\ x_1 & & & +x_4 & & & = & b_2 \\ x_1 & +x_2 & +x_3 & & +x_5 & +2x_6 & = & b_3 \\ x_1 & +2x_2 & +2x_3 & -2x_4 & +x_5 & +2x_6 & = & b_4 \end{array}$$

考慮 augmented matrix $[A\,|\,\mathbf{b}]$, 利用 Example 4.4.5 相同的 elementary row operations 我們得

$$\left[\begin{array}{cccccc|c} 2 & 1 & 1 & 0 & 0 & 0 & b_1 \\ 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 1 & 1 & 1 & 0 & 1 & 2 & b_3 \\ 1 & 2 & 2 & -2 & 1 & 2 & b_4 \end{array}\right] \sim \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & -2 & 0 & 0 & b_1-2b_2 \\ 0 & 1 & 1 & -1 & 1 & 2 & b_3-b_2 \\ 0 & 2 & 2 & -3 & 1 & 2 & b_4-b_2 \end{array}\right] \sim$$

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & -2 & 0 & 0 & b_1-2b_2 \\ 0 & 0 & 0 & 1 & 1 & 2 & b_3+b_2-b_1 \\ 0 & 0 & 0 & 1 & 1 & 2 & b_4+3b_2-2b_1 \end{array}\right] \sim \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & -2 & 0 & 0 & b_1-2b_2 \\ 0 & 0 & 0 & 1 & 1 & 2 & b_3+b_2-b_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_4-b_3+2b_2-b_1 \end{array}\right].$$

由解聯立方程組的方法 (即 2.1 節 (2)(a)(b) 的情形) 知, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解若且唯若 $b_4-b_3+2b_2-b_1=0$. 換言之, 由所有 $b_1-2b_2+b_3-b_4=0$ 的解, 所得的 $\mathbf{b}$ 所成的

集合便是 A 的 column space. 所以我們回到求矩陣 $B=\begin{bmatrix} 1 & -2 & 1 & -1 \end{bmatrix}$ 的 nullspace. 由於 x_1 為 pivot variable, x_2,x_3,x_4 為 free variable. 利用前面求 nullspace 的 basis 的方法, 令 $x_4=1,x_3=0,x_2=0$ 解出 $x_1=1$, 而令 $x_4=0,x_3=1,x_2=0$ 解出 $x_1=-1$, 最後令 $x_4=0,x_3=0,x_2=1$ 解出 $x_1=2$. 故得

$$\mathbf{v}_1=\begin{bmatrix} 1\\0\\0\\1 \end{bmatrix},\mathbf{v}_2=\begin{bmatrix} -1\\0\\1\\0 \end{bmatrix},\mathbf{v}_3=\begin{bmatrix} 2\\1\\0\\0 \end{bmatrix}$$

為 B 的 nullspace 的一組 basis, 也就是 A 的 column space 的一組 basis..

注意用這個方法, 若 $m\times n$ matrix A 化成 echelon form 後沒有一個 row 全為 0, 就表示所有的 $\mathbf{b}\in\mathbb{R}^m$ 皆會使聯立方程組有解, 故此時 A 的 column space 為 $\mathbb{R}^m$.

Example 4.4.6 找 column space 所用的方法缺點就是當我們找出 constrain equations 後, 還要再求另一個矩陣的 nullspace 才能找到 column space 的 basis. 接下來我們介紹一個更簡捷的方法.

首先注意當我們利用 elementary row operations 將 A 化為 echelon form A' 後, homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 和 $A'\mathbf{x}=\mathbf{O}$ 有相同的解集合. 現假設 $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 為 A 的 column vectors, 而 $\mathbf{a}'_1,\ldots,\mathbf{a}'_n$ 為 A' 的 column vectors. 若 $x_1=c_1,\ldots,x_n=c_n$ 為 $A'\mathbf{x}=\mathbf{O}$ 的一組解, 表示 $c_1\mathbf{a}'_1+\cdots+c_n\mathbf{a}'_n=\mathbf{O}$, 此時由於 $x_1=c_1,\ldots,x_n=c_n$ 亦為 $A\mathbf{x}=\mathbf{O}$ 的一組解故我們亦有 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{O}$. 同理若 $c_1,\cdots,c_n\in\mathbb{R}$ 使得 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{O}$, 我們亦會有 $c_1\mathbf{a}'_1+\cdots+c_n\mathbf{a}'_n=\mathbf{O}$. 這告訴我們存在不全為 0 的 c_i 使得 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{O}$ 若且唯若存在不全為 0 的 c_i 使得 $c_1\mathbf{a}'_1+\cdots+c_n\mathbf{a}'_n=\mathbf{O}$. 換言之, $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 為 linearly dependent 若且唯若 $\mathbf{a}'_1,\ldots,\mathbf{a}'_n$ 為 linearly dependent. 這也等價於 $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 為 linearly independent 若且唯若 $\mathbf{a}'_1,\ldots,\mathbf{a}'_n$ 為 linearly independent. 簡單來說當我們利用 elementary row operations 將一個矩陣變換到另一個矩陣, 兩個矩陣 column vectors 之間的線性關係是會被保持的. 我們看以下的例子.

Example 4.4.7. 考慮 Example 4.4.5 中的 4×6 matrix A, 且利用 elementary row operation 將之化為 reduced echelon form A'. 也就是將 Example 4.4.5 中的 echelon form 的 3-rd row 乘上 2 加到 echelon form 的 2-nd row, 再將 echelon form 的 3-rd row 乘上 -1 加到 echelon form 的 1-st row 得

$$A=\begin{bmatrix} 2&1&1&0&0&0\\1&0&0&1&0&0\\1&1&1&0&1&2\\1&2&2&-2&1&2 \end{bmatrix}\sim\begin{bmatrix} 1&0&0&1&0&0\\0&1&1&-2&0&0\\0&0&0&1&1&2\\0&0&0&0&0&0 \end{bmatrix}\sim A'=\begin{bmatrix} 1&0&0&0&-1&-2\\0&1&1&0&2&4\\0&0&0&1&1&2\\0&0&0&0&0&0 \end{bmatrix}$$

我們很容易看出 A' 的 3 個 pivot 所在的 column vectors $\mathbf{a}'_1,\mathbf{a}'_2,\mathbf{a}'_4$ 為 linearly independent. 事實上 $\mathbf{a}'_1,\mathbf{a}'_2,\mathbf{a}'_4$ 每一個都符合有一個非 0 entry (即 pivot 之 entry) 而其他向量在該 entry 為 0. 我們考慮相對應到 A 的 column vectors $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_4$. 它們也會是 linearly independent. 這是因為若我們考慮新的 4×3 matrix $[\mathbf{a}_1\ \ \mathbf{a}_2\ \ \mathbf{a}_4]$ 經由將 A 換成 A' 一樣步驟的 elementary row operation 我們會得到 $[\mathbf{a}'_1\ \ \mathbf{a}'_2\ \ \mathbf{a}'_4]$. 所以依前面的討論知, 因為 $\mathbf{a}'_1,\mathbf{a}'_2,\mathbf{a}'_4$ 為 linearly independent, 所以 $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_4$ 也會是 linearly independent. 另一方面, 在 A' 中我們很容易看出 $\mathbf{a}'_3=\mathbf{a}'_2$, $\mathbf{a}'_5=-\mathbf{a}'_1+2\mathbf{a}'_2+\mathbf{a}'_4$ 以及 $\mathbf{a}'_6=-2\mathbf{a}'_1+4\mathbf{a}'_2+2\mathbf{a}'_4$. 所以和剛才同樣理由, 依 elementary row operations 保持線性關係的性質, 我們有 $\mathbf{a}_3=\mathbf{a}_2$, $\mathbf{a}_5=-\mathbf{a}_1+2\mathbf{a}_2+\mathbf{a}_4$ 以及 $\mathbf{a}_6=-2\mathbf{a}_1+4\mathbf{a}_2+2\mathbf{a}_4$. 事實上直接檢查得

$$\mathbf{a}_2=\mathbf{a}_3=\begin{bmatrix}1\\0\\1\\2\end{bmatrix},\quad -\mathbf{a}_1+2\mathbf{a}_2+\mathbf{a}_4=-\begin{bmatrix}2\\1\\1\\1\end{bmatrix}+2\begin{bmatrix}1\\0\\1\\2\end{bmatrix}+\begin{bmatrix}0\\1\\0\\-2\end{bmatrix}=\begin{bmatrix}0\\0\\1\\1\end{bmatrix}=\mathbf{a}_5,$$

$$-2\mathbf{a}_1+4\mathbf{a}_2+2\mathbf{a}_4=-2\begin{bmatrix}2\\1\\1\\1\end{bmatrix}+4\begin{bmatrix}1\\0\\1\\2\end{bmatrix}+2\begin{bmatrix}0\\1\\0\\-2\end{bmatrix}=\begin{bmatrix}0\\0\\2\\2\end{bmatrix}=\mathbf{a}_6.$$

換言之, $\mathbf{a}_3,\mathbf{a}_5,\mathbf{a}_6\in\operatorname{Span}(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_4)$. 故由 Lemma 4.2.1 知 A 的 column space 為

$$\operatorname{Span}(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3,\mathbf{a}_4,\mathbf{a}_5,\mathbf{a}_6)=\operatorname{Span}(\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_4).$$

再加上 $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_4$ 為 linearly independent, 得證 $\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_4$ 是 A 的 column space 的一組 basis.

注意在 Example 4.4.7 中我們為了方便說明將 A 化為 reduced echelon form. 事實上既然我們知道 column space 的 basis 是由對應到 pivot 所在位置 A 的 column vectors 所組成, 所以化成 echelon form 知道 pivot 在那些 column 就可以找到 basis 了. 因此除非我們想要將 A 的其他 column vectors 用這組 basis 來表示, 一般是不需要進一步化成 reduced echelon form. 另外我們要強調的是 column space 的 basis 是要回到 A 的 column vectors 所組成, 而不是由 A 的 echelon form (或 reduced echelon form) pivot 所在的 column vectors 所組成. 這是因為一般我們在做 elementary row operations 已將 column vectors 各個 entry 做了調動, 所以 echelon form 的 column space 已不再是原來 A 的 column space 了.

我們將這個求 column space 的 basis 的方法做一個總結. 首先將 $m\times n$ matrix A 利用 elementary row operation 化為 echelon form A'. 假設 A' 的 pivot variables 為 $x_{i_1},\dots,x_{i_r}$, 則由於 A' 的 pivot 所在的 column vectors $\mathbf{a}'_{i_1},\dots,\mathbf{a}'_{i_r}$ 為 linearly independent 且 elementary row operations 會保持各 column vectors 之間的線性關係, 我們知對應到 A 的 column vectors $\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r}$ 亦為 linearly independent. 同理, 由於 A' 的其他 column vectors $\mathbf{a}'_j$ 皆符合 $\mathbf{a}'_j\in\operatorname{Span}(\mathbf{a}'_{i_1},\dots,\mathbf{a}'_{i_r})$, 我們得 A 的其他 column vectors $\mathbf{a}_j$ 也符合 $\mathbf{a}_j\in\operatorname{Span}(\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r})$. 因此得 $\operatorname{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)=\operatorname{Span}(\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r})$. 我們證得了 $\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r}$ 為 A 的 column space 的 spanning vectors 且為 linearly independent, 故 $\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r}$ 為 A 的 column space 的一組 basis. 我們有以下的定理.

Proposition 4.4.8. *假設 A 為 $m\times n$ matrix 且 $\mathbf{a}_1,\dots,\mathbf{a}_n\in\mathbb{R}^m$ 為 A 的 column vectors. 若利用 elementary row operations 將 A 化為 echelon form A' 後, A' 的 pivot 個數為 r, 則 A 的 column space 的 dimension 為 r. 假設 A' 的 pivot variables 為 $x_{i_1},\dots,x_{i_r}$, 則 $\mathbf{a}_{i_1},\dots,\mathbf{a}_{i_r}$ 為 A 的 column space 的一組 basis.*

接著我們看 row space 的 basis. 回顧一下, 若 $A=\begin{bmatrix} — & {}_1\mathbf{a} & — \\ & \vdots & \\ — & {}_m\mathbf{a} & — \end{bmatrix}$ 為 $m\times n$, 則 A 的 row space 為 $\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})$. 要求 A 的 row space 的 basis, 我們可以考慮 A 的 transpose A^{T}. 因為 A^{T} 的 column vectors 就是 A 的 row vectors, 求出 A^{T} 的 column space 的 basis 就等同於求 A 的 row space 的 basis. 所以我們可以用求 column space 的 basis 方法求出 A^{T} 的 column space 的 basis, 便得到 A 的 row space 的 basis. 不過這個方法有個缺點, 因為我們是換了一個矩陣 A^{T} 做 row operations, 因此就無法得到和原來 A 的 column space 之間的關係了. 以下介紹的方法, 便是直接對 A 做 elementary row operations 來求得 A 的 row space 的 basis, 所以我們可以得到 A 的 row space 和 column space 之間的關係.

這個方法的主要概念是 A 經過 elementary row operations 變換成 A' 後, A 和 A' 的 row space 是相同的. 這是因為若 ${}_1\mathbf{a},\ldots,{}_m\mathbf{a}$ 為 A 的 row vectors, ${}_1\mathbf{a}',\ldots,{}_m\mathbf{a}'$ 為 A' 的 row vectors, 則每個 ${}_i\mathbf{a}'$ 其實是 ${}_1\mathbf{a},\ldots,{}_m\mathbf{a}$ 中的向量互相交換, 或是乘上某個非 0 實數, 或是乘上某個實數後加到另一個向量. 也就是說每個 ${}_i\mathbf{a}'$ 其實是 ${}_1\mathbf{a},\ldots,{}_m\mathbf{a}$ 的線性組合, 所以對所有 $i=1,\ldots,m$ 皆有 ${}_i\mathbf{a}'\in\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})$. 因此由 $\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})$ 是 $\mathbb{R}^n$ 的 subspace 知 $\mathrm{Span}({}_1\mathbf{a}',\ldots,{}_m\mathbf{a}')\subseteq\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})$. 同理, 因 elementary row operation 是可以還原的, A' 也可經由 elementary row operations 換成 A, 所以我們也有 $\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})\subseteq\mathrm{Span}({}_1\mathbf{a}',\ldots,{}_m\mathbf{a}')$. 得證 $\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})=\mathrm{Span}({}_1\mathbf{a}',\ldots,{}_m\mathbf{a}')$, 亦即 A 和 A' 有相同的 row space. 我們看以下的例子.

Example 4.4.9. 考慮 Example 4.4.5 中的 4×6 matrix A, 且利用 elementary row operation 將之化為 reduced echelon form A' (參見 Example 4.4.7), 令 ${}_1\mathbf{a},{}_2\mathbf{a},{}_3\mathbf{a},{}_4\mathbf{a}$ 為 A 的 row vectors, ${}_1\mathbf{a}',{}_2\mathbf{a}',{}_3\mathbf{a}',{}_4\mathbf{a}'$ 為 A' 的 row vectors. 亦即

$$ {}_1\mathbf{a}=[2\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0],\ {}_2\mathbf{a}=[1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0],\ {}_3\mathbf{a}=[1\ \ 1\ \ 1\ \ 0\ \ 1\ \ 2],\ {}_4\mathbf{a}=[1\ \ 2\ \ 2\ \ -2\ \ 1\ \ 2], $$

$$ {}_1\mathbf{a}'=[1\ \ 0\ \ 0\ \ 0\ \ -1\ \ -2],\ {}_2\mathbf{a}'=[0\ \ 1\ \ 1\ \ 0\ \ 2\ \ 4],\ {}_3\mathbf{a}'=[0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 2],\ {}_4\mathbf{a}'=[0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0]. $$

利用 Example 4.4.5 的 elementary row operations, 我們知 A' 的 3-rd row ${}_3\mathbf{a}'$ 是由 A 的 3-rd row 減去 A 的 2-nd row 後再減去 A 的 2-nd row 乘上 -2 加到 1-st row 的向量, 亦即

$$ ({}_3\mathbf{a}-{}_2\mathbf{a})-({}_1\mathbf{a}-2\,{}_2\mathbf{a})={}_3\mathbf{a}+{}_2\mathbf{a}-{}_1\mathbf{a}=[1\ \ 1\ \ 1\ \ 0\ \ 1\ \ 2]+[1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0]-[2\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0]={}_3\mathbf{a}'. $$

而利用 Example 4.4.7 的 elementary row operations, A' 的 2-rd row ${}_2\mathbf{a}'$ 是由 A 的 2-nd row 乘上 -2 加到 A 的 1-st row 後再加上 2 倍的 A' 的 3-rd row 的向量, 亦即

$$ ({}_1\mathbf{a}-2\,{}_2\mathbf{a})+2({}_3\mathbf{a}+{}_2\mathbf{a}-{}_1\mathbf{a})=2\,{}_3\mathbf{a}-{}_1\mathbf{a}=[2\ \ 2\ \ 2\ \ 0\ \ 2\ \ 4]-[2\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0]={}_2\mathbf{a}'. $$

而 A' 的 1-st row ${}_1\mathbf{a}'$ 是由 A 的 2-nd row 減去 A' 的 3-rd row 的向量, 亦即

$$ {}_2\mathbf{a}-({}_3\mathbf{a}+{}_2\mathbf{a}-{}_1\mathbf{a})={}_1\mathbf{a}-{}_3\mathbf{a}=[2\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0]-[1\ \ 1\ \ 1\ \ 0\ \ 1\ \ 2]={}_1\mathbf{a}'. $$

從這裡我們得 $\mathrm{Span}({}_1\mathbf{a}',{}_2\mathbf{a}',{}_3\mathbf{a}',{}_4\mathbf{a}')\subseteq\mathrm{Span}({}_1\mathbf{a},{}_2\mathbf{a},{}_3\mathbf{a},{}_4\mathbf{a})$. 同理得 $\mathrm{Span}({}_1\mathbf{a},{}_2\mathbf{a},{}_3\mathbf{a},{}_4\mathbf{a})\subseteq\mathrm{Span}({}_1\mathbf{a}',{}_2\mathbf{a}',{}_3\mathbf{a}',{}_4\mathbf{a}')$ (此處略去不檢查了). 故得 $\mathrm{Span}({}_1\mathbf{a},{}_2\mathbf{a},{}_3\mathbf{a},{}_4\mathbf{a})=\mathrm{Span}({}_1\mathbf{a}',{}_2\mathbf{a}',{}_3\mathbf{a}',{}_4\mathbf{a}')$, 亦即 ${}_1\mathbf{a}',{}_2\mathbf{a}',{}_3\mathbf{a}',{}_4\mathbf{a}'$ 為 A 的 row space 的 spanning vectors. 在 echelon form 中, 沒有 pivot 的 row 必為零向量. 現 A' 的 pivot 個數為 3, 即 pivot 發生於前 3 個 row ${}_1\mathbf{a}',{}_2\mathbf{a}',{}_3\mathbf{a}'$, 而

${}_4\mathbf{a}'$ 為零向量, 所以僅 pivot 所在的 row ${}_1\mathbf{a}', {}_2\mathbf{a}', {}_3\mathbf{a}'$ 就可以成為 A 的 row space 的 spanning vectors. 現又由於 A' 為 reduced echelon form, 每一個 row 中 pivot 所在的位置其他的 row 在該位置皆為 0, 所以 ${}_1\mathbf{a}', {}_2\mathbf{a}', {}_3\mathbf{a}'$ 為 linearly independent. 得證 ${}_1\mathbf{a}', {}_2\mathbf{a}', {}_3\mathbf{a}'$ 為 A 的 row space 的一組 basis.

注意在 Example 4.4.9 中我們為了方便說明將 A 化為 reduced echelon form. 事實上既然 A 的 echelon form 和 reduced echelon form 有相同的 row space, 而它們 pivot 的個數又相同, 所以由 dimension 的性質 (即 Corollary 4.3.5), 知 echelon form 中 pivot 所在的 row vectors 也會是 A 的 row space 的一組 basis. 化成 reduced echelon form 的好處是比較容易讓我們將 row space 中的向量寫成這組 basis 的線性組合. 因此因此除非我們想要將 A 的 row space 中的 vectors 用這組 basis 來表示, 若僅想找到 row space 的 basis 一般是不需要進一步化成 reduced echelon form. 另外我們要強調的是 row space 的 basis 不可以回到 A 的 row vectors 去找. 這是因為一般我們在做 elementary row operations 已將 row vectors 所在的位置做了調動, 所以 row operation 並沒有保持 row vectors 之間的線性關係.

我們將這個求 row space 的 basis 的方法做一個總結. 首先將 $m\times n$ matrix A 利用 elementary row operation 化為 echelon form A'. 假設 A' 的 pivot 個數為 r, 則由於 A' 為 echelon form, A' 前 r 個 row vectors ${}_1\mathbf{a}', \dots, {}_r\mathbf{a}'$ 為 nonzero vectors. A' 其餘的 row vectors 皆為 zero vectors. 由於 elementary row operations 會保持 row space, 我們得 ${}_1\mathbf{a}', \dots, {}_r\mathbf{a}'$ 為 A 的 row space 的 spanning vectors. 又由化為 reduced echelon form 的情形我們知 A 的 row space 的 dimension 為 r, 故由 Corollary 4.3.5 知 ${}_1\mathbf{a}', \dots, {}_r\mathbf{a}'$ 為 A 的 row space 的一組 basis. 我們有以下的定理.

Proposition 4.4.10. *假設 A 為 $m\times n$ matrix. 若利用 elementary row operations 將 A 化為 echelon form A' 後, A' 的 pivot 個數為 r, 則 A 的 row space 的 dimension 為 r 且 A' 的前 r 個 row vectors ${}_1\mathbf{a}', \dots, {}_r\mathbf{a}'$ (即 A' 中的 nonzero row vectors) 為 A 的 row space 的一組 basis.*

我們可以利用找 column space 和 row space 的 basis 的方法找一般 $\mathbb{R}^m$ 的 subspace V 的 basis. 首先我們先找出 V 的一組 spanning vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$, 然後造一個以 $\mathbf{v}_1, \dots, \mathbf{v}_n$ 為 column vectors 的矩陣 $A = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix}$. 然後再利用找 A 的 column space 的 basis 的方法得到 V 的一組 basis. 我們也可造一個以 $\mathbf{v}_1^{\mathrm{T}}, \dots, \mathbf{v}_n^{\mathrm{T}}$ 為 row vectors 的 $n\times m$ matrix $B = \begin{bmatrix} — & \mathbf{v}_1^{\mathrm{T}} & — \\ & \vdots & \\ — & \mathbf{v}_n^{\mathrm{T}} & — \end{bmatrix}$ 然後再利用找 B 的 row space 的 basis 的方法得到 V 的一組 basis. 兩種方法都有它們的好處. 利用 column vectors 的方法, 由於最後找出的 basis 是從原來的 spanning vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ 中的向量所組成, 所以適合處理希望 basis 的 vectors 是由原來 spanning vectors 中選出的問題. 而利用 row vectors 的方法, 由於可以化為 reduced echelon form, 而 basis 是由此 reduced echelon form 中的 nonzero vectors 所組成, 所以雖然和來的

spanning vectors 無關, 不過很適合拿來判斷哪些向量在此 subspace 以及處理將 subspace 中的向量用此 basis 表示的問題. 例如以下的例子.

Example 4.4.11. 考慮 $\mathbb{R}^6$ 中的向量

$$\mathbf{v}_1=\begin{bmatrix}2\\1\\1\\0\\0\\0\end{bmatrix},\mathbf{v}_2=\begin{bmatrix}1\\0\\0\\1\\0\\0\end{bmatrix},\mathbf{v}_3=\begin{bmatrix}1\\1\\1\\0\\1\\2\end{bmatrix},\mathbf{v}_4=\begin{bmatrix}1\\2\\2\\-2\\1\\2\end{bmatrix}\quad\text{and}\quad\mathbf{w}=\begin{bmatrix}1\\-2\\-2\\3\\-2\\4\end{bmatrix}.$$

令 $V=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4)$, 試找出 V 的一組 basis, 並用之判斷 $\mathbf{w}$ 是否在 V 中.

依前面的討論我們知此問題適合用 row space 的方式處理, 所以考慮以 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4$ 為 row vectors 的矩陣 A. 此時 A 就是 Example 4.4.9 中的矩陣 A. 利用 Example 4.4.9 的結果, 我們得

$$\mathbf{u}_1=\begin{bmatrix}1\\0\\0\\0\\-1\\-2\end{bmatrix},\mathbf{u}_2=\begin{bmatrix}0\\1\\1\\0\\2\\4\end{bmatrix},\mathbf{u}_3=\begin{bmatrix}0\\0\\0\\1\\1\\2\end{bmatrix}$$

為 V 的一組 basis. 注意因為原來 V 中的向量為 column vector 的形式, 為了一致性這裡我們將 A 的 row space 的 basis 寫回成 column vectors. 依此我們知 $\mathbf{w}\in V$ 若且唯若存在 $c_1,c_2,c_3\in\mathbb{R}$ 使得 $\mathbf{w}=c_1\mathbf{u}_1+c_2\mathbf{u}_2+c_3\mathbf{u}_3$. 然而

$$c_1\mathbf{u}_1+c_2\mathbf{u}_2+c_3\mathbf{u}_3=\begin{bmatrix}c_1\\c_2\\c_2\\c_3\\-c_1+2c_2+c_3\\-2c_1+4c_2+2c_3\end{bmatrix},$$

我們發現要使得 $\mathbf{w}=c_1\mathbf{u}_1+c_2\mathbf{u}_2+c_3\mathbf{u}_3$, 在 1-st, 2-nd 和 3-rd entry 的地方 (這些地方就是 reduced echelon form 的 pivot 的位置) 需有 $c_1=1,c_2=-2,c_3=3$. 故將 $c_1=1,c_2=-2,c_3=3$ 代入, 發現每個 entry 皆吻合, 故有 $\mathbf{w}=\mathbf{u}_1-2\mathbf{u}_2+3\mathbf{u}_3$, 得知 $\mathbf{w}\in V$.

當我們對要找的 basis 沒有特殊要求時, 我們可以選擇兩種方法中可以使得矩陣的 row 的個數較少的那一種方法處理. 因為如此所需的 elementary row operations 相對起來會較少, 可以較快找到一組 basis. 例如前一個例子 Example 4.4.11 中的 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4$ 若考慮成 column vectors, 所得的矩陣為 6×4 matrix, 而考慮成 row vectors, 所得的矩陣為 4×6 matrix. 所以此時若僅想找出 $\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4)$ 的一組 basis, 用 row vectors 的情況處理會比較快.

給定一個矩陣 A, 從 Proposition 4.4.8 和 Proposition 4.4.10 我們知道 A 的 column space 和 row space 有相同的 dimension, 亦即 A 利用 elementary row operations 化為 echelon form 後其 pivot 的個數, 即 A 的 rank. A 的 column space 的維度就是 A 的 rank, 至於 A 的 nullspace 的維度, 我們也給予一特殊的名稱, 即以下的定義.

Definition 4.4.12. 假設 A 為 $m\times n$ matrix. A 的 nullspace 的 dimension 稱為 A 的 *nullity*, 記為 $\mathrm{null}(A)$, 亦即 $\mathrm{null}(A)=\dim(N(A))$.

依此定義, 由 Proposition 4.4.8 和 Proposition 4.4.10 我們知道 $\mathrm{rank}(A)$ 即為 A 利用 elementary row operations 化為 echelon form 後其 pivot 的個數, 而由 Proposition 4.4.4 我們知道 $\mathrm{null}(A)$ 就是 homogeneous linear system $A\mathbf{x}=\mathbf{O}$ 的 free variables 的個數, 即 A 的 column 個數減去 echelon form 的 pivot 個數, 因此我們有以下的 *Dimension Theorem* (或稱為 *rank equation*).

Theorem 4.4.13 (Dimension Theorem)**.** 假設 A 為 $m\times n$ *matrix.* 則

$$\mathrm{rank}(A)+\mathrm{null}(A)=n.$$

Question 4.4. 假設 A 為 $n\times n$ *invertible matrix.* 試求 $\mathrm{rank}(A)$ 以及 $\mathrm{null}(A)$.

Question 4.5. 假設 A 為 $m\times n$ *matrix.*

(1) 若對於任意 $\mathbf{v}\in\mathbb{R}^m$ 聯立方程組 $A\mathbf{x}=\mathbf{v}$ 皆有解, 試求 $\mathrm{rank}(A)$ 以及 $\mathrm{null}(A)$.

(2) 若存在 $\mathbf{v}\in\mathbb{R}^m$ 使得聯立方程組 $A\mathbf{x}=\mathbf{v}$ 有唯一解, 試求 $\mathrm{rank}(A)$ 以及 $\mathrm{null}(A)$.

考慮矩陣 A 的 transpose A^{T}. 由於 A 的 column space 就是 A^{T} 的 row space, 而 A 的 row space 就是 A^{T} 的 column space, 所以我們有以下的性質.

Proposition 4.4.14. 假設 A 為 $m\times n$ *matrix.* 則

$$\mathrm{rank}(A)=\mathrm{rank}(A^{\mathrm{T}}).$$

亦即利用 *elementary row operations* 將 A 以及 A^{T} 化為 *echelon form* 後, 它們的 *pivot* 個數相同.

注意, 若要直接證明利用 elementary row operations 將 A 以及 A^{T} 化為 echelon form 後, 它們的 pivot 個數相同, 會有相當的困難度. 但將這個問題轉為 column space 以及 row space 的 dimension 問題, 就很容易解決. 其實許多數學問題都是這樣, 有時只要換個角度看問題就能迎刃而解. 這也是當我們介紹一個新的概念後常常會去探討這個概念還有甚麼等價的條件的原因.

Question 4.6. 假設 A 為 $m\times n$ *matrix.* 試證明 $\mathrm{null}(A^{\mathrm{T}})=m-\mathrm{rank}(A)$.

4.5. 結論

在本章中我們了解了 $\mathbb{R}^m$ 的 subspace 以及其 basis 的意義. 一組 vectors 要成為一個 R^m 的 subspace 的 basis 的充要條件就是這組 vectors 必須是此 subspace 的 spanning vectors 以及必須為 linearly independent. 利用 spanning vectors 以及 linearly independent 相互間

的關係, 我們知道了一個 R^m 的 subspace 其 basis 是存在的, 而且組成 basis 的向量個數是唯一的. 因此有了 dimension (維度) 的概念. 最後我們由求一個矩陣的 column space, row space 以及 nullspace 的 basis 的方法, 得知 column space 和 row space 的 dimension 會相等, 即為此矩陣化為 echelon form 後, 其 pivot 個數, 也就是此矩陣的 rank. 而且定義 nullspace 的 dimension 為此矩陣的 nullity, 並因此推得了 Dimension Theorem.

Chapter 5

Linear Transformations of $\mathbb{R}^n$

在本章中我們介紹在 $\mathbb{R}^n$ 中重要的函數, 所謂的 linear transformation. 我們會介紹 linear transformation 相關的基本性質. 然後引進其矩陣表示法, 將 linear transformation 與矩陣相連結.

5.1. Basic Properties

在數學中, 函數是我們常常利用來了解所要探討的結構的重要工具. 在線性代數中, 我們要探討的結構就是 vector space, 而 linear transformation 就是幫助我們探討及理解 vector spaces 相互之間的關係的重要函數與工具.

5.1.1. Function. 給定 $\mathbb{R}^n$, $\mathbb{R}^m$. 若有一個從 $\mathbb{R}^n$ 的向量對應到 $\mathbb{R}^m$ 的向量的對應關係, 即對任意 $\mathbf{v} \in \mathbb{R}^n$, 此對應會將 $\mathbf{v}$ 對應到 $\mathbb{R}^m$ 中一個向量 $\mathbf{w}$. 若對每一個 $\mathbf{v} \in \mathbb{R}^n$, 所對應到的 $\mathbf{w}$ 我們用 $T(\mathbf{v})$ 來表示, 我們就用 $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(\mathbf{v}) = \mathbf{w} \in \mathbb{R}^m$, $\forall \mathbf{v} \in \mathbb{R}^n$ 來表示這一個對應關係, 而稱 T 是一個從 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 *function* (函數). 這裡 $\mathbb{R}^n$ 稱為 T 的 *domain* (定義域), 而 $\mathbb{R}^m$ 就稱為 T 的 *codomain* (對應域). 注意依函數的定義, 若 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是一個函數, 則對任意 $\mathbf{v} \in \mathbb{R}^n$, $T(\mathbf{v})$ 一定要是 $\mathbb{R}^m$ 中的一個確定的向量. 也就是說 $T(\mathbf{v}) \in \mathbb{R}^m$, 而且不能一下子令 $T(\mathbf{v}) = \mathbf{w}$, 一下子又改變成 $T(\mathbf{v}) = \mathbf{w}'$, 但 $\mathbf{w} \neq \mathbf{w}'$, 造成不一致的情形發生. 所以當我們在建構一個新的函數時, 一定要確認這一點. 也就是說, 我們必須說明定出來的函數是 *well-defined*.

另外要注意, 函數的定義中並沒有要求定義域中每一個元素都要對應到不同的元素去. 也就是說若 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是一個函數, 是容許在 $\mathbb{R}^n$ 中有兩個向量 $\mathbf{v} \neq \mathbf{v}'$, 但 $T(\mathbf{v}) = T(\mathbf{v}')$. 若我們多要求定義域中每一個元素都要對應到不同的元素, 即 $\mathbb{R}^n$ 中任意兩相異向量 $\mathbf{v} \neq \mathbf{v}'$, 所對應的 $T(\mathbf{v})$ 和 $T(\mathbf{v}')$ 要相異 (即 $T(\mathbf{v}) \neq T(\mathbf{v}')$), 則我們給這樣的函數一個特殊的名稱, 稱這樣的函數為 *one-to-one* (一對一), 有時也稱為 *injective*. 另外, 函數的定義中也沒有要求對應域中每一個元素都要被對應到. 也就是說若 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是一個函數, 是容許在 $\mathbb{R}^m$ 中

有向量 $\mathbf{w}$, 沒有任何 $\mathbb{R}^n$ 的向量會對應到 $\mathbf{w}$ (即不存在 $\mathbf{v}\in\mathbb{R}^n$ 使得 $T(\mathbf{v})=\mathbf{w}$). 若我們多要求對應域中每一個元素都要被對應到, 即 $\mathbb{R}^m$ 中任意向量 $\mathbf{w}$, 皆可找到 $\mathbf{v}\in\mathbb{R}^n$ 使得 $T(\mathbf{v})=\mathbf{w}$, 則我們給這樣的函數一個特殊的名稱, 稱這樣的函數為 *onto* (映成), 有時也稱為 *surjective*.

當一個函數是 one-to-one 且 onto (此時一般稱為 *bijective*), 那就更特別了. 這時候一定可以找到一個從原來函數的對應域送到原來函數的定義域的反向函數 (我們稱為原函數的 *inverse* (反函數)), 使其合成後會是將每一個元素自己映射到自己的函數 (即所謂的 identity function). 所以此時我們也稱這樣的函數為 *invertible* (可逆函數).

5.1.2. Linear Transformation. 要了解 $\mathbb{R}^n$ 中的向量, 若僅是考慮一般的函數, 並無法利用向量之間的運算, 幫助我們了解 $\mathbb{R}^n$ 中向量的結構. 我們需要的函數是能保持向量運算的, 所以有以下之定義.

Definition 5.1.1. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為一個函數, 若 T 滿足對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in\mathbb{R}^n$ 以及 $c_1,\ldots,c_k\in\mathbb{R}$ 皆有

$$T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k)=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k).$$

則稱 T 為一個 *linear transformation.* 有時我們簡稱 T 為 *linear.*

要注意這裡 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_1$ 是 $\mathbb{R}^n$ 中向量的線性組合, 而 $c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)$ 是 $\mathbb{R}^m$ 中向量的線性組合, 要區分清楚. 尤其在 $n\neq m$ 時要特別注意. 特別是當 $\mathbf{O}\in\mathbb{R}^n$ 時, 依 linear transformation 的定義, 我們有 $T(\mathbf{O})=T(\mathbf{O}+\mathbf{O})=T(\mathbf{O})+T(\mathbf{O})$. 此時兩邊加上 $T(\mathbf{O})$ 的加法反元素, 得 $T(\mathbf{O})$ 應為 $\mathbb{R}^m$ 中的零向量. 也就是說一個 linear transformation $T:\mathbb{R}^n\to\mathbb{R}^m$, 會將 $\mathbb{R}^n$ 中的零向量映射到 $\mathbb{R}^m$ 中的零向量. 雖然當 $n\neq m$ 時, $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 的零向量是不同的, 不過一般我們都用 $\mathbf{O}$ 來表示, 而不區分它. 所以我們仍用 $T(\mathbf{O})=\mathbf{O}$ 來表示 linear transformation 會將 $\mathbb{R}^n$ 中的零向量映射到 $\mathbb{R}^m$ 中的零向量. 這個性質雖然簡單, 但相當有用, 我們特別將此性質敘述如下.

Lemma 5.1.2. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為一個 *linear transformation.* 則 T 會將 $\mathbb{R}^n$ 中的零向量映射到 $\mathbb{R}^m$ 中的零向量, 亦即 $T(\mathbf{O})=\mathbf{O}$.

再次提醒, 這裡兩個 $\mathbf{O}$ 哪一個是 $\mathbb{R}^n$ 的零向量, 哪一個是 $\mathbb{R}^m$ 的零向量, 一定要區分清楚.

依定義要檢查 $T:\mathbb{R}^n\to\mathbb{R}^m$ 是否為 linear transformation, 我們必須考慮 $\mathbb{R}^n$ 中任意有限多個向量的線性組合代入 T 中是否符合 linear transformation 的要求, 感覺起來很麻煩. 事實上, 如同檢查 subspace 的方法 (參見 Proposition 4.1.2), 下一個定理告訴我們, 只要檢查任兩個向量的線性組合即可.

Proposition 5.1.3. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為一個函數, 則 T 為 *linear transformation* 若且唯若對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, $r\in\mathbb{R}$ 皆有 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$.

Proof. $(\Rightarrow)$: 依 T 為 linear 的定義, 對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{R}$ 皆有 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$.

($\Leftarrow$) : 我們要利用對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$, $r\in\mathbb{R}$ 皆有 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$ 這個性質來證明對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in\mathbb{R}^n$ 以及 $c_1,\ldots,c_k\in\mathbb{R}$ 皆有 $T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k)=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)$. 我們對向量的個數 k 作數學歸納法. 首先考慮只有一個向量的情形 (即 $k=1$), 我們要證明若 $\mathbf{v}_1\in\mathbb{R}^n$, $c_1\in\mathbb{R}$ 則 $T(c_1\mathbf{v}_1)=c_1T(\mathbf{v}_1)$. 此時考慮 $\mathbf{u}=\mathbf{O}$, $\mathbf{v}=\mathbf{v}_1$, $r=c_1$, 依 Lemma 5.1.2, 我們有 $T(c_1\mathbf{v}_1)=T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})=\mathbf{O}+rT(\mathbf{v})=c_1T(\mathbf{v}_1)$. 得證 $k=1$ 的情形成立. 現假設有 k 個向量時成立, 亦即對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in\mathbb{R}^n$ 以及 $c_1,\ldots,c_k\in\mathbb{R}$ 皆有 $T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k)=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)$. 我們要證明對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1}\in\mathbb{R}^n$ 以及 $c_1,\ldots,c_k,c_{k+1}\in\mathbb{R}$ 皆有 $T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k+c_{k+1}\mathbf{v}_{k+1})=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)+c_{k+1}T(\mathbf{v}_{k+1})$. 然而對此時令 $\mathbf{u}=c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k$, $\mathbf{v}=\mathbf{v}_{k+1}$ 以及 $r=c_{k+1}$. 依歸納假設我們有 $T(\mathbf{u})=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)$, 故

$$\begin{aligned}&T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k+c_{k+1}\mathbf{v}_{k+1})=\\&\quad T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)+c_{k+1}T(\mathbf{v}_{k+1}).\end{aligned}$$

故由數學歸納法知 T 為 linear transformation. $\square$

Example 5.1.4. (1) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix})=\begin{bmatrix}x_1+x_2\\x_1-x_3\end{bmatrix}$. 我們驗證 T 是一個 linear transformation. 任取 $\mathbf{u}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix},\mathbf{v}=\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}\in\mathbb{R}^3$, 以及 $r\in\mathbb{R}$. 我們有 $\mathbf{u}+r\mathbf{v}=\begin{bmatrix}a_1+rb_1\\a_2+rb_2\\a_3+rb_3\end{bmatrix}$. 故依 T 的定義, 我們有

$$T(\mathbf{u}+r\mathbf{v})=T(\begin{bmatrix}a_1+rb_1\\a_2+rb_2\\a_3+rb_3\end{bmatrix})=\begin{bmatrix}(a_1+rb_1)+(a_2+rb_2)\\(a_1+rb_1)-(a_3+rb_3)\end{bmatrix}=\begin{bmatrix}a_1+a_2+rb_1+rb_2\\a_1-a_3+rb_1-rb_3\end{bmatrix}.$$

另一方面我們有 $T(\mathbf{u})=T(\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix})=\begin{bmatrix}a_1+a_2\\a_1-a_3\end{bmatrix}$, $T(\mathbf{v})=T(\begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix})=\begin{bmatrix}b_1+b_2\\b_1-b_3\end{bmatrix}$, 故

$$T(\mathbf{u})+rT(\mathbf{v})=\begin{bmatrix}a_1+a_2\\a_1-a_3\end{bmatrix}+r\begin{bmatrix}b_1+b_2\\b_1-b_3\end{bmatrix}=\begin{bmatrix}a_1+a_2+rb_1+rb_2\\a_1-a_3+rb_1-rb_3\end{bmatrix}.$$

得證 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$, 故 T 為 linear transformation.

(2) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix})=\begin{bmatrix}x_1+x_2+1\\x_1-x_3\end{bmatrix}$. 我們說明 T 不是 linear transformation. 依 T 的定義, 我們有 $T(\mathbf{O})=\begin{bmatrix}1\\0\end{bmatrix}\neq\mathbf{O}$, 故由 Lemma 5.1.2 知, T 不是 linear transformation.

(3) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} x_1^2+x_2 \\ x_1-x_3 \end{bmatrix}$. 我們說明 T 不是 linear transformation. 雖然此時 $T(\mathbf{O})=\mathbf{O}$, 但 $T(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, 而 $T(\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}) = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$ 得

$$T(\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}) = T(2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}) \neq 2T(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}).$$

此與 linear transformation 的條件不符, 故 T 不是 linear transformation.

接下來我們來看一個最常見的 linear transformation, 事實上以後我們會知道所有 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation 都是這樣的形式.

Lemma 5.1.5. 令 A 為一個 $m\times n$ *matrix.* 考慮 $T:\mathbb{R}^n\to\mathbb{R}^m$ 定義為*:* $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$*.* 則 T 為一個 *linear transformation.*

Proof. 首先我們先檢查 T 是 well-defined, 也就是說 T 確實是一個從 $\mathbb{R}^n$ 映到 $\mathbb{R}^m$ 的函數. 任取 $\mathbf{v}\in\mathbb{R}^n$, 依定義 $T(\mathbf{v})=A\mathbf{v}$. 然而 A 為 $m\times n$ matrix, 依矩陣乘法定義 $A\mathbf{v}$ 是一個 $m\times 1$ matrix (注意這裡向量都視為 column vector, 所以 $\mathbf{v}\in\mathbb{R}^n$ 為 $n\times 1$ matrix), 故 $A\mathbf{v}\in\mathbb{R}^m$. 得 T 確實是一個從 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 function.

現要證明 T 為 linear, 亦即對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 以及 $r\in\mathbb{R}$, 我們要證明 $T(\mathbf{u}+r\mathbf{v}) = T(\mathbf{u})+rT(\mathbf{v})$. 不過依 T 的定義 $T(\mathbf{u})=A\mathbf{u}$, $T(\mathbf{v})=A\mathbf{v}$, 而 $T(\mathbf{u}+r\mathbf{v})=A(\mathbf{u}+r\mathbf{v})$. 故依矩陣乘法加法的分配律 (Proposition 3.1.8 和 Proposition 3.1.9) 我們得

$$T(\mathbf{u}+r\mathbf{v}) = A(\mathbf{u}+r\mathbf{v}) = A\mathbf{u}+A(r\mathbf{v}) = A\mathbf{u}+rA\mathbf{v} = T(\mathbf{u})+rT(\mathbf{v}).$$

□

從 Lemma 5.1.5 我們知道可以造出許多的 linear transformations. 事實上, 我們可以利用現有的 linear transformations 造出更多的 linear transformations. 首先若 T_1,T_2 皆為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation, 我們可以利用 T_1,T_2 造出新的 linear transformation, T_1+T_2. 前面已說過, 要造出新的函數需先說明定義域和對應域是甚麼. 這裡我們定義 T_1+T_2 仍為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的函數. 對於任意 $\mathbf{v}\in\mathbb{R}^n$, 我們定義 $(T_1+T_2)(\mathbf{v}) = T_1(\mathbf{v})+T_2(\mathbf{v})$. 依此定義, T_1+T_2 確實將 $\mathbb{R}^n$ 的向量映射到 $\mathbb{R}^m$ 中. 這是因為依假設, 對任意 $\mathbf{v}\in\mathbb{R}^n$, 我們有 $T_1(\mathbf{v})\in\mathbb{R}^m$ 以及 $T_2(\mathbf{v})\in\mathbb{R}^m$, 所以自然有 $(T_1+T_2)(\mathbf{v}) = T_1(\mathbf{v})+T_2(\mathbf{v})\in\mathbb{R}^m$. 所以 $T_1+T_2:\mathbb{R}^n\to\mathbb{R}^m$ 確實是 well-defined function. 接下來我們要說明若 $T_1:\mathbb{R}^n\to\mathbb{R}^m$, $T_2:\mathbb{R}^n\to\mathbb{R}^m$ 皆為 linear transformation, 則 $T_1+T_2:\mathbb{R}^n\to\mathbb{R}^m$ 亦為 linear transformation. 也就是說對於任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 以及 $r\in\mathbb{R}$, 我們要證明 $(T_1+T_2)(\mathbf{u}+r\mathbf{v}) = (T_1+T_2)(\mathbf{u})+r(T_1+T_2)(\mathbf{v})$. 首先依定義我們有

$$(T_1+T_2)(\mathbf{u}+r\mathbf{v}) = T_1(\mathbf{u}+r\mathbf{v})+T_2(\mathbf{u}+r\mathbf{v}).$$

接著利用 T_1,T_2 為 linear 我們得

$$T_1(\mathbf{u}+r\mathbf{v})+T_2(\mathbf{u}+r\mathbf{v}) = T_1(\mathbf{u})+rT_1(\mathbf{v})+T_2(\mathbf{u})+rT_2(\mathbf{v}).$$

另外, 依定義

$$(T_1+T_2)(\mathbf{u})+r(T_1+T_2)(\mathbf{v})=T_1(\mathbf{u})+T_2(\mathbf{u})+r(T_1(\mathbf{v})+T_2(\mathbf{v})),$$

故由向量運算性質, 得證 $(T_1+T_2)(\mathbf{u}+r\mathbf{v})=(T_1+T_2)(\mathbf{u})+r(T_1+T_2)(\mathbf{v})$, 亦即 T_1+T_2 為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation.

Question 5.1. 若 A_1,A_2 皆為 $m\times n$ *matrix.* 考慮 $T_1:\mathbb{R}^n\to\mathbb{R}^m$, $T_2:\mathbb{R}^n\to\mathbb{R}^m$, 分別定義為 $T_1(\mathbf{v})=A_1\mathbf{v}$, $T_2(\mathbf{v})=A_2\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$. 則 T_1+T_2 是怎樣的函數?

給定一個 linear transformation $T:\mathbb{R}^n\to\mathbb{R}^m$, 以及 $c\in\mathbb{R}$, 我們也可定義函數 $cT:\mathbb{R}^n\to\mathbb{R}^m$, 其定義為 $(cT)(\mathbf{v})=c(T(\mathbf{v}))$, $\forall\mathbf{v}\in\mathbb{R}^n$ (也就是說它把每一個 $\mathbb{R}^n$ 的向量 $\mathbf{v}$ 對應到 c 倍的 $T(\mathbf{v})$). 很容易看出 $rT:\mathbb{R}^n\to\mathbb{R}^m$ 確實是一個 function. 事實上, 它也是 linear transformation. 這是因為對於任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 以及 $r\in\mathbb{R}$, 我們有

$$(cT)(\mathbf{u}+r\mathbf{v})=c(T(\mathbf{u}+r\mathbf{v}))=cT(\mathbf{u})+rcT(\mathbf{v}).$$

而 $(cT)(\mathbf{u})+r(cT)(\mathbf{v})=c(T(\mathbf{u}))+rc(T(\mathbf{v}))$, 故知 $(cT)(\mathbf{u}+r\mathbf{v})=(cT)(\mathbf{u})+r(cT)(\mathbf{v})$, 得證 $cT:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation.

Question 5.2. 設 A 為 $m\times n$ *matrix.* 考慮 $T:\mathbb{R}^n\to\mathbb{R}^m$ 定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$. 則對於 $c\in\mathbb{R}$, cT 是怎樣的函數?

設 $T_1,\cdots,T_k$ 為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformations. $c_1,\ldots,c_k\in\mathbb{R}$, 則由前知 $c_1T_1,\ldots,c_kT_k$ 皆為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformations. 所以 $c_1T_2+c_2T_2$ 為 linear transformation. 再利用數學歸納法, 得 $c_1T_1+\cdots+c_kT_k$ 為 linear transformation. 因此我們有下面之結果.

Proposition 5.1.6. 設 $T_1,\cdots,T_k$ 為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 *linear transformations,* $c_1,\ldots,c_k\in\mathbb{R}$ 則 $c_1T_1+\cdots+c_kT_k$ 為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 *linear transformation.*

另一個產生 linear transformation 的方法就是利用 "合成函數". 若 $T:\mathbb{R}^n\to\mathbb{R}^m$ 和 $T':\mathbb{R}^m\to\mathbb{R}^k$ 為函數, 由於對任意 $\mathbf{v}\in\mathbb{R}^n$, 依定義 $T(\mathbf{v})\in\mathbb{R}^m$, 也就是說 $T(\mathbf{v})$ 會落在 T' 的定義域中. 所以我們可以將 $T(\mathbf{v})$ 代入 T' 中, 亦即得 $T'(T(\mathbf{v}))\in\mathbb{R}^k$. 這樣的方法幫我們定義出一個從 $\mathbb{R}^n$ 到 $\mathbb{R}^k$ 的函數, 稱之為 T,T' 的 *composite function* (合成函數), 我們用 $T'\circ T$ 來表示. 也就是說 $T'\circ T:\mathbb{R}^n\to\mathbb{R}^k$ 的定義為 $T'\circ T(\mathbf{v})=T'(T(\mathbf{v}))$, $\forall\mathbf{v}\in\mathbb{R}^n$. 我們有下面之結果.

Proposition 5.1.7. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 和 $T':\mathbb{R}^m\to\mathbb{R}^k$ 為 *linear transformation,* 則 $T'\circ T:\mathbb{R}^n\to\mathbb{R}^k$ 亦為 *linear transformation.*

Proof. 已知 $T'\circ T$ 為 function, 我們僅要證明 $T'\circ T$ 為 linear, 亦即對於任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 以及 $r\in\mathbb{R}$, 我們有 $(T'\circ T)(\mathbf{u}+r\mathbf{v})=(T'\circ T)(\mathbf{u})+r(T'\circ T)(\mathbf{v})$. 依定義 $(T'\circ T)(\mathbf{u}+r\mathbf{v})=T'(T(\mathbf{u}+r\mathbf{v}))$. 然而因為 T, T' 為 linear, 故有

$$T'(T(\mathbf{u}+r\mathbf{v}))=T'(T(\mathbf{u})+rT(\mathbf{v}))=T'(T(\mathbf{u}))+rT'(T(\mathbf{v})).$$

再由 $T'(T(\mathbf{u}))=(T'\circ T)(\mathbf{u})$ 以及 $T'(T(\mathbf{v}))=(T'\circ T)(\mathbf{v})$ 得證 $T'\circ T$ 為 linear transformation. □

Question 5.3. 設 A 為 $m\times n$ *matrix*, B 為 $k\times m$ *matrix*. 考慮 $T:\mathbb{R}^n\to\mathbb{R}^m$, $T':\mathbb{R}^m\to\mathbb{R}^k$, 分別定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$ 且 $T'(\mathbf{w})=B\mathbf{w}$, $\forall\mathbf{w}\in\mathbb{R}^m$. 則 $T'\circ T$ 是怎樣的函數?

當 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是 $\mathbb{R}^n$ 的一組 basis 時, 對任意 $\mathbf{v}\in\mathbb{R}^n$, 皆存在唯一的一組 $c_1,\dots,c_n\in\mathbb{R}$, 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 現若 $T:\mathbb{R}^n\to\mathbb{R}^m$ 是 linear transformation, 則由 linear transformation 定義知, 此時 $T(\mathbf{v})=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)$. 也就是說, 只要我們知道 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)$ 是 $\mathbb{R}^m$ 中的哪些向量, 則對於任意 $\mathbf{v}\in\mathbb{R}^n$, 我們都可以知道 $T(\mathbf{v})$ 為何. 因此我們有以下的定理.

Theorem 5.1.8. 假設 $\mathbf{v}_1,\dots,\mathbf{v}_n\in\mathbb{R}^n$, 為 $\mathbb{R}^n$ 的一組 *basis*. 令 $\mathbf{w}_1,\dots,\mathbf{w}_n\in\mathbb{R}^m$, 則存在唯一的 *linear transformation* $T:\mathbb{R}^n\to\mathbb{R}^m$, 滿足 $T(\mathbf{v}_1)=w_1,\dots,T(\mathbf{v}_i)=w_i,\dots,T(\mathbf{v}_n)=w_n$.

Proof. 首先證明存在性. 定義 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 $T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$, $\forall c_1,\dots,c_n\in\mathbb{R}$. 我們需說明這是 well-defined function. 也就是說對任意 $\mathbf{v}\in\mathbb{R}^n$, $T(\mathbf{v})$ 皆有定義且 $T(\mathbf{v})\in\mathbb{R}^m$. 然而因 $\mathbf{v}_1,\dots,\mathbf{v}_n\in\mathbb{R}^n$, 為 $\mathbb{R}^n$ 的一組 basis, 故存在唯一的一組 $c_1,\dots,c_n\in\mathbb{R}$, 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 故此時得 $T(\mathbf{v})=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n\in\mathbb{R}^m$. 接著我們要說明 T 為 linear transformation, 也就是說對任意 $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$ 以及 $r\in\mathbb{R}$, 我們要證明 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$. 由於 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 $\mathbb{R}^n$ 的一組 basis, 存在 $c_1,\dots,c_n$ 以及 $d_1,\dots,d_n\in\mathbb{R}$ 使得 $\mathbf{u}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 且 $\mathbf{v}=d_1\mathbf{v}_1+\cdots+d_n\mathbf{v}_n$. 故此時 $\mathbf{u}+r\mathbf{v}=(c_1+rd_1)\mathbf{v}_1+\cdots+(c_n+rd_n)\mathbf{v}_n$. 依 T 的定義得

$$T(\mathbf{u}+r\mathbf{v})=(c_1+rd_1)T(\mathbf{v}_1)+\cdots+(c_n+rd_n)T(\mathbf{v}_n)=(c_1+rd_1)\mathbf{w}_1+\cdots+(c_n+rd_n)\mathbf{w}_n.$$

另一方面

$$T(\mathbf{u})+rT(\mathbf{v})=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)+rT(d_1\mathbf{v}_1+\cdots+d_n\mathbf{v}_n)=$$
$$(c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n)+r(d_1\mathbf{w}_1+\cdots+d_n\mathbf{w}_n)=(c_1+rd_1)\mathbf{w}_1+\cdots+(c_n+rd_n)\mathbf{w}_n.$$

得證 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$.

接著證明唯一性, 我們用反證法. 也就是說若 $T':\mathbb{R}^n\to\mathbb{R}^m$ 是另一個 linear transformation 滿足 $T'(\mathbf{v}_1)=\mathbf{w}_1,\dots,T'(\mathbf{v}_n)=\mathbf{w}_n$, 且 $T'\neq T$, 則會造成矛盾. 依定義, $T'\neq T$ 表示存在 $\mathbf{v}\in\mathbb{R}^n$ 使得 $T'(\mathbf{v})\neq T(\mathbf{v})$. 此時因存在 $c_1,\dots,c_n$, 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 故依 T,T' 皆為 linear 的假設, 我們有

$$T'(\mathbf{v})=T'(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1T'(\mathbf{v}_1)+\cdots+c_nT'(\mathbf{v}_n)=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n=T(\mathbf{v}).$$

此與 $T'(\mathbf{v})\neq T(\mathbf{v})$ 相矛盾, 證得唯一性. □

要注意 Theorem 5.1.8 中的 $\mathbf{w}_1,\dots,\mathbf{w}_n\in\mathbb{R}^m$ 是可以任意選取的, 不需要是一組 basis 或是 linear independent. 這個定理, 再次讓我們確定 basis 的重要性. 它告訴我們給定 $\mathbb{R}^n$ 的一組 basis 後, 我們可以將這組 basis 裡的向量對應到 $\mathbb{R}^m$ 中任意的向量, 就會得到一個 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation. 更重要的是, 一般來講兩個函數要說明它們是相等的, 我們必須檢查定義域裡的每個元素是否被這兩個函數對應到對應域裡相同的元素. 這個過程是很複雜的, 因為一般來說定義域裡的元素有無窮多個, 我們無法一個一個檢查. 但是 linear transformation 就有這個好處, Theorem 5.1.8 告訴我們僅要檢查兩個 linear transformations

在一組 basis 裡中的那些有限多個向量是一致的, 那麼這兩個 linear transformation 事實上就會是相同的函數.

5.2. Kernel and Range

Linear transformation 由於有保持 linear combination 的特點, 所以它會保持定義域與對應域中的 subspaces. 在這一節中我們便是要探討一個 linear transformation 所得到的兩個重要的 subspaces, "kernel" 和 "range", 並利用這兩個 subspace 來探討 linear transformation 本身的特點.

首先對於一般的函數 $T:\mathbb{R}^n \to \mathbb{R}^m$, 任取定義域 $\mathbb{R}^n$ 中的子集合 S, 我們很自然的會考慮 $T(S) = \{T(\mathbf{v}) \in \mathbb{R}^m \mid \mathbf{v} \in S\}$ 這一個集合, 它就是將所有 S 中的元素利用 T 映射到 $\mathbb{R}^m$ 後的元素收集起來所得的集合. 同樣的, 對於對應域 $\mathbb{R}^m$ 中的子集合 S', 我們也會考慮 $T^{-1}(S') = \{\mathbf{v} \in \mathbb{R}^n \mid T(\mathbf{v}) \in S'\}$ 這樣的集合, 它就是收集所有在定義域中會映射到 S' 的元素所成的集合. 很容易知道 $T(S)$ 會是對應域 $\mathbb{R}^m$ 中的子集合, 而 $T^{-1}(S')$ 會是定義域 $\mathbb{R}^n$ 中的子集合. 注意當 $\mathbf{w} \in T(S)$ 時, 依定義這表示 $\mathbf{w}$ 是 S 中某個元素經 T 映射所得, 亦即存在 $\mathbf{v} \in V$ 使得 $\mathbf{w} = T(\mathbf{v})$. 所以 $T(S)$ 也可表示成 $T(S) = \{\mathbf{w} \in \mathbb{R}^m \mid \mathbf{w} = T(\mathbf{v}), \text{ for some } \mathbf{v} \in S\}$, 有時為了強調 $T(S)$ 為 $\mathbb{R}^m$ 的子集合, 我們也會用這種表示法.

由於 linear transformation 的特色, 當 $T:\mathbb{R}^n \to \mathbb{R}^m$ 為 linear transformation 時, 我們專注於考慮 V 為 $\mathbb{R}^n$ 的 subspace 時的情況. 也就是說我們要了解

$$T(V) = \{T(\mathbf{v}) \in \mathbb{R}^m \mid \mathbf{v} \in V\} = \{\mathbf{w} \in \mathbb{R}^m \mid \mathbf{w} = T(\mathbf{v}), \text{ for some } \mathbf{v} \in V\}$$

的特性. 同樣的若 W 為 $\mathbb{R}^m$ 的 subspace, 我們也要了解

$$T^{-1}(W) = \{\mathbf{v} \in \mathbb{R}^n \mid T(\mathbf{v}) \in W\}$$

的特性. 事實上, 我們有以下之結果.

Proposition 5.2.1. 假設 $T:\mathbb{R}^n \to \mathbb{R}^m$ 為 *linear transformation.* 若 V 為 $\mathbb{R}^n$ 的 *subspaces,* 則 $T(V)$ 是 $\mathbb{R}^m$ 的 *subspace.* 另外*,* 若 W 為 $\mathbb{R}^m$ 的 *subspaces,* 則 $T^{-1}(W)$ 是 $\mathbb{R}^n$ 的 *subspace.*

Proof. 依定義我們知 $T(V)$ 會是 $\mathbb{R}^m$ 的子集合, 而 $T^{-1}(W)$ 會是 $\mathbb{R}^n$ 的子集合. 故現僅需證明它們為 subspaces, 即利用 Proposition 4.1.2 我們要證明, 若 $\mathbf{w}, \mathbf{w}' \in T(V)$ 且 $r \in \mathbb{R}$, 則 $\mathbf{w} + r\mathbf{w}' \in T(V)$ 以及若 $\mathbf{v}, \mathbf{v}' \in T^{-1}(W)$ 且 $r \in \mathbb{R}$, 則 $\mathbf{v} + r\mathbf{v}' \in T^{-1}(W)$.

首先再強調一次, 當我們說 $\mathbb{R}^m$ 中的一個向量 $\mathbf{w}$ 在 $T(V)$ 時, 表示存在 $\mathbf{v} \in V$ 使得 $\mathbf{w} = T(\mathbf{v})$. 因此若 $\mathbf{w}, \mathbf{w}' \in T(V)$, 則存在 $\mathbf{v}, \mathbf{v}' \in V$ 使得 $T(\mathbf{v}) = \mathbf{w}, T(\mathbf{v}') = \mathbf{w}'$. 此時對於 $r \in \mathbb{R}$, 我們有 $\mathbf{w} + r\mathbf{w}' = T(\mathbf{v}) + rT(\mathbf{v}')$. 再利用 T 為 linear, 得 $\mathbf{w} + r\mathbf{w}' = T(\mathbf{v} + r\mathbf{v}')$. 然而依假設 V 為 $\mathbb{R}^n$ 的 subspace, 故由 $\mathbf{v}, \mathbf{v}' \in V$ 知 $\mathbf{v} + r\mathbf{v}' \in V$, 得證 $\mathbf{w} + r\mathbf{w}' = T(\mathbf{v} + r\mathbf{v}') \in T(V)$.

另一方面, 若 $\mathbf{v}, \mathbf{v}' \in T^{-1}(W)$, 表示 $T(\mathbf{v}), T(\mathbf{v}') \in W$. 此時對於 $r \in \mathbb{R}$, 由於 T 為 linear, 我們有 $T(\mathbf{v} + r\mathbf{v}') = T(\mathbf{v}) + rT(\mathbf{v}')$. 然而依假設 W 為 $\mathbb{R}^m$ 的 subspace, 故由 $T(\mathbf{v}), T(\mathbf{v}') \in W$ 知 $T(\mathbf{v}) + rT(\mathbf{v}') \in W$. 因此由 $T(\mathbf{v} + r\mathbf{v}') = T(\mathbf{v}) + rT(\mathbf{v}') \in W$, 得證 $\mathbf{v} + r\mathbf{v}' \in T^{-1}(W)$. □

特別的, 在 $V=\mathbb{R}^n$ 和 $W=\{\mathbf{O}\}$ 這兩個特殊情況時, 即

$$T(\mathbb{R}^n)=\{\mathbf{w}\in\mathbb{R}^m \mid \mathbf{w}=T(\mathbf{v}) \text{ for some } \mathbf{v}\in\mathbb{R}^n\} \quad \text{and} \quad T^{-1}(\{\mathbf{O}\})=\{\mathbf{v}\in\mathbb{R}^n \mid T(\mathbf{v})=\mathbf{O}\}$$

這兩個 subspaces, 對我們了解 T 這個 linear transformation 非常有幫助. 我們先給這兩個 subspace 特殊的名稱.

Definition 5.2.2. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation. 我們稱 $\mathbb{R}^m$ 的 subspace $T(\mathbb{R}^n)$ 為 T 的 *range* (有時也稱為 *image*). 我們稱 $\mathbb{R}^n$ 的 subspace $T^{-1}(\{\mathbf{O}\})$ 為 T 的 *kernel*, 通常用 $\ker(T)$ 來表示.

首先我們來看 linear transformation 的 range. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation. 由 Proposition 5.2.1 我們知 T 的 range $T(\mathbb{R}^n)$ 是 $\mathbb{R}^m$ 的 subspace, 故知 $\dim(T(\mathbb{R}^n))\leq\dim(\mathbb{R}^m)=m$. 若 $\dim(T(\mathbb{R}^n))=m$, 則依 Corollary 4.3.6 知此時 $T(\mathbb{R}^n)=\mathbb{R}^m$. 也就是說對於任意的 $\mathbf{w}\in\mathbb{R}^m$, 由於 $\mathbf{w}\in T(\mathbb{R}^n)$ 依定義存在 $\mathbf{v}\in\mathbb{R}^n$ 使得 $\mathbf{w}=T(\mathbf{v})$. 也就是說此時 T 為 onto. 另一方面, 若 T 為 onto, 則依定義對於任意 $\mathbf{w}\in\mathbb{R}^m$, 存在 $\mathbf{v}\in\mathbb{R}^n$ 使得 $T(\mathbf{v})=\mathbf{w}$ 故得 $\mathbf{w}\in T(\mathbb{R}^n)$, 得證 $\mathbb{R}^m\subseteq T(\mathbb{R}^n)$. 再利用已知 $T(\mathbb{R}^n)\subseteq\mathbb{R}^m$, 得證 $T(\mathbb{R}^n)=\mathbb{R}^m$. 我們有以下的性質.

Proposition 5.2.3. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation.* 則 T 為 *onto* 若且唯若 $\dim(T(\mathbb{R}^n))=m$.

一般來說, 我們要判斷一個函數是否為 onto 便是要確認其 range 是否就是 codomain (對應域). 對於一般的函數要確認是否為 onto 有時並不容易. 不過 Proposition 5.2.3 告訴我們對於 linear transformation, 可以直接由它的 range 的 dimension 來判斷是否為 onto. 至於如何知道一個 linear transformation 的 range 呢? 我們有以下的性質.

Proposition 5.2.4. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation* 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbb{R}^n$ 為 $\mathbb{R}^n$ 的一組 *spanning vectors.* 則

$$T(\mathbb{R}^n)=\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)).$$

Proof. 設 $\mathbf{w}\in T(\mathbb{R}^n)$, 表示存在 $\mathbf{v}\in\mathbb{R}^n$ 使得 $\mathbf{w}=T(\mathbf{v})$. 又因 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 $\mathbb{R}^n$ 的一組 spanning vectors, 故存在 $c_1,\cdots,c_n\in\mathbb{R}$, 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 因此利用 T 為 linear 得

$$\mathbf{w}=T(\mathbf{v})=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)\in\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)),$$

得證 $T(\mathbb{R}^n)\subseteq\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n))$.

另一方面, 設 $\mathbf{w}\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 表示存在 $c_1,\cdots,c_n\in\mathbb{R}$, 使得 $\mathbf{w}=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)$. 因此利用 T 為 linear 得

$$\mathbf{w}=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)\in T(\mathbb{R}^n),$$

得證 $\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n))\subseteq T(\mathbb{R}^n)$. 因此證明了 $T(\mathbb{R}^n)=\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n))$. $\square$

Example 5.2.5. (1) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix})=\begin{bmatrix}x_1+x_2\\x_1-x_3\end{bmatrix}$. 在 Example 5.1.4 中我們已知 T 是一個 linear transformation. 考慮定義域 $\mathbb{R}^3$ 的 standard basis $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$, 我們得 $T(\begin{bmatrix}1\\0\\0\end{bmatrix})=\begin{bmatrix}1\\1\end{bmatrix}$, $T(\begin{bmatrix}0\\1\\0\end{bmatrix})=\begin{bmatrix}1\\0\end{bmatrix}$, $T(\begin{bmatrix}0\\0\\1\end{bmatrix})=\begin{bmatrix}0\\-1\end{bmatrix}$. 由於 $\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\-1\end{bmatrix}$ 為 $\mathbb{R}^2$ 的一組 spanning vectors, 由 Proposition 5.2.4 我們有 $T(\mathbb{R}^3)=\mathrm{Span}(\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\-1\end{bmatrix})=\mathbb{R}^2$. 故得 T 為 onto.

(2) 考慮 $T:\mathbb{R}^2\to\mathbb{R}^3$ 定義為 $T(\begin{bmatrix}x_1\\x_2\end{bmatrix})=\begin{bmatrix}x_1\\x_1+x_2\\x_1-x_2\end{bmatrix}$. 很容易驗證 T 是一個 linear transformation. 考慮定義域 $\mathbb{R}^2$ 的 standard basis $\{\mathbf{e}_1,\mathbf{e}_2\}$, 我們得 $T(\begin{bmatrix}1\\0\end{bmatrix})=\begin{bmatrix}1\\1\\1\end{bmatrix}$, $T(\begin{bmatrix}0\\1\end{bmatrix})=\begin{bmatrix}0\\1\\-1\end{bmatrix}$. 由 Proposition 5.2.4 我們有 $T(\mathbb{R}^2)=\mathrm{Span}(\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\-1\end{bmatrix})$. 由於 $\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\-1\end{bmatrix}$ 為 linearly independent, 故得 $\dim(T(\mathbb{R}^2))=2$. 由 Proposition 5.2.3 知 T 不是 onto.

Question 5.4. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation.* 若 $m>n$, 則 T 有可能是 *onto* 嗎?

要注意是有可能一個 linear transformation T 的 range 為 $\{\mathbf{O}\}$. 此表示 T 將所有定義域的向量皆映射到 $\mathbf{O}$, 亦即 $T(\mathbf{v})=\mathbf{O}$, $\forall\mathbf{v}\in\mathbb{R}^n$. 這樣的 linear transformation, 我們依慣例, 仍稱之為 *zero mapping.*

Question 5.5. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation* 且 $\mathbf{v}_1,\dots,\mathbf{v}_n\in\mathbb{R}^n$ 為 $\mathbb{R}^n$ 的一組 *basis.* 試證明 T 為 *zero mapping* 若且唯若 $T(\mathbf{v}_1)=\cdots=T(\mathbf{v}_n)=\mathbf{O}$.

接下來我們看 kernel 與 linear transformation 的關係. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation. 若 T 為 one-to-one, 由於已知 $T(\mathbf{O})=\mathbf{O}$, 故知不可能有非零的向量 $\mathbf{v}$ 使得 $T(\mathbf{v})=\mathbf{O}$. 因此可得 $\ker(T)=\{\mathbf{O}\}$. 其實反過來 $\ker(T)=\{\mathbf{O}\}$, 也會使得 T 為 one-to-one, 我們有以下的結果.

Proposition 5.2.6. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation.* 則 T 為 *one-to-one* 若且唯若 $\dim(\ker(T))=0$, 亦即 $\ker(T)=\{\mathbf{O}\}$.

Proof. 我們已知當 T 為 one-to-one 時, 不會有非零向量映射到 $\mathbf{O}$, 故知 $\ker(T)=\{\mathbf{O}\}$, 得 $\dim(\ker(T))=0$. 反之, 當 $\dim(\ker(T))=0$, 即 $\ker(T)=\{\mathbf{O}\}$, 此時若 T 不是 one-to-one, 表示存在 $\mathbf{v},\mathbf{v}'\in\mathbb{R}^n$ 滿足 $\mathbf{v}\neq\mathbf{v}'$ 但 $T(\mathbf{v})=T(\mathbf{v}')$. 由於 T 為 linear, 得 $T(\mathbf{v}-\mathbf{v}')=T(\mathbf{v})-T(\mathbf{v}')=\mathbf{O}$, 亦即 $\mathbf{v}-\mathbf{v}'\in T^{-1}(\{\mathbf{O}\})=\ker(T)=\{\mathbf{O}\}$. 得到 $\mathbf{v}=\mathbf{v}'$ 之矛盾, 故證得 T 為 one-to-one. $\square$

Example 5.2.7. 我們探討 Example 5.2.5 中的 linear transformation 是否為 one-to-one.

(1) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix})=\begin{bmatrix}x_1+x_2\\x_1-x_3\end{bmatrix}$. 若 $\mathbf{v}=\begin{bmatrix}a\\b\\c\end{bmatrix}\in\ker(T)$, 表示 $T(\mathbf{v})=T(\begin{bmatrix}a\\b\\c\end{bmatrix})=\begin{bmatrix}a+b\\a-c\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}$, 故得 $a+b=0$ 以及 $a-c=0$, 因此可得 $\begin{bmatrix}1\\-1\\1\end{bmatrix}\in\ker(T)$, 知 $\ker(T)\neq\{\mathbf{O}\}$ (事實上 $\ker(T)=\mathrm{Span}(\begin{bmatrix}1\\-1\\1\end{bmatrix})$). 所以知 T 不是 one-to-one.

(2) 考慮 $T:\mathbb{R}^2\to\mathbb{R}^3$ 定義為 $T(\begin{bmatrix}x_1\\x_2\end{bmatrix})=\begin{bmatrix}x_1\\x_1+x_2\\x_1-x_2\end{bmatrix}$. 若 $\mathbf{v}=\begin{bmatrix}a\\b\end{bmatrix}\in\ker(T)$, 表示 $T(\mathbf{v})=T(\begin{bmatrix}a\\b\end{bmatrix})=\begin{bmatrix}a\\a+b\\a-b\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$. 故得 $a=0$, $a+b=0$ 以及 $a-b=0$, 即 $a=b=0$. 因此可得 $\ker(T)=\{\mathbf{O}\}$, 所以由 Proposition 5.2.6 知 T 是 one-to-one.

要注意 Proposition 5.2.6 需 T 為 linear transformation 才適用. 例如 $f(x)=x^2$ 的情形, 雖然 $f^{-1}(0)=\{0\}$ (因為只有當 $x=0$ 才會使得 $x^2=0$) 但 $f(x)$ 不是一對一 (例如 $f(1)=f(-1)=1$). 事實上我們知道 $f(x)$ 不是 linear. 所以當 f 不是 linear transformation 時, 不能由 $f^{-1}(\{0\})$ 來判斷是否為 one-to-one.

Question 5.6. *Proposition 5.2.6* 中是哪一個部分需用到 T 為 *linear* 的假設? 是由 T 為 *one-to-one* 推得 $\ker(T)=\{\mathbf{O}\}$ 還是由 $\ker(T)=\{\mathbf{O}\}$ 推得 T 為 *one-to-one?*

Question 5.7. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation.* 在 *Question 5.5* 中我們知道 T 為 *zero mapping* 和 T 的 *range* 的等價關係. 你知道 T 為 *zero mapping* 和 T 的 *kernel* 的等價關係嗎?

對於一般的函數, 要判斷其是否為 one-to-one 並不容易, 但當 T 為 linear transformation 時, Proposition 5.2.6 提供我們一個簡便的方法判斷 T 是否為 one-to-one. 也就是僅要確認是否有非零向量會映射到 $\mathbf{O}$ 即可. 所以求一個 linear transformation 的 kernel 是一個重要的課題. 在下一節中, 我們將介紹求得 kernel 的方法.

5.3. Matrix Representation

給定一個 $m\times n$ matrix A, 前面我們已知可以定義出一個 linear transformation $T:\mathbb{R}^n\to\mathbb{R}^m$, 其定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$. 在這一節中, 我們要說明所有的 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformations 都可以寫成這樣的形式. 並利用此將 linear transformation 和 matrix 相連結, 來推得一些重要的性質.

前面 Theorem 5.1.8 告訴我們給定一個 linear transformation, 只要知道此 linear transformation 將一組 $\mathbb{R}^n$ 的 basis 對應到哪些向量, 就可以唯一確定這一個 linear transformation. 在 $\mathbb{R}^n$ 中, 我們有一個最簡單的 basis, 即 standard basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$. 若 $T:\mathbb{R}^n\to\mathbb{R}^m$

是一個 linear transformation, 由前面所述, 我們僅要知道 $T(\mathbf{e}_1),\ldots,T(\mathbf{e}_n)$ 是哪些 $\mathbb{R}^m$ 的 vectors, 就可以知道任意 $\mathbf{v}\in\mathbb{R}^n$, $T(\mathbf{v})$ 為何了.

事實上對於每一個 $\mathbf{v}\in\mathbb{R}^n$, 都可以找到 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $\mathbf{v}=c_1\mathbf{e}_1+\cdots+c_n\mathbf{e}_n$, 也就是說此時 $\mathbf{v}$ 的坐標表示法為 $\mathbf{v}=\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}$. 因此由 T 為 linear, 得 $T(\mathbf{v})=T(c_1\mathbf{e}_1+\cdots+c_n\mathbf{e}_n)=c_1T(\mathbf{e}_1)+\cdots+c_nT(\mathbf{e}_n)$. 現若考慮 $m\times n$ matrix A, 其中 A 的 i-th column 為 $T(\mathbf{e}_i)$, 則

$$A\mathbf{v}=\begin{bmatrix}| & | & & |\\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n)\\ | & | & & |\end{bmatrix}\begin{bmatrix}c_1\\c_2\\\vdots\\c_n\end{bmatrix}=c_1T(\mathbf{e}_1)+\cdots+c_nT(\mathbf{e}_n)=T(\mathbf{v}).$$

也就是說對任意 $\mathbf{v}\in\mathbb{R}^n$, 皆有 $T(\mathbf{v})=A\mathbf{v}$. 因此 T 就等同於將 $\mathbf{v}$ 左邊乘上 A 這一個矩陣這樣的 linear transformation. 我們有以下這一個重要的定理.

Theorem 5.3.1. 給定一個 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 *function* T. 則 T 為 *linear transformation* 若且唯若存在一個 $m\times n$ *matrix* A 使得 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$. 此 $m\times n$ *matrix* A 是唯一的, 事實上對任意 $i=1,\ldots,n$, A 的 *i-th column* 為 $T(\mathbf{e}_i)$, 其中 $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ 為 $\mathbb{R}^n$ 的 *standard basis.*

Proof. 由 Lemma 5.1.5 我們知道, 若 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$, 則 T 為 linear transformation. 反之, 若 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation, 如前面所討論的, 我們可以考慮 A 為 i-th column 為 $T(\mathbf{e}_i)$ 的 $m\times n$ matrix, 則由矩陣乘法性質知 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$.

現若 B 為 $m\times n$ matrix 亦滿足 $T(\mathbf{v})=B\mathbf{v}$, 依矩陣乘法定義知對任意 $i=1,\ldots,n$, $B\mathbf{e}_i$ 為 B 的 i-th column. 但由假設 $B\mathbf{e}_i=T(\mathbf{e}_i)$, 亦即 B 的 i-th column 為 $T(\mathbf{e}_i)$. 因此 B 的所有 column 皆與前述 A 的 column 相一致, 證得唯一性. □

簡單來說 Theorem 5.3.1 告訴我們從 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformations 和 $m\times n$ matrices 之間有一個一對一的對應關係 (注意矩陣階數與定義域, 對應域之間的關係). 由於一個 linear transformation 和其對應的 $m\times n$ matrix 關係特別密切, 我們有以下的定義.

Definition 5.3.2. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 linear transformation 且 $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ 為 $\mathbb{R}^n$ 的 standard basis. 則對於 $i=1,\ldots,n$, 其 i-th column 為 $T(\mathbf{e}_i)$ 的 $m\times n$ matrix 稱為 T 的 *standard matrix representation.*

由於 T 的 standard matrix representation 是唯一的且和 T 有關, 以後我們都用 $[T]$ 來表示 T 的 standard matrix representation. 也就是說對任意 $\mathbf{v}\in\mathbb{R}^n$, 我們有 $T(\mathbf{v})=[T]\mathbf{v}$.

Example 5.3.3. 我們探討 Example 5.2.5 中的 linear transformation 其 standard matrix representation.

(1) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix} x_1\\ x_2\\ x_3\end{bmatrix})=\begin{bmatrix} x_1+x_2\\ x_1-x_3\end{bmatrix}$. 由於

$$T(\mathbf{e}_1)=T(\begin{bmatrix}1\\0\\0\end{bmatrix})=\begin{bmatrix}1\\1\end{bmatrix},T(\mathbf{e}_2)=T(\begin{bmatrix}0\\1\\0\end{bmatrix})=\begin{bmatrix}1\\0\end{bmatrix},T(\mathbf{e}_3)=T(\begin{bmatrix}0\\0\\1\end{bmatrix})=\begin{bmatrix}0\\-1\end{bmatrix},$$

故得

$$[T]=\begin{bmatrix} | & | & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & T(\mathbf{e}_3) \\ | & | & | \end{bmatrix}=\begin{bmatrix} 1 & 1 & 0\\ 1 & 0 & -1\end{bmatrix}.$$

事實上我們有

$$[T]\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix} 1 & 1 & 0\\ 1 & 0 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix} x_1+x_2\\ x_1-x_3\end{bmatrix}=T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}).$$

(2) 考慮 $T:\mathbb{R}^2\to\mathbb{R}^3$ 定義為 $T(\begin{bmatrix} x_1\\ x_2\end{bmatrix})=\begin{bmatrix} x_1\\ x_1+x_2\\ x_1-x_2\end{bmatrix}$. 由於

$$T(\mathbf{e}_1)=T(\begin{bmatrix}1\\0\end{bmatrix})=\begin{bmatrix}1\\1\\1\end{bmatrix},T(\mathbf{e}_2)=T(\begin{bmatrix}0\\1\end{bmatrix})=\begin{bmatrix}0\\1\\-1\end{bmatrix},$$

故得

$$[T]=\begin{bmatrix} | & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) \\ | & | \end{bmatrix}=\begin{bmatrix} 1 & 0\\ 1 & 1\\ 1 & -1\end{bmatrix}.$$

事實上我們有

$$[T]\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix} 1 & 0\\ 1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix} x_1\\ x_1+x_2\\ x_1-x_2\end{bmatrix}=T(\begin{bmatrix}x_1\\x_2\end{bmatrix}).$$

有了 standard matrix representation, 我們就可以利用以下的定理幫助我們找出一個 linear transformation 的 range 和 kernel.

Proposition 5.3.4. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation* 且令 $[T]\in\mathscr{M}_{m\times n}$ 為其 *standard matrix representation.* 則 T 的 *range* 等於 $[T]$ 的 *column space,* 而 T 的 *kernel* 等於 $[T]$ 的 *nullspace.*

Proof. 由於 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 為 $\mathbb{R}^n$ 的一組 basis, 由 Proposition 5.2.4 我們知 T 的 range, 即 $T(\mathbb{R}^n)=\mathrm{Span}(T(\mathbf{e}_1),\ldots,T(\mathbf{e}_n))$. 然而 $T(\mathbf{e}_1),\ldots,T(\mathbf{e}_n)$ 剛好就是 $[T]$ 的 n 個 column, 故由定義 $\mathrm{Span}(T(\mathbf{e}_1),\ldots,T(\mathbf{e}_n))$ 就是 $[T]$ 的 column space. 得證 T 的 range 就是 $[T]$ 的 column space.

另一方面, 若 $\mathbf{v}\in\ker(T)$, 表示 $T(\mathbf{v})=\mathbf{O}$. 然而依 standard matrix representation 之定義 $T(\mathbf{v})=[T]\mathbf{v}$, 故得 $[T]\mathbf{v}=\mathbf{O}$, 亦即 $\mathbf{v}$ 屬於 $[T]$ 的 nullspace. 得證 $\ker(T)$ 包含於 $[T]$ 的 nullspace. 反之, 若 $\mathbf{v}$ 屬於 $[T]$ 的 nullspace, 表示 $[T]\mathbf{v}=\mathbf{O}$, 亦即 $T(\mathbf{v})=\mathbf{O}$, 得證 $\mathbf{v}\in\ker(T)$. 證明了 $[T]$ 的 nullspace 包含於 $\ker(T)$, 因此 T 的 kernel 等於 $[T]$ 的 nullspace. □

回顧一個矩陣 A 的 column space 的維度, 我們稱為 A 的 rank, 用 $\mathrm{rank}(A)$ 來表示. 而 A 的 nullspace 的維度稱為 A 的 nullity, 用 $\mathrm{null}(A)$ 來表示 (參見 Definition 2.3.1). 由 Proposition 5.3.4, 我們知道 T 的 range 的維度等於 $\mathrm{rank}([T])$, 而 T 的 kernel 的維度等於 $\mathrm{null}([T])$, 也就是說我們有以下的結果.

Corollary 5.3.5. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation* 且令 $[T]\in\mathscr{M}_{m\times n}$ 為其 *standard matrix representation.* 則

$$\dim(T(\mathbb{R}^n))=\mathrm{rank}([T])\quad\text{and}\quad\dim(\ker(T))=\mathrm{null}([T]).$$

因為這個原因一般我們也稱 T 的 range 的維度為 T 的 *rank*, 而 T 的 kernel 的維度稱為 T 的 *nullity*. 這更進一步的強調了 linear transformation 以及 matrix 之間的關係, 因為這一層關係我們也很容易推得 linear transformation 的 Dimension Theorem.

Theorem 5.3.6 (Dimension Theorem for Linear Transformation)**.** 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation.* 則

$$\dim(T(\mathbb{R}^n))+\dim(\ker(T))=n.$$

Proof. 由於 T 的 standard matrix representation $[T]$ 為 $m\times n$ matrix, 由 Theorem 4.4.13 我們知 $\mathrm{rank}([T])+\mathrm{null}([T])=n$. 故套用 Corollary 5.3.5, 得證本定理. □

Example 5.3.7. 我們利用 Example 5.3.3 中的 linear transformation 及其 standard matrix representation 探討其 range 和 kernel.

(1) 考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix})=\begin{bmatrix}x_1+x_2\\x_1-x_3\end{bmatrix}$, 以及其 standard matrix representation $[T]=\begin{bmatrix}1&1&0\\1&0&-1\end{bmatrix}$. 由於 $[T]$ 的 column space 為 $\mathrm{Span}(\begin{bmatrix}1\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\-1\end{bmatrix})=\mathbb{R}^2$. 因此由 Proposition 5.3.4 我們有 $T(\mathbb{R}^3)=\mathbb{R}^2$ (此與 Example 5.2.5(1) 一致). 另一方面, $[T]$ 的 nullspace 為聯立方程組 $[T]\mathbf{x}=\mathbf{O}$, 即

$$\left\{\begin{array}{rrrcl}x_1&+x_2&&=&0\\x_1&&-x_3&=&0\end{array}\right.\quad\sim\quad\left\{\begin{array}{rrrcl}x_1&+x_2&&=&0\\&-x_2&-x_3&=&0\end{array}\right.$$

的解集合. 因此由 Proposition 5.3.4 我們有 $\ker(T)=\mathrm{Span}(\begin{bmatrix}1\\-1\\1\end{bmatrix})$ (此與 Example 5.2.7(1) 一致). 檢查 Dimension Theorem, 我們確實有 $\dim(T(\mathbb{R}^3))+\dim(\ker(T))=2+1=3$.

(2) 考慮 $T:\mathbb{R}^2\to\mathbb{R}^3$ 定義為 $T(\begin{bmatrix}x_1\\x_2\end{bmatrix})=\begin{bmatrix}x_1\\x_1+x_2\\x_1-x_2\end{bmatrix}$, 以及其 standard matrix representation $[T]=\begin{bmatrix}1&0\\1&1\\1&-1\end{bmatrix}$. 由於 $[T]$ 的 column space 為 $\mathrm{Span}(\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\-1\end{bmatrix})$. 因此由 Proposition 5.3.4 我們有 $T(\mathbb{R}^3)=\mathrm{Span}(\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\-1\end{bmatrix})$ (此與 Example 5.2.5(2) 一致). 另一方面, $[T]$ 的

nullspace 為聯立方程組 $[T]\mathbf{x}=\mathbf{O}$, 即

$$\begin{cases} x_1 & & = & 0 \\ x_1 & +x_2 & = & 0 \\ x_1 & -x_2 & = & 0 \end{cases} \quad \sim \quad \begin{cases} x_1 & & = & 0 \\ & x_2 & = & 0 \end{cases}$$

的解集合. 因此由 Proposition 5.3.4 我們有 $\ker(T)=\{\begin{bmatrix}0\\0\end{bmatrix}\}=\{\mathbf{O}\}$ (此與 Example 5.2.7(2) 一致). 檢查 Dimension Theorem, 我們確實有 $\dim(T(\mathbb{R}^2))+\dim(\ker(T))=2+0=2$.

當 T_1,T_2 皆為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation 時, 對任意 $c_1,c_2\in\mathbb{R}$, 我們可以利用它們得到一個新的 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation $c_1T_1+c_2T_2$ (參見 Proposition 5.1.6). 我們自然會想知道 $c_1T_1+c_2T_2$ 的 standard matrix representation 和 T_1,T_2 的 standard matrix representation 是否有關. 另外, 若 T 為 $\mathbb{R}^m$ 到 $\mathbb{R}^k$ 的 linear transformation, 我們可得合成函數 $T\circ T_1$ 為 $\mathbb{R}^n$ 到 $\mathbb{R}^k$ 的 linear transformation (參見 Proposition 5.1.7). 同樣的, 我們要探討 $T\circ T_1$ 的 standard matrix representation 和 T_1,T 的 standard matrix representation 是否有關.

Lemma 5.3.8. 設 T_1,T_2 為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 *linear transformations,* 而 T 為 $\mathbb{R}^m$ 到 $\mathbb{R}^k$ 的 *linear transformation.* 令 $[T_1],[T_2]$ 以及 $[T]$ 分別為 T_1,T_2 和 T 的 *standard matrix representation.*

(1) 對任意 $c_1,c_2\in\mathbb{R}$, 皆有 $c_1T_1+c_2T_2:\mathbb{R}^n\to\mathbb{R}^m$ 的 *standard matrix representation* 為 $c_1[T_1]+c_2[T_2]$, 亦即

$$[c_1T_1+c_2T_2]=c_1[T_1]+c_2[T_2].$$

(2) $T\circ T_1:\mathbb{R}^n\to\mathbb{R}^k$ 的 *standard matrix representation* 為 $[T][T_1]$, 亦即

$$[T\circ T_1]=[T][T_1].$$

Proof. (1) 依定義對任意 $\mathbf{v}\in\mathbb{R}^n$, 我們有 $(c_1T_1+c_2T_2)(\mathbf{v})=c_1T_1(\mathbf{v})+c_2T_2(\mathbf{v})$. 又依 standard matrix representation 的定義 $T_1(\mathbf{v})=[T_1]\mathbf{v},T_2(\mathbf{v})=[T_2]\mathbf{v}$, 故依矩陣乘法的分配律得

$$(c_1T_1+c_2T_2)(\mathbf{v})=c_1[T_1]\mathbf{v}+c_2[T_2]\mathbf{v}=(c_1[T_1]+c_2[T_2])\mathbf{v}.$$

換言之, $c_1[T_1]+c_2[T_2]$ 是一個 $m\times n$ matrix 且滿足 $c_1T_1+c_2T_2:\mathbb{R}^n\to\mathbb{R}^m$ 的 standard matrix representation 之要求, 故由 standard matrix representation 的唯一性 (Theorem 5.3.1) 知 $[c_1T_1+c_2T_2]=c_1[T_1]+c_2[T_2]$.

(2) 依定義對任意 $\mathbf{v}\in\mathbb{R}^n$, 我們有 $(T\circ T_1)(\mathbf{v})=T(T_1(\mathbf{v}))$. 又依 standard matrix representation 的定義 $T_1(\mathbf{v})=[T_1]\mathbf{v}$, 故得 $(T\circ T_1)(\mathbf{v})=T([T_1]\mathbf{v})$. 又依定義, 對任意 $\mathbf{w}\in\mathbb{R}^m$ 皆有 $T(\mathbf{w})=[T]\mathbf{w}$, 故得 $(T\circ T_1)(\mathbf{v})=T([T_1]\mathbf{v})=[T]([T_1]\mathbf{v})$. 再依矩陣乘法的結合律得 $[T]([T_1]\mathbf{v})=([T][T_1])\mathbf{v}$. 換言之, $[T][T_1]$ 是一個 $k\times n$ matrix 且滿足 $T\circ T_1:\mathbb{R}^n\to\mathbb{R}^k$ 的 standard matrix representation 之要求 $(T\circ T_1)(\mathbf{v})=([T][T_1])\mathbf{v}$, 故由 standard matrix representation 的唯一性 (Theorem 5.3.1) 知 $[T\circ T_1]=[T][T_1]$. □

Example 5.3.9. 我們利用 Example 5.3.3 中的 linear transformations 及其 standard matrix representations 探討它們的合成函數.

考慮 $T:\mathbb{R}^3\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix} x_1\\ x_2\\ x_3\end{bmatrix})=\begin{bmatrix} x_1+x_2\\ x_1-x_3\end{bmatrix}$. 我們知 T 的 standard matrix representation 為 $[T]=\begin{bmatrix} 1 & 1 & 0\\ 1 & 0 & -1\end{bmatrix}$. 另外考慮 $T':\mathbb{R}^2\to\mathbb{R}^3$ 定義為 $T'(\begin{bmatrix} x_1\\ x_2\end{bmatrix})=\begin{bmatrix} x_1\\ x_1+x_2\\ x_1-x_2\end{bmatrix}$. 我們知 T' 的 standard matrix representation 為 $[T']=\begin{bmatrix} 1 & 0\\ 1 & 1\\ 1 & -1\end{bmatrix}$. 依合成函數定義 $T'\circ T:\mathbb{R}^3\to\mathbb{R}^3$ 滿足

$$(T'\circ T)(\begin{bmatrix} x_1\\ x_2\\ x_3\end{bmatrix})=T'(\begin{bmatrix} x_1+x_2\\ x_1-x_3\end{bmatrix})=\begin{bmatrix} x_1+x_2\\ (x_1+x_2)+(x_1-x_3)\\ (x_1+x_2)-(x_1-x_3)\end{bmatrix}=\begin{bmatrix} x_1+x_2\\ 2x_1+x_2-x_3\\ x_2+x_3\end{bmatrix}.$$

依此結果, 我們得 $T'\circ T$ 的 standard matrix representation 為 $[T'\circ T]=\begin{bmatrix} 1 & 1 & 0\\ 2 & 1 & -1\\ 0 & 1 & 1\end{bmatrix}$. 另一方面, 考慮矩陣乘法, 我們有 $[T'][T]=\begin{bmatrix} 1 & 0\\ 1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix} 1 & 1 & 0\\ 1 & 0 & -1\end{bmatrix}=\begin{bmatrix} 1 & 1 & 0\\ 2 & 1 & -1\\ 0 & 1 & 1\end{bmatrix}$. 的確得到 $[T'\circ T]=[T'][T]$.

我們利用 matrix 來幫助我們了解 linear transformation. 反過來, 我們也可以利用 linear transformation 來幫助我們了解 matrix 的性質. 例如當 $T:\mathbb{R}^n\to\mathbb{R}^m$, $T':\mathbb{R}^m\to\mathbb{R}^k$ 為 linear transformations. 由於 T 的 range 是 $\mathbb{R}^m$ 的 subspace, 即 $T(\mathbb{R}^n)\subseteq\mathbb{R}^m$, 我們有 $(T'\circ T)(\mathbb{R}^n)=T'(T(\mathbb{R}^n))\subseteq T'(\mathbb{R}^m)$. 也就是說 $T'\circ T:\mathbb{R}^n\to\mathbb{R}^k$ 這一個 linear transformation 的 range 是包含於 T' 的 range 的 subspace. 利用 subspace 之間 dimension 的關係 (Corollary 4.3.6), 我們得 $\dim((T'\circ T)(\mathbb{R}^n))\leq\dim(T'(\mathbb{R}^m))$. 利用此我們可以推得以下關於矩陣的性質.

Proposition 5.3.10. 假設 A 為 $m\times n$ *matrix,* B 為 $k\times m$ *matrix,* 則 $\operatorname{rank}(BA)\leq\operatorname{rank}(A)$ 且 $\operatorname{rank}(BA)\leq\operatorname{rank}(B)$.

Proof. 考慮 $T:\mathbb{R}^n\to\mathbb{R}^m$ 定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$, 以及 $T':\mathbb{R}^m\to\mathbb{R}^k$ 定義為 $T'(\mathbf{w})=B\mathbf{w}$, $\forall\mathbf{w}\in\mathbb{R}^m$. 依此定義, 我們有 $[T]=A,[T']=B$ 以及由 Lemma 5.3.8 知 $[T'\circ T]=[T'][T]=BA$. 由前面的討論我們又知 $\dim((T'\circ T)(\mathbb{R}^n))\leq\dim(T'(\mathbb{R}^m))$, 因此利用 Corollary 5.3.5 知 $\dim((T'\circ T)(\mathbb{R}^n))=\operatorname{rank}([T'\circ T])=\operatorname{rank}(BA)$ 以及 $\dim(T'(\mathbb{R}^m))=\operatorname{rank}([T'])=\operatorname{rank}(B)$, 得證 $\operatorname{rank}(BA)\leq\operatorname{rank}(B)$.

另一方面, 套用剛剛的結果, 我們有 $\operatorname{rank}(A^{\mathrm{T}}B^{\mathrm{T}})\leq\operatorname{rank}(A^{\mathrm{T}})$. 由於 $\operatorname{rank}(A^{\mathrm{T}})=\operatorname{rank}(A)$ 以及 $\operatorname{rank}((BA)^{\mathrm{T}})=\operatorname{rank}(BA)$ (Proposition 4.4.14), 我們得

$$\operatorname{rank}(BA)=\operatorname{rank}((BA)^{\mathrm{T}})=\operatorname{rank}(A^{\mathrm{T}}B^{\mathrm{T}})\leq\operatorname{rank}(A^{\mathrm{T}})=\operatorname{rank}(A).$$

□

一般來說, 當一個函數是 invertible (即 one-to-one 且 onto) 時, 並不容易將其 inverse (反函數) 具體的寫下來. 不過對於 invertible linear transformation, 利用 standard matrix representation 我們可以很容易的將其 inverse 寫下. 首先我們來探討何時一個 linear transformation 會是 invertible. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 invertible linear transformation. 由於 T 為 onto, 我們需要 $\dim(T(\mathbb{R}^n))=m$ (Proposition 5.2.3). 然而 T 為 one-to-one, 故知 $\dim(\ker(T))=0$ (Proposition 5.2.6). 利用 Dimension Theorem for linear transformation (Theorem 5.3.5) 我們得 $m=\dim(T(\mathbb{R}^n))=\dim(T(\mathbb{R}^n))+\dim(\ker(T))=n$. 這告訴我們只有在 $m=n$ 時, T 才有可能為 invertible. 現假設 $T:\mathbb{R}^n\to\mathbb{R}^n$ 為 one-to-one, 由 $\dim(\ker(T))=0$, 我們得 $\dim(T(\mathbb{R}^n))=\dim(T(\mathbb{R}^n))+\dim(\ker(T))=n$, 亦即 T 為 onto. 同樣的, 若 T 為 onto, 則由 $\dim(T(\mathbb{R}^n))=n$ 得 $\dim(\ker(T))=n-\dim(T(\mathbb{R}^n))=0$, 亦即 T 為 one-to-one. 這告訴我們當 T 是 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 的 linear transformation 時, T 為 one-to-one 和 T 為 onto 是等價的. 因而只要其中一個是對的, 就可以得到 T 為 invertible.

Lemma 5.3.11. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation.* 則僅有當 $m=n$ 時, T 才有可能是 *invertible.* 又當 $m=n$ 時, T 為 *invertible* 和 T 為 *onto* 是等價的也和 T 為 *one-to-one* 等價.

由 Lemma 5.3.11 我們知道一個 linear transformation T 的 standard matrix representation $[T]$ 必須是 square matrix, T 才有可能是 invertible. 而且此時, 若 $[T]$ 為 $n\times n$ matrix (即 T 為 $\mathbb{R}^n$ 到 $\mathbb{R}^n$), 則 T 是 invertible 若且唯若 $\dim(T(\mathbb{R}^n))=n$, 亦即 $\operatorname{rank}([T])=n$. 另外這也等價於 $\dim(\ker(T))=0$, 亦即 $\operatorname{null}([T])=0$. 這都表示 $[T]$ 為 invertible matrix (參見 Theorem 3.5.2). 因此我們知 T 為 invertible 若且唯若 $[T]$ 為 invertible. 現若 T 為 invertible, 則因 $[T]$ 為 invertible matrix, 我們知 $[T]$ 的反矩陣 $[T]^{-1}$ 存在. 現考慮 $T':\mathbb{R}^n\to\mathbb{R}^n$ 其定義為 $T'(\mathbf{v})=[T]^{-1}\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{R}^n$. 則我們有 $T'\circ T:\mathbb{R}^n\to\mathbb{R}^n$ 的 standard matrix representation 為 $[T'\circ T]=[T'][T]=[T]^{-1}[T]=I_n$. 同理 $T\circ T':\mathbb{R}^n\to\mathbb{R}^n$ 的 standard matrix representation 為 $[T\circ T']=[T][T']=[T][T]^{-1}=I_n$. $T'\circ T$ 和 $T\circ T'$ 的 standard matrix representation 皆等於 I_n 表示 $T'\circ T=T\circ T'$ 而且對任意 $\mathbf{v}\in\mathbb{R}^n$ 皆有 $(T'\circ T)(\mathbf{v})=(T\circ T')(\mathbf{v})=I_n\mathbf{v}=\mathbf{v}$, 得知 T' 為 T 的 inverse (反函數). 因此我們證得了以下之定理.

Theorem 5.3.12. 假設 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為 *linear transformation* 且令 $[T]\in\mathscr{M}_{n\times n}$ 為其 *standard matrix representation.* 則 T 為 *invertible function* 若且唯若 $[T]$ 為 *invertible matrix.* 又若 T 為 *invertible,* 則 T 的 *inverse,* 亦為 *linear transformation* 且其 *standard matrix representation* 為 $[T]^{-1}$.

一般來講, 我們會將 invertible function f 的 inverse 用 f^{-1} 來表示. 所以當 $T:\mathbb{R}^n\to\mathbb{R}^n$ 為 invertible linear transformation, 我們也用 T^{-1} 來表示其 inverse. Theorem 5.3.12 告訴我們, 當 $T:\mathbb{R}^n\to\mathbb{R}^n$ 為 invertible linear transformation 時, $T^{-1}:\mathbb{R}^n\to\mathbb{R}^n$ 亦為 invertible linear transformation, 而且其 standard matrix representation 會是 $[T^{-1}]=[T]^{-1}$.

Example 5.3.13. 考慮 $T:\mathbb{R}3\to\mathbb{R}^3$ 定義為 $T(\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix})=\begin{bmatrix}2x_2\\x_1-x_2\\2x_2+x_3\end{bmatrix}$. 我們可得 T 的 standard matrix representation 為 $[T]=\begin{bmatrix}0&2&0\\1&-1&0\\0&2&1\end{bmatrix}$. 在 Example 3.5.8 中我們算出

$[T]^{-1} = \begin{bmatrix} \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$. 故得 $T^{-1} : \mathbb{R}^3 \to \mathbb{R}^3$ 的定義為 $T^{-1}(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = \begin{bmatrix} \frac{1}{2}x_1 + x_2 \\ \frac{1}{2}x_1 \\ -x_1 + x_3 \end{bmatrix}$. 我們驗證

$$(T^{-1} \circ T)(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = T'(\begin{bmatrix} 2x_2 \\ x_1 - x_2 \\ 2x_2 + x_3 \end{bmatrix}) = \begin{bmatrix} \frac{1}{2}(2x_2) + (x_1 - x_2) \\ \frac{1}{2}(2x_2) \\ -(2x_2) + (2x_2 + x_3) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$
$$(T \circ T^{-1})(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}) = T(\begin{bmatrix} \frac{1}{2}x_1 + x_2 \\ \frac{1}{2}x_1 \\ -x_1 + x_3 \end{bmatrix}) = \begin{bmatrix} 2(\frac{1}{2}x_1) \\ (\frac{1}{2}x_1 + x_2) - (\frac{1}{2}x_1) \\ 2(\frac{1}{2}x_1) + (-x_1 + x_3) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

得知 T^{-1} 確為 T 的 inverse.

5.4. 結論

我們學習了 $\mathbb{R}^n$ 上重要的函數, linear transformation. 一個定義在 $\mathbb{R}^n$ 的 linear transformation 可以由一組 $\mathbb{R}^n$ 的 basis 所映得的向量唯一確定, 所以我們得到所謂的 standard matrix representation. 利用 standard matrix representation, 可以幫助我們了解 linear transformation. 因此我們可以利用矩陣的性質推得許多有關 linear transformation 的性質.

由於 linear transformation 可以由一組 basis 所決定. 因此我們可以依 linear transformation 的特性選定一組較易掌握的 basis 來決定此 linear transformation. 利用這樣的概念, 我們也可以由一組 basis 經由此 linear transformation 的變化情形將一個較複雜的 linear transformation 拆解成一些較容易掌握的 linear transformations 的合成.

Part III

General Vector Space

這一部分我們將探討更一般的 vectors space. 過去從 $\mathbb{R}^m$ 中的 vector space 及其相關概念中我們發現只要符合向量的運算規則, vector space 的性質與我們探討的元素是否在 $\mathbb{R}^m$ 無關. 因此我們引進了一般 vector space 的概念, 也推廣至如 basis 及 dimension 等概念. 接著我們介紹一般 finite dimensional vector space 上的 linear transformation. 由於 basis 的存在, 我們也可將 vector space 中的元素定出坐標, 因此我們很自然得到一般 linear transformation 的 matrix representation, 並利用它幫助我們進一步了解如何掌握抽象的 linear transformation. 我們也介紹所謂的 inner product space, 利用大家熟悉的內積, 我們將投影的概念推廣到更一般的情形, 以便將來做更多方面的應用.

Chapter 6

Linear Transformation of General Vector Space

在這一章中, 我們將推廣前兩章的結果到一般的 vector space. 我們將介紹 vector space 並探討 vector space 上的 linear transformation. 由於前兩章的結果, 僅利用 vector space 的特性便可推得, 而不必侷限在 $\mathbb{R}^n$ 的情況. 所以我們很容易推得和前面一致的結果.

6.1. Vector Space and Subspace

我們曾經提過像 $\mathbb{R}^m$ 這樣, 裡面任意有限多個向量的線性組合仍在 $\mathbb{R}^m$ 中, 而向量的運算又符合 Proposition 1.2.3 的 8 項規則, 我們便稱之為 vector space. 在這一節中我們將正式定義 vector space 並探討 vector space 以及其 subspaces 的相關性質.

給定一非空集合 V, 我們說 V 中有加法運算 $+$, 表示對於任意 V 中兩個元素 $\mathbf{u},\mathbf{v}\in V$, 經由這個運算所得的結果 $\mathbf{u}+\mathbf{v}$ 仍然是 V 中的元素 (此為加法封閉性). 又我們說 $\mathbb{R}$ 對 V 有係數積表示對任意 $r\in\mathbb{R}$ 以及 $\mathbf{v}\in V$, 皆有 r 對 $\mathbf{v}$ 運算所得 $r\mathbf{v}$ 仍然在 V 中 (此為係數積封閉性).

Definition 6.1.1. 假設非空集合 V 中有加法運算 $+$, 以及 $\mathbb{R}$ 對 V 的係數積. 若這兩種運算符合以下 8 項性質, 則稱 V 為一個 *vector space*.

(1) 對任意 $\mathbf{u},\mathbf{v}\in V$, 皆有 $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$.

(2) 對任意 $\mathbf{u},\mathbf{v},\mathbf{w}\in V$, 皆有 $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.

(3) 存在一向量 $\mathbf{O}\in V$ 滿足對任意 $\mathbf{u}\in V$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$.

(4) 對任意 $\mathbf{u}\in V$ 皆可找到 $\mathbf{u}'\in V$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$.

(5) 對任意 $r,s\in\mathbb{R}$ 以及 $\mathbf{u}\in V$, 皆有 $r(s\mathbf{u})=(rs)\mathbf{u}$.

(6) 對任意 $r,s\in\mathbb{R}$ 以及 $\mathbf{u}\in V$, 皆有 $(r+s)\mathbf{u}=r\mathbf{u}+s\mathbf{u}$.

(7) 對任意 $r\in\mathbb{R}$ 以及 $\mathbf{u},\mathbf{v}\in V$ 皆有 $r(\mathbf{u}+\mathbf{v})=r\mathbf{u}+r\mathbf{v}$.

(8) 對任意 $\mathbf{u}\in V$, 皆有 $1\mathbf{u}=\mathbf{u}$.

在此要說明一下, 一般來說我們不能說一個集合是 vector space, 一定要附帶說明它的加法及係數積為何. 不過當我們談到一般抽象的 vector space 時, 我們說 V 是一個 vector space 時就隱含其中有加法運算且直接用 $+$ 表示, 同時也隱含其中有實數的係數積, 而不再去強調其中有加法及係數積. 同樣的對於常見的 vector space, 例如 $\mathbb{R}^n$, 由於我們已經有常用的加法及係數積, 所以不會再次強調其加法及係數積為何. 不過當我們要介紹一個新的具體的 vector space 時, 就一定要說明如何定出其加法及係數積. 另外 vector space 不一定要具有實數的係數積, 具一般的 "field" 的係數積也可. 所以一般來說我們必須說是 over 哪一個 field 的 vector space (因為 over 不同的 field 影響很大). 不過由於許多同學可能不知道甚麼是 field 而且我們這裡談論的都是實數 $\mathbb{R}$ 的情形, 所以即使依上面的定義我們應該說是 over $\mathbb{R}$ 的 vector space, 不過我們都省略僅說是一個 vector space 即可.

在 Corollary 1.2.4 我們證明了在 $\mathbb{R}^n$ 的情況, 上述性質 (3) 中 $\mathbf{O}$ 是唯一的而 (4) 中若給定 $\mathbf{u}$, 則 $\mathbf{u}'$ 也是唯一的. 由於當時我們的證明故意避開用 $\mathbb{R}^n$ 的加法與係數積的定義處理, 而僅利用上述 8 項性質來證明, 所以知道這個結果對一般的 vector space 也成立.

Proposition 6.1.2. 假設 V 為 *vector space,* 則在 V 中存在唯一的向量 $\mathbf{O}$ 滿足對任意 $\mathbf{u}\in V$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$. 另外, 給定 $\mathbf{u}\in V$, 存在唯一的 $\mathbf{u}'\in V$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$.

既然 $\mathbf{O}$ 是唯一的, 以後就用 $\mathbf{O}$ 這個專屬的符號來表示 V 中唯一符合 $\mathbf{O}+\mathbf{u}=\mathbf{u}, \forall\mathbf{u}\in V$ 的這個元素, 且稱之為 V 的 *additive identity* 或依慣例稱之為 *zero element.* 又給定 $\mathbf{u}\in V$, 存在唯一的 $\mathbf{u}'$ 使得 $\mathbf{u}+\mathbf{u}'=\mathbf{O}$, 依慣例我們以後就用 $-\mathbf{u}$ 來表示這一個唯一的 $\mathbf{u}'$, 且稱之為 $\mathbf{u}$ 的 *additive inverse.*

接著我們要再強調的是, 雖然在性質 (3) 中提到 $\mathbf{O}$ 是必須滿足對所有 $\mathbf{u}\in V$ 皆有 $\mathbf{O}+\mathbf{u}=\mathbf{u}$ 才可以 (事實上 Proposition 6.1.2 的證明需用到對所有 $\mathbf{u}$ 皆對才可以), 也就是說依定義要驗證對所有 $\mathbf{u}\in V$ 皆有 $\mathbf{w}+\mathbf{u}=\mathbf{u}$, 才能確定 $\mathbf{w}=\mathbf{O}$. 不過有了 Proposition 6.1.2 後, 我們便可推得只要有一個 $\mathbf{u}\in V$, 會使得 $\mathbf{w}+\mathbf{u}=\mathbf{u}$, 就可以認定 $\mathbf{w}=\mathbf{O}$ 了. 利用這個性質我們便可推得許多有關於 $\mathbf{O}$ 的性質. 例如對任意實數 r, 我們皆會有 $r\mathbf{O}=\mathbf{O}$. 這是因為利用性質 (7) 可得 $r\mathbf{O}+r\mathbf{O}=r(\mathbf{O}+\mathbf{O})=r\mathbf{O}$. 也就是說 $r\mathbf{O}$ 加了 $r\mathbf{O}$ 仍為 $r\mathbf{O}$, 所以由前面所提可知 $r\mathbf{O}=\mathbf{O}$. 為了方便起見, 我們列出以下的結果, 至於證明請參見 Corollary 1.2.6 以及 Corollary 1.2.7.

Proposition 6.1.3. 假設 V 為 *vector space,* 我們有以下之結果.

(1) 若對於 $\mathbf{w}\in V$ 存在 $\mathbf{u}\in V$ 使得 $\mathbf{w}+\mathbf{u}=\mathbf{u}$, 則 $\mathbf{w}=\mathbf{O}$.

(2) 對任意 $\mathbf{v}\in V$ 皆有 $0\mathbf{v}=\mathbf{O}$.

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

(4) 對任意 $r\in\mathbb{R},\mathbf{v}\in V$ 皆有 $(-1)(r\mathbf{v})=-(r\mathbf{v})=r(-\mathbf{v})$.

由 Proposition 6.1.3 我們知當 $r=0$ 或 $\mathbf{v}=\mathbf{O}$ 時會有 $r\mathbf{v}=\mathbf{O}$, 但若 $r\neq 0$ 且 $\mathbf{v}\neq\mathbf{O}$, 是否有可能 $r\mathbf{v}=\mathbf{O}$ 呢? 答案是不可能. 這是因為若 $r\neq 0$ 且 $\mathbf{v}\neq\mathbf{O}$, 我們可以考慮 $1/r$ 乘上 $r\mathbf{v}$ 由 vector space 運算性質 (5) 得 $\frac{1}{r}(r\mathbf{v})=\frac{r}{r}\mathbf{v}=1\mathbf{v}$, 再由性質 (8) 知 $1\mathbf{v}=\mathbf{v}$, 故得 $\frac{1}{r}(r\mathbf{v})=\mathbf{v}\neq\mathbf{O}$. 現若 $r\mathbf{v}=\mathbf{O}$, 由 Proposition 6.1.3 (3) 會得到 $\frac{1}{r}(r\mathbf{v})=\frac{1}{r}\mathbf{O}=\mathbf{O}$, 造成矛盾. 故知此時 $r\mathbf{v}$ 絕對不會是 $\mathbf{O}$.

Question 6.1. *假設 V 為 vector space, $\mathbf{v}\in V$ 且 $\mathbf{v}\neq\mathbf{O}$, 試證明若 $r,s\in\mathbb{R}$ 且 $r\neq s$, 則 $r\mathbf{v}\neq s\mathbf{v}$, 並利用這個結果證明若 V 是 vector space over $\mathbb{R}$ 且 V 中有非零元素, 則 V 有無窮多個元素.*

總而言之, vector space 中所要求加法及係數積的 8 項性質, 就是要確保一個 vector space 中的元素運算都可像實數一般處理. 例如, 我們可以如實數一樣引用"減法"的符號, 也就是說將 $\mathbf{w}+(-\mathbf{v})$ 寫成 $\mathbf{w}-\mathbf{v}$. 如此一來以後我們在一些等式的推演時就直接沿用大家習慣的「移項」的說法. 例如 $2\mathbf{u}+\mathbf{v}=\mathbf{w}$, 我們就直接移項且乘以 $1/2$ 得 $\mathbf{u}=\frac{1}{2}(\mathbf{w}-\mathbf{v})$.

接下來我們看一些常見的 vector space 的例子.

Example 6.1.4. (A) 考慮 $\mathscr{M}_{m\times n}$ 為所有 $m\times n$ matrices 所成的集合. 利用我們一般定義矩陣的加法與係數積 (參見 Definition 3.1.2), 我們可以證得這兩種運算符合 vector space 運算的 8 項性質 (參見 Proposition 3.1.3). 所以在這個加法運算與係數積之下 $\mathscr{M}_{m\times n}$ 是一個 vector space.

(B) 考慮 $P(\mathbb{R})$ 為所有以 x 為變數的實係數多項式. 考慮一般多項式的加法與係數積, 我們證明 $P(\mathbb{R})$ 為 vector space. 首先回顧若 $f(x)=a_nx^n+\cdots+a_1x+a_0, g(x)=b_mx^m+\cdots+b_1x+b_0\in P(\mathbb{R})$, 其中 $m\leq n$, 則我們可以將 $g(x)$ 寫成 $g(x)=b_nx^n+\cdots+b_mx^m+\cdots+b_1x+b_0$, 其中 $b_n=b_{n-1}=\cdots=b_{m+1}=0$. 為了方便起見雖然多項式的次數可能不同, 以後我們都用這種方式將它們補成相同次數再相加. 所以我們可以將 $f(x)+g(x)$ 的定義寫成 $f(x)+g(x)=\sum_{i=0}^{n}(a_i+b_i)x^i$. 而對於 $r\in\mathbb{R}$, 係數積 $rf(x)$ 的定義為 $rf(x)=\sum_{i=0}^{n}(ra_i)x^i$. 利用這個定義, 我們知道兩多項式相加仍為多項式以及實數對多項式的係數積仍為多項式, 所以在此定義之下加法和係數積確為 $P(\mathbb{R})$ 中的運算. 接著我們要一一檢查是否符合 vectors space 的 8 項運算規則.

(1) 對任意 $f(x)=\sum_{i=0}^{n}a_ix^i$, $g(x)=\sum_{i=0}^{n}b_ix^i\in P(\mathbb{R})$ 我們有

$$f(x)+g(x)=\sum_{i=0}^{n}(a_i+b_i)x^i=\sum_{i=0}^{n}(b_i+a_i)x^i=g(x)+f(x).$$

(2) 對任意 $f(x)=\sum_{i=0}^{n}a_ix^i$, $g(x)=\sum_{i=0}^{n}b_ix^i, h(x)=\sum_{i=0}^{n}c_ix^i\in P(\mathbb{R})$ 我們有

$$(f(x)+g(x))+h(x)=\sum_{i=0}^{n}(a_i+b_i)x^i+\sum_{i=0}^{n}c_ix^i=\sum_{i=0}^{n}((a_i+b_i)+c_i)x^i,$$

$$f(x)+(g(x)+h(x))=\sum_{i=0}^{n}a_ix^i+\sum_{i=0}^{n}(b_i+c_i)x^i=\sum_{i=0}^{n}(a_i+(b_i+c_i))x^i.$$

由於 $(a_i+b_i)+c_i=a_i+(b_i+c_i)$, 故得證 $(f(x)+g(x))+h(x)=f(x)+(g(x)+h(x))$.

(3) 考慮零多項式 $g(x)=0=\sum_{i=0}^{n}b_ix^i\in P(\mathbb{R})$, 其中 $b_i=0$, $\forall i=0,1,\ldots,n$. 此時對任意 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{R})$, 可得

$$f(x)+g(x)=\sum_{i=0}^{n}(a_i+b_i)x^i=\sum_{i=0}^{n}a_ix^i=f(x).$$

(4) 給定 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{R})$, 我們考慮 $h(x)=\sum_{i=0}^{n}(-a_i)x^i\in P(\mathbb{R})$, 則

$$f(x)+h(x)=\sum_{i=0}^{n}(a_i-a_i)x^i=\sum_{i=0}^{n}0x^i=0.$$

(5) 對任意 $r,s\in\mathbb{R}$ 以及 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{R})$, 我們有

$$r(sf(x))=r(\sum_{i=0}^{n}(sa_i)x^i)=\sum_{i=0}^{n}(r(sa_i))x^i=\sum_{i=0}^{n}((rs)a_i)x^i=(rs)\sum_{i=0}^{n}a_ix^i=(rs)f(x).$$

(6) 對任意 $r,s\in\mathbb{R}$ 以及 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{R})$, 皆有

$$(r+s)f(x)=\sum_{i=0}^{n}((r+s)a_i)x^i=\sum_{i=0}^{n}(ra_i+sa_i)x^i=\sum_{i=0}^{n}(ra_i)x^i+\sum_{i=0}^{n}(sa_i)x^i=rf(x)+sf(x).$$

(7) 對任意 $r\in\mathbb{R}$ 以及 $f(x)=\sum_{i=0}^{n}a_ix^i$, $g(x)=\sum_{i=0}^{n}b_ix^i\in P(\mathbb{R})$ 皆有

$$r(f(x)+g(x))=r(\sum_{i=0}^{n}(a_i+b_i)x^i)=\sum_{i=0}^{n}(r(a_i+b_i))x^i=\sum_{i=0}^{n}(ra_i+rb_i)x^i=rf(x)+rg(x).$$

(8) 對任意 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{R})$, 皆有

$$1f(x)=\sum_{i=0}^{n}(1a_i)x^i=\sum_{i=0}^{n}a_ix^i=f(x).$$

因為 $P(\mathbb{R})$ 的加法與係數積符合 vector space 的 8 項運算規則, 所以在這個加法與係數積的運算之下 $P(\mathbb{R})$ 是一個 vector space.

(C) 考慮 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 為所有 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformations 所成的集合. 我們定義過 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的加法與係數積, 並知在加法與係數積之下 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 是封閉的 (參見 Proposition 5.1.6). 由於 linear transformation 的加法及係數積的定義是對定義域裡的每一個元素代入後定義的, 所以由 $\mathbb{R}^m$ 是 vector space, 我們很容易證明 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的加法與係數積符合 vector space 的 8 項運算規則. 例如對於 (2), 假設 $T_1,T_2,T_3\in\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$. 對於所有 $\mathbf{v}\in\mathbb{R}^n$, 我們有

$$((T_1+T_2)+T_3)(\mathbf{v})=(T_1+T_2)(\mathbf{v})+T_3(\mathbf{v})=(T_1(\mathbf{v})+T_2(\mathbf{v}))+T_3(\mathbf{v}),$$

而

$$(T_1+(T_2+T_3))(\mathbf{v})=T_1(\mathbf{v})+(T_2+T_3)(\mathbf{v})=T_1(\mathbf{v})+(T_2(\mathbf{v})+T_3(\mathbf{v})).$$

故利用 $\mathbb{R}^m$ 中的向量加法符合結合律, 得 $((T_1+T_2)+T_3)(\mathbf{v})=(T_1+(T_2+T_3))(\mathbf{v})$. 由於這是對所有 $\mathbf{v}\in\mathbb{R}^n$ 皆成立的, 故由函數相等的定義得 $(T_1+T_2)+T_3=T_1+(T_2+T_3)$. 另外要說明的是 (3) 中的 zero element 指的就是 zero mapping, 即 $T:\mathbb{R}^n\to\mathbb{R}^m$, 滿足 $T(\mathbf{v})=\mathbf{O}$, $\forall\mathbf{v}\in\mathbb{R}^n$ 這樣的 linear transformation. 由於 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的加法與係數積符合 vector space 的 8 項運算規則, 因此在這個加法與係數積的運算之下 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 是一個 vector space.

在 $\mathbb{R}^n$ 的情況, 我們介紹過 $\mathbb{R}^n$ 的非空子集合若在 $\mathbb{R}^n$ 的加法及係數積運算之下為 vector space, 便稱為 $\mathbb{R}^n$ 的 subspace. 同樣的對於一般 vector space 的非空子集合, 如果在此 vector space 的加法及係數積運算之下這個子集合亦為 vector space, 則稱為此 vector space 的 *subspace*.

Definition 6.1.5. 假設 V 為 vector space, W 為 V 的 nonempty subset. 若在 V 的加法及係數積運算之下 W 亦為 vector space, 則稱 W 為 V 的 *subspace*.

雖然一個 vector space 的 subspace 仍為 vector space, 但要檢查是否為 subspace 不必像檢查 vector space 一樣要去檢查 8 項的運算規則. 這是因為原本 vector space 的 8 項運算規則中除了 (3)(4) 兩項會和所在的集合有關外, 其他各項僅是元素間的運算規則, 和所在的集合無關. 所以我們有和 $\mathbb{R}^n$ 的 subspace 同樣的性質 (參見 Proposition 4.1.2).

Proposition 6.1.6. 假設 V 為 *vector space* 且 W 為 V 的子集合. 則 W 為 V 的 *subspace* 若且唯若 $\mathbf{O} \in W$ 且對任意 $\mathbf{u},\mathbf{v} \in W$, $r \in \mathbb{R}$ 皆有 $\mathbf{u}+r\mathbf{v} \in W$.

Proof. $(\Rightarrow)$: 依 subspace 的定義, 加法及係數積皆有封閉性, 故對任意 $\mathbf{u},\mathbf{v} \in W$, $r \in \mathbb{R}$ 皆有 $\mathbf{u}+r\mathbf{v} \in W$. 又依定義 W 為非空集合, 故必存在一向量 $\mathbf{w} \in W$. 現考慮 $0\mathbf{w}$, 依封閉性 $0\mathbf{w} \in W$. 又因 V 為 vector space, 我們知 $0\mathbf{w} = \mathbf{O}$ (Proposition 6.1.3(2)). 故得證 $\mathbf{O} = 0\mathbf{w} \in W$.

$(\Leftarrow)$: 由 $\mathbf{O} \in W$, 我們知 W 為 V 的非空子集合. 現由對任意 $\mathbf{u},\mathbf{v} \in W$, $r \in \mathbb{R}$ 皆有 $\mathbf{u}+r\mathbf{v} \in W$, 表示 V 的加法和係數積對 W 的元素有封閉性, 亦即這兩個運算也可視為 W 上的加法與係數積. 所以要證明 W 為 V 的 subspace, 我們須證明此加法與係數積在 W 上符合 vector space 的 8 項的運算規則. 然而除了 (3)(4) 兩項以外, 其他各項對於所有 V 的元素都成立. 而 W 為 V 的子集合, W 的元素皆為 V 的元素, 因此知這幾項對於所有 W 的元素皆成立. 又已知 $\mathbf{O} \in W$, 所以 (3) 成立. 而對於任意 $\mathbf{w} \in W$, 由係數積封閉性知 $(-1)\mathbf{w} \in W$, 又因 $\mathbf{w}+(-1)\mathbf{w} = 0\mathbf{w} = \mathbf{O}$, 得證 (4) 也成立. □

由 Proposition 6.1.6, 我們知道要檢查一個 vector space V 中的子集合 W 是否為 V 的 subspace, 我們僅要檢查

(1) $\mathbf{O} \in V$

(2) $\mathbf{u},\mathbf{v} \in V$, $r \in \mathbb{R} \Rightarrow \mathbf{u}+r\mathbf{v} \in V$.

是否成立即可. 我們看以下的例子.

Example 6.1.7. (A) 考慮 $\mathscr{M}_{n\times n}$, 即所有 $n\times n$ 方陣所成的 vector space. 我們想知道 $\mathscr{M}_{n\times n}$ 中所有的 symmetric matrices (對稱矩陣) 所成的集合是否為 $\mathscr{M}_{n\times n}$ 的 subspace. 首先回顧 $A \in \mathscr{M}_{n\times n}$ 為 symmetric matrix, 表示 $A^{\mathrm{T}} = A$. 很明顯的 $n\times n$ 階零矩陣 $\mathbf{O}$ 為 symmetric matrix. 而若 $A,B \in \mathscr{M}_{n\times n}$ 滿足 $A^{\mathrm{T}} = A, B^{\mathrm{T}} = B$, 則對任意 $r \in \mathbb{R}$, 我們有 $(A+rB)^{\mathrm{T}} = A^{\mathrm{T}} + (rB)^{\mathrm{T}} = A+rB$ (參見 Proposition 3.2.4). 亦即 $A+rB$ 亦為 symmetric matrix, 得證 $\mathscr{M}_{n\times n}$ 中所有的 symmetric matrices 所成的集合為 $\mathscr{M}_{n\times n}$ 的 subspace.

$\mathscr{M}_{n\times n}$ 中所有的 invertible matrices (可逆矩陣) 所成的集合是否為 $\mathscr{M}_{n\times n}$ 的 subspace 呢? 答案是否定的. 很明顯的零矩陣 $\mathbf{O}$ 就不是 invertible, 所以由 $\mathbf{O}$ 不在其中就可得 $\mathscr{M}_{n\times n}$ 中所有的 invertible matrices 所成的集合不是 $\mathscr{M}_{n\times n}$ 的 subspace. 其實即使我們考慮 invertible matrices 所成的集合與 $\{\mathbf{O}\}$ 的聯集, 仍不會是 $\mathscr{M}_{n\times n}$ 的 subspace. 因為即使此時 $\mathbf{O}$ 在其中, 但仍有可能兩個 invertible matrices 相加後就不是 invertible. 例如在 2×2 的情形, $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 和 $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 皆為 invertible, 但是 $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ 不是 invertible.

(B) 考慮 $P(\mathbb{R})$, 即所有以 x 為變數的實係數多項式所成的 vector space. 給定一自然數 $n\in\mathbb{N}$, 我們說明所有次數小於等於 n 的多項式所成的集合 $P_n(\mathbb{R})$ 是 $P(\mathbb{R})$ 的 subspace. 首先我們可以將 $P_n(\mathbb{R})$ 寫成 $P_n(\mathbb{R}) = \{\sum_{i=0}^{n} a_i x^i \mid a_i \in \mathbb{R}\}$. 很明顯的零多項式屬於 $P_n(\mathbb{R})$ (注意一般數學上定義零多項式的次數為 $-\infty$, 而不是 0. 這個部分以後代數課程會去談論, 這裡就不多談). 又若 $f(x) = \sum_{i=0}^{n} a_i x^i, g(x) = \sum_{i=0}^{n} b_i x^i \in P_n(\mathbb{R})$, 則對任意 $r\in\mathbb{R}$, 我們有 $f(x)+rg(x) = \sum_{i=0}^{n}(a_i+rb_i)x^i \in P_n(\mathbb{R})$. 故知 $P_n(\mathbb{R})$ 為 $P(\mathbb{R})$ 的 subspace. 要注意, 若僅考慮次數等於 n 的多項式所成的集合, 那麼就不會是 $P(\mathbb{R})$ 的 subspace 了. 很明顯的零多項式就不會在裡面. 又即使加入零多項式, 但仍有可能兩個次數為 n 的多項式相加之後其次數變小了, 例如 $(x^2+x+1)+(-x^2+x+1) = 2x+2$. 也就是說在這情況之下加法是不封閉的, 所以無法成為一個 vector space.

(C) 考慮 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 為所有 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformations 所成的 vector space. 給定 $\mathbb{R}^n$ 的一個 subspace V, 我們考慮

$$N(V) = \{T \in \mathscr{L}(\mathbb{R}^n,\mathbb{R}^m) \mid V \subseteq \ker(T)\}.$$

亦即 $N(V)$ 裡的元素皆為滿足對任意 $\mathbf{v}\in V$, $T(\mathbf{v}) = \mathbf{O}$ 的 linear transformation T. 很明顯的, $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 裡的 additive identity, 即 zero mapping, 因其 kernel 為 $\mathbb{R}^n$ 故由 $V\subseteq\mathbb{R}^n$ 知 zero mapping 屬於 $N(V)$. 現若 $T_1,T_2\in N(V)$, 則對於任意 $r\in\mathbb{R}$, 由於對所有 $\mathbf{v}\in V$ 皆有

$$(T_1+rT_2)(\mathbf{v}) = T_1(\mathbf{v}) + r(T_2(\mathbf{v})) = \mathbf{O} + r\mathbf{O} = \mathbf{O}.$$

得證 $V\subseteq\ker(T_1+rT_2)$, 因此 $T_1+rT_2\in N(V)$. 得證 $N(V)$ 為 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的 subspace.

Question 6.2. 考慮 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 為所有 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 *linear transformations* 所成的 *vector space*. 給定 $\mathbb{R}^n$ 的一個 *subspace* V, 我們考慮

$$S(V) = \{T \in \mathscr{L}(\mathbb{R}^n,\mathbb{R}^m) \mid \ker(T) \subseteq V\}.$$

是否 $S(V)$ 為 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的 *subspace?*

給定 $\mathbb{R}^m$ 的一個 *subspace* W, 我們考慮

$$R(W) = \{T \in \mathscr{L}(\mathbb{R}^n,\mathbb{R}^m) \mid T(\mathbb{R}^n) \subseteq W\}.$$

是否 $R(W)$ 為 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的 *subspace?* 另外若考慮

$$I(W) = \{T \in \mathscr{L}(\mathbb{R}^n,\mathbb{R}^m) \mid W \subseteq T(\mathbb{R}^n)\}.$$

是否 $I(W)$ 為 $\mathscr{L}(\mathbb{R}^n,\mathbb{R}^m)$ 的 *subspace?*

6.2. Basis and Dimension

在這一節中, 我們將介紹 vector space 中 basis 的概念, 並介紹 finitely generated vector space 中 dimension 之概念及性質.

我們在介紹 $\mathbb{R}^n$ 的 subspace 時, 引進了 basis 的概念. 有了 basis, 表示此 subspace 中的元素, 都可以用這一組 basis 的線性組合唯一表示. 這又等同於這組 basis 是此 subspace 的 spanning vectors 且為 linearly independent (參見 Proposition 4.1.8). 我們可將此概念推廣到一般的 vector space.

Definition 6.2.1. 假設 V 為 vector space 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$. 對於任意 $c_1,\ldots,c_n\in\mathbb{R}$, 我們稱 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 為 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 *linear combination*. 所有 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 linear combination 所成的集合, 我們用 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 來表示, 亦即 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=\{\sum_{i=1}^n c_i\mathbf{v}_i\mid c_1,\ldots,c_n\in\mathbb{R}\}$.

我們知道 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 會是 V 的 subspace (參見 Proposition 1.3.2). 事實上它是包含 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 最小的 subspace. 這是因為若 W 是 V 的 subspace 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in W$, 則由 subspace 的加法與係數積的封閉性得 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)\subseteq W$. 特別的, 當 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=V$ 時, 我們給它一個特別的名稱.

Definition 6.2.2. 假設 V 為 vector space. 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 滿足 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=V$, 則稱 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 為 V 的一組 *spanning set*. 此時我們稱 V 為 *finitely generated vector space*.

為何特別稱之為 finitely generated vector space 呢? 這表示此 vector space 中所有的元素都可表示成固定有限多個元素的線性組合. 例如 $\mathbb{R}^n$ 就是 finitely generated (可由 $\mathbf{e}_1,\ldots\mathbf{e}_n$ 所展成). 由於當初我們僅談 $\mathbb{R}^n$ 及其 subspaces, 它們都是 finitely generated vector space 所以不特別去強調. 現在我們談論一般的 vector space, 有可能不是 finitely generated, 所以要區分出來, 我們看以下的例子.

Example 6.2.3. 我們討論前面提的一些 vector space 哪些是 finitely generated vector space.

(A) $\mathscr{M}_{m\times n}$ 是 finitely generated. 因為考慮 $M_{i,j}\in\mathscr{M}_{m\times n}$, 表示 (i,j)-th entry 為 1, 其他 entry 為 0 的 $m\times n$ matrix. 很容易看出所有的 $m\times n$ matrix 皆可寫成 $M_{i,j}$ 其中 $1\leq i\leq m$, $1\leq j\leq n$ 的 linear combination. 所以 $\{M_{i,j}\mid 1\leq i\leq m, 1\leq j\leq n\}$ 是 $\mathscr{M}_{m\times n}$ 的 spanning set, 也因此 $\mathscr{M}_{m\times n}$ 是 finitely generated vector space.

(B) $P(\mathbb{R})$ 不是 finitely generated vector space. 這是因為如果 $\{f_1(x),\ldots,f_n(x)\}$ 為 $P(\mathbb{R})$ 的 spanning set, 假設 $f_1(x),\ldots,f_n(x)$ 的最高次為 m, 則任何 $f_1(x),\ldots,f_n(x)$ 的 linear combination $c_1f_1(x)+\cdots+c_nf_n(x)$ 的次數皆不可能大於 m. 也就是說 $\mathrm{Span}(f_1(x),\ldots,f_n(x))$ 不可能包含次數大於 m 的多項式. 此與 $\{f_1(x),\ldots,f_n(x)\}$ 為 $P(\mathbb{R})$ 的 spanning set 明顯不合, 故知 $P(\mathbb{R})$ 不可能是 finitely generated. 不過次數小於等於 n 的多項式所成的集合 $P_n(\mathbb{R})$ 就是 finitely generated vector space. 很容易看出 $\{x^n,\ldots,x,1\}$ 就是 $P_n(\mathbb{R})$ 的 spanning set.

大家或許直覺會認為 finite generated vector space 的 subspace 一定也是 finitely generated. 這是對的, 不過證明卻不是如直覺那麼簡單 (大家不妨現在試著證明看看). 它的證明等到介紹完 linearly independence 的概念, 我們就可以處理.

Spanning set 的概念是處理 linear combination 的存在性, 而 linear independence 的概念便是處理 linear combination 的唯一性. 我們可推廣 $\mathbb{R}^n$ 中向量 linearly independent 的定義到一般的 vector space.

Definition 6.2.4. 假設 V 為 vector space 且 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$. 若存在 $c_1,\dots,c_n\in\mathbb{R}$ 不全為 0 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$, 則稱 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 *linearly dependent*. 反之, 若只有在 $c_1,\dots,c_n$ 皆為 0 的情況才會使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$, 則稱 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 *linearly independent*.

如同在 $\mathbb{R}^n$ 的情形, $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly independent, 可以確保選取不同的 $c_1,\dots,c_n$ 就會得到不同的 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 也就是說每一個不同的 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 的 linear combination 都會有不同的結果. 反之, 如果 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly dependent, 假設 $d_1,\dots,d_n$ 不全為 0 使得 $d_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$. 則對任意 $c_1,\dots,c_n\in\mathbb{R}$, $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 和 $(c_1+d_1)\mathbf{v}_1+\cdots+(c_n+d_n)\mathbf{v}_n$ 是 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 不同的 linear combination, 但它們是表示同樣的元素. 這是因為我們有

$$(c_1+d_1)\mathbf{v}_1+\cdots+(c_n+d_n)\mathbf{v}_n=$$
$$(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)+(d_1\mathbf{v}_1+\cdots+d_n\mathbf{v}_n)=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n+\mathbf{O}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n.$$

Example 6.2.5. 很容易看出在 $P(\mathbb{R})$ 中 $x^n,x^{n-1},\dots,x,1$ 為 linearly independent. 這是因為如果 $c_n,\dots,c_1,c_0\in\mathbb{R}$ 使得 $c_nx^n+\cdots+c_1x+c_01$ 為零多項式, 則依零多項式的定義, $c_n,\dots,c_1,c_0$ 必全為 0. 現在我們介紹 $P(\mathbb{R})$ 中另一種重要的 linearly independent 的多項式的建構方法, 稱為 *Lagrange interpolation polynomials*. 我們僅舉出一個例子, 一般狀況請大家自行推廣.

給定 a,b,c 三相異實數, 我們希望找到三個二次多項式 $p_1(x),p_2(x),p_3(x)$ 滿足

$$p_1(a)=1,p_1(b)=p_1(c)=0,\quad p_2(b)=1,p_2(a)=p_2(c)=0\quad\text{and}\quad p_3(c)=1,p_3(a)=p_3(b)=0.$$

由於 $p_1(b)=p_1(c)=0$, 我們知 $p_1(x)$ 應為 $(x-b)(x-c)$ 的倍式, 也就是存在實數 r 使得 $p_1(x)=r(x-b)(x-c)$. 但又要求 $p_1(a)=1$, 故代入 $x=a$ 得 $r=1/(a-b)(a-c)$. 同理可求出 $p_2(x),p_3(x)$ 因此我們有

$$p_1(x)=\frac{(x-b)(x-c)}{(a-b)(a-c)},\quad p_2(x)=\frac{(x-a)(x-c)}{(b-a)(b-c)}\quad\text{and}\quad p_3(x)=\frac{(x-a)(x-b)}{(c-a)(c-b)}.$$

我們要說明 $p_1(x),p_2(x),p_3(x)$ 為 linearly independent. 首先觀察, 若 $f(x)=c_1p_1(x)+c_2p_2(x)+c_3p_3(x)$, 則代入 $x=a$ 時可由 $p_1(a)=1,p_2(a)=p_3(a)=0$, 得 $f(a)=c_1$. 同理知 $f(b)=c_2,f(c)=c_3$. 因此現若 $f(x)$ 為零多項式, 由 $f(a)=f(b)=f(c)=0$, 可得 $c_1=c_2=c_3=0$. 也就是說只有當 $c_1=c_2=c_3=0$ 時才會使得 $c_1p_1(x)+c_2p_2(x)+c_3p_3(x)$ 為零多項式, 得證 $p_1(x),p_2(x),p_3(x)$ 為 linearly independent.

我們可以將上面的討論推廣. 對於任意 n 個相異實數 $a_1,\dots,a_n$, 我們都可以得到 n 個 $n-1$ 次的多項式 $p_1(x),\dots,p_n(x)$ 滿足 $p_i(a_i)=1$ 且當 $j\neq i$ 時, $p_i(a_j)=0$. 利用前面的論述, 我們可以得到 $p_1(x),\dots,p_n(x)$ 為 linearly independent.

Spanning set 和 linearly independent 性質看似無關, 不過它們之間卻存有某種程度的互補關係. 我們將它們的關係列出如下.

Lemma 6.2.6. 假設 V 為 *vector space* 且令 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$.

(1) $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{n-1}) \neq \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{n-1},\mathbf{v}_n)$ 若且唯若 $\mathbf{v}_{n+1} \notin \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$.

(2) 若已知 $\mathbf{v}_1,\ldots,\mathbf{v}_{n-1}$ 為 *linearly independent,* 則 $\mathbf{v}_1,\ldots,\mathbf{v}_{n-1},\mathbf{v}_n$ 為 *linearly independent* 若且唯若 $\mathbf{v}_n \notin \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{n-1})$.

(3) 假設 $\mathbf{w}_1,\ldots,\mathbf{w}_m \in V$. 若 $\mathbf{w}_1,\ldots,\mathbf{w}_m \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 且 $m>n$, 則 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 為 *linearly dependent.*

Proof. 其實 (1), (2), (3) 分別就是 Lemma 4.2.1, Lemma 4.2.4 和 Lemma 4.2.5 在一般 vector space 的情形. 由於 (1), (2) 的證明和 Lemma 4.2.1, Lemma 4.2.4 一致. 所以我們僅證明 (3) 即可. 在 Lemma 4.2.5 的證明方式用到了 $\mathbb{R}^m$ 向量的坐標表示法, 我們現在談論的是一般 vector space, 所以要稍加修改.

由於 $\mathbf{w}_1,\ldots,\mathbf{w}_m \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 因此對任意 $j=1,\ldots,m$, $\mathbf{w}_j$ 都可以寫成 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 linear combination. 也就是說, 存在 $a_{1,j},\ldots,a_{i,j},\ldots,a_{n,j} \in \mathbb{R}$ 使得

$$\mathbf{w}_j = a_{1,j}\mathbf{v}_1 + \cdots + a_{i,j}\mathbf{v}_i + \cdots + a_{n,j}\mathbf{v}_n.$$

現在我們要找到 $c_1,\ldots,c_m \in \mathbb{R}$ 不全為 0 使得 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m = \mathbf{O}$, 便證得 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 為 linearly dependent. 現將 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m$ 中每一個 $\mathbf{w}_j$ 換成 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 linear combination 後會等於

$$(c_1a_{1,1}+\cdots+c_ma_{1,m})\mathbf{v}_1+\cdots+(c_1a_{i,1}+\cdots+c_ma_{i,m})\mathbf{v}_i+\cdots+(c_1a_{m,1}+\cdots+c_ma_{m,m})\mathbf{v}_m. \quad (6.1)$$

因此若我們能找到 $c_1,\ldots,c_m \in \mathbb{R}$ 使得式子 (6.1) 中每個 $\mathbf{v}_i$ 的係數等於 0, 便可得到 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m = \mathbf{O}$. 因此我們只要找到聯立方程組

$$\begin{cases} a_{1,1}x_1+\cdots+a_{1,m}x_m &= 0 \\ \quad\vdots & \\ a_{i,1}x_1+\cdots+a_{i,m}x_m &= 0 \\ \quad\vdots & \\ a_{n,1}x_1+\cdots+a_{n,m}x_m &= 0 \end{cases}$$

的一組解 $x_1=c_1,\ldots,x_m=c_m$, 就可以使得 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m=\mathbf{O}$. 然而這個 homogeneous linear system 的方程式個數 n 少於未知數個數 m, 由 Corollary 3.4.7 知必存在不全為 0 的 $c_1,\ldots,c_m \in \mathbb{R}$ 使得 $x_1=c_1,\ldots,x_m=c_m$ 為其一組解. 故得證 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 為 linearly dependent. □

現在我們可以回答 finitely generated vector space 的 subspace 也是 finitely generated.

Proposition 6.2.7. 假設 V 為 *finitely generated vector space.* 若 W 為 V 的 *subspace,* 則 W 為 *finitely generated vector space.*

Proof. 依 V 為 finitely generated 的假設, 存在 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$ 滿足 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=V$. 由於 $\{\mathbf{O}\}=\mathrm{Span}(\mathbf{O})$ 為 finitely generated, 我們僅需要考慮 $W\neq\{\mathbf{O}\}$ 的情況. 我們用反證法, 假設 W 不是 finitely generated. 現任取 $\mathbf{w}_1\in W$ 其中 $\mathbf{w}_1\neq\mathbf{O}$. 由於 W 不是 finitely generated, 我們知 $\mathrm{Span}(\mathbf{w}_1)\neq W$, 亦即存在 $\mathbf{w}_2\in W$ 且 $\mathbf{w}_2\notin\mathrm{Span}(\mathbf{w}_1)$. 由 Lemma 6.2.6 (2) 知 $\mathbf{w}_1,\mathbf{w}_2$ 為 linearly independent. 同理, 因 W 不是 finitely generated, 我們知 $\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)\neq W$ 亦即存在 $\mathbf{w}_3\in W$ 且 $\mathbf{w}_3\notin\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2)$. 由 Lemma 6.2.6 (2) 知 $\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3$ 為 linearly independent. 這樣一直下去, 利用數學歸納法假設我們得 $\mathbf{w}_1,\ldots,\mathbf{w}_k\in W$ 為 linearly independent. 由於 $\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)\neq W$, 存在 $\mathbf{w}_{k+1}\in W$ 且 $\mathbf{w}_{k+1}\notin\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)$. 因此再由 Lemma 6.2.6 (2) 知 $\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+}$ 為 linearly independent. 我們利用數學歸納法證明了, 若 W 不是 finitely generated, 則對任意 $m\in\mathbb{N}$, 皆存在 $\mathbf{w}_1,\ldots,\mathbf{w}_m\in W$ 為 linearly independent. 然而這在 $m>n$ 是會造成矛盾的. 因為此時由於 $\mathbf{w}_1,\ldots,\mathbf{w}_m\in W\subseteq V=\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, Lemma 6.2.6 (3) 告訴我們 $\mathbf{w}_1,\ldots,\mathbf{w}_m$ 必為 linearly dependent. 因此知 W 為 finitely generated vector space. □

Question 6.3. 利用在 $P(\mathbb{R})$ 中對於任意 $n\in\mathbb{N}$, $x^n,x^{n-1},\ldots,x,1$ 為 *linearly independent,* 證明 $P(\mathbb{R})$ 不是 *finitely generated vector space.*

有了 spanning set 的存在性以及 linearly independent 的唯一性, 我們知道若 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 為 vector space V 的 spanning set 且為 linearly independent, 則所有 V 中的元素都可以唯一寫成 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 linear combination, 也因此我們有以下的定義.

Definition 6.2.8. 假設 V 為 vector space. 若 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 為 V 的 spanning set 且為 linearly independent, 則稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 *basis.*

接下來我們要說明非零的 finitely generated vector space 都會有 basis. 其實這個性質對於不是 finitely generated 的 vector space 也對, 不過由於這會牽涉到較抽象的邏輯概念而且我們以後談論的 vector space 都是 finitely generated, 所以我們不去談論它.

Theorem 6.2.9. 假設 $V\neq\{\mathbf{O}\}$ 為 *finitely generated vector space.* 則存在 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 為 V 的一組 *basis.* 又若 $\mathbf{w}_1,\ldots,\mathbf{w}_m\in V$ 亦為 V 的一組 *basis,* 則 $m=n$.

Proof. 這個定理是 Theorem 4.3.1 和 Theorem 4.3.2 在一般 vector space 的推廣. 我們可以利用 Theorem 4.3.1 的方法處理 basis 的存在性. 不過這裡我們利用 finitely generated 的性質, 直接用數學歸納法處理.

我們對 vector space 的 spanning set 的元素個數做數學歸納法. 首先假設 V 可由一個元素展成. 也就是說 $V=\mathrm{Span}(\mathbf{u})$. 此時由於 $V\neq\{\mathbf{O}\}$, 知 $\mathbf{u}\neq\mathbf{O}$. 故 $\mathbf{u}$ 本身是 linearly independent 且 $\{\mathbf{u}\}$ 是 V 的 spanning set 故 $\mathbf{u}$ 是 V 的 basis. 利用納法假設當一個 vector space 可由 k 個元素展成時, basis 是存在的. 現假設 V 是一個可由 $k+1$ 個元素展成的 vector space, 我們假設 $V=\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_k,\mathbf{u}_{k+1})$. 現考慮 $W=\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_k)$. 由於 W 是一個可由 k 個元素展成的 vector space, 依歸納法假設存在 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in W$ 為 W 的一組 basis. 亦即 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=W$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent. 現若 $W=V$, 當然 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 就是 V 的一組 basis. 而若 $W\neq V$, 表示 $\mathbf{u}_{k+1}\notin W$ (否則會造成 $V=\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_k,\mathbf{u}_{k+1})\subseteq W$ 之矛

盾). 因此由 $\mathbf{u}_{k+1} \notin W = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 以及 Lemma 6.2.6 (2) 知 $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}_{k+1}$ 為 linearly independent. 又因為 $V = \mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1})$ 而 $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k) = W = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 故得 $V = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}_{k+1})$. 得證 $\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{u}_{k+1}$ 為 V 的一組 basis.

對於組成一組 basis 的元素個數是唯一的證明, 由於 Lemma 4.2.5 以及 Theorem 4.3.2 的證明只需用到 vector space 的假設, 並不需用到 $\mathbb{R}^m$ 的假設. 因此可以直接套用其證明, 我們就不再重複了. □

Theorem 6.2.9 告訴我們組成 V 的一組 basis 的元素個數是固定的. 也就是說若找到 n 個元素形成 V 的 basis, 則 V 其他的 basis 一定也會是由 n 個元素所組成. 由於這個重要的結果我們有以下的定義.

Definition 6.2.10. 假設 V 是一個 finitely generated 的 vector space. 組成 V 的一組 basis 的元素個數稱為 V 的 *dimension* (維度), 用 $\dim(V)$ 來表示.

由於組成 finitely generated vector space 的一組 basis 的元素個數是有限的, 所以一般我們也稱 finitely generated vector space 為 *finite dimensional vector space.*

Example 6.2.11. 我們探討在 Example 6.2.3 中的 finite dimensional vector space 的維度為多少.

(A) 在 Example 6.2.3 (A) 我們知道 $\{M_{i,j} \in \mathscr{M}_{m\times n} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ 是 $\mathscr{M}_{m\times n}$ 的 spanning set. 很容易檢查它們為 linearly independent, 所以 $\{M_{i,j} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ 是 $\mathscr{M}_{m\times n}$ 的一組 basis. 因此 $\dim(\mathscr{M}_{m\times n}) = m \times n$.

(B) 我們知道 $\{x^n, \ldots, x, 1\}$ 是 $P_n(\mathbb{R})$ 的 spanning set 且為 linearly independent. 故知 $x^n, \ldots, x, 1$ 為 $P_n(\mathbb{R})$ 的一組 basis, 因此 $\dim(P_n(\mathbb{R})) = n+1$.

對於 finite dimensional vector space 有關於 dimension 的性質, 我們匯集如下. 相關的證明請參考 Proposition 4.3.4, Corollary 4.3.6.

Proposition 6.2.12. 假設 V 為 *finite dimensional vector space.*

(1) 若 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 V 的 *spanning set,* 則 $\dim(V) \leq n$. 特別的, 若此時 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 *linearly dependent,* 則 $\dim(V) < n$.

(2) 若 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 為 *linearly independent,* 則 $\dim(V) \geq n$. 特別的, 若此時 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 不是 V 的 *spanning set,* 則 $\dim(V) > n$.

(3) 若 W 為 V 的 *subspace,* 則 $\dim(W) \leq \dim(V)$. 此時 $\dim(W) = \dim(V)$ 若且唯若 $V = W$.

(4) 假設 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$. 下列的敘述為等價.
  (a) $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的一組 *basis.*
  (b) $\dim(V) = n$ 且 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 V 的 *spanning set.*
  (c) $\dim(V) = n$ 且 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 *linearly independent.*

強調一下, Proposition 6.2.12 (4) 告訴我們當要檢查 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是否為 V 的一組 basis 時, 若我們知道 $\dim(V)$ 恰好是 n, 則僅要檢查它們是否為 spanning set 或 linearly independent 其中一項就可.

Example 6.2.13. 在 Example 6.2.5 中給定 a,b,c 三相異實數, 我們找到三個二次多項式

$$p_1(x)=\frac{(x-b)(x-c)}{(a-b)(a-c)},\quad p_2(x)=\frac{(x-a)(x-c)}{(b-a)(b-c)}\quad\text{and}\quad p_3(x)=\frac{(x-a)(x-b)}{(c-a)(c-b)}.$$

滿足

$$p_1(a)=1, p_1(b)=p_1(c)=0;\quad p_2(b)=1, p_2(a)=p_2(c)=0;\quad p_3(c)=1, p_3(a)=p_3(b)=0.$$

我們知道 $p_1(x),p_2(x),p_3(x)\in P_2(\mathbb{R})$ 為 linearly independent 且 $\dim(P_2(\mathbb{R}))=3$, 故由 Proposition 6.2.12 (4) 知 $p_1(x),p_2(x),p_3(x)$ 為 $P_2(\mathbb{R})$ 的一組 basis.

同理對於任意 n 個相異實數 $a_1,\ldots,a_n$, 我們有 n 個 $n-1$ 次的多項式 $p_1(x),\ldots,p_n(x)$ 滿足 $p_i(a_i)=1$ 且當 $j\neq i$ 時, $p_i(a_j)=0$. 由於 $p_1(x),\ldots,p_n(x)\in P_{n-1}(\mathbb{R})$ 且為 linearly independent, 故由 $\dim(P_{n-1}(\mathbb{R}))=n$ 知 $p_1(x),\ldots,p_n(x)$ 為 $P_{n-1}(\mathbb{R})$ 的一組 basis.

6.3. Linear Transformation

我們將介紹在一般 vector space 之間的 linear transformation 及其相關性質. 如同在 $\mathbb{R}^n$ 的情形, 要探討兩個 vector spaces 之間的關係, 我們需要的函數是能保持 vector space 的加法及係數積運算的, 所以有以下之定義.

Definition 6.3.1. 假設 V,W 為 vector spaces, $T:V\to W$ 為一個函數. 若 T 滿足對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in V$ 以及 $c_1,\ldots,c_k\in\mathbb{R}$ 皆有

$$T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k)=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k).$$

則稱 T 為一個 *linear transformation*. 有時我們簡稱 T 為 *linear*.

要注意這裡 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_1$ 是 V 中的 linear combination, 而 $c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)$ 是 W 中的 linear combination, 要區分清楚. 特別是當 $\mathbf{O}\in V$ 為 V 的 zero element 時, 依 linear transformation 的定義, 我們有 $T(\mathbf{O})=T(\mathbf{O}+\mathbf{O})=T(\mathbf{O})+T(\mathbf{O})$. 此時兩邊加上 $T(\mathbf{O})$ 的加法反元素, 得 $T(\mathbf{O})$ 應為 W 中的 zero element. 也就是說一個 linear transformation $T:V\to W$, 會將 V 中的 zero element 映射到 W 中的 zero element. 雖然當 $V\neq W$ 時, 它們的 zero element 是不同的, 不過一般我們都還是用 $\mathbf{O}$ 來表示, 而不區分它. 所以我們仍用 $T(\mathbf{O})=\mathbf{O}$ 來表示.

依定義要檢查 $T:V\to W$ 是否為 linear transformation, 我們必須考慮 V 中任意的 linear combination 代入 T 中是否符合 linear transformation 的要求. 不過就如同 $\mathbb{R}^n$ 的情況 (參見 Proposition 5.1.3), 我們只要檢查任兩個元素的 linear combination 即可.

Proposition 6.3.2. 假設 V,W 為 *vector spaces* 且 $T:V\to W$ 為一個函數. 則 T 為 *linear transformation* 若且唯若對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{R}$ 皆有 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$.

當 T_1,T_2 皆為 V 到 W 的 linear transformation 時, 我們可以定義 T_1,T_2 之間的加法 $T_1+T_2: V \to W$, 為 $(T_1+T_2)(\mathbf{v}) = T_1(\mathbf{v})+T_2(\mathbf{v}), \forall \mathbf{v} \in V$. 我們也可定義 linear transformation $T: V \to W$ 的係數積. 即對於 $r \in \mathbb{R}$, 定義 $rT: V \to W$ 為 $(rT)(\mathbf{v}) = r(T(\mathbf{v})), \forall \mathbf{v} \in V$. 在此定義之下 T_1+T_2 和 rT 仍為 V 到 W 的 linear transformation (參見 Proposition 5.1.6). 事實上這樣的加法及係數積運算會符合 vector space 的 8 項運算規則, 其中 (3) 的 additive identity 就是 V 到 W 的 zero mapping (就是所有 V 中的元素映射至 W 的 $\mathbf{O}$). 所以我們有以下的結論 (證明從略).

Proposition 6.3.3. 假設 V,W 為 *vector spaces,* 令 $\mathscr{L}(V,W)$ 為所有 V 到 W 的 *linear transformation* 所成的集合. 則利用函數的加法及係數積的定義, $\mathscr{L}(V,W)$ 是一個 *vector space.*

另一個產生 linear transformation 的方法就是利用“合成函數”. 假設 U,V,W 為 vector spaces, 若 $T: V \to W$ 和 $T': W \to U$ 為 linear transformations, 我們定義 $T' \circ T: V \to U$ 為 $T' \circ T(\mathbf{v}) = T'(T(\mathbf{v})), \forall \mathbf{v} \in V$. 我們有下面之結果 (證明請參見 Proposition 5.1.7).

Proposition 6.3.4. 假設 U,V,W 為 *vector spaces.* 若 $T: V \to W$ 和 $T': W \to U$ 為 *linear transformations,* 則 $T' \circ T: V \to U$ 亦為 *linear transformation.*

對於 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation 會有 standard matrix representation 一個主要的關鍵是 linear transformation 可以由定義域的一組 basis 唯一確定. 這對於一般的 linear transformation 也是對的 (證明請參見 Theorem 5.1.8).

Theorem 6.3.5. 假設 V,W 為 *vector spaces* 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$, 為 V 的一組 *basis.* 任取 $\mathbf{w}_1,\ldots,\mathbf{w}_n \in W$, 則存在唯一的 *linear transformation* $T: V \to W$, 滿足對於任意 $i = 1,\ldots,n$ 皆有 $T(\mathbf{v}_i) = w_i$.

由於 linear transformation 保持加法及係數積的特性, 它會保持 subspace 之間的關係. 也就是說當 $T: V \to W$ 為 linear transformation 時, 對於 V 的 subspace V' 我們考慮

$$T(V') = \{T(\mathbf{v}) \in W \mid \mathbf{v} \in V'\} = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}), \text{ for some } \mathbf{v} \in V'\}.$$

同樣的若 W' 為 W 的 subspace, 我們也考慮

$$T^{-1}(W') = \{\mathbf{v} \in V \mid T(\mathbf{v}) \in W'\}.$$

我們會有以下之結果 (證明請參見 Proposition 5.2.1).

Proposition 6.3.6. 假設 V,W 為 *vector spaces* 且 $T: V \to W$ 為 *linear transformation.* 若 V' 為 V 的 *subspaces,* 則 $T(V')$ 是 W 的 *subspace.* 另外, 若 W' 為 W 的 *subspaces,* 則 $T^{-1}(W')$ 是 V 的 *subspace.*

特別的, 在 $V' = V$ 和 $W' = \{\mathbf{O}\}$ 這兩個特殊情況時, 即

$$T(V) = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\} \quad \text{and} \quad T^{-1}(\{\mathbf{O}\}) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{O}\}$$

這兩個 subspaces, 對我們了解 T 這個 linear transformation 非常有幫助. 如同在 $\mathbb{R}^n$ 的情形我們有以下之定義.

Definition 6.3.7. 假設 V,W 為 vector spaces 且 $T:V\to W$ 為 linear transformation. 我們稱 W 的 subspace $T(V)$ 為 T 的 *range* (有時也稱為 *image*). 我們稱 V 的 subspace $T^{-1}(\{\mathbf{O}\})$ 為 T 的 *kernel*, 通常用 $\ker(T)$ 來表示.

依定義, 一個 linear transformation $T:V\to W$ 是 onto 就表示其 range $T(V)$ 是整個 W. 要如何得知 T 的 range 呢? 我們有以下的結果 (證明請參見 Proposition 5.2.4).

Proposition 6.3.8. 假設 V,W 為 *vector spaces* 且 $T:V\to W$ 為 *linear transformation.* 若 $\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ 為 V 的 *spanning set.* 則

$$T(V)=\mathrm{Span}(T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)).$$

特別的, 我們有 T 為 *onto* 若且唯若 $W=\mathrm{Span}(T(\mathbf{v}_1),\dots,T(\mathbf{v}_n))$.

至於 kernel 就和是否為 one-to-one 有關了. 同樣的我們有和 Proposition 5.2.6 一致的結果.

Proposition 6.3.9. 假設 V,W 為 *vector spaces* 且 $T:V\to W$ 為 *linear transformation.* 則 T 為 *one-to-one* 若且唯若 $\ker(T)=\{\mathbf{O}\}$.

在 Proposition 6.3.8 我們知道了 $T:V\to W$ 是 onto 時會保持 spanning set. 也就是說 T 會將 V 的一組 spanning set 送到 W 的一組 spanning set. 然而怎樣的 linear transformation 會保持 linearly independent 的關係呢? 我們有以下之定理.

Proposition 6.3.10. 假設 V,W 為 *vector spaces* 且假設 $T:V\to W$ 為 *one-to-one linear transformation.* 則 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 *linearly independent* 若且唯若 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 *linearly independent.*

Proof. ($\Rightarrow$) 假設 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 linearly independent, 我們希望說明 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 linearly independent. 用反證法, 假設 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 linearly dependent, 亦即存在 $c_1,\dots,c_n\in\mathbb{R}$ 不全為 0 使得 $c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=\mathbf{O}$. 現因 T 為 linear, 我們有 $\mathbf{O}=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)$. 因此利用 T 為 one-to-one, 知僅有在 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$ 時才有可能使得 $T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=\mathbf{O}$. 故由 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 linearly independent 得知 $c_1=\cdots=c_n=0$. 此和 $c_1,\dots,c_n\in\mathbb{R}$ 不全為 0 的假設相矛盾, 故知 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 linearly independent.

($\Leftarrow$) 這個方向的證明不需要 T 為 one-to-one, 僅需要 T 為 linear transformation. 現假設 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 linearly independent. 我們用反證法假設 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 linearly dependent, 亦即存在 $c_1,\dots,c_n\in\mathbb{R}$ 不全為 0 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$. 現因 T 為 linear, 我們有 $c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=T(\mathbf{O})=\mathbf{O}$. 但由於 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 linearly independent, 我們得到 $c_1=\cdots=c_n=0$ 之矛盾. 故得證 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 linearly independent. $\square$

當 $T:V\to W$ 是 one-to-one 且 onto 時, 我們知道它是 invertible, 亦即存在 $T^{-1}:W\to V$, 滿足對所有 $\mathbf{v}\in V$ 皆有 $T^{-1}\circ T(\mathbf{v})=T^{-1}(T(\mathbf{v}))=\mathbf{v}$ 以及對任意 $\mathbf{w}\in W$ 皆有 $T\circ T^{-1}(\mathbf{w})=T(T^{-1}(\mathbf{w}))=\mathbf{w}$. 一般來說我們稱 T^{-1} 為 T 的 *inverse*. 在 Theorem 5.3.12 中我們知道 $\mathbb{R}^n$ 上的 linear transformation 如果是 invertible, 則其 inverse 亦為 linear transformation. 這在一般的情形也對, 不過由於在 Theorem 5.3.12 我們是用 standard matrix representation 來證明. 這裡我們尚未知道有沒有 matrix representation, 所以必須有另外的證明.

Theorem 6.3.11. *假設 V,W 為 vector spaces 且 $T:V\to W$ 為 linear transformation. 若 T 為 invertible, 則 T 的 inverse $T^{-1}:W\to V$ 亦為 linear transformation.*

Proof. 對於任意 $\mathbf{w}\in W$, 由於 T 是 one-to-one 且 onto, 故存在唯一的 $\mathbf{v}\in V$ 滿足 $T(\mathbf{v})=\mathbf{w}$. 依反函數定義此時 $T^{-1}(\mathbf{w})=\mathbf{v}$. 所以 T^{-1} 確實定出一個從 W 到 V 的函數. 我們要證明 $T^{-1}:W\to V$ 是一個 linear transformation. 也就是說任取 $\mathbf{w}_1,\mathbf{w}_2\in V$, $r\in R$, 我們要證明 $T^{-1}(\mathbf{w}_1+r\mathbf{w}_2)=T^{-1}(\mathbf{w}_1)+rT^{-1}(\mathbf{w}_2)$. 現假設 $T^{-1}(\mathbf{w}_1)=\mathbf{v}_1$, $T^{-1}(\mathbf{w}_2)=\mathbf{v}_2$. 亦即 $T(\mathbf{v}_1)=\mathbf{w}_1$ 且 $T(\mathbf{v}_2)=\mathbf{w}_2$. 要檢查 $T^{-1}(\mathbf{w}_1+r\mathbf{w}_2)$ 是否等於 $T^{-1}(\mathbf{w}_1)+rT^{-1}(\mathbf{w}_2)=\mathbf{v}_1+r\mathbf{v}_2$ 就等同於檢查是否 $T(\mathbf{v}_1+r\mathbf{v}_2)=\mathbf{w}_1+r\mathbf{w}_2$. 然而已知 T 為 linear transformation, 我們有 $T(\mathbf{v}_1+r\mathbf{v}_2)=T(\mathbf{v}_1)+rT(\mathbf{v}_2)=\mathbf{w}_1+r\mathbf{w}_2$. 故得證 $T^{-1}(\mathbf{w}_1+r\mathbf{w}_2)=\mathbf{v}_1+r\mathbf{v}_2=T^{-1}(\mathbf{w}_1)+rT^{-1}(\mathbf{w}_2)$, 亦即 T^{-1} 是一個 linear transformation. □

當 $T:V\to W$ 是 invertible 的 linear transformation 時, 我們也時常稱 T 為一個 *isomorphism*. 意思是此時 V 和 W 看成 vector space 時有相似的結構而 T 就是保持這個結構的函數. 事實上由 T 為 onto, 我們知道 T 會保持 V 和 W 的 spanning set, 而由 T 為 one-to-one 我們知道 T 會保持 linearly independent 的關係, 所以我們有以下的結果.

Theorem 6.3.12. *設 V,W 為 vector spaces 且 $T:V\to W$ 為 isomorphism. 則 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 V 的一組 basis 若且唯若 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)$ 為 W 的一組 basis.*

Proof. ($\Rightarrow$) 假設 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 V 的一組 basis. 因為 T 為 onto 且 $\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ 為 V 的 spanning set, 利用 Proposition 6.3.8 知 $\{T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\}$ 為 W 的 spanning set. 又因為 T 為 one-to-one 且 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 linearly independent, 利用 Proposition 6.3.10 知 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)\in W$ 為 linearly independent. 故得 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)$ 為 W 的一組 basis.

($\Leftarrow$) 假設 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)$ 為 W 的一組 basis. 由於 $T^{-1}:W\to V$ 為 linear transformation (Theorem 6.3.11), 且 $T:V\to W$ 為其 inverse, 故知 T^{-1} 為 one-to-one 且 onto, 也就是說 $T^{-1}:W\to V$ 為 isomorphism. 故由 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)$ 為 W 的一組 basis, 套用前面所證可得 $T^{-1}(T(\mathbf{v}_1)),\dots,T^{-1}(T(\mathbf{v}_n))$ 為 V 的一組 basis. 由於對於 $i=1,\dots,n$ 皆有 $T^{-1}(T(\mathbf{v}_i))=\mathbf{v}_i$, 故知 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 V 的一組 basis. □

當 V,W 為 finite dimensional vector space 時, 由 Theorem 6.3.12, 我們知若 $T:V\to W$ 為 isomorphism, 則 V 和 W 的 basis 有相同個數的組成元素, 也就是說 $\dim(V)=\dim(W)$. 換句話說若 $\dim(V)\neq\dim(W)$, 則不可能存在一個 V 到 W 的 isomorphism. 這個性質的反向也是對的, 我們有以下的定理.

Corollary 6.3.13. 設 V,W 為 *finite dimensional vector spaces.* 則存在 $T:V\to W$ 為 *isomorphism* 若且唯若 $\dim(V)=\dim(W)$.

Proof. ($\Rightarrow$) 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis, 即 $\dim(V)=n$. 此時若 $T:V\to W$ 為 isomorphism 由 Theorem 6.3.12 知 $T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)$ 為 W 的一組 basis. 故得 $\dim(W)=n=\dim(V)$.

($\Leftarrow$) 假設 $\dim(V)=\dim(W)$, 且令 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 和 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 分別為 V 的一組 basis 和 W 的一組 basis. 考慮一 linear transformation $T:V\to W$ 滿足 $T(\mathbf{v}_i)=\mathbf{w}_i$, $\forall i=1,\ldots,n$ (參見 Theorem 6.3.5). 我們要說明 T 為 isomorphism. 首先由 6.3.8 知 $T(V)=\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n))=\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)$. 又由 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis 知 $\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)=W$. 故得證 $T(V)=W$, 即 T 為 onto. 現假設 $\mathbf{v}\in\ker(T)$. 由於 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 V 的一組 basis, 故存在 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 依假設 $T(\mathbf{v})=\mathbf{O}$, 亦即

$$\mathbf{O}=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n.$$

然而因 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 linearly independent, 知 $c_1=\cdots=c_n=0$. 故得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{O}$. 我們證得 $\ker(T)=\{\mathbf{O}\}$, 故由 Proposition 6.3.9 知 T 為 one-to-one, 得證 T 為 isomorphism. □

6.4. Coordinatization

這一節中, 我們將介紹一種很重要的 linear transformation, 就是將一個 vector space 裡的元素坐標化. 利用坐標化我們可以將抽象的 vector space 的問題, 化成具體的 $\mathbb{R}^n$ 空間的問題處理.

假設 V 是 finite dimensional vector space, 選定 V 的一組 basis, 我們可以將此組 basis 裡的元素排序, 並固定這個順序不變, 那麼這樣的一組有順序的 basis, 我們稱之為 *ordered basis.* 這裡要特別強調, 即使 basis 裡的元素相同但排序不同, 我們也視為相異的 ordered basis. 所以一般在談論 ordered basis 時, 我們會用 $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 來表示, 以強調其順序. 舉例來說 $\left(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right)$ 和 $\left(\begin{bmatrix}0\\1\end{bmatrix},\begin{bmatrix}1\\0\end{bmatrix}\right)$ 就是 $\mathbb{R}^2$ 中兩組不同的 ordered basis.

有時為了方便起見, 給了一組 ordered basis 後, 我們會用一個符號來表示這一組 ordered basis. 例如給定 $\mathscr{B}=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 為 V 的一組 ordered basis, 我們就會 $\mathscr{B}$ 來表示這一組 ordered basis $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$. 對於 $\mathbb{R}^n$ 的 standard basis, 我們通常用 $\mathscr{E}$ 來表示 $(\mathbf{e}_1,\ldots,\mathbf{e}_n)$ 這一組 ordered basis.

有了 vector space V 的一組 ordered basis $\mathscr{B}=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 後, 我們就可以將 V 中的元素 "坐標化" (*coordinatization*). 意思就是對任意 $\mathbf{v}\in V$, 我們利用 $\mathscr{B}$ 這一組 ordered basis 將 $\mathbf{v}$ 寫成 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 後, $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}$ 就是利用 $\mathscr{B}$ 將 $\mathbf{v}$ 坐標化後所得的坐標表示法. 為了方便, 我們就用 ${}_{\mathscr{B}}[\mathbf{v}]$ 來表示利用 $\mathscr{B}$ 將 $\mathbf{v}$ 坐標化後所得的坐標. 坐標化的好處是, 我們可

以將 ${}_{\mathscr{B}}[\mathbf{v}]$ 看成是 $\mathbb{R}^n$ 中的一個向量. 這樣我們就可以將較抽象的 vector space 中的元素, 看成是 $\mathbb{R}^n$ 中的向量來處理.

Example 6.4.1. 我們看看前面提過的幾個 vector space 坐標化的情形.

(A) 考慮 $\mathscr{M}_{2\times 2}$ 及其 ordered basis

$$\mathscr{E} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

(通常我們稱這一組 basis 為 $\mathscr{M}_{2\times 2}$ 的 standard basis). 對於任意 $\mathscr{M}_{2\times 2}$ 中的元素 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 由於

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

我們得 $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 利用 $\mathscr{E}$ 所得的坐標表示法為

$${}_{\mathscr{E}}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}.$$

例如我們有

$${}_{\mathscr{E}}\begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ -3 \\ 4 \end{bmatrix}.$$

(B) 在 $P_2(\mathbb{R})$ 中通常我們會稱 $x^2, x, 1$ 這組 basis 為 standard basis. 考慮 $\mathscr{E} = (x^2, x, 1)$ 這組 ordered basis. 很容易看出 $2x^2 - 3x + 4$ 用 $\mathscr{E}$ 所得的坐標為 $\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}$, 所以我們有

$${}_{\mathscr{E}}[2x^2 - 3x + 4] = \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}.$$

我們也可考慮 ordered basis $\mathscr{B} = (p_1(x), p_2(x), p_3(x))$ 其中

$$p_1(x) = -(x-1)(x+1), \quad p_2(x) = (1/2)x(x+1) \quad \text{and} \quad p_3(x) = (1/2)x(x-1)$$

(參見 Example 6.2.5, Example 6.2.13). 由於

$$p_1(0) = 1, p_1(1) = p_1(-1) = 0; p_2(1) = 1, p_2(0) = p_2(-1) = 0; p_3(-1) = 1, p_3(0) = p_3(1) = 0,$$

若 $2x^2 - 3x + 4 = c_1p_1(x) + c_2p_2(x) + c_3p_3(x)$, 則分別代 $x = 0, 1, -1$, 可得 $c_1 = 4, c_2 = 3, c_3 = 9$. 故

$${}_{\mathscr{B}}[2x^2 - 3x + 4] = \begin{bmatrix} 4 \\ 3 \\ 9 \end{bmatrix}.$$

(C) 我們也可將 $\mathbb{R}^n$ 中的向量用不同的 ordered basis 坐標化. 例如在 $\mathbb{R}^3$ 中考慮 ordered basis $\mathscr{B} = (\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\end{bmatrix}, \begin{bmatrix}1\\0\\1\end{bmatrix})$. 若要求向量 $\begin{bmatrix}1\\2\\3\end{bmatrix}$ 以 $\mathscr{B}$ 為 ordered basis 的坐標表示, 我們要求出 $c_1, c_2, c_3 \in \mathbb{R}$ 滿足

$$\begin{bmatrix}1\\2\\3\end{bmatrix} = c_1\begin{bmatrix}1\\1\\1\end{bmatrix} + c_2\begin{bmatrix}0\\1\\1\end{bmatrix} + c_3\begin{bmatrix}1\\0\\1\end{bmatrix} = \begin{bmatrix}1 & 0 & 1\\1 & 1 & 0\\1 & 1 & 1\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix}.$$

解聯立方程組得, $c_1 = 0, c_2 = 2, c_3 = 1$, 故得

$${}_{\mathscr{B}}\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}0\\2\\1\end{bmatrix}.$$

要注意這裡我們有 ${}_{\mathscr{E}}\begin{bmatrix}1\\2\\3\end{bmatrix} = \begin{bmatrix}1\\2\\3\end{bmatrix}$, 其中 $\mathscr{E}$ 為 $\mathbb{R}^3$ 的 standard ordered basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. 這是因為我們原來就是用 standard ordered basis $\mathscr{E}$ 來將所有 $\mathbb{R}^3$ 的向量的坐標化.

給定 V 的一組 ordered basis $\mathscr{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$, 將 V 中的元素利用 $\mathscr{B}$ 坐標化, 其實就定出了一個從 V 到 $\mathbb{R}^n$ 的函數 $T_{\mathscr{B}} : V \to \mathbb{R}^n$, 其中對任意 $\mathbf{v} \in V$, 我們有 $T_{\mathscr{B}}(\mathbf{v}) = {}_{\mathscr{B}}[\mathbf{v}]$. 為什麼這是一個函數呢? 因為 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 V 的一組 basis, 所以由 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 是 V 的 spanning set, 可得任意的 $\mathbf{v} \in V$ 確實都存在 $c_1, \ldots, c_n \in \mathbb{R}$ 使得 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. 所以 $T_{\mathscr{B}}$ 確實可以將每個定義域中的元素 $\mathbf{v}$ 對應到對應域 $\mathbb{R}^n$ 中的向量 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}$. 而且這個對應關係是 well-defined, 也就是說不會有將同一個 $\mathbf{v}$ 對應到 $\mathbb{R}^n$ 中兩個不同向量的情況. 這是因為 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly independent, 所以每個 $\mathbf{v} \in V$, 僅有一組 $c_1, \ldots, c_n \in \mathbb{R}$ 會使得 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$.

既然 $T_{\mathscr{B}} : V \to \mathbb{R}^n$ 是一個 well-defined 的函數, 那它會是 linear transformation 嗎? 答案是肯定的. 考慮 $\mathbf{v}, \mathbf{w} \in V$ 且假設 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, $\mathbf{w} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$, 其中 $c_1, \ldots, c_n$ 與 $d_1, \ldots, d_n$ 皆為實數. 依定義我們有

$$T_{\mathscr{B}}(\mathbf{v}) = {}_{\mathscr{B}}[\mathbf{v}] = \begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix} \quad \text{and} \quad T_{\mathscr{B}}(\mathbf{w}) = {}_{\mathscr{B}}[\mathbf{w}] = \begin{bmatrix}d_1\\\vdots\\d_n\end{bmatrix}.$$

對於任意 $r \in \mathbb{R}$, 由於

$$\mathbf{v} + r\mathbf{w} = (c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) + r(d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n) = (c_1 + rd_1)\mathbf{v}_1 + \cdots + (c_n + rd_n)\mathbf{v}_n,$$

我們有

$$T_{\mathscr{B}}(\mathbf{v} + r\mathbf{w}) = {}_{\mathscr{B}}[\mathbf{v} + r\mathbf{w}] = \begin{bmatrix}c_1 + rd_1\\\vdots\\c_n + rd_n\end{bmatrix} = \begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix} + r\begin{bmatrix}d_1\\\vdots\\d_n\end{bmatrix} = T_{\mathscr{B}}(\mathbf{v}) + rT_{\mathscr{B}}(\mathbf{w}).$$

得證 $T_{\mathscr{B}} : V \to \mathbb{R}^n$ 為 linear transformation.

$T_{\mathscr{B}}: V \to \mathbb{R}^n$ 不只是 linear transformation, 事實上它是 V 和 $\mathbb{R}^n$ 之間的 isomorphism. 也就是說 $T_{\mathscr{B}}: V \to \mathbb{R}^n$ 是 one-to-one 且 onto. 要檢查 $T_{\mathscr{B}}$ 為 one-to-one, 我們僅要檢查 $\ker(T_{\mathscr{B}}) = \{\mathbf{O}\}$ 即可 (參見 Proposition 6.3.9). 然而若 $\mathbf{v} \in \ker(T_{\mathscr{B}})$, 表示 $\mathbf{v}$ 用 $\mathscr{B}$ 的坐標表示法是 $\mathbb{R}^n$ 中的零向量, 亦即 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 其中 $c_1 = \cdots = c_n = 0$. 很自然的, 這表示 $\mathbf{v} = \mathbf{O}$, 故知 $\ker(T_{\mathscr{B}}) = \{\mathbf{O}\}$. 要檢查 $T_{\mathscr{B}}$ 為 onto, 我們可以利用 $T_{\mathscr{B}}(\mathbf{v}_i) = \mathbf{e}_i$, $\forall i = 1, \ldots, n$, 故得證 $T_{\mathscr{B}}(V) = \mathrm{Span}(T_{\mathscr{B}}(\mathbf{v}_1), \ldots, T_{\mathscr{B}}(\mathbf{v}_n)) = \mathrm{Span}(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \mathbb{R}^n$ (參見 Proposition 6.3.8), 即 $T_{\mathscr{B}}$ 為 onto. 我們證得了以下的定理.

Theorem 6.4.2. 假設 V 為 *vector space,* $\dim(V) = n$ 且 $\mathscr{B}$ 為 V 的一組 *ordered basis.* 考慮 $T_{\mathscr{B}}: V \to \mathbb{R}^n$ 定義為 $T_{\mathscr{B}}(\mathbf{v}) = {}_{\mathscr{B}}[\mathbf{v}]$, $\forall \mathbf{v} \in V$. 則 $T_{\mathscr{B}}$ 為 V 到 $\mathbb{R}^n$ 的 *isomorphism.*

知道 $T_{\mathscr{B}}: V \to \mathbb{R}^n$ 為 isomorphism 的好處就是, 以後我們要探討 V 中元素的性質, 我們可以利用 $T_{\mathscr{B}}$, 將問題轉換成大家熟悉的 $\mathbb{R}^n$ 中的向量的性質. 例如我們要判斷 V 中的元素 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 是否為 linearly independent. 我們可以先找到一組 V 的 ordered basis $\mathscr{B}$, 然後考慮 ${}_{\mathscr{B}}[\mathbf{w}_1], \ldots, {}_{\mathscr{B}}[\mathbf{w}_k]$, 這一組 $\mathbb{R}^n$ 中的向量. 利用我們熟悉的判斷 $\mathbb{R}^n$ 中向量是否為 linearly independent 的方法判斷 ${}_{\mathscr{B}}[\mathbf{w}_1], \ldots, {}_{\mathscr{B}}[\mathbf{w}_k]$ 是否為 linearly independent. 由於對於 $i = 1, \ldots, k$, ${}_{\mathscr{B}}[\mathbf{w}_i] = T_{\mathscr{B}}(\mathbf{w}_i)$, 因此由 $T_{\mathscr{B}}$ 為 isomorphism 知 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 為 linearly independent 若且唯若 ${}_{\mathscr{B}}[\mathbf{w}_1], \ldots, {}_{\mathscr{B}}[\mathbf{w}_k]$ 為 linearly independent (參見 Proposition 6.3.10). 因此我們可以由 ${}_{\mathscr{B}}[\mathbf{w}_1], \ldots, {}_{\mathscr{B}}[\mathbf{w}_k]$ 是否為 linearly independent, 來決定 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 是否為 linearly independent. 我們看以下的例子.

Example 6.4.3. 我們看利用坐標化來處理一般 vector space 是否 linear independent 的問題.

(A) 考慮 $P_2(\mathbb{R})$ 中 3 個非零多項式 $f_2(x), f_1(x), f_0(x)$, 其 $f_2(x), f_1(x), f_0(x)$ 分別為次數為 2, 1, 0 的多項式. 假設 $f_2(x) = ax^2 + bx + c$, $f_1(x) = dx + e$, $f_0(x) = r$ 其中 a, d, r 皆不等於 0. 我們利用 $P_2(\mathbb{R})$ 的 standard ordered basis $\mathscr{E} = (x^2, x, 1)$, 可得

$$
{}_{\mathscr{E}}[f_2(x)] = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad {}_{\mathscr{E}}[f_1(x)] = \begin{bmatrix} 0 \\ d \\ e \end{bmatrix} \quad \text{and} \quad {}_{\mathscr{E}}[f_0(x)] = \begin{bmatrix} 0 \\ 0 \\ r \end{bmatrix}.
$$

由於 a, d, r 皆不等於 0, 很容易看出矩陣

$$
\begin{bmatrix} a & 0 & 0 \\ b & d & 0 \\ c & e & r \end{bmatrix}
$$

的 rank 為 3, 亦即 ${}_{\mathscr{E}}[f_2(x)], {}_{\mathscr{E}}[f_1(x)], {}_{\mathscr{E}}[f_0(x)]$ 為 linearly independet. 因此得證 $f_2(x)$, $f_1(x)$, $f_0(x)$ 為 linearly independent. 再由 $\dim(P_2(\mathbb{R})) = 3$, 得證 $f_0(x), f_1(x), f_2(x)$ 為 $P_2(\mathbb{R})$ 的一組 basis.

我們可以將這個結果推廣到 $P_n(\mathbb{R})$. 也就是說考慮 $P_n(\mathbb{R})$ 中 $n+1$ 個非零多項式 $f_0(x), \ldots, f_n(x)$, 其中對於 $i = 0, \ldots, n$, $f_i(x)$ 是次數為 i 的多項式. 利用對 standard ordered basis $\mathscr{E} = (x^n, \ldots, x, 1)$ 坐標化, 我們可得 $f_0(x), \ldots, f_n(x)$ 為 $P_n(\mathbb{R})$ 的一組 basis.

(B) 假設 V 為 vector space, 且 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4 \in V$ 為 linearly independent. 令

$$\begin{array}{rrrrr} \mathbf{w}_1 = & \mathbf{v}_1 & -\mathbf{v}_2 & +2\mathbf{v}_3 & +\mathbf{v}_4 \\ \mathbf{w}_2 = & 2\mathbf{v}_1 & -2\mathbf{v}_2 & +4\mathbf{v}_3 & +2\mathbf{v}_4 \\ \mathbf{w}_3 = & \mathbf{v}_1 & & +\mathbf{v}_3 & +\mathbf{v}_4 \\ \mathbf{w}_4 = & & \mathbf{v}_2 & & +\mathbf{v}_4 \\ \mathbf{w}_5 = & -\mathbf{v}_1 & +\mathbf{v}_2 & -\mathbf{v}_3 & \end{array}.$$

我們要找出 $W = \mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,\mathbf{w}_4,\mathbf{w}_5)$ 的一組 basis.

考慮 $U = \mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4)$, 因為 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4$ 為 linearly independent, 我們知 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4$ 為 U 的一組 basis. 由於 $\mathbf{w}_1,\dots,\mathbf{w}_5 \in U$, 我們知 W 為 U 的 subspace. 我們的想法是利用 $\mathscr{B} = (\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4)$ 這組 U 的 ordered basis 將 U 的元素坐標化. 由於 $\mathbf{w}_1,\dots,\mathbf{w}_5 \in U$, 我們可以將 $\mathbf{w}_1,\dots,\mathbf{w}_5$ 坐標化, 得 ${}_{\mathscr{B}}[\mathbf{w}_1],\dots,{}_{\mathscr{B}}[\mathbf{w}_5]$ 這 5 個 $\mathbb{R}^4$ 中的向量. 利用過去我們知道求 $\mathbb{R}^4$ 中 $\mathrm{Span}({}_{\mathscr{B}}[\mathbf{w}_1],\dots,{}_{\mathscr{B}}[\mathbf{w}_5])$ 的 basis 的方法求出 $\mathrm{Span}({}_{\mathscr{B}}[\mathbf{w}_1],\dots,{}_{\mathscr{B}}[\mathbf{w}_5])$ 的一組 basis. 再將它們還原成 U 中的元素, 就得到 W 的一組 basis.

現由於

$${}_{\mathscr{B}}[\mathbf{w}_1] = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, {}_{\mathscr{B}}[\mathbf{w}_2] = \begin{bmatrix} 2 \\ -2 \\ 4 \\ 2 \end{bmatrix}, {}_{\mathscr{B}}[\mathbf{w}_3] = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, {}_{\mathscr{B}}[\mathbf{w}_4] = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, {}_{\mathscr{B}}[\mathbf{w}_5] = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

我們考慮以它們為 column 的 4×5 matrix 並利用 elementary row operations 將之化為 echelon form 得

$$\begin{bmatrix} 1 & 2 & 1 & 0 & -1 \\ -1 & -2 & 0 & 1 & 1 \\ 2 & 4 & 1 & 0 & -1 \\ 1 & 2 & 1 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

由於 echelon form 的 1-st, 3-rd, 4-th column 為 pivot 所在位置, 故知 ${}_{\mathscr{B}}[\mathbf{w}_1], {}_{\mathscr{B}}[\mathbf{w}_3], {}_{\mathscr{B}}[\mathbf{w}_4]$ 為 $\mathrm{Span}({}_{\mathscr{B}}[\mathbf{w}_1],\dots,{}_{\mathscr{B}}[\mathbf{w}_5])$ 的一組 basis (參見 Proposition 4.4.8). 由於 $T_{\mathscr{B}}$ 為 isomprphism 保持 spanning set 以及 linearly independent 的性質, 我們得 $\mathbf{w}_1,\mathbf{w}_3,\mathbf{w}_4$ 為 W 的一組 basis.

Question 6.4. 考慮 *Example 6.4.3 (B)* 中的 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3,\mathbf{v}_4$ 以及 $\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,\mathbf{w}_4,\mathbf{w}_5$. 試問 $\dim(\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3,\mathbf{w}_4,\mathbf{w}_5))$ 為何? 並將 $\mathbf{w}_2,\mathbf{w}_5$ 寫成 $\mathbf{w}_1,\mathbf{w}_3,\mathbf{w}_4$ 的 *linear combination.*

6.5. Matrix Representation

當 V,W 分別為 dimension 為 n,m 的 vector space. 我們可以透過 V,W 的 ordered basis, 將 V,W 的元素轉換成 $\mathbb{R}^n$ 和 $\mathbb{R}^m$ 的 vector. 因此我們可以將 V 到 W 的 linear transformation T 視為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation, 而談論 T 的 matrix representation.

分別給定 V,W 的一組 ordered basis $\mathscr{B} = (\mathbf{v}_1,\dots,\mathbf{v}_n)$ 以及 $\beta = (\mathbf{w}_1,\dots,\mathbf{w}_m)$. 令 $T_{\mathscr{B}} : V \to \mathbb{R}^n$ 與 $T_\beta : W \to \mathbb{R}^m$ 分別為利用 $\mathscr{B}$ 以及 β 將 V, W 的元素坐標化的 linear transformation. 回顧一下, 這裡 $T_{\mathscr{B}}, T_\beta$ 皆為 isomorphism. 對於任意 V 到 W 的 linear transformation T, 我們考慮合成函數 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1} : \mathbb{R}^n \to \mathbb{R}^m$. 由於 T_β, T 以及 $T_{\mathscr{B}}^{-1}$ 皆為 linear transformation, 所以 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}$ 是一個 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation (Proposition 6.3.4). 因此 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}$

有一個 standard matrix representation, 我們定義這個 standard matrix representation 為 T 相對於 $\mathscr{B}, \beta$ 這兩組 ordered basis 所得的 *matrix representation*, 並用 ${}_{\beta}[T]_{\mathscr{B}}$ 來表示. 到底 ${}_{\beta}[T]_{\mathscr{B}}$ 是怎樣的矩陣呢? 依定義它是一個 $m\times n$ matrix, 且對於 $i=1,\ldots,n$, ${}_{\beta}[T]_{\mathscr{B}}$ 的 i-th column 應為 $T_{\beta}\circ T\circ T_{\mathscr{B}}^{-1}(\mathbf{e}_i)$. 由於 $T_{\mathscr{B}}(\mathbf{v}_i)=\mathbf{e}_i$, 所以 $T_{\mathscr{B}}^{-1}(\mathbf{e}_i)=\mathbf{v}_i$. 因此得

$$T_{\beta}\circ T\circ T_{\mathscr{B}}^{-1}(\mathbf{e}_i)=T_{\beta}(T(T_{\mathscr{B}}^{-1}(\mathbf{e}_i)))=T_{\beta}(T(\mathbf{v}_i))=[T(\mathbf{v}_i)]_{\beta}.$$

也就是說 ${}_{\beta}[T]_{\mathscr{B}}$ 的 i-th column 就是將 ordered basis $\mathscr{B}=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 的第 i 個元素 $\mathbf{v}_i$ 代入 T 中所得的 $T(\mathbf{v}_i)\in W$, 再利用 β 將其坐標化所得的 $\mathbb{R}^m$ 中的向量. 我們大致上有以下的表示法

$${}_{\beta}[T]_{\mathscr{B}}=\begin{bmatrix} | & | & & | \\ [{}_{\beta}T(\mathbf{v}_1)] & {}_{\beta}[T(\mathbf{v}_2)] & \cdots & {}_{\beta}[T(\mathbf{v}_n)] \\ | & | & & | \end{bmatrix}.$$

Example 6.5.1. 考慮 $P_2(\mathbb{R})$ 上的 standard ordered basis $\mathscr{E}=(x^2,x,1)$ 以及 $P_3(\mathbb{R})$ 上的 standard ordered basis $\varepsilon=(x^3,x^2,x,1)$. 考慮函數 $T:P_2(\mathbb{R})\to P_3(\mathbb{R})$ 定義為, 對任意 $p(x)\in P_2(\mathbb{R})$, $T(p(x))=(x+1)p(x-1)$. 我們先驗證 T 為 linear transformation. 對任意 $p(x),q(x)\in P_2(\mathbb{R})$ 以及 $r\in\mathbb{R}$, 我們有

$$\begin{aligned}&T(p(x)+rq(x))=\\&\qquad(x+1)(p(x-1)+rq(x-1))=(x+1)p(x-1)+r(x+1)q(x-1)=T(p(x))+rT(q(x)).\end{aligned}$$

得證 T 為 linear transformation. 接下來我們要求 T 對於 $\mathscr{E},\varepsilon$ 的 matrix representation ${}_{\varepsilon}[T]_{\mathscr{E}}$. 依照前面的探討, 我們知 ${}_{\varepsilon}[T]_{\mathscr{E}}$ 的 1-st column 應該就是將 x^2 代入 T, 得 $T(x^2)=(x+1)(x-1)^2$ 再利用 ε 將 $(x+1)(x-1)^2$ 坐標化寫成 $\mathbb{R}^3$ 的向量. 由於 $(x+1)(x-1)^2=x^3-x^2-x+1$, 所以得 ${}_{\varepsilon}[T]_{\mathscr{E}}$ 的 1-st column 為

$${}_{\varepsilon}[T(x^2)]={}_{\varepsilon}[(x+1)(x-1)^2]={}_{\varepsilon}[x^3-x^2-x+1]=\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}.$$

同理我們有 ${}_{\varepsilon}[T]_{\mathscr{E}}$ 的 2-nd, 3-rd column 分別為

$${}_{\varepsilon}[T(x)]={}_{\varepsilon}[(x+1)(x-1)]={}_{\varepsilon}[x^2-1]={}_{\varepsilon}\begin{bmatrix}0\\1\\0\\-1\end{bmatrix}, {}_{\varepsilon}[T(1)]={}_{\varepsilon}[x+1]=\begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

因此得

$${}_{\varepsilon}[T]_{\mathscr{E}}=\begin{bmatrix}1&0&0\\-1&1&0\\-1&0&1\\1&-1&1\end{bmatrix}.$$

我們更換 $P_2(\mathbb{R})$ 以及 $P_3(\mathbb{R})$ 的 ordered basis. 回顧在 Example 6.2.5 以及 Example 6.2.13 中利用 Lagrange interpolation polynomial 在 $-1,0,1$ 的情形, 我們可以考慮 $P_2(\mathbb{R})$

的一組 basis $p_1(x),p_2(x),p_3(x)$, 其中

$$\begin{array}{lll} p_1(-1)=1 & p_1(0)=0 & p_1(1)=0 \\ p_2(-1)=0 & p_2(0)=1 & p_2(1)=0 \\ p_3(-1)=0 & p_3(0)=0 & p_3(1)=1 \end{array}$$

令 $\mathscr{B}=(p_1(x),p_2(x),p_3(x))$ 為 $P_2(\mathbb{R})$ 的 ordered basis. 同樣的利用 Lagrange interpolation polynomial 在 $-1,0,1,2$ 的情形我們考慮 $P_3(\mathbb{R})$ 的一組 basis $q_1(x),q_2(x),q_3(x),q_4(x)$, 其中

$$\begin{array}{llll} q_1(-1)=1 & q_1(0)=0 & q_1(1)=0 & q_1(2)=0 \\ q_2(-1)=0 & q_2(0)=1 & q_2(1)=0 & q_2(2)=0 \\ q_3(-1)=0 & q_3(0)=0 & q_3(1)=1 & q_3(2)=0 \\ q_4(-1)=0 & q_4(0)=0 & q_4(1)=0 & q_4(2)=1. \end{array}$$

令 $\beta=(q_1(x),q_2(x),q_3(x),q_4(x))$ 為 $P_3(\mathbb{R})$ 的 ordered basis. 我們要得到 T 對於 $\mathscr{B},\beta$ 的 matrix representation ${}_\beta[T]_\mathscr{B}$. 首先 ${}_\beta[T]_\mathscr{B}$ 的 1-st column 為 $T(p_1(x))=(x+1)p_1(x-1)$ 利用 β 坐標化所得 $\mathbb{R}^3$ 的向量. 現若 $(x+1)p_1(x-1)=c_1q_1(x)+c_2q_2(x)+c_3q_3(x)+c_4q_4(x)$. 將 x 分別代 $-1,0,1,2$, 我們得到

$$c_1=(-1+1)p_1(-2)=0, c_2=(0+1)p_1(-1)=1, c_3=(1+1)p_1(0)=0, c_4=(2+1)p_1(1)=0.$$

也就是說 ${}_\beta[T]_\mathscr{B}$ 的 1-st column 為

$${}_\beta[T(p_1(x))]={}_\beta[(x+1)p_1(x-1)]=\begin{bmatrix}0\\1\\0\\0\end{bmatrix}.$$

同樣的方法我們可以得到 ${}_\beta[T]_\mathscr{B}$ 的 2-nd, 3-rd column 分別為

$${}_\beta[T(p_2(x))]={}_\beta[(x+1)p_2(x-1)]=\begin{bmatrix}0\\0\\2\\0\end{bmatrix}, {}_\beta[T(p_3(x))]={}_\beta[(x+1)p_3(x-1)]=\begin{bmatrix}0\\0\\0\\3\end{bmatrix}.$$

因此得

$${}_\beta[T]_\mathscr{B}=\begin{bmatrix}0&0&0\\1&0&0\\0&2&0\\0&0&3\end{bmatrix}.$$

由 Example 6.5.1 我們知道同樣的 linear transformation 用不同的 ordered basis 會有不同的 matrix representation. 也因此要注意當要寫下 matrix representation 時一定要表明定義域和對應域的 ordered basis 為何.

到底 matrix representation 有何用處呢? 就如同在 $\mathbb{R}^n$ 上的 standard matrix representation, 利用 matrix representation 可以很快的幫我們求出 linear transformation 在定義域的每個元素的取值. 通常我們會利用圖示來幫助我們了解較複雜的函數合成問題. 給定 V,W 的 ordered basis $\mathscr{B},\beta$, 以及一個 V 到 W 的 linear transformation T, 我們可以圖示如下:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_{\mathscr{B}}"'] & W \arrow[d, shift left, "T_\beta"] \\ \mathbb{R}^n \arrow[r] \arrow[u, shift right, "T_{\mathscr{B}}^{-1}"'] & \mathbb{R}^m \arrow[u, shift left, "T_\beta^{-1}"] \end{tikzcd}$$

這樣的圖示一般稱為 *commutative diagram*. 它表示底下 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的函數為 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}$, 亦即先利用 $T_{\mathscr{B}}^{-1}$ 將 $\mathbb{R}^n$ 映射到 V, 再利用 T 將 V 映射到 W, 最後利用 T_β 將 W 映射到 $\mathbb{R}^m$. Commutative diagram 的好處是幫助我們看出這些函數合成後如何取值. 事實上 commutative diagram 指的就是圖形上任一點到另一點若有不同的可行路徑, 經由這兩種路徑所得的結果會相同. 例如在上圖中, 從 V 到 W 有兩個路徑: 一個是直接利用 T; 另一個是從 V 先經由 $T_{\mathscr{B}}$ 到 $\mathbb{R}^n$, 接著藉由 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}$ 從 $\mathbb{R}^n$ 到 $\mathbb{R}^m$, 最後從 $\mathbb{R}^m$ 藉由 T_β^{-1} 到達 W. 也就是說對任意 $\mathbf{v} \in V$, 我們可以由 T 得到 $T(\mathbf{v})$. 也可先由 $T_{\mathscr{B}}$ 得到 $T_{\mathscr{B}}(\mathbf{v})$, 接著利用 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}$ 將 $T_{\mathscr{B}}(\mathbf{v})$ 送至 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}(T_{\mathscr{B}}(\mathbf{v})) = T_\beta(T(\mathbf{v}))$, 最後再利用 T_β^{-1} 將 $T_\beta(T(\mathbf{v}))$ 送至 $T_\beta^{-1}(T_\beta(T(\mathbf{v}))) = T(\mathbf{v})$.

利用 $V \to \mathbb{R}^n$ 接 $\mathbb{R}^n \to \mathbb{R}^m$ 再接 $\mathbb{R}^m \to W$ 這樣的路徑來表示原本 T 從 V 到 W 這樣的路徑到底有何好處呢? 主要的原因是 $\mathbb{R}^n \to \mathbb{R}^m$ 這一段的路徑, 有 standard matrix representation, 即 ${}_\beta[T]_{\mathscr{B}}$. 也就是說對於任意 $\mathbb{R}^n$ 的向量, 我們只要左邊乘上 ${}_\beta[T]_{\mathscr{B}}$ 就可以知道會被映射到哪一個 $\mathbb{R}^m$ 的向量. 所以對任意 $\mathbf{v} \in V$, 我們可以先利用 $\mathscr{B}$ 將 $\mathbf{v}$ 坐標化得 $T_{\mathscr{B}}(\mathbf{v}) = {}_{\mathscr{B}}[\mathbf{v}]$. 接著由於 ${}_{\mathscr{B}}[\mathbf{v}] \in \mathbb{R}^n$, 故將之代入 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}$ 就是將 ${}_{\mathscr{B}}[\mathbf{v}]$ 左邊乘上 ${}_\beta[T]_{\mathscr{B}}$ 這一個 matrix. 也就是說 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}({}_{\mathscr{B}}[\mathbf{v}])$ 就是 ${}_\beta[T]_{\mathscr{B}}\, {}_{\mathscr{B}}[\mathbf{v}]$. 最後再利用 W 的 ordered basis β 將 ${}_\beta[T]_{\mathscr{B}}\, {}_{\mathscr{B}}[\mathbf{v}]$ 這個 $\mathbb{R}^m$ 中的向量還原回 W 的元素 $T_\beta^{-1}({}_\beta[T]_{\mathscr{B}}\, {}_{\mathscr{B}}[\mathbf{v}])$, 就是 $T(\mathbf{v})$ 之值. 因此我們有以下之結果.

Proposition 6.5.2. *假設 V,W 為 vector space 且 $\mathscr{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$, $\beta = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ 分別為 V,W 的 ordered basis. 設 $T : V \to W$ 為 linear transformation 且 ${}_\beta[T]_{\mathscr{B}} \in \mathscr{M}_{m\times n}$ 為 T 相對於 $\mathscr{B},\beta$ 的 matrix representation. 對於任意 $\mathbf{v} \in V$, 若 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ 且*

$$ {}_\beta[T]_{\mathscr{B}} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix}, $$

則 $T(\mathbf{v}) = d_1\mathbf{w}_1 + \cdots + d_m\mathbf{w}_m$. 亦即

$$ {}_\beta[T]_{\mathscr{B}}\, {}_{\mathscr{B}}[\mathbf{v}] = {}_\beta[T(\mathbf{v})]. $$

Proof. 因 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 依定義 $\mathbf{v}$ 利用 $\mathscr{B}$ 坐標化所得 $\mathbb{R}^n$ 的向量 $T_{\mathscr{B}}(\mathbf{v})$ 為 $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. 故由前面所述 $T_\beta \circ T \circ T_{\mathscr{B}}^{-1}(T_{\mathscr{B}}(\mathbf{v})) = T_\beta(T(\mathbf{v}))$ 就是將 $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 左邊乘上 ${}_\beta[T]_{\mathscr{B}}$ 所得的 $\mathbb{R}^m$ 中向量 $\begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix}$. 故由 $T_\beta(T(\mathbf{v})) = \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix}$ 可得 $T(\mathbf{v}) = d_1\mathbf{w}_1 + \cdots + d_m\mathbf{w}_m$. $\square$

Example 6.5.3. 我們考慮 Example 6.5.1 的例子, 即考慮 $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$, 其中對於任意 $p(x) \in P_2(\mathbb{R})$, $T(p(x)) = (x+1)p(x-1)$. 考慮 $p(x) = x^2 - 1$ 的情形. 因 $p(x-1) = (x-1)^2 - 1$, 依 T 的定義得

$$T(p(x)) = (x+1)p(x-1) = (x+1)((x-1)^2 - 1) = x^3 - x^2 - 2x.$$

當考慮 $P_2(\mathbb{R})$ 的 ordered basis $\mathscr{E} = (x^2, x, 1)$ 以及 $P_3(\mathbb{R})$ 的 ordered basis $\varepsilon = (x^3, x^2, x, 1)$, 我們知道 T 對於 $\mathscr{E}, \varepsilon$ 的 matrix representation 為 ${}_{\varepsilon}[T]_{\mathscr{E}} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}$ 故由 ${}_{\mathscr{E}}[p(x)] = {}_{\mathscr{E}}[x^2 - 1] = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ 得

$${}_{\varepsilon}[T(p(x))] = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -2 \\ 0 \end{bmatrix},$$

即 $T(p(x)) = x^3 - x^2 - 2x$.

當然我們也可以利用 Example 6.5.1 中 V 的 ordered basis $\mathscr{B} = (p_1(x), p_2(x), p_3(x))$ 以及 W 的 ordered basis $\beta = (q_1(x), q_2(x), q_3(x), q_4(x))$ 求出 $T(x^2 - 1)$. 此時若 $p(x) = x^2 - 1 = c_1 p_1(x) + c_2 p_2(x) + c_3 p_3(x)$, 則代 $x = -1, 0, 1$ 得 $c_1 = 0$, $c_2 = -1$, $c_3 = 0$, 亦即 ${}_{\mathscr{B}}[p(x)] = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$. 故將 ${}_{\mathscr{B}}[p(x)]$ 左邊乘上 T 對於 $\mathscr{B}, \beta$ 的 representation matrix ${}_{\beta}[T]_{\mathscr{B}}$ 得

$${}_{\beta}[T(p(x))] = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -2 \\ 0 \end{bmatrix}.$$

得知 $T(p(x)) = -2q_3(x)$. 由於 $q_3(-1) = q_3(0) = q_3(2) = 0$ 以及 $q_3(1) = 1$, 我們有 $q_3(x) = (x+1)x(x-2)/(-2)$, 故得 $T(p(x)) = (x+1)x(x-2) = x^3 - x^2 - 2x$.

當 $T : \mathbb{R}^n \to \mathbb{R}^m$ 是 linear transformation 時, 我們可以利用 T 的 standard matrix representation $[T]$ 的 nullspace 來決定 T 的 kernel, 也可利用 $[T]$ 的 column space 來決定 T 的 range. 同樣的, 當 $T : V \to W$, 為 linear transformation, 我們也可利用 T 的 matrix representation 來決定 T 的 kernel 和 range. 分別選定 V 和 W 的 ordered basis $\mathscr{B} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 和 $\beta = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$. 由前面的 commutative diagram, 利用 $V \to W$ 接著 $W \to \mathbb{R}^m$ 的路徑以及 $V \to \mathbb{R}^n$ 接著 $\mathbb{R}^n \to \mathbb{R}^m$ 的路徑, 對於任意 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in V$, 我們有

$${}_{\beta}[T(\mathbf{v})] = {}_{\beta}[T]_{\mathscr{B}} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}. \tag{6.2}$$

現若 $\mathbf{v}\in\ker(T)$, 表示 $T(\mathbf{v})=\mathbf{O}$, 因此由 ${}_\beta[T(\mathbf{v})]={}_\beta[\mathbf{O}]=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$, 得 ${}_\beta[T]_{\mathscr{B}}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$, 亦即 $\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}={}_{\mathscr{B}}[\mathbf{v}]$ 為 ${}_\beta[T]_{\mathscr{B}}$ 的 nullspace 的向量. 反之, 若 $\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}\in\mathbb{R}^n$ 為 ${}_\beta[T]_{\mathscr{B}}$ 的 nullspace 的向量, 表示 ${}_\beta[T]_{\mathscr{B}}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$. 故由式子 (6.2) 知, 當 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 時, 我們有 ${}_\beta[T(\mathbf{v})]=\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$. 此即表示 $T(\mathbf{v})=\mathbf{O}$, 得證 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in\ker(T)$.

另一方面, 若 $\mathbf{w}=d_1\mathbf{w}_1+\cdots+d_m\mathbf{w}_m\in T(V)$, 表示存在 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in V$, 使得 $T(\mathbf{v})=\mathbf{w}$. 因此由式子 (6.2) 知, $\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}={}_\beta[T(\mathbf{v})]={}_\beta[T]_{\mathscr{B}}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$, 亦即 $\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}$ 是 ${}_\beta[T]_{\mathscr{B}}$ 的 column space 的向量. 反之, 若 $\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}\in\mathbb{R}^m$ 為 ${}_\beta[T]_{\mathscr{B}}$ 的 column space 的向量, 表示存在 $\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}\in\mathbb{R}^n$ 使得 $\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}={}_\beta[T]_{\mathscr{B}}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$. 故由式子 (6.2) 知, 若令 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 我們有 ${}_\beta[T(\mathbf{v})]=\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}$, 得證 $d_1\mathbf{w}_1+\cdots+d_m\mathbf{v}_m=T(\mathbf{v})\in T(V)$. 我們證得了以下的結果.

Proposition 6.5.4. 假設 V,W 為 *vector space* 且 $\mathscr{B}=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, $\beta=(\mathbf{w}_1,\ldots,\mathbf{w}_m)$ 分別為 V,W 的 *ordered basis*. 設 $T:V\to W$ 為 *linear transformation* 且 ${}_\beta[T]_{\mathscr{B}}\in\mathscr{M}_{m\times n}$ 為 T 相對於 $\mathscr{B},\beta$ 的 *matrix representation*. 則 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in\ker(T)$ 若且唯若 $\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$ 屬於 ${}_\beta[T]_{\mathscr{B}}$ 的 *nullspace*. 而 $d_1\mathbf{w}_1+\cdots+d_m\mathbf{w}_m\in T(V)$ 若且唯若 $\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}$ 屬於 ${}_\beta[T]_{\mathscr{B}}$ 的 *column space*.

分別給定 $\mathscr{B}=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, $\beta=(\mathbf{w}_1,\ldots,\mathbf{w}_m)$ 為 V,W 的 ordered basis, 我們知道 $T_{\mathscr{B}}:V\to\mathbb{R}^n$, $T_\beta:W\to\mathbb{R}^m$ 為 isomorphism, 因此由 Proposition 6.5.4 我們知道 $T_{\mathscr{B}}$ 會將 $\ker(T)$ 的一組 basis 送至 ${}_\beta[T]_{\mathscr{B}}$ 的 nullspace 的一組 basis 且 T_β 會將 T 的 range $T(V)$ 的一組 basis 送至 ${}_\beta[T]_{\mathscr{B}}$ 的 column space 的一組 basis (參見 Theorem 6.3.12). 所以我們有 $\dim(\ker(T))=\mathrm{null}({}_\beta[T]_{\mathscr{B}})$ 以及 $\dim(T(V))=\mathrm{rank}({}_\beta[T]_{\mathscr{B}})$. 因為這個原因, 一般我們也稱 T 的 range 的維度為 T 的 *rank*, 而 T 的 kernel 的維度稱為 T 的 *nullity*. 利用此一結果, 我們推廣了 Theorem 5.3.6.

Theorem 6.5.5 (Dimension Theorem for Linear Transformation)**.** 假設 V,W 為 *finite dimensional vector space* 且 $T:V\to W$ 為 *linear transformation.* 則

$$\dim(T(V))+\dim(\ker(T))=\dim(V).$$

Proof. 考慮 V,W 的 ordered basis $\mathscr{B}=(\mathbf{v}_1,\ldots,\mathbf{v}_n),\beta=(\mathbf{w}_1,\ldots,\mathbf{w}_m)$. 由於 T 對於 $\mathscr{B},\beta$ 的 matrix representation ${}_{\beta}[T]_{\mathscr{B}}$ 為 $m\times n$ matrix, 由 Theorem 4.4.13 我們知 $\operatorname{rank}({}_{\beta}[T]_{\mathscr{B}})+\operatorname{null}({}_{\beta}[T]_{\mathscr{B}})=n$. 故套用前述 $\dim(\ker(T))=\operatorname{null}({}_{\beta}[T]_{\mathscr{B}})$ 以及 $\dim(T(V))=\operatorname{rank}({}_{\beta}[T]_{\mathscr{B}})$, 得證本定理. □

Example 6.5.6. 我們考慮 Example 6.5.1 的例子, 當考慮 $P_2(\mathbb{R})$ 的 ordered basis $\mathscr{E}=(x^2,x,1)$ 以及 $P_3(\mathbb{R})$ 的 ordered basis $\varepsilon=(x^3,x^2,x,1)$, 我們知道 T 對於 $\mathscr{E},\varepsilon$ 的 matrix representation ${}_{\varepsilon}[T]_{\mathscr{E}}$ 利用 elementary row operations 化為 echelon form 可得

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

由於 pivot 的個數等於 column 的個數, 我們知 ${}_{\varepsilon}[T]_{\mathscr{E}}$ 的 nullspace 為 $\{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\}$, 故知 $\ker(T)=\{\mathbf{O}\}$, 亦即 T 為 one-to-one. 另一方面 ${}_{\varepsilon}[T]_{\mathscr{E}}$ 的 rank 為 3, 故 $\{\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix},\begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix},\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}\}$ 為 column space 的一組 basis. 因此得 $\{x^3-x^2-x+1,x^2-1,x+1\}$ 為 $T(P_2(\mathbb{R}))$ 的一組 basis. 由於 $\dim(P_3(\mathbb{R}))=4\neq\dim(T(P_2(\mathbb{R})))=3$, 我們知 $T(P_2(\mathbb{R}))\neq P_3(\mathbb{R})$, 故 T 不是 onto. 最後驗證 Dimension Theorem, 我們有

$$\dim(T(P_2(\mathbb{R})))+\dim(\ker(T))=3+0=3=\dim(P_2(\mathbb{R})).$$

Example 6.5.7. 考慮 $\mathscr{M}_{2\times2}$ 所形成的 vector space (參見 Example 6.1.4 (A)). 考慮函數 $T:\mathscr{M}_{2\times2}\to\mathscr{M}_{2\times2}$ 定義為 $T(A)=A-A^{\mathrm{T}}$, $\forall A\in\mathscr{M}_{2\times2}$. 我們可得 T 為 linear transformation. 這是因為對任意 $A,B\in\mathscr{M}_{2\times2}$ 以及 $r\in\mathbb{R}$, 我們有

$$T(A+rB)=(A+rB)-(A+rB)^{\mathrm{T}}=A+rB-A^{\mathrm{T}}-rB^{\mathrm{T}}=(A-A^{\mathrm{T}})+r(B-B^{\mathrm{T}})=T(A)+rT(B).$$

考慮 $\mathscr{M}_{2\times2}$ 的 ordered basis $\mathscr{E}=\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right)$. 由於

$$T(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix})=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},\quad T(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix})=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

$$T(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix})=\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},\quad T(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix})=\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

我們得 T 對於 $\mathscr{E},\mathscr{E}$ 的 matrix representation 為 ${}_{\mathscr{E}}[T]_{\mathscr{E}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. 利用 elementary row operation 將 ${}_{\mathscr{E}}[T]_{\mathscr{E}}$ 化為 echelon form $\begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. 得到 ${}_{\mathscr{E}}[T]_{\mathscr{E}}$ 的 nullspace 的一組 basis 為 $\{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\}$, 因此得 $\{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\}$ 為 $\ker(T)$ 的一組 basis. 又 ${}_{\mathscr{E}}[T]_{\mathscr{E}}$ 的 column space 的一組 basis 為 $\{\begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}\}$, 故得 $\{\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\}$ 為 T 的 range $T(\mathscr{M}_{2\times 2})$ 的一組 basis. 注意我們有 $\dim(T(\mathscr{M}_{2\times 2})) + \dim(\ker(T)) = 1+3 = 4 = \dim(\mathscr{M}_{2\times 2})$. 另外若 $A \in \ker(T)$ 表示 $T(A) = A - A^{\mathrm{T}} = \mathbf{O}$, 亦即 $A = A^{\mathrm{T}}$. 反之亦然, 也就是說 $A \in \ker(T)$ 若且唯若 A 為 symmetric matrix. 因此由 $\dim(\ker(T)) = 3$, 我們知所有 2×2 的 symmetric matrices 所成的 subspace 的維度為 3.

Question 6.5. 試求所有 3×3 的 *symmetric matrices* 所成的 *subspace* 的維度為何?

當 T_1, T_2 皆為 $\mathbb{R}^n$ 到 $\mathbb{R}^m$ 的 linear transformation 時, 對任意 $c_1, c_2 \in \mathbb{R}$, 我們知道 $c_1T_1 + c_2T_2$ 的 standard matrix representation $[c_1T_1 + c_2T_2]$ 和 T_1, T_2 的 standard matrix representations $[T_1], [T_2]$ 的關係為 $[c_1T_1 + c_2T_2] = c_1[T_1] + c_2[T_2]$ (參見 Lemma 5.3.8). 這對於一般的 linear transformations $T_1 : V \to W$ 以及 $T_2 : V \to W$ 的 matrix representations 也是對的. 不過要特別注意, 一般的 linear transformation 的 matrix representation 是和定義域以及對應域的 ordered basis 有關, 所以只有當 T_1, T_2 都考慮對應相同的 ordered basis 所得的 matrix representation, 這樣的矩陣運算才有意義. 也就是說當分別給定 V, W 的 ordered basis, $\mathscr{B}, \beta$, 我們會有

$$ {}_{\beta}[c_1T_1 + c_2T_2]_{\mathscr{B}} = c_1\,{}_{\beta}[T_1]_{\mathscr{B}} + c_2\,{}_{\beta}[T_2]_{\mathscr{B}}. $$

對於合成函數也有類似的情況, 若 $T : V \to W$, $T' : W \to U$ 為 linear transformations. 若分別給定 V, W, U 的 ordered basis $\mathscr{B}, \beta, \gamma$, 我們有以下的圖示

$$
\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, shift right, "T_{\mathscr{B}}"'] & W \arrow[d, shift right, "T_{\beta}"'] \\
\mathbb{R}^n \arrow[r] \arrow[u, shift right, "T_{\mathscr{B}}^{-1}"'] & \mathbb{R}^m \arrow[u, shift right, "T_{\beta}^{-1}"']
\end{tikzcd}
\qquad
\begin{tikzcd}
W \arrow[r, "T'"] \arrow[d, shift right, "T_{\beta}"'] & U \arrow[d, shift right, "T_{\gamma}"'] \\
\mathbb{R}^m \arrow[r] \arrow[u, shift right, "T_{\beta}^{-1}"'] & \mathbb{R}^k \arrow[u, shift right, "T_{\gamma}^{-1}"']
\end{tikzcd}
$$

這裡由於 T 的對應域和 T' 的定義域相同, 所以我們可以考慮合成函數 $T' \circ T$. 又由於 W 都用固定的 ordered basis β, 所以兩邊 W 到 $\mathbb{R}^m$ 的之間的函數相同 (皆為 T_β). 因此我們可以將上面兩個 commutative diagrams 合併成一個 commutative diagram 如下:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_{\mathscr{B}}"'] & W \arrow[r, "T'"] \arrow[d, shift left, "T_\beta"] & U \arrow[d, shift left, "T_\gamma"] \\ \mathbb{R}^n \arrow[r] \arrow[u, shift right, "T_{\mathscr{B}}^{-1}"'] & \mathbb{R}^m \arrow[r] \arrow[u, shift left, "T_\beta^{-1}"] & \mathbb{R}^k \arrow[u, shift left, "T_\gamma^{-1}"] \end{tikzcd}$$

由於底部 $\mathbb{R}^n \to \mathbb{R}^m$ 的 matrix representation 為 ${}_\beta[T]_{\mathscr{B}}$, 而 $\mathbb{R}^m \to \mathbb{R}^k$ 的 matrix representation 為 ${}_\gamma[T']_\beta$, 因此由 Lemma 5.3.8 知, 它們的合成所對應的 matrix representation 為 ${}_\gamma[T']_\beta\,{}_\beta[T]_{\mathscr{B}}$. 因此我們有

$$ {}_\gamma[T' \circ T]_{\mathscr{B}} = {}_\gamma[T']_\beta\,{}_\beta[T]_{\mathscr{B}}.$$

特別的, 當 $T : V \to W$ 為 isomorphism 時, T 的 inverse $T^{-1} : W \to V$ 亦為 linear transformation 且 $\dim(V) = \dim(W)$. 此時當 V 和 W 分別選定固定的 ordered basis $\mathscr{B}, \beta$ 後, 我們有以下的 commutative diagram.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_{\mathscr{B}}"'] & W \arrow[r, "T^{-1}"] \arrow[d, shift left, "T_\beta"] & V \arrow[d, shift left, "T_{\mathscr{B}}"] \\ \mathbb{R}^n \arrow[r] \arrow[u, shift right, "T_{\mathscr{B}}^{-1}"'] & \mathbb{R}^n \arrow[r] \arrow[u, shift left, "T_\beta^{-1}"] & \mathbb{R}^n \arrow[u, shift left, "T_{\mathscr{B}}^{-1}"] \end{tikzcd}$$

由於兩端 V 所用的 ordered basis 是一致的, 所以由 $T^{-1} \circ T$ 是 V 到 V 的 identity map 知, 底部 $\mathbb{R}^n \to \mathbb{R}^n$ 再接 $\mathbb{R}^n \to \mathbb{R}^n$ 這個路徑是將 $\mathbb{R}^n$ 的向量固定不動, 也就是說它們的合成所對應的 matrix representation 為 identity matrix. 亦即 ${}_{\mathscr{B}}[T^{-1}]_\beta\,{}_\beta[T]_{\mathscr{B}} = I_n$. 同理, 由於 $T \circ T^{-1}$ 為 W 到 W 的 identity map, 我們得 ${}_\beta[T]_{\mathscr{B}}\,{}_{\mathscr{B}}[T^{-1}]_\beta = I_n$. 得證 ${}_{\mathscr{B}}[T^{-1}]_\beta$ 為 ${}_\beta[T]_{\mathscr{B}}$ 的反矩陣. 綜合以上的討論, 我們有以下有關 Lemma 5.3.8 和 Theorem 5.3.12 的推廣.

Theorem 6.5.8. 假設 V, W, U 為 *finite dimensional vector space* 且令 $\mathscr{B}, \beta, \gamma$ 分別為 V, W, U 的 *ordered basis.*

(1) 假設 T_1, T_2 為 V 到 W 的 *linear transformations.* 則對任意 $c_1, c_2 \in \mathbb{R}$, 我們有

$$ {}_\beta[c_1T_1 + c_2T_2]_{\mathscr{B}} = c_1\,{}_\beta[T_1]_{\mathscr{B}} + c_2\,{}_\beta[T_2]_{\mathscr{B}}.$$

(2) 設 $T : V \to W$ 及 $T' : W \to U$ 為 *linear transformation.* 則

$$ {}_\gamma[T' \circ T]_{\mathscr{B}} = {}_\gamma[T']_\beta\,{}_\beta[T]_{\mathscr{B}}.$$

(3) 設 $T : V \to W$ 為 *linear transformation.* 則 T 為 *isomorphism* 若且唯若 ${}_\beta[T]_{\mathscr{B}}$ 為 *invertible matrix.* 又此時 $T^{-1} : W \to V$ 對應於 $\beta, \mathscr{B}$ 的 *matrix representation* 為

$$ {}_{\mathscr{B}}[T^{-1}]_\beta = \left({}_\beta[T]_{\mathscr{B}}\right)^{-1}.$$

我們知道一個 linear transformation, 當我們用不同的 ordered bases 所得的 matrix representation 會不同. 假設 $\mathscr{B}, \mathscr{B}'$ 為 V 的兩組 ordered bases, 而 β, β' 為 W 的兩組 ordered basis. 對於 linear transformation $T : V \to W$, 其對應於這兩對 ordered bases 的 matrix representations ${}_\beta[T]_{\mathscr{B}}$ 和 ${}_{\beta'}[T]_{\mathscr{B}'}$ 之間會有甚麼關係呢? 首先我們考慮 identity map $\mathrm{id} : V \to V$. 注意雖然是 identity map, 但其 matrix representation 未必會是 identity matrix. 事實上, 當我們定義域和對應域都選同一組 ordered basis $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, 則由於 $\mathrm{id}(\mathbf{v}_i) = \mathbf{v}_i$, 故其 matrix representation 是 identity matrix. 但若定義域是使用 $\mathscr{B}$ 這一組

ordered basis, 而對應域選的是 $\mathscr{B}'=\{\mathbf{v}'_1,\dots,\mathbf{v}'_n\}$ 這一組 ordered basis, identity map 對應於 $\mathscr{B},\mathscr{B}'$ 的 matrix representation ${}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}}$ 其 i-th column 雖然仍和 $\mathrm{id}(\mathbf{v}_i)=\mathbf{v}_i$ 有關, 不過卻是要將 $\mathbf{v}_i$ 寫成以 $\{\mathbf{v}'_1,\dots,\mathbf{v}'_n\}$ 為 ordered basis 的坐標表示法 ${}_{\mathscr{B}'}[\mathbf{v}_i]$. 所以當 $\mathscr{B}$ 和 $\mathscr{B}'$ 相異時, ${}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}}$ 不是 identity matrix. 現對任意 $\mathbf{v}\in V$, 因 $\mathbf{v}$ 對於 $\mathscr{B}$ 的坐標表示法為 ${}_{\mathscr{B}}[\mathbf{v}]$, 依 matrix representation 的性質 (Proposition 6.5.2) 可得

$$ {}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}}\,{}_{\mathscr{B}}[\mathbf{v}] = {}_{\mathscr{B}'}[\mathrm{id}(\mathbf{v})] = {}_{\mathscr{B}'}[\mathbf{v}]. $$

也就是說, 矩陣 ${}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}}$ 可以將 V 中元素對於 $\mathscr{B}$ 的坐標表示轉換成對於 $\mathscr{B}'$ 的坐標表示, 也因此我們稱 ${}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}}$ 為 *change-of-basis matrix.*

要注意 $\mathrm{id}:V\to V$ 是 isomorphism, 所以由 Theorem 6.5.8 (3), 我們得 ${}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}}$ 為 invertible 且

$$ ({}_{\mathscr{B}'}[\mathrm{id}]_{\mathscr{B}})^{-1} = {}_{\mathscr{B}}[\mathrm{id}^{-1}]_{\mathscr{B}'} = {}_{\mathscr{B}}[\mathrm{id}]_{\mathscr{B}'} \tag{6.3} $$

也就是說將 $\mathscr{B}$ 的坐標表示轉換成對於 $\mathscr{B}'$ 的坐標表示的 change-of-basis matrix 的 inverse 就是 $\mathscr{B}'$ 的坐標表示轉換成對於 $\mathscr{B}$ 的坐標表示的 change-of-basis matrix.

我們回到原先的問題, 假設 $T:V\to W$ 為 linear transformation 且 $\mathscr{B},\mathscr{B}'$ 為 V 的兩組 ordered bases, 而 β,β' 為 W 的兩組 ordered basis. 我們要探討 ${}_{\beta}[T]_{\mathscr{B}}$ 和 ${}_{\beta'}[T]_{\mathscr{B}'}$ 之間的關係. 由於 $\mathrm{id}_V:V\to V$, $T:V\to W$ 和 $\mathrm{id}_W:W\to W$ 之合成 $\mathrm{id}_W\circ T\circ\mathrm{id}_V:V\to W$ 仍為 $T:V\to W$, 所以由 Theorem 6.5.8 (2) 得

$$ {}_{\beta'}[\mathrm{id}_W]_{\beta}\,{}_{\beta}[T]_{\mathscr{B}}\,{}_{\mathscr{B}}[\mathrm{id}_V]_{\mathscr{B}'} = {}_{\beta'}[T]_{\mathscr{B}'}. $$

這就是所謂的 change-of basis formula, 我們將之完整敘述如下.

Theorem 6.5.9 (Change-of-basis Formula)**.** 假設 $T:V\to W$ 為 *linear transformation* 且 $\mathscr{B},\mathscr{B}'$ 為 V 的兩組 *ordered bases,* 而 β,β' 為 W 的兩組 *ordered basis,* 則存在 *invertible matrix* P,Q 使得 ${}_{\beta'}[T]_{\mathscr{B}'}=Q({}_{\beta}[T]_{\mathscr{B}})P$, 其中 P 為將 $\mathscr{B}'$ 的坐標表示轉換成 $\mathscr{B}$ 的坐標表示的 *change-of-basis matrix* ${}_{\mathscr{B}}[\mathrm{id}_V]_{\mathscr{B}'}$*,* 而 Q 為將 β 的坐標表示轉換成 β' 的坐標表示的 *change-of-basis matrix* ${}_{\beta'}[\mathrm{id}_W]_{\beta}$*.*

Example 6.5.10. 在 Example 6.5.1 中我們考慮 linear transformation $T:P_2(\mathbb{R})\to P_3(\mathbb{R})$, 其中 $T(p(x))=(x+1)p(x-1)$, $\forall p(x)\in P_2(\mathbb{R})$. 另外我們考慮 $P_2(\mathbb{R})$ 的兩組 ordered bases $\mathscr{E}=(x^2,x,1)$, $\mathscr{B}=(p_1(x),p_2(x),p_3(x))$ 其中

$$ p_1(x)=\frac{1}{2}(x^2-x),\quad p_2(x)=-x^2+1,\quad p_3(x)=\frac{1}{2}(x^2+x) $$

以及 $P_3(\mathbb{R})$ 的兩組 ordered bases $\varepsilon=(x^3,x^2,x,1)$, $\beta=(q_1(x),q_2(x),q_3(x),q_4(x))$ 其中

$$ q_1(x)=\frac{-x^3+3x^2-2x}{6},q_2(x)=\frac{x^3-2x^2-x+2}{2},q_3(x)=\frac{-x^3+x^2+2x}{2},q_4(x)=\frac{x^3-x}{6}. $$

在 Example 6.5.1 中我們得到

$$ {}_{\varepsilon}[T]_{\mathscr{E}}=\begin{bmatrix} 1 & 0 & 0\\ -1 & 1 & 0\\ -1 & 0 & 1\\ 1 & -1 & 1 \end{bmatrix},\quad {}_{\beta}[T]_{\mathscr{B}}=\begin{bmatrix} 0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 2 & 0\\ 0 & 0 & 3 \end{bmatrix}. $$

因 ${}_{\mathscr{E}}[p_1(x)] = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix}$, ${}_{\mathscr{E}}[p_2(x)] = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, ${}_{\mathscr{E}}[p_3(x)] = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$ 依定義 $\mathscr{B}$ 到 $\mathscr{E}$ 的 change-of-basis matrix 為 ${}_{\mathscr{E}}[\mathrm{id}_{P_2(\mathbb{R})}]_{\mathscr{B}} = \begin{bmatrix} 1/2 & -1 & 1/2 \\ -1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$. 另外若 $x^3 = c_1q_1(x)+c_2q_2(x)+c_3q_3(x)+c_4q_4(x)$, 則因 $q_1(-1)=1, q_2(-1)=q_3(-1)=q_4(-1)=0$, 將 $x=-1$ 代入前式得 $c_1=-1$, 同理我們可得 $c_2=0, c_3=1, c_4=8$, 亦即 ${}_{\beta}[x^3] = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 8 \end{bmatrix}$. 用同樣方法求 $x^2, x, 1$ 對於 β 的坐標表示法, 我們得 ε 到 β 的 change-of-basis matrix 為 ${}_{\beta}[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{bmatrix}$. 我們也可以先寫下 β 到 ε 的 change-of-basis matrix ${}_{\varepsilon}[\mathrm{id}_{P_3(\mathbb{R})}]_{\beta} = \begin{bmatrix} -1/6 & 1/2 & -1/2 & 1/6 \\ 1/2 & -1 & 1/2 & 0 \\ -1/3 & -1/2 & 1 & -1/6 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ 再取 inverse 得 ${}_{\beta}[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon}$. 最後我們驗算

$$
{}_{\beta}[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon}\,{}_{\varepsilon}[T]_{\mathscr{E}}\,{}_{\mathscr{E}}[\mathrm{id}_{P_2(\mathbb{R})}]_{\mathscr{B}} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{-1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} = {}_{\beta}[T]_{\mathscr{B}}.
$$

在線性代數中, 我們經常談論的一種 linear transformation 是其定義域及對應域為相同的 vector space. 這樣的 linear transformation 我們特別稱之為 *linear operator*. 關於 linear operator 我們通常對於定義域及對應域會選同樣的一組 ordered basis. 此時利用 Theorem 6.5.9, 我們得以下之結果.

Corollary 6.5.11. 假設 $T: V \to V$ 為 *linear transformation* 且 $\mathscr{B}, \mathscr{B}'$ 為 V 的兩組 *ordered bases.* 則存在 *invertible matrix* P 使得 ${}_{\mathscr{B}'}[T]_{\mathscr{B}'} = P^{-1}({}_{\mathscr{B}}[T]_{\mathscr{B}})P$, 其中 P 為將 $\mathscr{B}'$ 的坐標表示轉換成 $\mathscr{B}$ 的坐標表示的 *change-of-basis matrix* ${}_{\mathscr{B}}[\mathrm{id}_V]_{\mathscr{B}'}$.

Proof. 考慮 Theorem 6.5.9 其中 $W = V$, $\beta = \mathscr{B}$ 以及 $\beta' = \mathscr{B}'$ 的情形. 此時 $Q = {}_{\mathscr{B}'}[\mathrm{id}_V]_{\mathscr{B}}$ 由式子 (6.3), 知 $Q = ({}_{\mathscr{B}}[\mathrm{id}_V]_{\mathscr{B}'})^{-1} = P^{-1}$, 得證本定理. □

給定一個 $n \times n$ matrix A 我們知道它可以代表某一個 dimension 為 n 的 vector space V 上的 linear operator $T: V \to V$, 對於 V 的某一組 ordered basis 的 matrix representation. 若 P 為 $n \times n$ invertible matrix, 則我們稱 $B = P^{-1}AP$ 和 A 為 *similar*. 意味著我們也可將 B 視為 $T: V \to V$ 的一個 matrix representation 只是選取 V 不同的 ordered basis 而已. 有時一個 linear operator, 若選取夠好的一組 ordered basis, 我們可以得到更好的 matrix representation 以至於更容易了解這個 linear transformation. 有關於這個課題, 等以後談到對角化時我們再進一步探討. 我們先看一個簡單的例子.

Example 6.5.12. 考慮 linear operator $T:\mathbb{R}^2\to\mathbb{R}^2$ 定義為 $T(\begin{bmatrix}x\\y\end{bmatrix})=\frac{1}{25}\begin{bmatrix}9x+12y\\12x+16y\end{bmatrix}$. 若利用標準基底 $\varepsilon=(\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix})$ 我們得 ${}_{\varepsilon}[T]_{\varepsilon}=\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix}$. 然而若用 $\beta=(\begin{bmatrix}3\\4\end{bmatrix},\begin{bmatrix}-4\\3\end{bmatrix})$ 這組 ordered basis 可由 $T(\begin{bmatrix}3\\4\end{bmatrix})=\begin{bmatrix}3\\4\end{bmatrix}$, $T(\begin{bmatrix}-4\\3\end{bmatrix})=\begin{bmatrix}0\\0\end{bmatrix}$, 得 ${}_{\beta}[T]_{\beta}=\begin{bmatrix}1&0\\0&0\end{bmatrix}$. 我們很容易由

$$ {}_{\beta}[T\circ T]_{\beta}=({}_{\beta}[T]_{\beta})({}_{\beta}[T]_{\beta})=\begin{bmatrix}1&0\\0&0\end{bmatrix}^2=\begin{bmatrix}1&0\\0&0\end{bmatrix}={}_{\beta}[T]_{\beta}$$

推得 $T\circ T=T$. 事實上從 β 這組 ordered basis 我們很容易看出 T 就是將 $\mathbb{R}^2$ 上的向量對 $\begin{bmatrix}3\\4\end{bmatrix}$ 的 projection. 另外令 $P={}_{\varepsilon}[\mathrm{id}]_{\beta}=\begin{bmatrix}3&-4\\4&3\end{bmatrix}$, 我們得

$$\begin{bmatrix}1&0\\0&0\end{bmatrix}={}_{\beta}[T]_{\beta}=({}_{\beta}[\mathrm{id}]_{\varepsilon})({}_{\varepsilon}[T]_{\varepsilon})({}_{\varepsilon}[\mathrm{id}]_{\beta})=P^{-1}(\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix})P,$$

所以 $\begin{bmatrix}1&0\\0&0\end{bmatrix}$ 和 $\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix}$ 為 similar.

6.6. 結論

我們介紹了一般的 finite dimensional vector spaces 以及它們之間的 linear transformations. 由於 finite dimensional vector space 也會有 basis 存在, 所以我們可以加它們坐標化以至於將它們如我們熟悉的 $\mathbb{R}^n$ 來處理. 也因此這樣的 linear transformation 都可以用矩陣來表示. 所以我們可以利用矩陣的理論來更進一步了解這些 linear transformations. 不過要注意一個 linear transformation 的 matrix representation 和 ordered basis 的選取有關. 有時選取好的 ordered basis 可以讓我們得到更好的 matrix representation 以至於更容易了解這個 linear transformation. 因此一個 linear transformation 選用不同 order basis 所得的 matrix representation 之間的關係分外重要. 希望大家能好好了解它們之間的關係.

Chapter 7

Inner Product Space

在 $\mathbb{R}^m$ 中有關於 dot product 的性質, 可以推廣到一般的 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. 對於 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 1.4.2 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) $\langle\mathbf{v},\mathbf{v}\rangle\geq 0$ 且 $\langle\mathbf{v},\mathbf{v}\rangle=0$ 若且唯若 $\mathbf{v}=\mathbf{O}$.

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

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

我們便稱 V 為 *inner product space*. 在這一章中我們將介紹 inner product space 的性質. 由於這些性質僅利用上述 inner product 的性質便可得到, 所以只要是 finite dimensional inner product space 都會有我們介紹的這些性質. 不過因為大家對較抽象的 inner product 較不熟悉, 所以這裡我們僅用大家熟悉 $\mathbb{R}^m$ 中一般的內積來處理. 另外, 由於我們將利用矩陣的乘法來處理內積, 對於 $\mathbf{v},\mathbf{w}\in\mathbb{R}^m$, 原來我們用 $\mathbf{v}\cdot\mathbf{w}$ 表示 $\mathbf{v},\mathbf{w}$ 的內積, 在這裡為了避免混淆, 我們將改用 $\langle\mathbf{v},\mathbf{w}\rangle$ 來表示 $\mathbf{v},\mathbf{w}$ 的內積. 也就是說, 若將 $\mathbf{v},\mathbf{w}$ 視為 column vectors 且用矩陣的乘法表示, 我們有 $\langle\mathbf{v},\mathbf{w}\rangle=\mathbf{v}^{\mathrm{T}}\mathbf{w}$.

7.1. Projection

我們曾經在介紹 $\mathbb{R}^2$ 上的 linear transformation 時介紹 projection. 簡單來說, 給定非零向量 $\mathbf{w}\in\mathbb{R}^2$, 我們定義 $\mathbf{v}\in\mathbb{R}^2$ 在 $\mathbf{w}$ 上的投影 $\mathrm{Proj}_{\mathbf{w}}(\mathbf{v})$ 必須在 $\mathbf{w}$ 所展成的空間中 (即 $\mathrm{Span}(\mathbf{w})$), 而且 $\mathbf{v}-\mathrm{Proj}_{\mathbf{w}}(\mathbf{v})$ 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 $\mathbb{R}^m$ 的一個 subspace W 上的投影. 也就是說, 當 W 為 $\mathbb{R}^m$ 的 subspace, 對於 $\mathbf{v}\in\mathbb{R}^m$, 我們定義 $\mathbf{v}$ 在 W 的 projection 為 W 上的一個向量 $\mathbf{w}$, 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 7.1.1. 給定 W 為 $\mathbb{R}^m$ 的 subspace. 令

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

也就是說 $W^\perp$ 為 $\mathbb{R}^m$ 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 7.1.2. 給定 W 為 $\mathbb{R}^m$ 的 subspace. 對於 $\mathbf{v}\in\mathbb{R}^m$, 若 $\mathbf{w}\in W$ 滿足 $\mathbf{v}-\mathbf{w}\in W^\perp$, 則稱 $\mathbf{w}$ 為 the *projection* of $\mathbf{v}$ on W.

等一下我們將說明對於任意 $\mathbf{v}\in\mathbb{R}^m$, 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, 將來我們要說明一個 $\mathbb{R}^m$ 的向量 $\mathbf{w}$ 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 $\mathbb{R}^m$ 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 (3)(4), 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 $\mathbb{R}^m$ 的 subspace.

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 $\mathbf{O}$ 一定在其中. 事實上我們有以下的結果.

Lemma 7.1.3. 假設 W 為 $\mathbb{R}^m$ 的 *subspace.* 則 $W\cap W^\perp=\{\mathbf{O}\}$.

Proof. 因 $W\cap W^\perp$ 為 subspace, 故知 $\mathbf{O}\in W\cap W^\perp$. 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 (2) 知 $\mathbf{w}=\mathbf{O}$, 得證 $W\cap W^\perp=\{\mathbf{O}\}$. □

現在我們可以證明 projection 的唯一性.

Proposition 7.1.4. 給定 W 為 $\mathbb{R}^m$ 的 *subspace.* 對於 $\mathbf{v}\in\mathbb{R}^m$, $\mathbf{v}$ 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 7.1.3 知 $\mathbf{w}-\mathbf{w}'=\mathbf{O}$, 即 $\mathbf{w}=\mathbf{w}'$, 證得唯一性. □

接下來我們要探討 projection 的存在性. 由於 $W^\perp$ 為 subspace, 我們更進一步的想要知道其 dimension. 由於 $W^\perp$ 與 W 有關, 我們可以預期 $\dim(W^\perp)$ 和 $\dim(W)$ 應該有關. 事實上假設 $\dim(W)=n$ 且令 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis. 考慮 $A=\begin{bmatrix} — & \mathbf{w}_1 & — \\ & \vdots & \\ — & \mathbf{w}_n & — \end{bmatrix}$, 即 A 的 i-th row 為 $\mathbf{w}_i$ (寫成 row vector). 由於 $\mathbf{w}_i\in\mathbb{R}^m$, 故知 A 為 $n\times m$ matrix. 對於任意 $\mathbf{v}\in W^\perp$, 由於 $\mathbf{w}_i\in W$, 我們有 $\langle\mathbf{w}_i,\mathbf{v}\rangle=0$, $\forall i=1,\ldots,n$. 因此由矩陣乘法定義知 $A\mathbf{v}=\mathbf{O}$, 也就是說 $W^\perp$ 包含於 A 的 nullspace. 反之, 若 $\mathbf{v}\in\mathbb{R}^m$ 為 A 的 nullspace 中一元素, 則由 $A\mathbf{v}=\mathbf{O}$

得 $\langle \mathbf{w}_i, \mathbf{v}\rangle = 0$, $\forall i = 1,\ldots,n$. 因此對任意 $\mathbf{w} \in W$, 由於 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis, 存在 $c_1,\ldots,c_n \in \mathbb{R}$ 使得 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$. 故得

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

也就是說 A 的 nullspace 包含於 $W^\perp$. 我們得證了以下的性質.

Lemma 7.1.5. *給定 W 為 $\mathbb{R}^m$ 的 subspace. 假設 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis. 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ (寫成 row vector) 為 row vector 的 $n \times m$ matrix. 則 $W^\perp$ 為 A 的 nullspace.*

Example 7.1.6. 令 $W = \mathrm{Span}(\begin{bmatrix}1\\2\\2\\1\end{bmatrix}, \begin{bmatrix}3\\4\\2\\3\end{bmatrix})$ 我們要找出 W 的 orthogonal complement $W^\perp$.

首先考慮 $A = \begin{bmatrix}1 & 2 & 2 & 1\\ 3 & 4 & 2 & 3\end{bmatrix}$. 由 Lemma 7.1.5, 我們知道 A 的 nullspace $\{\mathbf{x} \in \mathbb{R}^4 \mid A\mathbf{x} = \mathbf{O}\}$ 即為 $W^\perp$. 利用 elementary row operation 將 A 的 1-st row 乘以 -3 加到 2-nd row 得 $\begin{bmatrix}1 & 2 & 2 & 1\\ 0 & -2 & -4 & 0\end{bmatrix}$. 再將 2-nd row 乘以 $-1/2$ 得 $\begin{bmatrix}1 & 2 & 2 & 1\\ 0 & 1 & 2 & 0\end{bmatrix}$. 解得 $\begin{bmatrix}-1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}2\\-2\\1\\0\end{bmatrix}$ 為 A 的 nullspace, 即 $W^\perp$ 的一組 basis.

Question 7.1. *若 Lemma 7.1.5 中 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 僅假設 $\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n) = W$ 但不假設 linearly independent, 是否仍可得 $W^\perp$ 為 A 的 nullspace?*

一般來說我們習慣將 $\mathbb{R}^m$ 的向量寫成 column vector, 所以若將 Lemma 7.1.5 中 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 寫成 column vector 且分別為 A 的 column vector (此時 A 為 $m \times n$ matrix), 我們可將 Lemma 7.1.5 改寫為以下定理.

Corollary 7.1.7. *給定 W 為 $\mathbb{R}^m$ 的 subspace. 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 $W^\perp$ 為 A^{T} 的 nullspace.*

在 Corollary 7.1.7 中依 column space 的定義 $W = C(A)$, 故 Corollary 7.1.7 又可改寫成以下定理.

Corollary 7.1.8. *假設 A 為 $m \times n$ matrix. 則 $C(A)^\perp = N(A^{\mathrm{T}})$.*

現在我們可以算出 $\dim(W^\perp)$ 和 $\dim(W)$ 之間的關係了. 事實上假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\dim(W) = n$. 利用 W 的一組 basis 為 row vectors 所形成的 $n \times m$ matrix A, 並套用 Lemma 7.1.5 我們知道 $W^\perp$ 會是 A 的 nullspace. 也因此 $\dim(W^\perp)$ 會是 A 的 nullspace 的維度 (即 A 的 nullity). 依定義 A 的 rank 會是 A 的 row space (即 A 的 row vectors 所展成的 space W) 的維度, 即 $\mathrm{rank}(A) = \dim(W) = n$. 故利用 Theorem 4.4.13 我們知道 A 的 nullity 為 A 的 column 的個數 m 減去 $\mathrm{rank}(A)$, 即 $\mathrm{null}(A) = m - \mathrm{rank}(A) = m - n$. 因此我們得到以下重要的性質.

Theorem 7.1.9. 假設 W 為 $\mathbb{R}^m$ 的 *subspace.* 則 $\dim(W^\perp) = m - \dim(W)$.

利用 Theorem 7.1.9 我們很快可以推出以下的結果.

Corollary 7.1.10. 假設 W 為 $\mathbb{R}^m$ 的 *subspace.* 則 $(W^\perp)^\perp = W$.

Proof. 由 Theorem 7.1.9 我們知道 $\dim((W^\perp)^\perp) = m - \dim(W^\perp) = m - (m - \dim(W)) = \dim(W)$. 然而給定 $\mathbf{w} \in W$, 由於對任意 $\mathbf{w}' \in W^\perp$, 皆有 $\langle \mathbf{w}, \mathbf{w}' \rangle = 0$, 故得 $\mathbf{w} \in (W^\perp)^\perp$. 因此知 $W \subseteq (W^\perp)^\perp$. 故由 $\dim(W) = \dim((W^\perp)^\perp)$ 利用 Corollary 4.3.6 得證 $(W^\perp)^\perp = W$. □

這裡我們要強調一下, Corollary 7.1.10 的證明是利用 dimension 來處理. 當 W 是 infinite dimensional vector space 的 subspace 時, $(W^\perp)^\perp = W$ 未必成立. 也就是說雖然我們證明了 $W \subseteq (W^\perp)^\perp$, 但另一方向 $(W^\perp)^\perp \subseteq W$ 未必正確. 而在 finite dimensional 的情形, 還好有 dimension 可以計算, 所以才可以藉由 Corollary 4.3.6 得證 $(W^\perp)^\perp = W$.

Question 7.2. 假設 A 為 $m \times n$ *matrix.* 試證明 $C(A) = N(A^{\mathrm{T}})^\perp$.

由 Theorem 7.1.9 我們知道若 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 W 的一組 basis, 則由於 $\dim(W^\perp) = m - n$, 我們可以找到一組 $W^\perp$ 的 basis $\mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$. 我們將證明此時 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 linearly independent, 因而得知以下的結果.

Corollary 7.1.11. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 W 的一組 *basis* 以及 $\mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $W^\perp$ 的一組 *basis.* 則 $\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $\mathbb{R}^m$ 的一組 *basis.*

Proof. 首先證明 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 linearly independent. 我們用反證法, 假設存在 $c_1, \ldots, c_n, c_{n+1}, \ldots, c_m \in \mathbb{R}$ 不全為 0 滿足 $c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n + c_{n+1}\mathbf{u}_1 + \cdots + c_m\mathbf{u}_{m-n} = \mathbf{O}$. 令 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$, 由於 W 為 vector space, 我們有 $\mathbf{w} \in W$. 又由於 $\mathbf{w} = -(c_{n+1}\mathbf{u}_1 + \cdots + c_m\mathbf{u}_{m-n})$ 以及 $W^\perp$ 為 vector space, 我們也有 $\mathbf{w} \in W^\perp$. 換言之, $\mathbf{w} \in W \cap W^\perp$. 因此由 Lemma 7.1.3 得證 $\mathbf{w} = \mathbf{O}$. 再由 $\mathbf{w}_1, \ldots, \mathbf{w}_n$ 為 linearly independent 以及 $c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n = \mathbf{w} = \mathbf{O}$ 得 $c_1 = \cdots = c_n = 0$. 同理得 $c_{n+1} = \cdots = c_m = 0$. 此與 $c_1, \ldots, c_n, c_{n+1}, \ldots, c_m$ 不全為 0 之假設相矛盾, 得證 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 linearly independent. 最後再由 $\dim(\mathbb{R}^m) = m$, 以及 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $\mathbb{R}^m$ 中 m 個 linearly independent vectors, 得知 $\mathbf{w}_1, \ldots, \mathbf{w}_n, \mathbf{u}_1, \ldots, \mathbf{u}_{m-n}$ 為 $\mathbb{R}^m$ 的一組 basis. □

既然 W 的一組 basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ 和 $W^\perp$ 的一組 basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_{m-n}\}$ 可形成 $\mathbb{R}^m$ 的一組 basis, 對任意 $\mathbf{v} \in \mathbb{R}^m$, 存在 $c_1, \ldots, c_n, d_1, \ldots, d_{m-n} \in \mathbb{R}$ 使得

$$\mathbf{v} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n + d_1\mathbf{u}_1 + \cdots + d_{m-n}\mathbf{u}_{m-n}.$$

此時若令 $\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$, 則 $\mathbf{w} \in W$ 且 $\mathbf{v} - \mathbf{w} = d_1\mathbf{u}_1 + \cdots + d_{m-n}\mathbf{u}_{m-n} \in W^\perp$. 因此依定義知 $\mathbf{w}$ 即為 the projection of $\mathbf{v}$ on W. 這也證明了 projection 的存在性. 綜合以上的結果, 我們有以下的定理.

Theorem 7.1.12. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *basis* 以及 $\mathbf{u}_1,\ldots,\mathbf{u}_{m-n}$ 為 $W^\perp$ 的一組 *basis.* 對於 $\mathbf{v}\in\mathbb{R}^m$, 若 $\mathbf{v}=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n+d_1\mathbf{u}_1+\cdots+d_{m-n}\mathbf{u}_{m-n}$, 則 $\mathbf{v}$ 在 W 的 *projection* 為

$$\mathrm{Proj}_W(\mathbf{v})=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n.$$

Example 7.1.13. 我們利用 Example 7.1.6 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 projection. 已知 $\begin{bmatrix}-1\\0\\0\\1\end{bmatrix},\begin{bmatrix}2\\-2\\1\\0\end{bmatrix}$ 為 $W^\perp$ 的一組 basis, 故將 $\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$ 寫成 $\begin{bmatrix}1\\2\\2\\1\end{bmatrix},\begin{bmatrix}3\\4\\2\\3\end{bmatrix},\begin{bmatrix}-1\\0\\0\\1\end{bmatrix},\begin{bmatrix}2\\-2\\1\\0\end{bmatrix}$ 的 linear combination 得

$$\begin{bmatrix}2\\0\\1\\4\end{bmatrix}=-\begin{bmatrix}1\\2\\2\\1\end{bmatrix}+\begin{bmatrix}3\\4\\2\\3\end{bmatrix}+2\begin{bmatrix}-1\\0\\0\\1\end{bmatrix}+\begin{bmatrix}2\\-2\\1\\0\end{bmatrix}.$$

因此求出 $\mathbf{v}$ 在 W 的 projection 為 $-\begin{bmatrix}1\\2\\2\\1\end{bmatrix}+\begin{bmatrix}3\\4\\2\\3\end{bmatrix}=\begin{bmatrix}2\\2\\0\\2\end{bmatrix}$.

給定 $\mathbb{R}^m$ 的一個 subspace W, 對於 $\mathbf{v}\in\mathbb{R}^m$, 既然 $\mathrm{Proj}_W(\mathbf{v})$ 是存在且唯一的, 我們可以定義 projection on W 這樣的函數. 也就是將 Proj_W 視為是一個 $\mathbb{R}^m\to\mathbb{R}^m$ 函數且對任意 $\mathbf{v}\in\mathbb{R}^m$ 的取值就是 $\mathrm{Proj}_W(\mathbf{v})$. 我們要說明 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 事實上是一個 linear transformation.

首先對於任意 $\mathbf{v},\mathbf{v}'\in\mathbb{R}^m$, 為了方便我們令 $\mathbf{w}=\mathrm{Proj}_W(\mathbf{v})$ 以及 $\mathbf{w}'=\mathrm{Proj}_W(\mathbf{v}')$. 也就是說 $\mathbf{w},\mathbf{w}'\in W$ 且滿足 $\mathbf{v}-\mathbf{w}\in W^\perp$ 以及 $\mathbf{v}'-\mathbf{w}'\in W^\perp$. 我們要證明 $\mathrm{Proj}_W(\mathbf{v}+\mathbf{v}')=\mathrm{Proj}_W(\mathbf{v})+\mathrm{Proj}_W(\mathbf{v}')=\mathbf{w}+\mathbf{w}'$. 換言之, 我們要證明 $\mathbf{w}+\mathbf{w}'\in W$ 且滿足 $(\mathbf{v}+\mathbf{v}')-(\mathbf{w}+\mathbf{w}')\in W^\perp$. 然而 $W,W^\perp$ 皆為 vector space, 我們自然由 $\mathbf{w},\mathbf{w}'\in W$ 以及 $\mathbf{v}-\mathbf{w},\mathbf{v}'-\mathbf{w}'\in W^\perp$ 可得 $\mathbf{w}+\mathbf{w}'\in W$ 以及 $(\mathbf{v}+\mathbf{v}')-(\mathbf{w}+\mathbf{w}')=(\mathbf{v}-\mathbf{w})+(\mathbf{v}'-\mathbf{w}')\in W^\perp$. 另外對任意 $r\in\mathbb{R}$, 我們也要證明 $\mathrm{Proj}_W(r\mathbf{v})=r\mathrm{Proj}_W(\mathbf{v})=r\mathbf{w}$. 也就是證明 $r\mathbf{w}\in W$ 以及 $r\mathbf{v}-r\mathbf{w}\in W^\perp$. 同樣的由 $W,W^\perp$ 皆為 vector space, 以及 $\mathbf{w}\in W,\mathbf{v}-\mathbf{w}\in W^\perp$, 很容易看出這是對的. 這也證明了

Proposition 7.1.14. 假設 W 為 $\mathbb{R}^m$ 的 *subspace,* 則 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 是一個 *linear transformation.*

Question 7.3. 試說明甚麼是 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 的 *image* 以及 *kernel.*

其實利用 Theorem 7.1.12 來求 $\mathbf{v}$ 在 W 上的投影有點麻煩. 因為我們需利用 W 的一組 basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 求出 $W^\perp$ 的一組 basis $\mathbf{u}_1,\ldots,\mathbf{u}_{m-n}$, 再將 $\mathbf{v}$ 寫成 $\mathbf{w}_1,\ldots,\mathbf{w}_n,\mathbf{u}_1,\ldots,\mathbf{u}_{m-n}$ 的線性組合, 才能求出 $\mathrm{Proj}_W(\mathbf{v})$. 我們希望將此步驟簡化. 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 $C(A)$. 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 $\mathbf{v}-\mathbf{w}\in W^{\perp}=(C(A))^{\perp}$ 的要求, 由 Corollary 7.1.8 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 $A^{\mathrm{T}}(\mathbf{v}-A\mathbf{x})=\mathbf{O}$. 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 $\mathrm{Proj}_W(\mathbf{v})$ 一定存在, 所以一定存在 $\mathbf{x}\in\mathbb{R}^n$ 使得 $A\mathbf{x}=\mathrm{Proj}_W(\mathbf{v})$, 也因此這個 $\mathbf{x}$ 必滿足 $\mathbf{v}-A\mathbf{x}\in W^{\perp}$, 所以聯立方程組 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 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. 問題是要如何解 $(A^{\mathrm{T}}A)\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 呢? 我們有以下的定理.

Lemma 7.1.15. 假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathrm{rank}(A)=n$, 則 $A^{\mathrm{T}}A$ 是一個 $n\times n$ *invertible matrix.*

Proof. 首先由 $A\in\mathscr{M}_{m\times n}$ 且 $A^{\mathrm{T}}\in\mathscr{M}_{n\times m}$, 我們知 $A^{\mathrm{T}}A\in\mathscr{M}_{n\times n}$. 因此可以利用 Theorem 3.5.9, 來證明 $A^{\mathrm{T}}A$ 為 invertible. 也就是說證明 $(A^{\mathrm{T}}A)\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution 就等同於證得 $A^{\mathrm{T}}A$ 為 invertible matrix. 假設 $\mathbf{x}=\mathbf{u}$ 為 $(A^{\mathrm{T}}A)\mathbf{x}=\mathbf{O}$ 的一個解, 亦即 $(A^{\mathrm{T}}A)\mathbf{u}=\mathbf{O}$. 由於 $(A^{\mathrm{T}}A)\mathbf{u}=A^{\mathrm{T}}(A\mathbf{u})$, 故得 $A\mathbf{u}\in N(A^{\mathrm{T}})$. 然而 $A\mathbf{u}\in C(A)$, 所以有 $A\mathbf{u}\in C(A)\cap N(A^{\mathrm{T}})=C(A)\cap C(A)^{\perp}$. 因此利用 Lemma 7.1.3 得證 $A\mathbf{u}=\mathbf{O}$. 然而依假設 $\mathrm{rank}(A)=n$, 故由 Corollary 2.3.5 知聯立方程組 $A\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution, 也就是說 $\mathbf{u}=\mathbf{O}$. 由此知 $(A^{\mathrm{T}}A)\mathbf{x}=\mathbf{O}$ 沒有 nontrivial solution, 故 $A^{\mathrm{T}}A$ 是一個 $n\times n$ invertible matrix. □

前面我們假設 A 是以 W 的一組 basis 為 column vector 所得的 $m\times n$ matrix, 所以 $\mathrm{rank}(A)=\dim(W)=n$. 故利用 Lemma 7.1.15 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.

Theorem 7.1.16. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 為 W 的一組 *basis.* 令 A 為以 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 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 7.1.17. 我們要利用 Theorem 7.1.16 的結果處理 Example 7.1.13 的情形, 也就是求 $\mathbf{v}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$ 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 Theorem 7.1.16 得

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

和 Example 7.1.13 的結果吻合.

對於 Theorem 7.1.16 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 $W^{\perp}=\{\mathbf{O}\}$, 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 $A^{-1}(A^{\mathrm{T}})^{-1}$. 也因為這樣 Theorem 7.1.16 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 7.1.16 確實簡化了 Theorem 7.1.12 求 projection 的程序. 我們只要求出 W 的一組 basis 即可, 不必再求 $W^{\perp}$ 的 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 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ 用 P_W 來表示, 且稱之為對於 W 的 *projection matrix*. 要注意, 依定義 P_W 事實上就是 $\mathrm{Proj}_W:\mathbb{R}^m\to\mathbb{R}^m$ 這一個 linear transformation 的 standard matrix representation. 由於一個 linear transformation 其 standard matrix representation 是唯一的, 所以雖然 W 選取不同的 basis 為 column vectors 所得的 $m\times n$ matrix A 會不同, 但由這些 A 所得的 projection matrix 都會相同.

Question 7.4. 在 *Example 7.1.17* 中對於 W 的 *projective matrix* 為何?

其實 Theorem 7.1.16 求 projection 的方法還是很繁瑣, 主要是在求 projection matrix 的過程中需求 $(A^{\mathrm{T}}A)^{-1}$ 這一個反矩陣. 將來我們會介紹另一個較簡捷可以幫助我們求 projection 的方法.

7.2. Inconsistent Systems

在這一節中我們將利用 projection 的概念處理聯立方程組無解的情況. 也就是當一個 linear system 為 inconsistent 時, 我們探討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統計所學二維資料的最適合直線.

內積的性質 (2) 告訴我們除了零向量 $\mathbf{O}$ 以外, 其餘向量 $\mathbf{v}\in V$ 皆需符合 $\langle\mathbf{v},\mathbf{v}\rangle>0$, 所以很自然地我們可依此定義向量的長度 $\|\mathbf{v}\|=\sqrt{\langle\mathbf{v},\mathbf{v}\rangle}$. 在我們熟悉的 $\mathbb{R}^3$ 空間中. 一個點在一個平面的投影會是該平面上距離該點最近的點. 這個性質對於推廣到一般 inner product space 的 projection 也是對的. 事實上我們有以下的結果

Proposition 7.2.1. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{v}\in\mathbb{R}^m$. 若 $\mathbf{w}$ 為 $\mathbf{v}$ 在 W 的投影, 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 $\|\mathbf{v}-\mathbf{w}'\|>\|\mathbf{v}-\mathbf{w}\|$. □

當 $A\in\mathscr{M}_{m\times n}$, 我們都知道聯立方程組 $A\mathbf{x}=\mathbf{b}$ 有解, 若且唯若 $\mathbf{b}\in C(A)$. 因此, 一般在日常應用中, 當我們要處理的聯立方程組 $A\mathbf{x}=\mathbf{b}$ 是 inconsistent 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 $C(A)$ 這一個 subspace, 所以依 Proposition 7.2.1 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 $C(A)$ 的 projection. 也因此由 projection 的定義知, $\mathbf{b}-\mathbf{b}_0\in C(A)^\perp=N(A^\mathrm{T})$. 此即表示 $A^\mathrm{T}(\mathbf{b}-\mathbf{b}_0)=\mathbf{O}$, 也就是

$$A^\mathrm{T}\mathbf{b}_0=A^\mathrm{T}\mathbf{b}. \tag{7.1}$$

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

$$A^\mathrm{T}A\mathbf{x}=A^\mathrm{T}\mathbf{b}. \tag{7.2}$$

注意式子 (7.2) 和式子 (7.1) 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 $A^\mathrm{T}A\mathbf{x}=A^\mathrm{T}\mathbf{b}$. 不過我們必須説明式子 (7.2) 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 $\mathbf{b}_0\in C(A)$ 且 $\mathbf{b}_0-\mathbf{b}\in N(A^\mathrm{T})=C(A)^\perp$. 因此得證 $\mathbf{b}_0$ 就是 $\mathbf{b}$ 在 $C(A)$ 的 projection, 故由 Proposition 7.2.1 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 (7.2) 一定有解嗎?

Proposition 7.2.2. *假設 $A\in\mathscr{M}_{m\times n}$ 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. 特別若 $\operatorname{rank}(A)=n$, 則聯立方程組有解且解唯一.*

Proof. 令 $\mathbf{b}_0$ 為 $\mathbf{b}$ 在 $C(A)$ 的 projection, 亦即 $\mathbf{b}_0\in C(A)$ 且 $\mathbf{b}-\mathbf{b}_0\in C(A)^\perp$. 因 $\mathbf{b}_0\in C(A)$ 故存在 $\mathbf{x}_0\in\mathbb{R}^n$ 使得 $A\mathbf{x}_0=\mathbf{b}_0$. 又 $\mathbf{b}-\mathbf{b}_0\in C(A)^\perp$ 而 $C(A)^\perp=N(A^\mathrm{T})$ (Corollary 7.1.8) 故得

$$\mathbf{O}=A^\mathrm{T}(\mathbf{b}-\mathbf{b}_0)=A^\mathrm{T}\mathbf{b}-A^\mathrm{T}\mathbf{b}_0=A^\mathrm{T}\mathbf{b}-A^\mathrm{T}(A\mathbf{x}_0),$$

故知 $\mathbf{x}_0$ 確實滿足 $A^\mathrm{T}A\mathbf{x}_0=A^\mathrm{T}\mathbf{b}$, 因此聯立方程組 $A^\mathrm{T}A\mathbf{x}=A^\mathrm{T}\mathbf{b}$ 一定有解.

當 $\operatorname{rank}(A)=n$, 由 Lemma 7.1.15 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. □

注意當 $\operatorname{rank}(A)<n$ 時, 聯立方程組 $A^\mathrm{T}A\mathbf{x}=A^\mathrm{T}\mathbf{b}$ 雖然有解不過其解因等同於聯立方程 $A\mathbf{x}=\operatorname{Proj}_{C(A)}(\mathbf{b})$ 的解, 故由 A 為 $m\times n$ matrix 以及 Theorem 3.4.6 知其解不唯一 (會有無窮多解). 我們特別有以下的定義.

Definition 7.2.3. 假設 $A\in\mathscr{M}_{m\times n}$ 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 equations*. 在 normal equations 的解集合中, 長度最短的解稱為 $A\mathbf{x}=\mathbf{b}$ 的 *least squares solution*.

特別地, 當 $\operatorname{rank}(A)=n$ 時, 聯立方程組 $A^\mathrm{T}A\mathbf{x}=A^\mathrm{T}\mathbf{b}$ 的唯一解 $\mathbf{x}=(A^\mathrm{T}A)^{-1}A^\mathrm{T}\mathbf{b}$, 就是 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution.

這裡要強調的是, 在解 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution, 我們是求 normal equation 的解, 而不是要求 $\mathbf{b}$ 在 $C(A)$ 的 projection. 所以在解出 least squares solution 後, 是不必再在解的左邊乘上 A 的. 另外即使聯立方程組 $A\mathbf{x}=\mathbf{b}$ 為 consistent, 即 $\mathbf{b}\in C(A)$, 此時 $\mathbf{b}$ 在 $C(A)$ 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. 因此在處理 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution 的問題時, 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 $A\mathbf{x}=\mathbf{b}$ 的解即可.

Example 7.2.4. 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 $\mathbf{b}=\begin{bmatrix}2\\0\\1\\4\end{bmatrix}$, 我們要找到此方程組的 least squares solution. 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

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

解得 $\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-1\\1\end{bmatrix}$ 為 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution. 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 $C(A)$ 的 projection (參見 Example 7.1.17).

Question 7.5. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $W=\mathrm{Span}(\mathbf{w}_1,\dots,\mathbf{w}_n)$. 令 A 為以 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 為 *column vector* 的 $m\times n$ *matrix*. 試說明對於任意 $\mathbf{v}\in\mathbb{R}^m$ 若 $\mathbf{x}=\mathbf{x}_0\in\mathbb{R}^n$ 為 $A\mathbf{x}=\mathbf{v}$ 的 *normal equations* $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{v}$ 的一組解, 則 $\mathbf{v}$ 在 W 的 *projection* 為 $A\mathbf{x}_0$. 又當 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 不是 *linearly independent* 時, *normal equations* 的解不唯一, 試問這會不會和 $\mathrm{Proj}_W(\mathbf{v})$ 的唯一性相衝突? 又此時對於 W 的 *projection matrix* P_W 是否仍為 $A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$?

有關 least squares solution 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 $(x_1,y_1),\dots,(x_m,y_m)$, 我們希望找到一條直線 $y=ax+b$, 希望這些點 (x_i,y_i) 盡可能地靠近這條直線. 注意這裡 $x_1,\dots,x_m$ 以及 $y_1,\dots,y_m$ 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 a 以及 y 截距 b. 依聯立方程組的觀點來看, 我們希望找到 a,b 使之符合

$$\begin{array}{rcl}ax_1+b&=&y_1\\ax_2+b&=&y_2\\&\vdots&\\ax_m+b&=&y_m.\end{array}$$

換句話說我們要解聯立方程組 $A\mathbf{x}=\mathbf{b}$, 其中 $A=\begin{bmatrix}x_1&1\\x_2&1\\\vdots&\vdots\\x_m&1\end{bmatrix}$, $\mathbf{x}=\begin{bmatrix}a\\b\end{bmatrix}$, $\mathbf{b}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix}$. 當然了, 這些給定的資料 $(x_1,y_1),\dots,(x_m,y_m)$ 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 least squares solution, 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為資料是要探討

x_i, y_i 之間的關係, 所以這些資料中 $x_1, \ldots, x_m$ 是不會全相同的 (否則 $(x_1,y_1),\ldots,(x_m,y_m)$ 會同時落在一鉛直線上, 也沒甚麼好探討的了), 所以這裡 $\operatorname{rank}(A)=2$, 因此由 Proposition 7.2.2 知 normal equations $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$ 一定有解. 接下來我們要說明的是, 這樣所得的直線其意義為何?

給定 $a,b\in\mathbb{R}$, 對於 $i=1,\ldots,m$, 令 $y'_i=ax_i+b$. 也就是說 (x_i,y_i) 會在直線 $y=ax+b$ 上. 為方便起見令 $\mathbf{b}'=\begin{bmatrix} y'_1\\ y'_2\\ \vdots\\ y'_m\end{bmatrix}$, 此時 $\mathbf{b}'=\begin{bmatrix} x_1 & 1\\ x_2 & 1\\ \vdots & \vdots\\ x_m & 1\end{bmatrix}\begin{bmatrix} a\\ b\end{bmatrix}\in C(A)$. 又令 $\varepsilon_i=y'_i-y_i$, 此即將直線 $y=ax+b$ 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 $\varepsilon_1^2+\cdots+\varepsilon_m^2$ 的值越小越好. 然而依定義 $\varepsilon_1^2+\cdots+\varepsilon_m^2$ 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 $A\mathbf{x}=\mathbf{b}$ 解出的 least squares solution $\mathbf{x}=\begin{bmatrix} a\\ b\end{bmatrix}$, 這裡的 a,b 就是讓直線 $y=ax+b$ 所估計得的 $\mathbf{b}'$ 會是 $\mathbf{b}$ 在 $C(A)$ 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 $\varepsilon_1^2+\cdots+\varepsilon_m^2$ 最小的要求. 也因此利用 least squares solution 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 7.2.5. 考慮二維資料 $(-1,0),(1,1),(2,3)$ 我們要找出此資料的 least square line $y=ax+b$.

原先要解的方程組是

$$\begin{aligned} -1a+b &= 0\\ 1a+b &= 1\\ 2a+b &= 3, \end{aligned}$$

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

$$\begin{bmatrix} -1 & 1 & 2\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix} -1 & 1\\ 1 & 1\\ 2 & 1\end{bmatrix}\begin{bmatrix} a\\ b\end{bmatrix}=\begin{bmatrix} -1 & 1 & 2\\ 1 & 1 & 1\end{bmatrix}\begin{bmatrix} 0\\ 1\\ 3\end{bmatrix},$$

即

$$\begin{aligned} 6a+2b &= 7\\ 2a+3b &= 4, \end{aligned}$$

解得 least squares solution 為 $a=13/14, b=5/7$ 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 7.2.6. 考慮二維資料 $(x_1,y_1),\ldots,(x_m,y_m)$. 令 $y=ax+b$ 為其 *least squares line* 且對於 $i=1,\ldots,m$ 令 $y'_i=ax_i+b$. 則有以下知性質:

(1) $y_1+\cdots+y_m=y'_1+\cdots+y'_m$.

(2) $x_1y_1+\cdots+x_my_m=x_1y'_1+\cdots+x_my'_m$.

(3) 令 $\overline{x}$ 和 $\overline{y}$ 分別為 $x_1,\ldots,x_m$ 以及 $y_1,\ldots,y_m$ 的平均數, 即 $\overline{x}=(x_1+\cdots+x_m)/m$, $\overline{y}=(y_1+\cdots+y_m)/m$. 則點 $(\overline{x},\overline{y})$ 在直線 $y=ax+b$ 上, 即 $\overline{y}=a\overline{x}+b$.

Proof. 為方便起見令 $\mathbf{b}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_m\end{bmatrix},\mathbf{b}'=\begin{bmatrix}y'_1\\y'_2\\\vdots\\y'_m\end{bmatrix}$ 又令 $A=\begin{bmatrix}x_1&1\\x_2&1\\\vdots&\vdots\\x_m&1\end{bmatrix}$, 則依 $\mathbf{x}=\begin{bmatrix}a\\b\end{bmatrix}$ 是 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution 得 $\mathbf{b}'=A\begin{bmatrix}a\\b\end{bmatrix}$ 為 $\mathbf{b}$ 在 $C(A)$ 的 projection. 即 $\mathbf{b}-\mathbf{b}'\in C(A)^{\perp}=N(A^{\mathrm{T}})$, 故得

$$A^{\mathrm{T}}(\mathbf{b}-\mathbf{b}')=\begin{bmatrix}x_1&\cdots&x_m\\1&\cdots&1\end{bmatrix}\begin{bmatrix}y_1-y'_1\\y_2-y'_2\\\vdots\\y_m-y'_m\end{bmatrix}=\begin{bmatrix}x_1(y_1-y'_1)+\cdots+x_m(y_m-y'_m)\\(y_1-y'_1)+\cdots+(y_m-y'_m)\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}.$$

得證 (1), (2).

令 $\overline{y'}=(y'_1+\cdots+y'_m)/m$, 由 (1) 知 $\overline{y}=\overline{y'}$. 又由於對所有 $i=1,\ldots,m$, 皆有 $y'_i=ax_i+b$, 故知 $\overline{y'}=a\overline{x}+b$. 得證 (3), 即 $\overline{y}=a\overline{x}+b$. □

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.

7.3. Orthogonal Basis and Gram-Schmidt Process

我們曾經介紹兩種求 $\mathbb{R}^m$ 中的向量 $\mathbf{v}$ 在一個 subspace W 的 projection 的方法. 這兩種方法其實都有一點繁雜. 第一種方法, 即 Theorem 7.1.12, 我們必須先由 W 的 basis 找到 $W^{\perp}$ 的 basis, 再將 $\mathbf{v}$ 寫成這兩組 basis 所形成的 $\mathbb{R}^m$ 的 basis 的 linear combination, 最後再刪除 $W^{\perp}$ 的部分, 就是 $\mathbf{v}$ 在 W 的 projection 了. 這裡最麻煩的是將 $\mathbf{w}$ 寫成這一組 basis 的 linear combination, 一般來說牽涉到解聯立方程組的問題. 另外好不容易寫成 linear combination 後又要將一部分捨去, 有點多餘做虛功的感覺. 另一種方法, 即 Theorem 7.1.16, 我們必須寫下以 W 的一組 basis 為 column vectors 的矩陣 A, 然後利用它得到對 W 投影的 projection matrix. 這裡較麻煩的地方就是, 我們必須求出 $A^{\mathrm{T}}A$ 的 inverse. 上述兩種方法都牽涉到 W 的一組 basis. 不過我們知道不管選哪一組 basis, 所得的結果都會一樣. 是否有可能找到一組特殊的 basis, 可以讓上述兩種方法較繁雜的部分變簡單呢? 在這一節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 $\mathbb{R}^n$ 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

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

因為 $\mathbf{w}_i\neq\mathbf{O}$, 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 7.3.1. 假設 W 為 $\mathbb{R}^n$ 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為其一組 *basis* 滿足 $\langle \mathbf{w}_i,\mathbf{w}_j\rangle = 0,\forall i\neq j$. 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 7.3.2. 假設 W 是 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 basis 滿足對於任意 $i\neq j$ 皆有 $\langle \mathbf{w}_i,\mathbf{w}_j\rangle=0$. 則稱 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthogonal basis*. 若又要求 $\langle \mathbf{w}_i,\mathbf{w}_i\rangle=1$, $\forall i=1,\ldots,n$, 則稱 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthonormal basis*.

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

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

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

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 Theorem 7.1.12 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 $\mathbf{v}\in\mathbb{R}^m$ 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 $c_i=\langle \mathbf{v},\mathbf{w}_i\rangle/\langle \mathbf{w}_i,\mathbf{w}_i\rangle$. 所以我們得到一個比 Theorem 7.1.12 更簡捷求 projection 的方法.

Theorem 7.3.3. 假設 W 為 $\mathbb{R}^m$ 的 *subspace* 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 *orthogonal basis*. 若 $\mathbf{v}\in\mathbb{R}^m$, 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

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

由 Theorem 7.3.3, 我們知道只要找到 W 的一組 orthogonal basis, 不只不必去解聯立方程組而且不必去求 $W^\perp$ 的一組 basis, 確實省去了許多麻煩. 同樣的, 我們也可利用 W 的一

組 orthogonal basis $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 更簡便的得到對於 W 的 projection matrix. 事實上令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vectors 的 $m\times n$ matrix, 則因

$$A^{\mathrm{T}}A=\begin{bmatrix} - & \mathbf{w}_1^{\mathrm{T}} & - \\ - & \mathbf{w}_2^{\mathrm{T}} & - \\ & \vdots & \\ - & \mathbf{w}_n^{\mathrm{T}} & - \end{bmatrix}\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}$$

其 (i,j)-th entry 為 $\mathbf{w}_i^{\mathrm{T}}\mathbf{w}_j=\langle\mathbf{w}_i,\mathbf{w}_j\rangle$, 故由 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 orthogonal, 我們得 $A^{\mathrm{T}}A$ 為 diagonal matrix, 也因此很容易求得 $(A^{\mathrm{T}}A)^{-1}$, 亦即

$$A^{\mathrm{T}}A=\begin{bmatrix} \|\mathbf{w}_1\|^2 & & & \\ & \|\mathbf{w}_2\|^2 & & \\ & & \ddots & \\ & & & \|\mathbf{w}_n\|^2 \end{bmatrix},\quad (A^{\mathrm{T}}A)^{-1}=\begin{bmatrix} \frac{1}{\|\mathbf{w}_1\|^2} & & & \\ & \frac{1}{\|\mathbf{w}_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|\mathbf{w}_n\|^2} \end{bmatrix}.$$

這也克服了利用 Theorem 7.1.16 求 projection 較複雜的部分, 我們有以下的結果.

Theorem 7.3.4. *假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 orthogonal basis. 則對於任意 $\mathbf{v}\in\mathbb{R}^m$, 皆有 $\mathrm{Proj}_W(\mathbf{v})=P_W\mathbf{v}$, 其中 P_W 為 $m\times m$ matrix*

$$P_W=\frac{1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}}+\frac{1}{\|\mathbf{w}_2\|^2}\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}}+\cdots+\frac{1}{\|\mathbf{w}_n\|^2}\mathbf{w}_n\mathbf{w}_n^{\mathrm{T}}.$$

Proof. 首先我們強調這裏我們將 $\mathbf{w}_i$ 視為 $m\times 1$ matrix, 所以 $\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}}$ 會是 $m\times m$ matrix. 令 A 為以 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 column vectors 的 $m\times n$ matrix. 利用 Theorem 7.1.16 我們知道 $P_W=A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. 然而因 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 orthogonal, 由前述結果知

$$A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}=\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \frac{1}{\|\mathbf{w}_1\|^2} & & & \\ & \frac{1}{\|\mathbf{w}_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|\mathbf{w}_n\|^2} \end{bmatrix}\begin{bmatrix} - & \mathbf{w}_1^{\mathrm{T}} & - \\ - & \mathbf{w}_2^{\mathrm{T}} & - \\ & \vdots & \\ - & \mathbf{w}_n^{\mathrm{T}} & - \end{bmatrix} \tag{7.3}$$

因此我們只要檢查 $m\times m$ matrix

$$\frac{1}{\|\mathbf{w}_1\|^2}\mathbf{w}_1\mathbf{w}_1^{\mathrm{T}}+\frac{1}{\|\mathbf{w}_2\|^2}\mathbf{w}_2\mathbf{w}_2^{\mathrm{T}}+\cdots+\frac{1}{\|\mathbf{w}_n\|^2}\mathbf{w}_n\mathbf{w}_n^{\mathrm{T}} \tag{7.4}$$

確實與 (7.3) 的 $m\times m$ matrix 為相同的矩陣.

對於 $i=1,\ldots,m$, 令 $\mathbf{e}_i\in\mathbb{R}^m$, 為 i-th entry 為 1 其他 entry 為 0 的 vector. 我們知道任何 $m\times m$ matrix 右邊乘上 $\mathbf{e}_i$ 便會得到此 matrix 的 i-th column. 所以若能證明 (7.3) 和 (7.4) 這兩個矩陣的右邊乘上 $\mathbf{e}_i$ 會相等, 便證得 (7.3) 和 (7.4) 為相同的矩陣, 因而得證本定理. 然而 (7.3) 的右邊乘上 $\mathbf{e}_i$ 會等於

$$\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \frac{1}{\|\mathbf{w}_1\|^2} & & & \\ & \frac{1}{\|\mathbf{w}_2\|^2} & & \\ & & \ddots & \\ & & & \frac{1}{\|\mathbf{w}_n\|^2} \end{bmatrix}\begin{bmatrix} \langle\mathbf{w}_1,\mathbf{e}_i\rangle \\ \langle\mathbf{w}_2,\mathbf{e}_i\rangle \\ \vdots \\ \langle\mathbf{w}_n,\mathbf{e}_i\rangle \end{bmatrix}=\begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} \frac{\langle\mathbf{w}_1,\mathbf{e}_i\rangle}{\|\mathbf{w}_1\|^2} \\ \frac{\langle\mathbf{w}_2,\mathbf{e}_i\rangle}{\|\mathbf{w}_2\|^2} \\ \vdots \\ \frac{\langle\mathbf{w}_n,\mathbf{e}_i\rangle}{\|\mathbf{w}_n\|^2} \end{bmatrix}$$

此恰等於 (7.4) 的右邊乘上 $\mathbf{e}_i$ 的結果
$$\frac{\langle \mathbf{w}_1, \mathbf{e}_i \rangle}{\|\mathbf{w}_1\|^2}\mathbf{w}_1 + \frac{\langle \mathbf{w}_2, \mathbf{e}_i \rangle}{\|\mathbf{w}_2\|^2}\mathbf{w}_2 + \cdots + \frac{\langle \mathbf{w}_n, \mathbf{e}_i \rangle}{\|\mathbf{w}_n\|^2}\mathbf{w}_n.$$
$\square$

Question 7.6. 試由 $\mathrm{Proj}_W : \mathbb{R}^m \to \mathbb{R}^m$ 是 *linear transformation* 以及此 *linear transformation* 的 *standard matrix representation* 就是 P_W, 利用 *Theorem 7.3.3* 證明 *Theorem 7.3.4.*

我們已經知道只要找到 W 的一組 orthogonal basis, 就可以輕易求得 $\mathbf{v} \in \mathbb{R}^m$ 在 W 的 projection. 對一般 $\mathbb{R}^m$ 的 nonzero subspace, 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 $\mathbb{R}^m$ 的 nonzero subspace 一定存在 orthogonal basis (以及 orthonormal basis). 這個方法就是所謂的 *Gram-Schmidt process.*

給定 $\mathbb{R}^m$ 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 $c_1\mathbf{w}_1 + \cdots + c_k\mathbf{w}_k = \mathbf{O}$. 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 $\dim(W) = n$, 故由 Corollary 4.3.5 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 $W \neq \{\mathbf{O}\}$, 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 $\mathbf{v}_2 \notin W_1$. 現在我們要利用 $\mathbf{v}_2$, 找到 $\mathbf{w}_2 \in W$ 滿足 $\mathbf{w}_2 \neq \mathbf{O}$ 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 $\mathbf{w}_2 \neq \mathbf{O}$. 這是因為若 $\mathbf{w}_2 = \mathbf{O}$, 會得到 $\mathbf{v}_2 = \mathrm{Proj}_{W_1}(\mathbf{v}_2) \in W_1$, 此與當初 $\mathbf{v}_2 \notin W_1$ 的假設相矛盾. 另一方面因為 $W = \mathrm{Span}(\mathbf{w}_1)$, 利用 Proposition 1.4.9 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 $\mathbf{v}_3 \notin W_2$. 現在我們要利用 $\mathbf{v}_3$, 找到 $\mathbf{w}_3 \in W$ 滿足 $\mathbf{w}_3 \neq \mathbf{O}$ 且 $\langle \mathbf{w}_1, \mathbf{w}_3 \rangle = \langle \mathbf{w}_2, \mathbf{w}_3 \rangle = 0$. 同前, 我們考慮 $\mathbf{w}_3 = \mathbf{v}_3 - \mathrm{Proj}_{W_1}(\mathbf{v}_2)$, 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 $\langle \mathbf{w}_1, \mathbf{w}_3 \rangle = \langle \mathbf{w}_2, \mathbf{w}_3 \rangle = \mathbf{O}$. 另外因 $\mathbf{v}_3 \notin W_2$, 同前面的理由我們有 $\mathbf{w}_3 \neq \mathbf{O}$. 另一方面因為 $\mathbf{w}_1, \mathbf{w}_2$ 為 W_2 的 orthogonal basis, 利用 Theorem 7.3.3

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

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

Theorem 7.3.5 (Gram-Schmidt Process)**.** *假設 W 為 $\mathbb{R}^m$ 的 subspace 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 W 的一組 basis. 令*

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

這樣一直下去, 即對於 $i = 1,\ldots,n-1$ 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 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 為 W 的一組 orthogonal basis. 而且

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

Example 7.3.6. 我們要用 orthogonal basis 來處理 projection 的問題. 回顧在 Example 7.1.17 我們要求 $\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 4 \end{bmatrix}$ 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 7.3.3 得

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

Example 7.3.7. 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. 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

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

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

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

假設 $A\in\mathscr{M}_{m\times n}$ 且 $\mathrm{rank}(A)=n$, 則 A 的 column vectors 形成 $C(A)$ 的一組 basis. 假設 A 的 column 分別為 $\mathbf{v}_1,\dots,\mathbf{v}_n$, 我們利用 Gram-Schmidt process 得到 $C(A)$ 的一組 orthonormal basis $\mathbf{u}_1,\dots,\mathbf{u}_n$. 由於 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 是 orthonormal basis, 我們有

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

然而在 Gram-Schmidt process, 對於 $i=1,\dots,n$, 我們都有 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_i)=\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_i)$, 所以由 $\mathbf{u}_{i+1},\dots,\mathbf{u}_n\in\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_i)^\perp$ 知

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

另外由 $\mathbf{v}_i\not\in\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_{i-1})=\mathrm{Span}(\mathbf{u}_1,\dots,\mathbf{u}_{i-1})$, 我們知 $\langle\mathbf{v}_i,\mathbf{u}_i\rangle\neq 0$. 因此若令 Q 為以 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 column vectors 的 $m\times n$ matrix, 依矩陣乘法定義我們有 $A=QR$, 其中 R 的 (i,j)-th entry 為 $\langle\mathbf{v}_i,\mathbf{u}_j\rangle$. 故由前面所述, 當 $j>i$ 時 $\langle\mathbf{v}_i,\mathbf{u}_j\rangle=0$, 我們知 R 為 $n\times n$ upper triangular matrix (上三角矩陣), 而且對角線的位置 (i,i)-th entry 為 $\langle\mathbf{v}_i,\mathbf{u}_i\rangle\neq 0$, 我們得 $\mathrm{rank}(R)=n$, 故 R 為 invertible. 這就是所謂 A 的 QR decomposition. 我們再用一個例子來說明.

Example 7.3.8. 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 7.3.7 的 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$, 我們直接套用其結果得

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

很容易檢查, 我們確有 $A=QR$.

QR decomposition 可以簡化我們處理 least squares solution 的問題. 假設我們要處理聯立方程組 $A\mathbf{x}=\mathbf{b}$, 其中 $A\in\mathscr{M}_{m\times n}$ 且 $\operatorname{rank}(A)=n$. 回顧意下, 要求 $A\mathbf{x}=\mathbf{b}$ 的 least squares solution, 就是要解 normal equations $A^{\mathrm{T}}A\mathbf{x}=A^{\mathrm{T}}\mathbf{b}$. 現若將 A 寫成 QR decomposition $A=QR$, 我們就是要解 $(QR)^{\mathrm{T}}QR\mathbf{x}=(QR)^{\mathrm{T}}\mathbf{b}$, 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 $n\times n$ diagonal matrix, 且其 (i,i)-th entry 為 $\|\mathbf{u}_i\|^2=1$. 也就是說 $Q^{\mathrm{T}}Q$ 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.

7.4. 結論

我們介紹了在一個 inner product space 中的 subspace W 其 orthogonal complement $W^{\perp}$ 的概念, 並利用它推廣了過去在 $\mathbb{R}^2$, $\mathbb{R}^3$ 投影的概念到一般的 inner product space. 有了投影的概念, 我們可以對於一個 inconsistent 的聯立方成組提出了最佳解的概念, 即 least square solution. 我們提出了三種得到 projection 的方法, 其中最有效的方法就是利用 Gram-Schmidt process 找到要投影的 subspace 的一組 orthogonal basis (或 orthonormal basis). 找到 subspace 的 orthogonal basis 的另一個好處是讓我們很容易將此 subspace 中任意的向量, 寫成這組 orthogonal basis 的 linear combination.

Part IV

Square Matrix

在這最後一部分我們將探討 square matrix 的特性. 其實 sqaure matrix 和所謂 linear operator 是息息相關的, 不過在此我們無法多作介紹, 而將相關課題置於較進階的“線性代數再探”再做討論. 首先我們介紹 square matrix 上特有的函數 determinant. 雖然 determinant 可以視作有向面積, 有向體積的推廣. 不過它對我們了解一個 square matrix 是否為 invertible 給了一個量化的測試方法. 這也幫助我們求得一個 square matrix 的 characteristic polynomial, 以便了解它有哪些 eigenvalues, 更進一步求出它的 eigenvectors. 經由最後一章 eigenvalue 以及 eigenvector 的介紹, 我們要學會哪些 square matrix 是可以對角化的. 我們還會知道 symmetric matrix 一定可以對角化, 並利用它來處理二次曲線, 二次曲面的問題.

Chapter 8

Determinant

這一章我們將介紹 determinant (行列式). 對於一個 $n\times n$ matrix, 我們將它的行列式視為其 row vectors 所張成的平行多面體的 "有向體積". 利用這個看法, 我們探討 determinant 應有的性質, 再利用這些性質來得到 determinant 的定義. 這一章中, 我們探討的 $\mathbb{R}^n$ 向量大多以 row vector 來表示, 除非與矩陣乘法有關才會用 column vector 來表示.

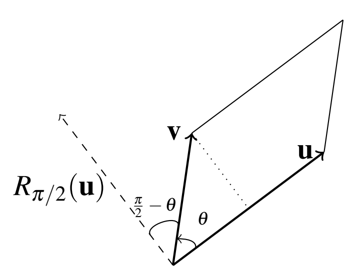
8.1. Signed Area in $\mathbb{R}^2$ and Properties of Determinant Function

我們都知道當 A 為 2×2 matrix, 其行列式 $\det(A)$ 的定義, 我們可以將它視為 A 的兩個 row vectors 所張出的平行四邊形的 *sign area*. 我們將用這個看法將之推廣到一般 $n\times n$ matrix 的情形並探討 determinant 應具有的性質.

考慮 $A=\begin{bmatrix} a & b\\ c & d\end{bmatrix}$, A 的 determinant 定義為 $\det(A)=ad-bc$. 考慮 $\mathbb{R}^2$ 中兩向量 $\mathbf{u}=(a,b)$, $\mathbf{v}=(c,d)$. 當 $\mathbf{u}$ 為零向量, 表示 $a=0,b=0$, 此時我們有 $\det(A)=0$. 而當 $\mathbf{u}\neq\mathbf{O}$, 此時將 $\mathbf{u}$ 逆時針轉 $\pi/2$ 角得 $R_{\pi/2}(\mathbf{u})=(-b,a)$. 因此我們有

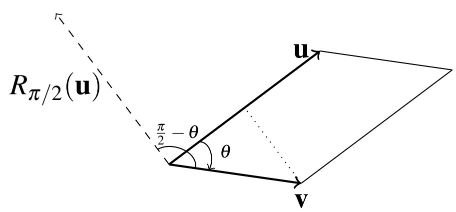
$$R_{\pi/2}(\mathbf{u})\cdot\mathbf{v}=(-b,a)\cdot(c,d)=ad-bc=\det(A). \tag{8.1}$$

假設 $\mathbf{u}$ 到 $\mathbf{v}$ 的角度為 θ 且 $0\leq\theta\leq\pi/2$ (如下圖所示). 此時 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形, 以 $\mathbf{u}$ 為底的話, 此平行四邊形的高為 $\|\mathbf{v}\|\sin\theta$. 故平行四邊形的面積為 $\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta$. 由於 $R_{\pi/2}(\mathbf{u})$ 是將 $\mathbf{u}$ 逆時針轉 $\pi/2$ 角, 我們有 $\|R_{\pi/2}(\mathbf{u})\|=\|\mathbf{u}\|$ 且 $R_{\pi/2}(\mathbf{u})$ 與 $\mathbf{v}$ 的夾角為 $(\pi/2)-\theta$. 因為 $\sin\theta=\cos((\pi/2)-\theta)$, 我們得 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形面積為 $\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta=\|\mathbf{u}\|\|\mathbf{v}\|\cos((\pi/2)-\theta)=\|R_{\pi/2}(\mathbf{u})\|\|\mathbf{v}\|\cos((\pi/2)-\theta)=R_{\pi/2}(\mathbf{u})\cdot\mathbf{v}=\det(A)$.



而若 $\mathbf{u}$ 到 $\mathbf{v}$ 的角度為 θ 且 $\pi/2<\theta\leq\pi$, 則 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形, 以 $\mathbf{u}$ 為底的話, 其高仍為 $\|\mathbf{v}\|\sin\theta$. 此時 $R_{\pi/2}(\mathbf{u})$ 與 $\mathbf{v}$ 的夾角為 $\theta-(\pi/2)$. 因此由 $\sin\theta=\cos(\theta-(\pi/2))$ 我們仍得 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形面積為 $\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta=\|\mathbf{u}\|\|\mathbf{v}\|\cos(\theta-(\pi/2))=R_{\pi/2}(\mathbf{u})\cdot\mathbf{v}=\det(A)$.

現假設 $\mathbf{u}$ 到 $\mathbf{v}$ 的角度為 θ 且 $-\pi/2\leq\theta<0$ (如下圖所示). 此時 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形, 以 $\mathbf{u}$ 為底的話, 此平行四邊形的高為 $-\|\mathbf{v}\|\sin\theta$. 故平行四邊形的面積為 $-\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta$. 由於此時 $R_{\pi/2}(\mathbf{u})$ 與 $\mathbf{v}$ 的夾角為仍為 $(\pi/2)-\theta$. 我們得 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形面積為 $-\|\mathbf{u}\|\|\mathbf{v}\|\sin\theta=-\|\mathbf{u}\|\|\mathbf{v}\|\cos((\pi/2)-\theta)=-R_{\pi/2}(\mathbf{u})\cdot\mathbf{v}=-\det(A)$.



同理當 $\mathbf{u}$ 到 $\mathbf{v}$ 的角度為 θ 且 $-\pi<\theta<-\pi/2$ 時, $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形面積為 $-\det(A)$.

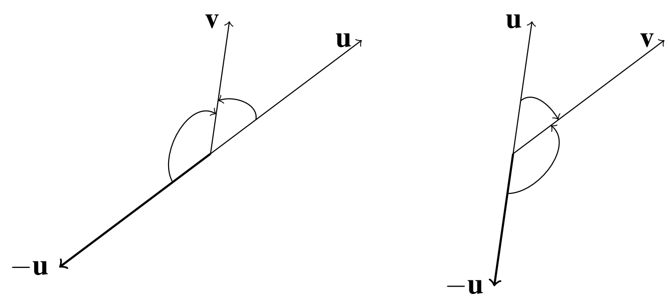
由以上的討論, 我們知當 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, 令 $\mathbf{u}=(a,b)$, $\mathbf{v}=(c,d)$. 此時 $\det(A)$ 的絕對值會是 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形的面積. 而 $\det(A)$ 為正的表示 $\mathbf{v}$ 在 $\mathbf{u}$ 的逆時針方向. 反之, $\det(A)$ 為負的表示 $\mathbf{v}$ 在 $\mathbf{u}$ 的順時針方向. 也就是說 $\det(A)$ 不只告訴我們有關於 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形的面積, 且告訴我們 $\mathbf{u},\mathbf{v}$ 之間的方向性. 因此我們稱 $\det(A)$ 為 $\mathbf{u},\mathbf{v}$ 所張成的平行四邊形的 *signed area*.

我們希望將此推廣到 $n\times n$ matrix. 也就是說當 A 為 $n\times n$ matrix, 令 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in\mathbb{R}^n$ 依序為 A 的 row vectors. 我們希望能定義 $\det(A)$ 使其值為 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 在 $\mathbb{R}^n$ 所張成的“平行多面體”的 *signed volume*. 也就是說, 希望 $\det(A)$ 的絕對值表示 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 在 $\mathbb{R}^n$ 所張成的“平行多面體”的體積, 而其正負號表示的是 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的方向性. 或許大家會疑惑? 當 $n\geq4$ 時, $\mathbb{R}^n$ 中 n 個向量所張的“平行多面體”是什麼樣子都不知道, 要如何去說它的體積呢? 沒錯, 我們就是希望能延伸 $\mathbb{R}^2$ 的平行四邊形面積的概念到 $\mathbb{R}^3$ 的平行六面體體積. 然後希望能一直延伸下去定義出一般 $\mathbb{R}^n$ 中 n 個向量所張的平行多面體的體積. 簡言之, 我們想利用預期一個體積應該符合哪些性質的方法, 來定義出體積. 所以接下來的工作就是列出幾個和 signed area 相關的性質, 希望能定義出 determinant (行列式) 這一個從 $\mathscr{M}_{n\times n}$ 到 $\mathbb{R}$ 的函數 (用 det 表示), 使得它符合這些性質.

首先我們要定義體積, 應該先定義單位體積為何才能確定體積. 在 $\mathbb{R}^2$ 中我們是因為定義了 $(1,0),(0,1)$ 所張的平行四邊形 (其實是正方形) 的面積為 1, 才得到其他平行四邊形的面積. 又 $(1,0)$ 到 $(0,1)$ 確實是逆時鐘轉 $\pi/2$, 依前面的方向性應為正向. 所以 $\det(I_2)=1$ 確實符合我們要求 $(1,0),(0,1)$ 所張的平行四邊形的面積為 1 且為正向的要求. 因此要決定 $\mathbb{R}^n$ 中的單位體積, 很自然的我們會定其 standard basis $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 所張的平行多面體的體積為 1, 且要求 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 這樣的方向性就是正向. 也就是說我們希望 $\mathbf{e}_1,\ldots,\mathbf{e}_n$ 所張的平行多面體的 signed volume 為 1, 亦即我們定 $\det(I_n)=1$.

至於一般 $\mathbb{R}^n$ 中 n 個向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的方向性怎麼定呢? 在 $\mathbb{R}^2$ 中, 若 $\mathbf{u},\mathbf{v}$ 為正向, 表示 $\mathbf{v}$ 在 $\mathbf{u}$ 的逆時鐘方向, 此時 $\mathbf{v},\mathbf{u}$ 就是負向, 因為 $\mathbf{u}$ 在 $\mathbf{v}$ 的順時鐘方向. 而 2×2 的 determinant $\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad-bc = -(cd-ad) = -\det\begin{bmatrix} c & d \\ a & b \end{bmatrix}$ 也符合這個性質. 因此我們認為 $\mathbb{R}^n$ 中 n 個向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 若將其中兩個相鄰向量 $\mathbf{v}_i,\mathbf{v}_{i+1}$ 變換順序就會改變方向性. 也就是說當 $A\in\mathscr{M}_{n\times n}$, 若將 A 的相鄰兩個 row 交換所得的矩陣為 A', 則我們要求 $\det(A') = -\det(A)$.

另一個改變 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的方向性的可能就是將其中一個 $\mathbf{v}_i$ 改為 $-\mathbf{v}_i$. 例如在 $\mathbb{R}^2$ 中, 若 $\mathbf{u},\mathbf{v}$ 為正向, 則 $-\mathbf{u},\mathbf{v}$ 就是負向. 反之, 若 $\mathbf{u},\mathbf{v}$ 為負向, 則 $-\mathbf{u},\mathbf{v}$ 就是正向. 如下圖所示:



而 2×2 的 determinant

$$\det\begin{bmatrix} -a & -b \\ c & d \end{bmatrix} = \det\begin{bmatrix} a & b \\ -c & -d \end{bmatrix} = -ad+bc = -(ad-bc) = -\det\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

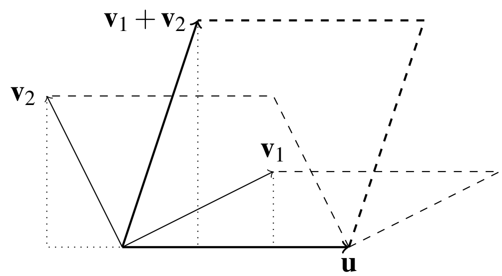
也符合這個性質. 因此我們認為 $\mathbb{R}^n$ 中 n 個向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 若將其中一個 $\mathbf{v}_i$ 改為 $-\mathbf{v}_i$ 就會改變方向性. 也就是說當 $A\in\mathscr{M}_{n\times n}$, 若將 A 的某個 row 乘上 -1 所得的矩陣為 A', 則我們要求 $\det(A') = -\det(A)$.

至於體積我們希望有怎樣的性質呢? 首先若 r 為一個正實數, 平行多邊形若有一邊為原來的 r 倍, 我們認為其體積應也會隨之改變為原來的 r 倍. 而 2×2 的 determinant

$$\det\begin{bmatrix} ra & rb \\ c & d \end{bmatrix} = \det\begin{bmatrix} a & b \\ rc & rd \end{bmatrix} = rad-rbc = r(ad-bc) = r\det\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

也符合這個性質. 因此我們認為當 $r>0$, $\mathbb{R}^n$ 中 n 個向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 若將其中一個 $\mathbf{v}_i$ 改為 $r\mathbf{v}_i$ 就會改變其體積為原來的 r 倍. 也就是說當 $r>0$, 若將 $A\in\mathscr{M}_{n\times n}$ 的某個 row 乘上 r 所得的矩陣為 A', 則我們要求 $\det(A') = r\det(A)$. 而若將 A 的某個 row 乘上 $-r$ 所得的矩陣為 A'', 我們可視為將該 row 先乘上 r 得到 A' 再在 A' 的該 row 乘上 -1, 所以我們要求 $\det(A'') = -\det(A') = -r\det(A)$. 換言之, 不管 r 是正實數或負實數, 若將 A 的某個 row 乘上 r 所得的矩陣為 A', 則我們都要求 $\det(A') = r\det(A)$.

最後如果 $\mathbb{R}^n$ 中 n 個向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 將其中一個向量 $\mathbf{v}_i$ 拆成兩個向量之和, 即 $\mathbf{v}_i = \mathbf{w}_i + \mathbf{w}'_i$, 則我們認為 $\mathbf{v}_1,\ldots,\mathbf{v}_i,\ldots,\mathbf{v}_n$ 所張成的平行多面體其有向體積應為 $\mathbf{v}_1,\ldots,\mathbf{w}_i,\ldots,\mathbf{v}_n$ 所形成的平行多面體和 $\mathbf{v}_1,\ldots,\mathbf{w}'_i,\ldots,\mathbf{v}_n$ 所形成的平行多面體的有向體積之和. 例如在 $\mathbb{R}^2$ 中下圖所示:



注意這裡以 $\mathbf{u}$ 為底, $\mathbf{u},\mathbf{v}_1+\mathbf{v}_2$ 所張的平行四邊形的高為 $\mathbf{u},\mathbf{v}_1$ 所張的平行四邊形和 $\mathbf{u},\mathbf{v}_2$ 所張的平行四邊形的高之和. 而 2×2 的 determinant

$$\det\begin{bmatrix} a+a' & b+b' \\ c & d \end{bmatrix} =$$
$$(a+a')d-(b+b')c=(ad-bc)+(a'd-b'c)=\det\begin{bmatrix} a & b \\ c & d \end{bmatrix}+\det\begin{bmatrix} a' & b' \\ c & d \end{bmatrix},$$

$$\det\begin{bmatrix} a & b \\ c+c' & d+d' \end{bmatrix} =$$
$$a(d+d')-b(c+c')=(ad-bc)+(ad'-bc')=\det\begin{bmatrix} a & b \\ c & d \end{bmatrix}+\det\begin{bmatrix} a & b \\ c' & d' \end{bmatrix}$$

也符合這個性質. 因此我們要求當 A,B,C 三個 $n\times n$ matrix, 其中 A 的 i-th row 是 B 和 C 的 i-th row 之和, 而 A,B,C 其他各 row 皆相等時, $\det(A)=\det(B)+\det(C)$.

我們將上述討論所希望 determinant 應具有的性質總結如下:

(1) $\det(I_n)=1$.

(2) 若將 $n\times n$ matrix A 某相鄰兩個 row 交換所得的矩陣為 A', 則 $\det(A')=-\det(A)$.

(3) 若將 $n\times n$ matrix A 某個 row 乘上非零實數 r 所得的矩陣為 A', 則 $\det(A')=r\det(A)$.

(4) 若 A,B,C 三個 $n\times n$ matrix, 其中 A 的 i-th row 是 B 和 C 的 i-th row 之和, 而 A,B,C 其他各 row 皆相等, 則 $\det(A)=\det(B)+\det(C)$.

注意 (3), (4) 兩個性質, 我們通稱為 determinant 的 *multi-linear* 性質. 千萬不要搞錯, 它並不是說若 $A=B+rC$ 則 $\det(A)=\det(B)+r\det(C)$, 而是說僅有一個 row 寫成線性組合而其他 row 固定不動的情況之下, determinant 可保持該 row 線性組合的關係. 其大致的圖示如下

$$\det\begin{bmatrix} -\mathbf{v}_1- \\ \vdots \\ -\mathbf{v}_i+r\mathbf{v}_i'- \\ \vdots \\ -\mathbf{v}_n- \end{bmatrix}=\det\begin{bmatrix} -\mathbf{v}_1- \\ \vdots \\ -\mathbf{v}_i- \\ \vdots \\ -\mathbf{v}_n- \end{bmatrix}+r\det\begin{bmatrix} -\mathbf{v}_1- \\ \vdots \\ -\mathbf{v}_i'- \\ \vdots \\ -\mathbf{v}_n- \end{bmatrix}.$$

Question 8.1. 試利用 *determinant multi-linear* 的性質將 $\det\begin{bmatrix} ra_1+sa_2 & rb_1+sb_2 \\ tc_1+uc_2 & td_1+ud_2 \end{bmatrix}$ 寫成 $\det\begin{bmatrix} a_i & b_i \\ c_j & d_j \end{bmatrix}$ 的線性組合.

8.2. Uniqueness of the Determinant Function

上一節中我們給了 det 這個函數預期應該擁有的性質, 但是我們不知道這樣的函數存不存在. 因為或許這些性質要求太多, 會互相抵觸造成符合這些性質的函數根本不存在. 也有可能這些性質要求太少, 以至於有很多函數可以符合這些性質. 這一節中我們將探討, 若符合這些性質的函數存在的話, 那它會是唯一的. 也就是說, 不管怎麼去定這個函數如果定出的函數真能符合我們要求的性質, 那它一定就是唯一的那一個.

或許大家會疑惑, 連這個函數存不存在都不知道, 為何要先探討它的唯一性呢? 其實我們在處理數學問題時經常是這樣做的. 例如在解方程式時, 我們都是先假設其解存在, 然後再利用等量公理等方法解出其解可能為那些, 再將這些可能的值代回原方程式看看是否符合, 然後才找到真正的解. 也就是說, 我們不可能將所有的數都代入方程式來找解, 而解方程式的過程是利用若有解的話其解需要具備的性質幫我們縮小範圍找出真正的解來. 現在我們也是一樣, 想先由前一節列出的性質去推導出更多的性質, 然後得到符合這些性質的函數若存在的話僅有一個, 再由此得到這個函數可能的形式, 然後回過來驗證它真的符合我們要的性質.

要注意在本節中, 由於尚未證明 det 是存在的, 所以我們推導出來的性質都是在 det 存在的假設情況才會成立. 從邏輯的角度來看, 這裡推導出來的每個敘述之前都要加上“若 det 存在”這樣的假設條件. 不過以後我們將會證明 det 確實存在, 所以這些敘述事實上是正確的. 因此為了方便起見, 我們都略去“若 det 存在”這樣的假設條件.

首先我們探討 det 在 elementary row operation 之下其取質如何改變. 在 det 要求的性質 (2) 中我們要求當 A 的某相鄰兩個 row 交換其行列式值要變號. 其實這對 A 的任兩個 row 交換也會成立. 這是因為我們可以利用相鄰兩個 row 互換的方法將 A 的 i-th row 和 j-th row 交換. 因為不失一般性, 假設 $i<j$, 我們可以先將 A 的 i-th row 和 $i+1$-th row 互換, 接著將 $i+1$-th row 和 $i+2$-th row 互換, 這樣一直下去直到將原本 i-th row 換到 $j-1$-th row. 注意此時原本 $i+1$-th 到 $j-1$-th 各 row 其實都只是往上移一個 row, 而我們共做了 $(j-1)-i$ 次的相鄰兩個 row 互換的動作. 接下來從 j-th row 開始, 先和 $j-1$-th row 交換 (此時原本的 i-th row 已換到 j-th row), 然後再依序往上用相鄰兩 row 互換的方法將原本的 j-th row 換到 i-th row. 這次由下往上互換的動作從 j-th row 和 $j-1$-th row 交換一直到 $i+1$-th row 和 i-th row 交換共做了 $j-i$ 次的相鄰兩個 row 互換的動作. 所以從上而下再從下而上完成將 i-th row 和 j-th row 互換共做了 $(j-1-i)+(j-i)=2(j-i)+1$ 次的相鄰兩個 row 互換的動作. 例如 3-th row 和 6-th row 互換的動作, 我們先從上而下的先將 3-rd 和 4-th row 交換, 然後 4-th 和 5-th row 交換, 將原本的 3-rd row 換到 5-th row 的位置. 此時共做了 $(6-1)-3=2$ 次的相鄰兩個 row 互換的動作圖示如下:

$$\begin{bmatrix} \vdots \\ -\mathbf{v}_3- \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_3- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix}$$

接著我們從下而上的將 6-th 和 5-th row 交換 (此時原本 3-rd 已到達 6-th row 的位置), 然後 5-th 和 4-th row 交換, 最後將 4-th 和 3-rd row 交換將原本的 6-th row 換到 3-rd row 的位置. 此時共做了 $6-3=3$ 次的相鄰兩個 row 互換的動作圖示如下:

$$\begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_6- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_6- \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix}$$

由於每做一次相鄰兩 row 互換 det 會變一次號, 而 $2(j-1)+1$ 為奇數, 故最後 det 還是要變號. 我們推得了以下的性質.

Lemma 8.2.1. *假設 A 為 $n\times n$ matrix. 若將 A 任兩個 row 交換所得的矩陣為 A', 則 $\det(A')=-\det(A)$.*

回顧一下, 將 A 的 i-th 和 j-th row 交換這樣的 elementary row operation 所得的矩陣其實是將 A 的左邊乘上一個 elementary matrix E. 而 E 就是將 identity matrix I_n 的 i-th 和 j-th row 交換. 所以依 Lemma 8.2.1, 我們有 $\det(E)=-\det(I_n)$. 而 det 的性質 (1) 告訴我們 $\det(I_n)=1$, 因此得 $\det(E)=-1$. 又 Lemma 8.2.1 說將 A 的 i-th 和 j-th row 交換所得的矩陣 EA 其行列式為 $-\det(A)$, 因此若 E 為將 i-th 和 j-th row 交換這樣的 elementary row operation 所對應的 elementary matrix, 則 $\det(EA)=-\det(A)=\det(E)\det(A)$.

利用 Lemma 8.2.1, 如果 det 這個函數存在的話, 我們可以推得以下簡單的性質.

Lemma 8.2.2. *假設 A 為 $n\times n$ matrix 且 A 中有兩個 row 是相等的. 則 $\det(A)=0$.*

Proof. 假設 A 的 i-th row 和 j-th row 是相等的. 此時若將 A 的 i-th 和 j-th row 交換所得的矩陣為 A', 則由 Lemma 8.2.1 可得 $\det(A')=-\det(A)$. 但又 $A'=A$, 所以依 det 是一個函數 (之假設) 知 $\det(A)=\det(A')$. 故由 $\det(A)=\det(A')=-\det(A)$ 得證 $\det(A)=0$. □

第二種 elementary row operation 是將矩陣的某個 row 乘上一個非零實數. 這一個 elementary row operation 對行列式的影響其實就是我們要求 det 的性質 (3). 同樣的, 利用這一個 elementary row operation 將 A 的 i-th row 乘上一個非零實數 r 所得的矩陣就是將 A 的左邊乘上一個 elementary matrix E. 而 E 就是將 identity matrix I_n 的 i-th row 乘上 r. 因此依 det 的性質 (1)(3), 我們有此時 $\det(E)=r\det(I_n)=r$. 而性質 (3) 又要求

$\det(EA) = r\det(A)$, 故此時我們依然有 $\det(EA) = r\det(A) = \det(E)\det(A)$. 利用這個性質以及 det 這個函數存在的假設, 我們可以推得以下簡單的性質.

Lemma 8.2.3. 假設 A 為 $n\times n$ *matrix* 且 A 中有一個 *row* 全為 0. 則 $\det(A) = 0$.

Proof. 假設 A 的 i-th row 全為 0. 此時若將 A 的 i-th 乘上 2 所得的矩陣為 A', 則由 det 的性質 (3) 得 $\det(A') = 2\det(A)$. 但又 $A' = A$, 所以依 det 是一個函數 (之假設) 知 $\det(A) = \det(A')$. 故由 $\det(A) = \det(A') = 2\det(A)$ 得證 $\det(A) = 0$. □

第三種 elementary row operation 是將矩陣的某個 row 乘上非零實數 r 加到另一個 row. 現假設 A 為 $n\times n$ matrix 且設 A 的 k-th row 為 $\mathbf{v}_k$, for $k = 1,\ldots,n$. 令將 A 的 i-th row 乘上 r 加到 j-th row 所得的矩陣為 A', 則 A' 的 j-th row 為 $\mathbf{v}_j + r\mathbf{v}_i$, 而 A' 的 k-th row 仍為 $\mathbf{v}_k$, for $k\neq j$. 另外令 B 為 $n\times n$ matrix 其 j-th row 為 $\mathbf{v}_i$, 而 k-th row 為 $\mathbf{v}_k$, for $k\neq j$. 則依 det multi-linear 的性質 (即性質 (3) (4)), 我們有 $\det(A') = \det(A) + r\det(B)$. 但 B 的 i-th row 和 j-th row 皆為 $\mathbf{v}_i$, 故由 Lemma 8.2.2 知 $\det(B) = 0$. 得知 $\det(A') = \det(A)$, 因此我們有以下的性質.

Lemma 8.2.4. 假設 A 為 $n\times n$ *matrix.* 若將 A 的 *i-th row* 乘上 r 加到 *j-th row* 所得的矩陣為 A', 則 $\det(A') = \det(A)$.

同樣的, 將 A 的 i-th row 乘上非零實數 r 加到 j-th row 這樣的 elementary row operation 所得的矩陣其實是將 A 的左邊乘上一個 elementary matrix E. 而 E 就是將 identity matrix I_n 的 i-th row 乘上非零實數 r 加到 j-th row. 所以依 Lemma 8.2.4, 我們有 $\det(E) = \det(I_n) = 1$ 且 $\det(EA) = \det(A)$. 因此若 E 為將 i-th row 乘上非零實數 r 加到 j-th row 這樣的 elementary row operation 所對應的 elementary matrix, 則 $\det(EA) = \det(A) = \det(E)\det(A)$.

結合上面三種 elementary row operations 對 det 的影響, 我們得到以下重要的定理.

Theorem 8.2.5. 假設 A 為 $n\times n$ *matrix.* 若 E 為 *elementary matrix,* 則

$$\det(EA) = \det(E)\det(A).$$

Theorem 8.2.5 是 determinant 一個非常重要的性質, 它可以幫我們推導出許多有關於 determinant 的性質. 首先要注意的是三種 elementary row operations 所對應的 elementary matrices 它們的 determinant 皆不為 0, 回顧一下它們的 determinant 分別如下:

(1) 對於兩 row 交換的 elementary matrix E, 我們有 $\det(E) = -1$.
(2) 對於某個 row 乘上非零實數 r 的 elementary matrix E, 我們有 $\det(E) = r$.
(3) 對於某個 row 乘上非零實數 r 加到另一個 row 的 elementary matrix E, 我們有 $\det(E) = 1$.

另外, 若 E_1, E_2 為 elementary matrices, 由 Theorem 8.2.5 我們有

$$\det(E_2E_1A) = \det(E_2(E_1A)) = \det(E_2)\det(E_1A) = \det(E_2)\det(E_1)\det(A).$$

依數學歸納法可得, 若 $E_1,\ldots,E_k$ 為 elementary matrices, 則

$$\det(E_k\cdots E_1A) == \det(E_k)\cdots\det(E_1)\det(A). \tag{8.2}$$

利用這些 elementary matrices 的 determinant, 我們有以下關於 determinant 的重要性質.

Theorem 8.2.6. 假設 A,B 為 $n\times n$ *matrices.*

(1) A 為 *invertible* 若且唯若 $\det(A)\neq 0$.

(2) $\det(AB)=\det(A)\det(B)$.

(3) $\det(A^{\mathrm{T}})=\det(A)$.

Proof. (1) 假設 A 不是 invertible 表示 A 經過 elementary row operations 所得 echelon form A' 其 pivot 的個數會小於 n, 即 $\operatorname{rank}(A)<n$ (參見 Theorem 3.5.2). 因為 A' 的 pivot 的個數小於 n, 所以 A' 的最後一個 row (即 n-th row) 全為 0. 故由 Lemma 8.2.3 知 $\det(A')=0$. 然而我們知存在 elementary matrices $E_1,\ldots,E_k$ 使得 $A'=E_k\cdots E_1A$, 故由式子 (8.2) 知 $\det(A')=\det(E_k)\cdots\det(E_1)\det(A)$. 又 elementary matrices 的 determinants $\det(E_1),\ldots,\det(E_k)$ 皆不為 0, 故由 $\det(A')=0$ 得證 $\det(A)=0$. 而若 A 為 invertible, 則 A 可以寫成 elementary matrices 的乘積 (參見 Proposition 3.5.7). 故存在 elementary matrices $E_1\ldots E_k$ 使得 $A=E_k\cdots E_1$. 因此由式子 (8.2) 知 $\det(A)=\det(E_k)\cdots\det(E_1)$. 再由 $\det(E_1),\ldots,\det(E_k)$ 皆不為 0 得證 $\det(A)\neq 0$.

(2) 若 A 不是 invertible 由 (1) 我們知 $\det(A)=0$. 又此時因 B 也是 $n\times n$ matrix, 我們有 AB 也不是 invertible (參見 Proposition 3.5.5(3)). 故再由 (1) 知 $\det(AB)=0$, 得證 $\det(AB)=0=\det(A)\det(B)$. 現假設 A 為 invertible. 我們知存在 elementary matrices $E_1,\ldots,E_k$ 使得 $A=E_k\cdots E_1$. 因此由式子 (8.2) 以及 $\det(A)=\det(E_k)\cdots\det(E_1)$ 知

$$\det(AB)=\det(E_k\cdots E_1B)=\det(E_k)\cdots\det(E_1)\det(B)=\det(A)\det(B).$$

(3) 若 A 不是 invertible 由 (1) 我們知 $\det(A)=0$ 且此時 A^{T} 也不是 invertible (參見 Proposition 3.5.5(2)). 故得證 $\det(A^{\mathrm{T}})=0=\det(A)$. 現假設 A 為 invertible. 我們知存在 elementary matrices $E_1,\ldots,E_k$ 使得 $A=E_k\cdots E_1$ 且 $A^{\mathrm{T}}=E_1^{\mathrm{T}}\cdots E_k^{\mathrm{T}}$ (參見 Proposition 3.2.4). 由於當 E 為兩 row 交換的 elementary matrix 或是將某個 row 乘上非零實數 r 的 elementary matrix 皆有 $E=E^{\mathrm{T}}$, 而當 E 為將 i-th row 乘上非零實數 r 加到 j-th row 的 elementary matrix 時, E^{T} 為將 j-th row 乘上非零實數 r 加到 i-th row 的 elementary matrix, 故對任意 elementary matrix E 我們皆有 $\det(E)=\det(E^{\mathrm{T}})$. 因此得

$$\det(A^{\mathrm{T}})=\det(E_1^{\mathrm{T}}\cdots E_k^{\mathrm{T}})=\det(E_1^{\mathrm{T}})\cdots\det(E_k^{\mathrm{T}})=\det(E_1)\cdots\det(E_k).$$

最後由於 $n\times n$ matrix 的 determinant 為實數且實數乘法有交換律, 我們有

$$\det(E_1)\cdots\det(E_k)=\det(E_k)\cdots\det(E_1)=\det(A),$$

得證 $\det(A^{\mathrm{T}})=\det(A)$. □

由於一個矩陣經由 row operation 後取 transpose 就等同於將原矩陣的 transpose 做 column operation. 因此 Theorem 8.2.6 (3) 告訴我們 elementary column operation 對 determinant 的影響和 elementary row operation 對 determinant 的影響是一致的. 換言之, Theorem 8.2.5 也可改寫成以下的形式.

Corollary 8.2.7. *假設 A 為 $n\times n$ matrix. 若 E 為 elementary matrix, 則*

$$\det(AE)=\det(E)\det(A).$$

接著, 我們證明若 det 這個函數存在, 則它會是唯一的. 這裡的證明有一個很關鍵的觀念大家要注意, 就是本節中所有的性質我們都僅用前一節中對 det 所要求的四個性質推導出來的. 因此不管任何的函數, 只要符合這四個性質就都會符合前面這幾個定理.

Theorem 8.2.8. *最多僅有唯一的函數 $\det:\mathscr{M}_{n\times n}\to\mathbb{R}$ 會滿足*

(1) $\det(I_n)=1$.

(2) *若將 $n\times n$ matrix A 某相鄰兩個 row 交換所得的矩陣為 A', 則 $\det(A')=-\det(A)$.*

(3) *若將 $n\times n$ matrix A 某個 row 乘上非零實數 r 所得的矩陣為 A', 則 $\det(A')=r\det(A)$.*

(4) *若 A,B,C 三個 $n\times n$ matrix, 其中 A 的 i-th row 是 B 和 C 的 i-th row 之和, 而 A,B,C 其他各 row 皆相等, 則 $\det(A)=\det(B)+\det(C)$.*

Proof. 假設 $\det:\mathscr{M}_{n\times n}\to\mathbb{R}$ 和 $\det':\mathscr{M}_{n\times n}\to\mathbb{R}$ 皆滿足這四項規則. 考慮 $n\times n$ matrix A, 若 A 不是 invertible, 則由 Theorem 8.2.6 (1) 知 $\det(A)=\det'(A)=0$. 而若 A 為 invertible, 則存在 elementary matrices $E_1,\ldots,E_k$ 使得 $A=E_k\cdots E_1$. 因此由 Theorem 8.2.6 (2) 得 $\det(A)=\det(E_k)\cdots\det(E_1)$ 以及 $\det'(A)=\det'(E_k)\cdots\det'(E_1)$. 最後又由於對任意 elementary matrix E 皆有 $\det(E)=\det'(E)$, 我們證得

$$\det(A)=\det(E_k)\cdots\det(E_1)=\det'(E_k)\cdots\det'(E_1)=\det'(A).$$

因為對任意 $A\in\mathscr{M}_{n\times n}$, 皆有 $\det(A)=\det'(A)$, 我們證得 det 和 det$'$ 為相同的函數. □

或許同學會疑惑, 這裡明明已求出了所有 $n\times n$ matrix 的 determinant, 為什麼不是證出了存在性呢? 主要的原因是, 每一個 invertible matrix 寫成 elementary matrix 的乘積其寫法並不唯一. 所以我們無法利用這個方法定出 det 這個函數來, 因為有可能會因為寫成 elementary matrix 乘積的方法不同而得到不同的 determinant. 如此一來就違背函數同一個元素不能有不同取值的要求. 所以也唯有以後我們證明了 det 確實存在後, 才能確保一個 invertible matrix 寫成不同 elementary matrix 的乘積, 仍可求出相同的 determinant.

最後我們介紹如何利用 elementary row operation 求 $n\times n$ matrix 的 determinant. 首先我們先用 elementary row operations 將矩陣變為 echelon form. 而且在化為 echelon form 的過程中只用 (1) 兩 row 交換 (此時 determinant 變號) 以及 (3) 將某個 row 乘上非零實數 r 加到另一個 row (此時 determinant 不會改變), 這兩種 row operations. 若發現 pivot 的個數小於 n, 則我們得矩陣的 determinant 為 0. 而若 pivot 的個數為 n, 則我們利用做了幾次兩 row 交換的 row operation, 就可由 echelon form 的 determinant 得到原矩陣的 determinant

(即做了奇數次變號, 偶數次不變號). 然而如何求一個 echelon form 的 determinant 呢? 由於一個 $n\times n$ matrix 的 echelon form 一定是一個 *upper triangular matrix* (上三角矩陣), 也就是說矩陣 diagonal (對角線) 的位置 (即 (i,i)-th entry) 以下的位置皆為 0 (即 $a_{ij}=0$, for $i>j$), 此時下一個定理告訴我們其 determinant 就是 diagonal entries 的乘積.

Proposition 8.2.9. 假設 $A=[a_{ij}]$ 為 $n\times n$ *upper triangular matrix* 則 $\det(A)=a_{11}\cdots a_{nn}$, 即 $\det(A)$ 為 A 的 *diagonal entries* 的乘積.

Proof. 假設 A 有一個 diagonal entry a_{ii} 為 0, 因為 A 為 upper triangular, 我們知 A 化為 echelon form 後其 pivot 的個數必小於 n. 因此 A 不是 invertible, 得知 $\det(A)=0$. 而此時 A 的 diagonal entries 的乘積 $a_{11}\cdots a_{ii}\cdots a_{nn}$ 亦為 0, 故得證 $\det(A)=0=a_{11}\cdots a_{nn}$.

現若 A 的 diagonal entry 皆不為 0, 此時將 A 的每個 row 做以下的 elementary row operation: 就是對每一個 $i\in\{1,\ldots,n\}$, 將 A 的 i-st row 乘上 $1/a_{ii}$. 令所得的矩陣為 A', 此時我們有 $\det(A)=a_{11}\cdots a_{nn}\det(A')$. 因為 A' 的 diagonal entry 皆為 1, 利用 echelon form 化為 reduced echelon form 的方法 (參見 Section 2.2), 我們從最後一個 row (即 n-th row) 開始由下往上的利用該 row 乘上非零實數加到另一個 row 的方法將 A' 化為 I_n. 因為這裡所用的 elementary row operations 都不會影響 determinant, 所以我們有 $\det(A')=\det(I_n)=1$. 因此得證 $\det(A)=a_{11}\cdots a_{nn}\det(A')=a_{11}\cdots a_{nn}$ $\square$

Question 8.2. 假設 $A=[a_{ij}]$ 為 $n\times n$ *lower triangular matrix,* 即 A 的 *diagonal* 以上的位置皆為 0 (即 $a_{ij}=0$, *for* $i<j$). 試證明 $\det(A)=a_{11}\cdots a_{nn}$, 即 $\det(A)$ 為 A 的 *diagonal entries* 的乘積.

Example 8.2.10. 我們利用 elementary row operation 求 $\begin{bmatrix}0&2&-1&1\\1&2&0&2\\1&4&-2&6\\2&6&-1&8\end{bmatrix}$ 的 determinant. 首先將 1-st, 2-nd row 交換得 $\begin{bmatrix}1&2&0&2\\0&2&-1&1\\1&4&-2&6\\2&6&-1&8\end{bmatrix}$ (注意此時 determinant 會變號). 接著將 1-st row 分別乘上 $-1,-2$ 加到 3-rd, 4-th row 得 $\begin{bmatrix}1&2&0&2\\0&2&-1&1\\0&2&-2&4\\0&2&-1&4\end{bmatrix}$ (注意此時 determinant 不會改變). 最後將 2-nd row 分別乘上 $-1,-1$ 加到 3-rd, 4-th row 得 echelon form $\begin{bmatrix}1&2&0&2\\0&2&-1&1\\0&0&-1&3\\0&0&0&3\end{bmatrix}$ (注意此時 determinant 不會改變). 利用 Proposition 8.2.9 我們知最後所得的 echelon form 其 determinant 為 -6, 又整個化為 echelon form 的過程中僅用了一次兩

row 交換的 row operation, 故 determinant 僅變號一次, 得知

$$\det\begin{bmatrix} 0 & 2 & -1 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 4 & -2 & 6 \\ 2 & 6 & -1 & 8 \end{bmatrix} = -\det\begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 2 & -1 & 1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \end{bmatrix} = 6.$$

8.3. Determinant of 3×3 Matrix

我們利用上一節有關 determinant 的性質, 寫出 3×3 matrix 的 determinant 可能的形式, 從而證明 3×3 matrix 的 determinant 確實存在. 同時我們利用此 determinant 定義出 $\mathbb{R}^3$ 中三個向量所張成的平行六面體的 *sign volume*.

在 Theorem 8.2.6 (3) 中, 我們知道 $\det(A^{\mathrm{T}}) = \det(A)$. 因此有關 determinant 和 row 有關的性質, 對於 column 也有相對應的性質. 例如我們對 determinant 要求的 (3)(4) 兩個所謂 multi-linear 的性質, 是和 row 有關的, 因此對於 column 也會有 multi-linear 的性質. 我們大致圖示如下 (所有向量寫成 column vector):

$$\det\begin{bmatrix} | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_i + r\mathbf{v}_i' & \cdots & \mathbf{v}_n \\ | & & | & & | \end{bmatrix} = \det\begin{bmatrix} | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_i & \cdots & \mathbf{v}_n \\ | & & | & & | \end{bmatrix} + r\det\begin{bmatrix} | & & | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_i' & \cdots & \mathbf{v}_n \\ | & & | & & | \end{bmatrix}.$$

考慮 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$. 利用 determinant 對於 column 的 multi-linear 的性質, 由於 A 的 1-st column 可以寫成 $a_{11}\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + a_{21}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + a_{31}\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, 我們有

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} =$$
$$a_{11}\det\begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} + a_{21}\det\begin{bmatrix} 0 & a_{12} & a_{13} \\ 1 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} + a_{31}\det\begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 1 & a_{32} & a_{33} \end{bmatrix}.$$

當我們計算 $\det\begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix}$ 時, 我們可以利用上一節最後所介紹的方法, 用 elementary row operation 將矩陣先化為 echelon form. 由於我們不必動到 1-st row, 事實上我們是將 $\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$ 化為 echelon form. 因此我們有 $\det\begin{bmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} = \det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$. 同理, 利用 det 的性質, 我們有

$$\det\begin{bmatrix} 0 & a_{12} & a_{13} \\ 1 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} = -\det\begin{bmatrix} 1 & a_{22} & a_{23} \\ 0 & a_{12} & a_{13} \\ 0 & a_{32} & a_{33} \end{bmatrix} = -\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix},$$

$$\det\begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 1 & a_{32} & a_{33} \end{bmatrix} = -\det\begin{bmatrix} 0 & a_{12} & a_{13} \\ 1 & a_{32} & a_{33} \\ 0 & a_{22} & a_{23} \end{bmatrix} = \det\begin{bmatrix} 1 & a_{32} & a_{33} \\ 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \end{bmatrix} = \det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}.$$

因此依照 determinant 的性質 $\det(A)$ “應該” 定義為

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{21}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} + a_{31}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}.$$

注意, 這裡我們是根據 det 應有的性質寫下的定義, 它也確實是一個從 $\mathscr{M}_{3\times 3}$ 到 $\mathbb{R}$ 的函數 (不會將同一個矩陣送至不同的值). 不過這並不表示這樣定義出來的函數會符合當初我們要求的四個性質. 所以接下來我們將驗證這樣的定義確實會符合當初要求的四個性質, 也因此證明了 3×3 matrix 的 determinant 確實存在 (且唯一).

首先檢查 $\det(I_3) = 1$. 依定義

$$\det\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = 1\det\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 0\det\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + 0\det\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = 1.$$

故 $\det(I_3) = 1$ 成立.

接著檢查相鄰兩個 row 互換後 determinant 會變號. 依定義

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{21}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} + a_{31}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix},$$

$$\det\begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{21}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix} - a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} + a_{31}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{12} & a_{13} \end{bmatrix},$$

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = a_{11}\det\begin{bmatrix} a_{32} & a_{33} \\ a_{22} & a_{23} \end{bmatrix} - a_{31}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix} + a_{21}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{bmatrix}.$$

由於 2×2 matrix 的兩個 row 互換後其 determinant 會變號, 我們有

$$\det\begin{bmatrix} a_{32} & a_{33} \\ a_{22} & a_{23} \end{bmatrix} = -\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}, \quad \det\begin{bmatrix} a_{22} & a_{23} \\ a_{12} & a_{13} \end{bmatrix} = -\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{bmatrix}.$$

因此得證

$$\det\begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = -\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} = -\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

至於性質 (3), (4) 我們合併檢查, 即檢查 multi-linear 性質. 依定義

$$\det\begin{bmatrix} a_{11}+rb_{11} & a_{12}+rb_{12} & a_{13}+rb_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = (a_{11}+rb_{11})\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} -$$
$$a_{21}\det\begin{bmatrix} a_{12}+rb_{12} & a_{13}+rb_{13} \\ a_{32} & a_{33} \end{bmatrix} + a_{31}\det\begin{bmatrix} a_{12}+rb_{12} & a_{13}+rb_{13} \\ a_{22} & a_{23} \end{bmatrix}. \tag{8.3}$$

而 2×2 matrix 的 determinant 已知有 multi-linear 的性質, 亦即

$$\det\begin{bmatrix} a_{12}+rb_{12} & a_{13}+rb_{13} \\ a_{32} & a_{33}\end{bmatrix}=\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33}\end{bmatrix}+r\det\begin{bmatrix} b_{12} & b_{13} \\ a_{32} & a_{33}\end{bmatrix},$$

$$\det\begin{bmatrix} a_{12}+rb_{12} & a_{13}+rb_{13} \\ a_{22} & a_{23}\end{bmatrix}=\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23}\end{bmatrix}+r\det\begin{bmatrix} b_{12} & b_{13} \\ a_{22} & a_{23}\end{bmatrix}.$$

因此式子 (8.3) 等式右邊可寫成

$$\left(a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{bmatrix}-a_{21}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{32} & a_{33}\end{bmatrix}+a_{31}\det\begin{bmatrix} a_{12} & a_{13} \\ a_{22} & a_{23}\end{bmatrix}\right)+$$
$$r\left(b_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{bmatrix}-a_{21}\det\begin{bmatrix} b_{12} & b_{13} \\ a_{32} & a_{33}\end{bmatrix}+a_{31}\det\begin{bmatrix} b_{12} & b_{13} \\ a_{22} & a_{23}\end{bmatrix}\right).$$

再利用定義還原回 3×3 matrix determinant 得證

$$\det\begin{bmatrix} a_{11}+rb_{11} & a_{12}+rb_{12} & a_{13}+rb_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}=\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}+r\det\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

同理對於 2-nd row 和 3-rd row 我們有

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21}+rb_{21} & a_{22}+rb_{22} & a_{23}+rb_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}=\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}+r\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ b_{21} & b_{22} & b_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}$$

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31}+rb_{31} & a_{32}+rb_{32} & a_{33}+rb_{33}\end{bmatrix}=\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}+r\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ b_{31} & b_{32} & b_{33}\end{bmatrix}$$

我們利用 2×2 matrix 的 determinant 存在來證明 3×3 matrix 的 determinant 存在, 這樣的方法稱為“降階”的方法. 下一節中, 我們要用降階的方法以及數學歸納法證明任意 $n\times n$ matrix 的 determinant 皆存在. 其實我們這裡定義出 determinant 方法稱為對 1-st column 展開的降階, 我們也可以對 2-nd column 以及 3-rd column 展開. 為了方便起見, 我們有以下的定義.

Definition 8.3.1. 假設 $A=[a_{ij}]$ 為 3×3 matrix. 將 A 的 i-th row 和 j-th column 除去所得的 2×2 matrix, 稱為 A 的 (i,j) *minor matrix*, 用 A_{ij} 表示. 令 $a'_{ij}=(-1)^{i+j}\det(A_{ij})$, 稱為 A 的 (i,j) cofactor.

依此定義, 當初 $\det(A)$ 的定義可以寫成

$$\det(A)=a_{11}a'_{11}+a_{21}a'_{21}+a_{31}a'_{31}.$$

如果我們一開始將 $\det(A)$ 定義為 $\det(A)=a_{12}a'_{12}+a_{22}a'_{22}+a_{32}a'_{32}$, 即對 2-nd column 展開, 得

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{bmatrix}=-a_{12}\det\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33}\end{bmatrix}+a_{22}\det\begin{bmatrix} a_{11} & a_{13} \\ a_{31} & a_{33}\end{bmatrix}-a_{32}\det\begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23}\end{bmatrix}.$$

如同前面的證明會發現這個定義仍符合我們對 det 四項要求 (注意 cofactor 的正負號變化, 確保 $\det(I_3)=1$). 因此由 det 的唯一性 (參見 Theorem 8.2.8), 我們知道這樣求出的

determinant 和對 1-st column 展開的結果是一樣的. 同樣的我們也可對 3-rd column 展開得

$$\det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{13}\det\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} - a_{23}\det\begin{bmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{bmatrix} + a_{33}\det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

又因為 $\det(A) = \det(A^{\mathrm{T}})$, 我們可以將 A 轉置後對 A^{T} 的 1-st column 展開得

$$\det\begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} = a_{11}\det\begin{bmatrix} a_{22} & a_{32} \\ a_{23} & a_{33} \end{bmatrix} - a_{12}\det\begin{bmatrix} a_{21} & a_{31} \\ a_{23} & a_{33} \end{bmatrix} + a_{13}\det\begin{bmatrix} a_{21} & a_{31} \\ a_{22} & a_{32} \end{bmatrix}.$$

然而 2×2 matrix 取轉置後其 determinant 也不變, 所以上式等號右邊可改寫為

$$a_{11}\det\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12}\det\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13}\det\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}.$$

因此得知 $\det(A) = \det(A^{\mathrm{T}}) = a_{11}a'_{11} + a_{12}a'_{12} + a_{13}a'_{13}$, 也就是說 determinant 也可對 1-st row 展開求得. 同理對 2-nd row 和 3-rd row 展開也可求得 determinant. 我們有以下的定理.

Theorem 8.3.2. *假設 $A = [a_{ij}]$ 為 3×3 matrix. 令 a'_{ij} 為 A 的 (i,j) cofactor, 則*

$$\begin{aligned} \det(A) &= a_{11}a'_{11} + a_{21}a'_{21} + a_{31}a'_{31} = a_{12}a'_{12} + a_{22}a'_{22} + a_{32}a'_{32} = a_{13}a'_{13} + a_{23}a'_{23} + a_{33}a'_{33} \\ &= a_{11}a'_{11} + a_{12}a'_{12} + a_{13}a'_{13} = a_{21}a'_{21} + a_{22}a'_{22} + a_{23}a'_{23} = a_{31}a'_{31} + a_{32}a'_{32} + a_{33}a'_{33} \end{aligned}$$

我們得到了 3×3 matrix 的 determinant, 也因此由此可定義出 $\mathbb{R}^3$ 中三個向量所張成的平行六面體的 sign volume. 也就是說若將 $\mathbb{R}^3$ 上的三個向量 $\mathbf{u},\mathbf{v},\mathbf{w}$ 寫成 row vectors, 令矩陣 A 為 1-st, 2-nd, 3-rd row 依序為 $\mathbf{u},\mathbf{v},\mathbf{w}$ 的 3×3 matrix, 則 $\det(A)$ 就是 $\mathbf{u},\mathbf{v},\mathbf{w}$ 三個向量所張成的平行六面體的 sign volume. 其中 $\det(A)$ 的絕對值, 就是這平行六面體的體積. 而 $\det(A)$ 的正負號告訴我們 $\mathbf{u},\mathbf{v},\mathbf{w}$ 這三個向量的方向性. 這裡 $\mathbf{u},\mathbf{v},\mathbf{w}$ 這三個向量正反向我們是用所謂的 *right hand rule* (右手定則) 來區分, 意即將右手大拇指指向 $\mathbf{u}$ 的方向, 其餘四個指頭併攏指向 $\mathbf{v}$ 的方向, 若 $\mathbf{w}$ 位於手掌正面的方向則 $\mathbf{u},\mathbf{v},\mathbf{w}$ 為正向, 反之為負向. 例如 $\mathbf{i} = (1,0,0), \mathbf{j} = (0,1,0), \mathbf{k} = (0,0,1)$ 我們定為正向 (因 $\det(I_3) = 1 > 0$). 利用 Section 8.1 我們定的方向性規則, 可以知道 $\det(A) > 0$ 時 $\mathbf{u},\mathbf{v},\mathbf{w}$ 這三個向量為正向, 而 $\det(A) < 0$ 時為負向.

給定 $\mathbf{u} = (a_1,a_2,a_3)$, $\mathbf{v} = (b_1,b_2,b_3) \in \mathbb{R}^3$, 我們定義 $\mathbf{u},\mathbf{v}$ 的 *cross product* (外積) $\mathbf{u}\times\mathbf{v}$ 為

$$\mathbf{u}\times\mathbf{v} = \left(\det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix}, \det\begin{bmatrix} a_3 & a_1 \\ b_3 & b_1 \end{bmatrix}, \det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}\right).$$

要注意, 千萬不要將內積和外積弄混了, 兩個向量之內積是一個實數, 而兩個向量之外積仍為向量. 另外 $\mathbf{u}\times\mathbf{v}$ 和 $\mathbf{v}\times\mathbf{v}$ 是不相等的, 除非 $\mathbf{u}\times\mathbf{v} = \mathbf{O}$. 這是因為兩個 row 交換其 determinant 會變號, 因此依定義 $\mathbf{u}\times\mathbf{v} = -\mathbf{v}\times\mathbf{u}$. 而 $\mathbf{u}\times\mathbf{v}$ 何時會是 $\mathbf{O}$ 呢? 依定義 $\mathbf{u}\times\mathbf{v} = \mathbf{O}$ 若且唯若 $\det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} = \det\begin{bmatrix} a_3 & a_1 \\ b_3 & b_1 \end{bmatrix} = \det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} = 0$, 很容易知道這等同於 $\mathbf{u} = (a_1,a_2,a_3)$, $\mathbf{v} = (b_1,b_2,b_3)$ 為 linearly dependent.

現考慮 $\mathbf{u} = (a_1,a_2,a_3)$, $\mathbf{v} = (b_1,b_2,b_3)$, $\mathbf{w} = (c_1,c_2,c_3)$ 我們有

$$\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v}) = c_1\det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} + c_2\det\begin{bmatrix} a_3 & a_1 \\ b_3 & b_1 \end{bmatrix} + c_3\det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \tag{8.4}$$

由於 $\det\begin{bmatrix} a_3 & a_1 \\ b_3 & b_1 \end{bmatrix} = -\det\begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix}$ 式子 (8.4) 的右式又等同於將 $\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$ 對 3-rd row 展開的 determinant, 故得

$$\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v}) = (\mathbf{u}\times\mathbf{v})\cdot\mathbf{w} = \det\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \det\begin{bmatrix} -\mathbf{u}- \\ -\mathbf{v}- \\ -\mathbf{w}- \end{bmatrix}. \tag{8.5}$$

也就是說 $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$ 就是 $\mathbf{u},\mathbf{v},\mathbf{w}$ 這三個向量所張成的平行六面體的 sign volume.

特別的, 當 $\mathbf{w}=\mathbf{u}$ 或 $\mathbf{w}=\mathbf{v}$ 時, 由於 $\mathbf{u},\mathbf{v},\mathbf{w}$ 為 row vector 所形成的矩陣有兩個 row 相同, 所以其 determinant 為 0 (Lemma 8.2.2). 因此由式子 (8.5) 知 $\mathbf{u}\cdot(\mathbf{u}\times\mathbf{v}) = \mathbf{v}\cdot(\mathbf{u}\times\mathbf{v}) = 0$. 也就是說當 $\mathbf{u},\mathbf{v}$ 為 linearly independent 時, $\mathbf{u}\times\mathbf{v}$ 同時會和 $\mathbf{u}$ 以及和 $\mathbf{v}$ 垂直. 而當 $\mathbf{w}=\mathbf{u}\times\mathbf{v}$, 我們有 $(\mathbf{u}\times\mathbf{v})\cdot(\mathbf{u}\times\mathbf{v}) = \|\mathbf{u}\times\mathbf{v}\|^2$. 也就是 $\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}$ 所張成的平行六面體的 sign volume 為 $\|\mathbf{u}\times\mathbf{v}\|^2$. 考慮 $\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}$ 所張成的平行六面體以 $\mathbf{u},\mathbf{v}$ 所張的平行四邊形為底, 此時由於 $\mathbf{u}\times\mathbf{v}$ 和 $\mathbf{u}$ 以及和 $\mathbf{v}$ 垂直, 我們得 $\|\mathbf{u}\times\mathbf{v}\|$ 就是此平行六面體的高. 因此由 $\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}$ 所張成的平行六面體的體積 $\|\mathbf{u}\times\mathbf{v}\|^2$ 為 $\mathbf{u},\mathbf{v}$ 所張的平行四邊形面積乘上高 $\|\mathbf{u}\times\mathbf{v}\|$, 得 $\mathbf{u},\mathbf{v}$ 所張的平行四邊形面積為 $\|\mathbf{u}\times\mathbf{v}\|$. 另外由於 $\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}$ 所張成的平行六面體的 sign volume 為 $\|\mathbf{u}\times\mathbf{v}\|^2>0$, 我們知 $\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}$ 利用右手定則為正向. 最後我們將外積的性質歸納如下.

Theorem 8.3.3. 給定 $\mathbf{u},\mathbf{v}\in\mathbb{R}^3$. 則 $\mathbf{u}\times\mathbf{w}\neq\mathbf{O}$ 若且唯若 $\mathbf{u},\mathbf{v}$ 為 *linearly independent.* 此時 $\mathbf{u}\times\mathbf{v}$ 的長度為 $\mathbf{u},\mathbf{v}$ 所張的平行四邊形面積, 且 $\mathbf{u},\mathbf{v}$ 同時與 $\mathbf{u}\times\mathbf{v}$ 垂直, 又 $\mathbf{u},\mathbf{v},\mathbf{u}\times\mathbf{v}$ 利用右手定則為正向.

又假設 $\mathbf{w}\in\mathbb{R}$, 則 $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}\neq 0$ 若且唯若 $\mathbf{u},\mathbf{v},\mathbf{w}$ 為 *linearly independent.* 此時 $(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}$ 就是 $\mathbf{u},\mathbf{v},\mathbf{w}$ 這三個向量所張成的平行六面體的 *sign volume.*

8.4. Existence of the Determinant Function

在上一節中, 我們利用降階的方法以及 2×2 matrix 的 determinant 的存在性建構了 3×3 matrix 的 determinant, 因而得到其存在性. 接著我們可利用 3×3 matrix 的 determinant 存在性得到 4×4 matrix 的 determinant 的存在性, 然後一直下去. 在本節中, 我們就是要用數學歸納法證明一般 $n\times n$ matrix 的 determinant 皆存在.

首先我們將 Definition 8.3.1 的定義推廣到一般的情形.

Definition 8.4.1. 假設 $A=[a_{ij}]$ 為 $n\times n$ matrix. 將 A 的 i-th row 和 j-th column 除去所得的 $(n-1)\times(n-1)$ matrix, 稱為 A 的 (i,j) *minor matrix*, 用 A_{ij} 表示. 當 $(n-1)\times(n-1)$ matrix 的 determinant 存在時, 令 $a'_{ij}=(-1)^{i+j}\det(A_{ij})$, 稱為 A 的 (i,j) cofactor.

現利用數學歸納法假設 $(n-1)\times(n-1)$ matrix 的 determinant 存在, 對於 $n\times n$ matrix $A=[a_{ij}]$, 固定 $k\in\{1,\ldots,n\}$, 我們考慮 A 對 k-th column 展開, 定義

$$\det(A) = a_{1k}a'_{1k} + a_{2k}a'_{2k} + \cdots + a_{nk}a'_{nk}.$$

我們要利用 $(n-1)\times(n-1)$ matrix 的 determinant 符合 determinant 所要求的四個性質來證明這樣定出 $n\times n$ matrix 的 determinant 也會符合這四個性質.

首先證明 $\det(I_n)=1$. 由於 I_n 的 k-th column 為 $\mathbf{e}_k$, 僅有在 k-th entry 為 1, 其餘位置為 0. 也就是說若令 $A=[a_{ij}]=I_n$, 則 $a_{ik}=0$ for $i\neq k$ 且 $a_{kk}=1$. 因此依定義我們有 $\det(I_n)=a_{kk}a'_{kk}=a'_{kk}$. 然而 $A=I_n$ 在 (k,k) 的 minor matrix 為 I_{n-1}, 因此得 $A=I_n$ 的 (k,k) cofactor 為 $a'_{kk}=(-1)^{k+k}\det(I_{n-1})=\det(I_{n-1})$. 但依 induction 的假設, $\det(I_{n-1})=1$, 故知 $a'_{kk}=1$, 得證 $\det(I_n)=1$.

接著檢查相鄰兩個 row 互換後 determinant 會變號. 假設 $A=[a_{ij}]$, 固定 $l\in\{1,\ldots,n-1\}$, 假設將 A 的 l-th row 和 $l+1$-th row 交換所得的矩陣為 $B=[b_{ij}]$. 也就是說當 $i\neq l,l+1$ 時, $b_{ij}=a_{ij}$ 而 $b_{lj}=a_{l+1\,j}$, $b_{l+1\,j}=a_{lj}$. 因而我們有當 $i<l$ 時, B 的 (i,k) minor matrix B_{ik} 就是將 A 的 (i,k) minor matrix A_{ik} 相鄰的 $l-1$-th, l-th 兩個 row 交換 (此時依歸納假設 $\det(B_{ik})=-\det(A_{ik})$). 而當 $i>l+1$ 時, B_{ik} 就是將 A_{ik} 相鄰的 l-th, $l+1$-th 兩個 row 交換 (此時依歸納假設 $\det(B_{ik})=-\det(A_{ik})$). 又 B_{lk} 就是 $A_{l+1\,k}$ 且 $B_{l+1\,k}$ 就是 A_{lk}. 因此我們有 B 的 (i,k) cofactor b'_{ik} 為

$$(-1)^{i+k}\det(B_{ik})=\begin{cases}(-1)^{i+k}(-\det(A_{ik}))=-a'_{ik}, & \text{if } i\neq l \text{ and } i\neq l+1;\\ (-1)^{l+k}\det(A_{l+1\,k})=-a'_{l+1\,k}, & \text{if } i=l;\\ (-1)^{i+1+k}\det(A_{lk})=-a'_{lk}, & \text{if } i=l+1;\end{cases}$$

由此得證

$$\begin{array}{rcccccccccc}\det(B) & = & b_{1k}b'_{1k} & +\cdots+ & b_{lk}b'_{lk} & + & b_{l+1\,k}b'_{l+1\,k} & +\cdots+ & b_{nk}b'_{nk}\\ & = & a_{1k}(-a'_{1k}) & +\cdots+ & a_{l+1\,k}(-a'_{l+1\,k}) & + & a_{lk}(-a'_{lk}) & +\cdots+ & a_{nk}(-a'_{nk})\\ & = & -\det(A) & & & & & & \end{array}$$

至於性質 (3), (4) 我們合併檢查, 即檢查 multi-linear 性質. 固定 $l\in\{1,\ldots,n-1\}$ 以及 $r\in\mathbb{R}$. 假設 $A=[a_{ij}]$, $B=[b_{ij}]$, $C=[c_{ij}]$ 為 $n\times n$ matrices 滿足當 $i\neq l$ 時 $a_{ij}=b_{ij}=c_{ij}$ 而 $a_{lj}=b_{lj}+rc_{lj}$. 當 $i<l$ 時, A 的 (i,k) minor matrix A_{ik} 的 $l-1$-th row 就是 B_{ik} 的 $l-1$-th row 加上 r 倍的 C_{ik} 的 $l-1$-th row (此時依歸納假設 $\det(A_{ik})=\det(B_{ik})+r\det(C_{ik})$). 而當 $i>l+1$ 時, A_{ik} 的 l-th row 就是 B_{ik} 的 l-th row 加上 r 倍的 C_{ik} 的 l-th row (此時依歸納假設 $\det(A_{ik})=\det(B_{ik})+r\det(C_{ik})$). 又 A_{lk} 等於 B_{lk} 且等於 C_{lk}. 因此我們有 A 的 (i,k) cofactor a'_{ik} 為

$$(-1)^{i+k}\det(A_{ik})=\begin{cases}(-1)^{i+k}(\det(B_{ik})+r\det(B_{ik})=b'_{ik}+rc'_{ik}, & \text{if } i\neq l;\\ (-1)^{l+k}\det(B_{lk})=(-1)^{l+k}\det(C_{lk})=b'_{lk}=c'_{lk}, & \text{if } i=l;\end{cases}$$

由此得證

$$\begin{array}{rcccccccc}\det(A) & = & a_{1k}a'_{1k} & +\cdots+ & a_{lk}a'_{lk} & +\cdots+ & a_{nk}a'_{nk}\\ & = & b_{1k}(b'_{1k}+rc'_{1k}) & +\cdots+ & (b_{lk}+rc_{lk})b'_{lk} & +\cdots+ & b_{nk}(b'_{nk}+rc'_{nk})\\ & = & b_{1k}b'_{1k}+rc_{1k}c'_{1k} & +\cdots+ & b_{lk}b'_{lk}+rc_{lk}c'_{lk} & +\cdots+ & b_{nk}b'_{nk}+rc_{nk}c'_{nk}\\ & = & \det(B)+r\det(C). & & & & \end{array}$$

我們證得了 det 的存在性, 再加上 Theorem 8.2.8 的唯一性, 我們有以下的結論.

Theorem 8.4.2. 存在唯一的函數 $\det:\mathscr{M}_{n\times n}\to\mathbb{R}$ 滿足

(1) $\det(I_n)=1$.

(2) *若將 $n\times n$ matrix A 某相鄰兩個 row 交換所得的矩陣為 A', 則 $\det(A')=-\det(A)$.*

(3) *若將 $n\times n$ matrix A 某個 row 乘上非零實數 r 所得的矩陣為 A', 則 $\det(A')=r\det(A)$.*

(4) 若 A,B,C 三個 $n\times n$ *matrix,* 其中 A 的 *i-th row* 是 B 和 C 的 *i-th row* 之和, 而 A,B,C 其他各 *row* 皆相等, 則 $\det(A)=\det(B)+\det(C)$.

由於我們證得了對任意的 column 展開所得的 determinant 皆符合上述四項性質, 因此由唯一性得到對任意 column 展開所得的 determinant 之值皆會相同. 另外和 3×3 的情形相同, 由於 $\det(A^{\mathrm{T}})=\det(A)$, 我們也得到對任意 row 展開所得的 determinant 之值皆會相同. 因此我們有以下的結果.

Theorem 8.4.3. 假設 $A=[a_{ij}]$ 為 $n\times n$ *matrix.* 令 a'_{ij} 為 A 的 (i,j) *cofactor,* 則對任意 $k\in\{1,\ldots,n\}$ 皆有 $\det(A)=a_{1k}a'_{1k}+a_{2k}a'_{2k}+\cdots+a_{nk}a'_{nk}=a_{k1}a'_{k1}+a_{k2}a'_{k2}+\cdots+a_{kn}a'_{kn}$.

Question 8.3. 對於 $n\times n$ *matrix* $A=[a_{ij}]$ 考慮 A 的 *diagonal entry* 展開, 即考慮
$$a_{11}a'_{11}+a_{22}a'_{22}+\cdots+a_{nn}a'_{nn}.$$
試問這樣的展開方法會符合我們要求 *determinant* 的四項規則的哪幾項?

雖然我們利用 elementary row operations 的方法證明了 determinant 的唯一性, 又用降階的方法證明了 determinant 的存在性, 不過在計算 determinant 時, 這兩種方法都可混用. 一般來說用 row operation 或 column operation 來求 determinant 較快, 不過若發現有的 row 或 column 僅有一個不是 0 的 entry, 則對該 row 或 column 降階, 也可幫助我們較快算出 determinant. 我們看以下的例子.

Example 8.4.4. 我們求 $A=\begin{bmatrix}2&-1&3&5\\0&2&1&2\\0&5&3&3\\4&-2&7&8\end{bmatrix}$ 的 determinant. 首先觀察 A 的 1-st column 僅有兩個 entry 不為 0, 所以利用 1-st row 乘上 -2 加到 4-th row 得 $B=\begin{bmatrix}2&-1&3&5\\0&2&1&2\\0&5&3&3\\0&0&1&-2\end{bmatrix}$ (此時 $\det(A)=\det(B)$). 現因 B 的 1-st column 僅有一個非 0 entry, 我們對 1-st column 降階展開得 $\det(B)=2\det(C)$ 其中 C 為 B 的 $(1,1)$ minor matrix, 即 $C=\begin{bmatrix}2&1&2\\5&3&3\\0&1&-2\end{bmatrix}$. 接著將 C 的 2-nd column 乘上 2 加到 3-rd column 得 $D=\begin{bmatrix}2&1&4\\5&3&9\\0&1&0\end{bmatrix}$ (此時 $\det(C)=\det(D)$). 最後對 D 的 3-rd row 展開得 $\det(D)=(-1)^{3+2}\det\begin{bmatrix}2&4\\5&9\end{bmatrix}=2$. 故知 $\det(A)=\det(B)=2\det(C)=2\det(D)=4$.

8.5. Cramer's Rule and Adjoint Matrix

Determinant 不只可以幫助我們計算平行多面體的有向體積, 其實它也可以幫助我們解聯立方程組以及找到 invertible matrix 的反矩陣. 在這一節中, 由於和矩陣乘法有關, 所以所有向量皆用 column vector 表示.

首先考慮 $n \times n$ matrix $A = [a_{ij}]$ 對於 $j \in \{1, \ldots, n\}$ 令 $\mathbf{a}_j$ 表示 A 的 j-th column. 現對於 $\mathbb{R}^n$ 的一個 vector $\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, 對於任意 $k \in \{1, \ldots, n\}$, 考慮 C_k 為將 identity matrix I_n 的 k-th column 用 $\mathbf{c}$ 取代的 $n \times n$ matrix. 亦即當 $j \neq k$ 時, C_k 的 j-th column 為 $\mathbf{e}_j$, 而 C_k 的 k-th column 為 $\mathbf{c}$. 現考慮 AC_k, 依矩陣乘法的定義, 我們有

$$AC_k = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix} \begin{bmatrix} | & & c_1 & & | \\ \mathbf{e}_1 & \cdots & \vdots & \cdots & \mathbf{e}_n \\ | & & c_n & & | \end{bmatrix} = \begin{bmatrix} | & & | & & | \\ \mathbf{v}_1 & \cdots & c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n & \cdots & \mathbf{v}_n \\ | & & | & & | \end{bmatrix}. \quad (8.6)$$

也就是說當 $j \neq k$ 時, AC_k 的 j-th column 為 $\mathbf{v}_j$, 而 AC_k 的 k-th column 為 $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. 現對於 $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ 若 $x_1 = c_1, \ldots, x_n = c_n$ 為聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解, 亦即

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \mathbf{b}. \quad (8.7)$$

因此若令 B_k 表示將 A 的 k-th column 用 $\mathbf{b}$ 取代的 $n \times n$ matrix, 則結合式子 (8.6) (8.7), 我們有 $AC_k = B_k$. 因此由 determinant 的乘法性質 (Theorem 8.2.6 (2)), 得 $\det(A)\det(C_k) = \det(B_k)$. 然而對 C_k 的 k-th row 展開, 我們有 $\det(C_k) = c_k(-1)^{k+k}\det(I_{n-1}) = c_k$. 因此得證以下之定理.

Theorem 8.5.1. *假設 A 為 $n \times n$ matrix 且 $\mathbf{b} = \in \mathbb{R}^n$ 為 column vector. 對於 $k \in \{1, \ldots, n\}$ 令 B_k 表示將 A 的 k-th column 用 $\mathbf{b}$ 取代的 $n \times n$ matrix. 若 $x_1 = c_1, \ldots, x_n = c_n$ 為聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解, 則 $c_k\det(A) = \det(B_k)$.*

我們可以利用 Theorem 8.5.1 得到許多和解聯立方程組有關的性質. 首先若 $\det(A) \neq 0$, 表示 A 為 invertible, 我們知聯立方程組 $A\mathbf{x} = \mathbf{b}$ 一定有解且解唯一. 事實上此時由 Theorem 8.5.1 我們可以將此組解具體寫出. 這就是所謂的 *Cramer's Rule*.

Corollary 8.5.2 (Cramer's Rule)**.** *假設 A 為 $n \times n$ invertible matrix 且 $\mathbf{b} = \in \mathbb{R}^n$ 為 column vector. 對於所有 $k \in \{1, \ldots, n\}$ 令 B_k 表示將 A 的 k-th column 用 $\mathbf{b}$ 取代的 $n \times n$ matrix. 則聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有唯一的一組解, 且其解為*

$$x_k = \frac{\det(B_k)}{\det(A)}, \forall k = 1, \ldots, n.$$

Proof. 由假設 A 為 invertible, 知 $\det(A) \neq 0$ 且 $A\mathbf{x} = \mathbf{b}$ 必有解 (且解唯一). 然而 Theorem 8.5.1 告訴我們如果聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解, 則其解 $x_1 = c_1, \ldots, x_n = c_n$ 需滿足 $c_k\det(A) = \det(B_k), \forall k = 1, \ldots, n$. 然而因 $\det(A) \neq 0$, 故由解的存在性知 $x_k = \det(B_k)/\det(A), \forall k = 1, \ldots, n$ 是唯一可能的一組解. $\square$

Example 8.5.3. 考慮 Example 8.4.4 中的矩陣 $A = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 0 & 2 & 1 & 2 \\ 0 & 5 & 3 & 3 \\ 4 & -2 & 7 & 8 \end{bmatrix}$. 令 $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, 前面已知 $\det(A) = 4 \neq 0$, 我們可用 Cramer's rule 解聯立方程組 $A\mathbf{x} = \mathbf{b}$. 此時將 $\mathbf{b}$ 置換於 A 的 1-st column, 得 $B_1 = \begin{bmatrix} 1 & -1 & 3 & 5 \\ 0 & 2 & 1 & 2 \\ 0 & 5 & 3 & 3 \\ 0 & -2 & 7 & 8 \end{bmatrix}$. 同理得 $B_2 = \begin{bmatrix} 2 & 1 & 3 & 5 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 3 & 3 \\ 4 & 0 & 7 & 8 \end{bmatrix}$, $B_3 = \begin{bmatrix} 2 & -1 & 1 & 5 \\ 0 & 2 & 0 & 2 \\ 0 & 5 & 0 & 3 \\ 4 & -2 & 0 & 8 \end{bmatrix}$, $B_4 = \begin{bmatrix} 2 & -1 & 3 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 5 & 3 & 0 \\ 4 & -2 & 7 & 0 \end{bmatrix}$. 利用降階, 我們得 $\det(B_1) = 42$, $\det(B_2) = 12$, $\det(B_3) = -16$, $\det(B_4) = -4$. 故由 Cramer's rule 知 $x_1 = 21/2$, $x_2 = 3$, $x_3 = -4$, $x_4 = -1$ 是聯立方程組 $A\mathbf{x} = \mathbf{b}$ 之唯一一組解. 若令 $C_1 = \begin{bmatrix} 21/2 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$ 我們會有

$$AC_1 = \begin{bmatrix} 2 & -1 & 3 & 5 \\ 0 & 2 & 1 & 2 \\ 0 & 5 & 3 & 3 \\ 4 & -2 & 7 & 8 \end{bmatrix} \begin{bmatrix} 21/2 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 3 & 5 \\ 0 & 2 & 1 & 2 \\ 0 & 5 & 3 & 3 \\ 0 & -2 & 7 & 8 \end{bmatrix} = B_1.$$

同理若令

$$C_2 = \begin{bmatrix} 1 & 21/2 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & -4 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}, C_3 = \begin{bmatrix} 1 & 0 & 21/2 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}, C_4 = \begin{bmatrix} 1 & 0 & 0 & 21/2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & -1 \end{bmatrix},$$

我們會有 $AC_2 = B_2$, $AC_3 = B_3$ 以及 $AC_4 = B_4$.

至於當 A 不是 invertible 時 (即 $\det(A) = 0$), Theorem 8.5.1 就無法幫助我們找出聯立方程組的解. 不過由於 $\det(A) = 0$, 故利用 Theorem 8.5.1 知若 $x_1 = c_1, \ldots, x_n = c_n$ 為聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解, 則 $\det(B_k) = c_k \det(A) = 0$, $\forall k = 1, \ldots, n$. 換言之, 若存在 $k \in \{1, \ldots, n\}$ 使得 $\det(B_k) \neq 0$, 則聯立方程組 $A\mathbf{x} = \mathbf{b}$ 就無解.

Corollary 8.5.4. 假設 A 為 $n \times n$ *non-invertible matrix* 且 $\mathbf{b} \in \mathbb{R}^n$ 為 *column vector.* 對於所有 $k \in \{1, \ldots, n\}$ 令 B_k 表示將 A 的 *k-th column* 用 $\mathbf{b}$ 取代的 $n \times n$ *matrix.* 若存在 $k \in \{1, \ldots, n\}$ 使得 $\det(B_k) \neq 0$, 則聯立方程組 $A\mathbf{x} = \mathbf{b}$ 無解.

要注意 Corollary 8.5.4 的反向並不成立. 也就是說當 A 不是 invertible 時, 若對所有 $k = 1, \ldots, n$, 皆有 $\det(B_k) = 0$, 那麼我們是無從判斷聯立方程組 $A\mathbf{x} = \mathbf{b}$ 是否有解的. 例如在 $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ 的情形很容易判斷 $A\mathbf{x} = \mathbf{b}$ 有解, 且 $\det(A) = \det(B_1) = \det(B_2) =$

$\det(B_3)=0$. 而當 $\mathbf{b}=\begin{bmatrix}1\\1\\1\end{bmatrix}$ 時, 很容易判斷 $A\mathbf{x}=\mathbf{b}$ 無解, 但此時仍有 $\det(A)=\det(B_1)=\det(B_2)=\det(B_3)=0$.

由上面這幾種情況可知, Cramer's rule 並不是有效處理聯立方程組的方法. 一般在處理特定的聯立方程組, 還是直接用 elementary row operations 處理較為快速. 不過在處理一般抽象的方程組問題時, Cramer's rule 因為可以具體描繪出解的形式, 所以是很有用的工具. 我們看以下的性質.

Proposition 8.5.5. 假設 $A=[a_{ij}]$ 為 $n\times n$ *matrix* 其中 $a_{ij}\in\mathbb{Z}$, $\forall i,j\in\{1,\ldots,n\}$. 若 $\det(A)=\pm 1$, 則對於任意 $\mathbf{b}=\begin{bmatrix}b_1\\\vdots\\b_n\end{bmatrix}$, 其中 $b_i\in\mathbb{Z}$, $\forall i\in\{1,\ldots,n\}$, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解皆為整數.

Proof. 由於 $\det(A)=\pm 1\neq 0$, 利用 Cramer's rule 我們有聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解為 $x_1=\pm\det(B_1),\ldots,x_n=\pm\det(B_n)$, 其中對任意 $k\in\{1,\ldots,n\}$, B_k 為將 A 的 k-th column 用 $\mathbf{b}$ 取代的 $n\times n$ matrix. 由於 $a_{ij}\in\mathbb{Z}$ 且 $b_i\in\mathbb{Z}$, $\forall i,j\in\{1,\ldots,n\}$. 我們知矩陣 B_k 的所有 entry 皆為整數. 利用 determinant 的定義, 我們知此時 $\det(B_k)$ 亦為整數, 得證聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解皆為整數. □

另外一個 Cramer's rule 的應用就是幫我們找到 invertible matrix 的 inverse. 假設 $A\in\mathscr{M}_{n\times n}$ 為 invertible 且 C 為 A 的 inverse, 則由 $AC=I_n$, 依矩陣乘法定義我們知 C 的 j-th column $\begin{bmatrix}c_{1j}\\\vdots\\c_{nj}\end{bmatrix}$ 需滿足 $A\begin{bmatrix}c_{1j}\\\vdots\\c_{nj}\end{bmatrix}$ 等於 I_n 的 j-th column $\mathbf{e}_j$. 也就是說 C 的 j-th column 為聯立方程組 $A\mathbf{x}=\mathbf{e}_j$ 的解. 因此 C 的 (i,j)-th entry c_{ij} 應為聯立方程組 $A\mathbf{x}=\mathbf{e}_j$ 的解中 x_i 之值. 故由 Cramer's rule 知 $c_{ij}=\det(A(j,i))/\det(A)$, 其中 $A(j,i)$ 表示將 A 的 i-th column 用 $\mathbf{e}_j$ 取代的 $n\times n$ matrix. 然而利用對 $A(j,i)$ 的 i-th column 展開求 $\det(A(j,i))$, 我們得 $\det(A(j,i))=(-1)^{j+1}\det(A_{ji})=a'_{ji}$. 也就是說 c_{ij} 就是 A 的 (j,i) cofactor (注意 i,j 位置交換) 除以 $\det(A)$. 為了方便起見我們有以下的定義.

Definition 8.5.6. 假設 $A=[a_{ij}]$ 為 $n\times n$ matrix, 對於任意 $i,j\in\{1,\ldots,n\}$ 令 a'_{ij} 為 A 的 (i,j) cofactor. 考慮 $n\times n$ matrix A' 其 (i,j)-th entry 為 a'_{ij}. 我們稱 A' 為 A 的 *cofactor matrix* 而稱 A' 的 transpose $(A')^{\mathrm{T}}$ 為 A 的 *adjoint matrix*, 用 $\mathrm{adj}(A)$ 來表示.

注意 $\mathrm{adj}(A)$ 是將 A 的 cofactor 所成的矩陣 A' 取轉置而得, 千萬不要忘記取轉置. 另外要注意不管一個 $n\times n$ matrix 是否為 invertible, 皆可定義其 adjoint matrix.

我們回到剛才 A 為 invertible 的情況. 假設 C 為其 inverse. 依 $\mathrm{adj}(A)$ 的定義, 我們得到 C 的 (i,j)-th entry 就是 $\mathrm{adj}(A)$ 的 (i,j)-th entry 除以 $\det(A)$. 因此依矩陣係數積的定義, 我們有 $C=\frac{1}{\det(A)}\mathrm{adj}(A)$. 得證了以下的定理.

Proposition 8.5.7. 假設 A 為 $n\times n$ *invertible matrix.* 則

$$A^{-1}=\frac{1}{\det(A)}\operatorname{adj}(A).$$

Example 8.5.8. 考慮 Example 8.4.4 中的矩陣 $A=\begin{bmatrix}2&-1&3&5\\0&2&1&2\\0&5&3&3\\4&-2&7&8\end{bmatrix}$. 在 Example 8.5.3 中我們解出 $A\mathbf{x}=\mathbf{e}_1$ 的解, 事實上就是 A 的反矩陣 A^{-1} 的 1-st column. 在 Example 8.5.3 中的 $B_1=\begin{bmatrix}1&-1&3&5\\0&2&1&2\\0&5&3&3\\0&-2&7&8\end{bmatrix}$ 就是將 A 的 1-st column 用 $\mathbf{e}_1$ 取代所得的矩陣 $A(1,1)$. 若我們將 B_1 的 1-st column 展開求 $\det(B_1)$ 得 $\det(B_1)=(-1)^{1+1}\det\begin{bmatrix}2&1&2\\5&3&3\\-2&7&8\end{bmatrix}=42$ 就是 A 的 $(1,1)$ cofactor. 同樣的 $B_2=\begin{bmatrix}2&1&3&5\\0&0&1&2\\0&0&3&3\\4&0&7&8\end{bmatrix}$ 就是將 A 的 2-nd column 用 $\mathbf{e}_1$ 取代所得的矩陣 $A(1,2)$. 若我們將 B_2 的 2-nd column 展開求 $\det(B_2)$ 得 $\det(B_2)=(-1)^{1+2}\det\begin{bmatrix}0&1&2\\0&3&3\\4&7&8\end{bmatrix}=12$ 就是 A 的 $(1,2)$ cofactor. 同理得 B_3 是將 A 的 3-rd column 用 $\mathbf{e}_1$ 取代所得的矩陣 $A(1,3)$ 且 $\det(B_3)=(-1)^{1+3}\det\begin{bmatrix}0&2&2\\0&5&3\\4&-2&8\end{bmatrix}=-16$ 就是 A 的 $(1,3)$ cofactor. 而 B_4 是將 A 的 4-th column 用 $\mathbf{e}_1$ 取代所得的矩陣 $A(1,4)$ 且 $\det(B_4)=(-1)^{1+4}\det\begin{bmatrix}0&2&1\\0&5&3\\4&-2&7\end{bmatrix}=-4$ 就是 A 的 $(1,4)$ cofactor. 注意這裡求出的 cofactor 其實對應到 A 的 cofactor 所成的矩陣 A' 會是 A' 的 1-st row. 我們求出 A 其他的 cofactor 會有

$$a'_{21}=(-1)^{2+1}\det\begin{bmatrix}-1&3&5\\5&3&3\\-2&7&8\end{bmatrix}=-64,\quad a'_{22}=(-1)^{2+2}\det\begin{bmatrix}2&3&5\\0&3&3\\4&7&8\end{bmatrix}=-18,$$

$$a'_{23}=(-1)^{2+3}\det\begin{bmatrix}2&-1&5\\0&5&3\\4&-2&8\end{bmatrix}=20,\quad a'_{24}=(-1)^{2+4}\det\begin{bmatrix}2&-1&3\\0&5&3\\4&-2&7\end{bmatrix}=10.$$

以及 $a'_{31}=26,a'_{32}=8,a'_{33}=-8,a'_{34}=-4,a'_{41}=-20,a'_{42}=-6,a'_{43}=8,a'_{44}=2$. 因此得

$$A'=\begin{bmatrix}42&12&-16&-4\\-64&-18&20&10\\-26&8&-8&-4\\-20&-6&8&2\end{bmatrix},\quad \operatorname{adj}(A)=\begin{bmatrix}42&-64&-26&-20\\12&-18&8&-6\\-16&20&-8&8\\-4&10&-4&2\end{bmatrix}$$

也因此得

$$A^{-1}=\frac{1}{\det(A)}\operatorname{adj}(A)=\frac{1}{4}\begin{bmatrix} 42 & -64 & -26 & -20 \\ 12 & -18 & 8 & -6 \\ -16 & 20 & -8 & 8 \\ -4 & 10 & -4 & 2 \end{bmatrix}.$$

由上面例子可以看出利用 adjoint matrix 求反矩陣非常複雜, 所以在實際求反矩陣的情況還是利用從前學的 elementary row operation 的方法會比較快. 不過在證明抽象理論時, 利用 adjoint matrix 求 inverse 還是很有用的. 例如當 A 的每一個 entry 皆為整數時, 由於 $\operatorname{adj}(A)$ 的每一個 entry 也皆為整數, 因此若 $\det(A)=\pm 1$, 則由 Proposition 8.5.7 我們有以下的結果.

Corollary 8.5.9. 假設 A 為 $n\times n$ *matrix* 其中 A 的每一個 *entry* 皆為整數. 若 $\det(A)=\pm 1$, 則 A^{-1} 的每一個 *entry* 也皆為整數.

接下來我們探討有關 adjoint 的性質. 首先當 A 為 $n\times n$ invertible matrix 時, 由 Proposition 8.5.7 我們知 $\operatorname{adj}(A)=\det(A)A^{-1}$. 因此由 A^{-1} 亦為 invertible 且 $(A^{-1})^{-1}=A$, 我們得 $\operatorname{adj}(A^{-1})=\det(A^{-1})A$. 利用 determinant 的乘法性質 (Theorem 8.2.6 (2)), 我們有

$$\operatorname{adj}(A)\operatorname{adj}(A^{-1})=\det(A)\det(A^{-1})A^{-1}A=\det(I_n)I_n=I_n.$$

因此得證以下性質.

Proposition 8.5.10. 假設 A 為 $n\times n$ *invertible matrix,* 則 $\operatorname{adj}(A)$ 也是 $n\times n$ *invertible matrix* 且 $\operatorname{adj}(A)^{-1}=\operatorname{adj}(A^{-1})$.

當我們考慮 A^{T} 的 (i,j) minor matrix $(A^{\mathrm{T}})_{ij}$ 其實是將 A 的 (j,i) minor matrix A_{ji} 取 transpose, 亦即 $(A^{\mathrm{T}})_{ij}=(A_{ji})^{\mathrm{T}}$. 因此 A^{T} 的 (i,j) cofactor $(-1)^{i+j}\det((A^{\mathrm{T}})_{ij})$ 等於 $(-1)^{i+j}\det((A_{ji})^{\mathrm{T}})=(-1)^{j+i}\det(A_{ji})$ (我們用到 $\det(A^{\mathrm{T}})=\det(A)$). 也就是說 A^{T} 的 (i,j) cofactor 就是 A 的 (j,i) cofactor. 因此將 A 的 cofactor matrix A' 取轉置就是 A^{T} 的 cofactor matrix. 換言之, 若我們先將 A 轉置得 A^{T} 再取 A^{T} 的 cofactor matrix 也會是 $\operatorname{adj}(A)$. 也因此若將 A^{T} 轉置得 A (因為 $(A^{\mathrm{T}})^{\mathrm{T}}=A$) 後取 A 的 cofactor matrix 就會是 $\operatorname{adj}(A^{\mathrm{T}})$. 然而將 A 的 cofactor matrix 取轉置就是 $\operatorname{adj}(A)$, 所以我們得證以下的結果.

Proposition 8.5.11. 假設 A 為 $n\times n$ *matrix,* 則 $\operatorname{adj}(A^{\mathrm{T}})=\operatorname{adj}(A)^{\mathrm{T}}$.

當 A 為 $n\times n$ invertible matrix 時, 由 Proposition 8.5.7 我們知 $\operatorname{adj}(A)=\det(A)A^{-1}$. 因此由矩陣乘法與係數積關係 (Proposition 3.1.9) 得

$$A(\operatorname{adj}(A))=A(\det(A)A^{-1})=\det(A)(AA^{-1})=\det(A)I_n.$$

這個性質, 其實不只對 invertible matrix 成立, 其實對一般的 $n\times n$ matrix 皆成立. 現考慮一般的 $n\times n$ matrix $A=[a_{ij}]$. 依定義 $\operatorname{adj}(A)$ 的 (i,j)-th entry 為 a'_{ji}. 因此依矩陣乘法的定義 $A(\operatorname{adj}(A))$ 的 (i,j)-th entry 為

$$\sum_{k=1}^{n}a_{ik}a'_{j,k}=\sum_{k=1}^{n}a_{ik}(-1)^{j+k}\det(A_{jk}). \tag{8.8}$$

當 $i=j$ 時, 式子 (8.8) 恰為對 A 的 i-th row 展開所得的 $\det(A)$. 因此得 $A(\mathrm{adj}(A))$ 的 (i,i)-th entry (即 diagonal entry) 為 $\det(A)$. 而當 $i \neq j$ 時, 考慮 B 為將 A 的 j-th row 用 A 的 i-th row 取代, 而其他 row 保持不變的 $n \times n$ matrix. 此時對任意 $k=1,\ldots,n$, 因為 B 的 (j,k) minor matrix 是將 B 的 j-th row 除去, 所以和 A 的 (j,k) minor matrix 相同, 也就是說 $B_{jk}=A_{jk}$. 然而此時 B 的 (j,k)-th entry 因為是在 j-th row, 所以和 A 的 (i,k)-th entry 是一樣的, 也就是說 $b_{jk}=a_{ik}$. 現考慮利用 B 的 j-th row 展開所得的 $\det(B)$, 依定義為

$$\det(B)=\sum_{k=1}^{n} b_{jk}b'_{j,k}=\sum_{k=1}^{n} b_{jk}(-1)^{j+k}\det(B_{jk})=\sum_{k=1}^{n} a_{ik}(-1)^{j+k}\det(A_{jk}).$$

此與式子 (8.8) 是一致的. 也就是說當 $i \neq j$ 時, $A(\mathrm{adj}(A))$ 的 (i,j)-th entry 恰為 $\det(B)$. 然而 B 的 i-th row 和 j-th row 是一樣的, 故知 $\det(B)=0$. 因此我們證得當 $i \neq j$ 時, $A(\mathrm{adj}(A))$ 的 (i,j)-th entry 等於 0, 也因此證得了 $A(\mathrm{adj}(A))=\det(A)I_n$. 利用類似的方法考慮 column 的展開, 我們也可證得 $(\mathrm{adj}(A))A=\det(A)I_n$. 不過我們這裡想用前面有關 adjoint 和 transpose 的關係 (Proposition 8.5.11) 證明以下的定理.

Theorem 8.5.12. 假設 A 為 $n \times n$ *matrix,* 則

$$A(\mathrm{adj}(A))=(\mathrm{adj}(A))A=\det(A)I_n.$$

Proof. 前面已證得 $A(\mathrm{adj}(A))=\det(A)I_n$. 因為這是對所有 $n \times n$ matrix 皆成立, 故將 A 以 A^{T} 代換得 $A^{\mathrm{T}}(\mathrm{adj}(A^{\mathrm{T}}))=\det(A^{\mathrm{T}})I_n$. 由 Proposition 8.5.11 以及 $\det(A^{\mathrm{T}})=\det(A)$, 我們得 $A^{\mathrm{T}}(\mathrm{adj}(A)^{\mathrm{T}})=\det(A)I_n$. 將此等式兩邊取 transpose 得證 $(\mathrm{adj}(A))A=\det(A)I_n$. □

再次強調, Proposition 8.5.7 的證明是利用 Cramer's rule 因此需用到 $\det(A) \neq 0$ 之假設. 而 Theorem 8.5.12 是對一般 $n \times n$ matrix 皆成立的. 然而我們可以利用 Theorem 8.5.12 推得 Proposition 8.5.7, 所以可以說 Theorem 8.5.12 是 Proposition 8.5.7 的推廣.

Example 8.5.13. 考慮 $A=\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$. 依定義, 我們有

$$\mathrm{adj}(A)=\begin{bmatrix} \det\begin{bmatrix} b_2 & b_3 \\ c_2 & c_3 \end{bmatrix} & -\det\begin{bmatrix} a_2 & a_3 \\ c_2 & c_3 \end{bmatrix} & \det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \\ -\det\begin{bmatrix} b_1 & b_3 \\ c_1 & c_3 \end{bmatrix} & \det\begin{bmatrix} a_1 & a_3 \\ c_1 & c_3 \end{bmatrix} & -\det\begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} \\ \det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} & -\det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix} & \det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \end{bmatrix}.$$

依矩陣乘法定義 $A(\mathrm{adj}(A))$ 的 $(1,1)$-th, $(2,2)$-th 和 $(3,3)$-th entry 分別為

$$a_1\det\begin{bmatrix} b_2 & b_3 \\ c_2 & c_3 \end{bmatrix}-a_2\det\begin{bmatrix} b_1 & b_3 \\ c_1 & c_3 \end{bmatrix}+a_3\det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix},$$

$$-b_1\det\begin{bmatrix} a_2 & a_3 \\ c_2 & c_3 \end{bmatrix}+b_2\det\begin{bmatrix} a_1 & a_3 \\ c_1 & c_3 \end{bmatrix}-b_3\det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix},$$

$$c_1\det\begin{bmatrix} a_2 & a_3\\ b_2 & b_3\end{bmatrix}-c_2\det\begin{bmatrix} a_1 & a_3\\ b_1 & b_3\end{bmatrix}+c_3\det\begin{bmatrix} a_1 & a_2\\ b_1 & b_2\end{bmatrix}.$$

它們為 A 分別依 1-st, 2-nd 和 3-rd row 展開所得的 determinant, 故皆等於 $\det(A)$. 而 $A(\mathrm{adj}(A))$ 的 $(1,2)$-th, $(3,2)$-th entry 分別為

$$-a_1\det\begin{bmatrix} a_2 & a_3\\ c_2 & c_3\end{bmatrix}+a_2\det\begin{bmatrix} a_1 & a_3\\ c_1 & c_3\end{bmatrix}-a_3\det\begin{bmatrix} a_1 & a_2\\ c_1 & c_2\end{bmatrix},$$

$$-c_1\det\begin{bmatrix} a_2 & a_3\\ c_2 & c_3\end{bmatrix}+c_2\det\begin{bmatrix} a_1 & a_3\\ c_1 & c_3\end{bmatrix}-c_3\det\begin{bmatrix} a_1 & a_2\\ c_1 & c_2\end{bmatrix},$$

它們分別是將 $\begin{bmatrix} a_1 & a_2 & a_3\\ a_1 & a_2 & a_3\\ c_1 & c_2 & c_3\end{bmatrix}$, $\begin{bmatrix} a_1 & a_2 & a_3\\ c_1 & c_2 & c_3\\ c_1 & c_2 & c_3\end{bmatrix}$ 依 2-nd row 展開所得的 determinant, 故皆為 0. 同樣的 $A(\mathrm{adj}(A))$ 的 $(2,1)$-th, $(3,1)$-th entry 分別為

$$b_1\det\begin{bmatrix} b_2 & b_3\\ c_2 & c_3\end{bmatrix}-b_2\det\begin{bmatrix} b_1 & b_3\\ c_1 & c_3\end{bmatrix}+b_3\det\begin{bmatrix} b_1 & b_2\\ c_1 & c_2\end{bmatrix},$$

$$c_1\det\begin{bmatrix} b_2 & b_3\\ c_2 & c_3\end{bmatrix}-c_2\det\begin{bmatrix} b_1 & b_3\\ c_1 & c_3\end{bmatrix}+c_3\det\begin{bmatrix} b_1 & b_2\\ c_1 & c_2\end{bmatrix},$$

它們分別是將 $\begin{bmatrix} b_1 & b_2 & b_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}$, $\begin{bmatrix} c_1 & c_2 & c_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{bmatrix}$ 依 1-st row 展開所得的 determinant, 故皆為 0. 最後 $A(\mathrm{adj}(A))$ 的 $(1,3)$-th, $(2,3)$-th entry 分別為

$$a_1\det\begin{bmatrix} a_2 & a_3\\ b_2 & b_3\end{bmatrix}-a_2\det\begin{bmatrix} a_1 & a_3\\ b_1 & b_3\end{bmatrix}+a_3\det\begin{bmatrix} a_1 & a_2\\ b_1 & b_2\end{bmatrix}.$$

$$b_1\det\begin{bmatrix} a_2 & a_3\\ b_2 & b_3\end{bmatrix}-b_2\det\begin{bmatrix} a_1 & a_3\\ b_1 & b_3\end{bmatrix}+b_3\det\begin{bmatrix} a_1 & a_2\\ b_1 & b_2\end{bmatrix}.$$

它們分別是將 $\begin{bmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ a_1 & a_2 & a_3\end{bmatrix}$, $\begin{bmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ b_1 & b_2 & b_3\end{bmatrix}$ 依 3-rd row 展開所得的 determinant, 所以也皆為 0. 我們得

$$A(\mathrm{adj}(A))=\begin{bmatrix} \det(A) & 0 & 0\\ 0 & \det(A) & 0\\ 0 & 0 & \det(A)\end{bmatrix}=\det(A)\begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}.$$

最後我們要探討 $\mathrm{adj}(A)$ 的 determinant. 注意, 因為 $\det(A)I_n$ 是 diagonal matrix, 所以也是 upper triangular matrix. 因此依 Proposition 8.2.9 我們知 $\det(A)I_n$ 的 determinant 為其 diagonal entry 的乘積 $\det(A)^n$. 因此由 $A(\mathrm{adj}(A))=\det(A)I_n$ 等式兩邊取 determinant 得到以下結果.

Corollary 8.5.14. 假設 A 為 $n\times n$ *matrix,* 則

$$\det(\mathrm{adj}(A))=\det(A)^{n-1}.$$

8.6. 結論

Determinant 是 square matrix 上特有的一種函數. 它是從 $\mathscr{M}_{n\times n}$ 映射到 $\mathbb{R}$ 的一種 multi linear function. 另一方面 determinant 也將 $\mathbb{R}^2$ 中兩個向量所展出平行四邊形的有向面積以及 $\mathbb{R}^3$ 中三個向量所展出平行六面體的有向積積推廣到 $\mathbb{R}^n$ 的情形. 利用 determinant 我們可以判斷一個 square matrix 是否為 invertible, 也可幫助我們找到一個 invertible matrix 的 inverse, 甚至將聯立方成組的解寫下. 不過一般求 determinant 的過程頗繁雜, 我們可以善用 determinant 的性質, 適時的利用 elementary row operations 以及降階的方法將它求出.

Chapter 9

Eigenvalues and Eigenvectors

在這一章中, 我們介紹有關於 $n\times n$ matrices 的重要概念 eigenvalue 和 eigenvector. 我們首先介紹如何找到 eigenvalue 和 eigenvector 接著才介紹它們的性質. 最後我們探討 $n\times n$ matrix 對角化的問題以及其相關應用.

9.1. Characteristic Polynomial

給定 $n\times n$ matrix A, 以及 $\mathbf{v}\in\mathbb{R}^n$, 在數學上常會探討 $A^k\mathbf{v}$ 其中 k 為任意正整數的問題. 例如 Fibonacci sequence $F_1,F_2,\ldots$. 是一組滿足 $F_{k+1}=F_k+F_{k-1}$ 的遞迴數列. 若我們令 $A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$ 且對任意 $k\geq 2$ 令 $\mathbf{v}_k=\begin{bmatrix}F_k\\F_{k-1}\end{bmatrix}$, 則

$$A\mathbf{v}_k=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}F_k\\F_{k-1}\end{bmatrix}=\begin{bmatrix}F_k+F_{k-1}\\F_k\end{bmatrix}=\begin{bmatrix}F_{k+1}\\F_k\end{bmatrix}=\mathbf{v}_{k+1}.$$

因此我們有 $\mathbf{v}_3=A\mathbf{v}_2$, $\mathbf{v}_4=A\mathbf{v}_3=A(A\mathbf{v}_2)=A^2\mathbf{v}_2,\ldots$, 這樣一直下去可得 $\mathbf{v}_{k+1}=A^{k-1}\mathbf{v}_2$. 也就是說對於任意 $k\geq 2$, 我們只要能算出 $A^{k-1}\mathbf{v}_2$, 就能求出 F_{k+1} 為何.

一搬來說當 k 越大時, 計算 $A^k\mathbf{v}$ 就越困難. 不過在一種特殊其況, 即當存在 $\lambda\in\mathbb{R}$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$ 時, 我們有 $A^2\mathbf{v}=A(A\mathbf{v})=A(\lambda\mathbf{v})=\lambda(A\mathbf{v})=\lambda^2\mathbf{v}$. 同理我們會有 $A^3\mathbf{v}=\lambda^3\mathbf{v},\ldots$, $A^k\mathbf{v}=\lambda^k\mathbf{v}$. 也就是說在這種情況之下, 就很容易計算出 $A^k\mathbf{v}$. 因此我們對於怎樣的 $\mathbf{v}$ 會存在 $\lambda\in\mathbb{R}$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$ 特別有興趣. 所以有以下的定義.

Definition 9.1.1. 假設 A 為 $n\times n$ matrix. 若對於非零向量 $\mathbf{v}\in\mathbb{R}^n$, 存在 $\lambda\in\mathbb{R}$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$, 則稱 $\mathbf{v}$ 為 A 的一個 *eigenvector*, 而此 λ 稱為 A 的一個 *eigenvalue*.

注意, 依定義 A 的 eigenvector 一定是非零向量. 又若 $\mathbf{v}$ 是 A 的一個 eigenvector, 且 $\lambda,\lambda'\in\mathbb{R}$ 滿足 $A\mathbf{v}=\lambda\mathbf{v}=\lambda'\mathbf{v}$, 則由 $(\lambda-\lambda')\mathbf{v}=\mathbf{O}$ 以及 $\mathbf{v}\neq\mathbf{O}$, 可得 $\lambda=\lambda'$. 因此對於 A 的一個 eigenvector $\mathbf{v}$ 一定有也僅有一個實數 λ 會滿足 $A\mathbf{v}=\lambda\mathbf{v}$. 此時我們稱 eigenvector $\mathbf{v}$ 所對應的 eigenvalue 為 λ.

Question 9.1. 假設 A 為 $n\times n$ *matrix,* $\mathbf{v}\in\mathbb{R}^n$ 為非零向量滿足 $A\mathbf{v}=\mathbf{O}$. 是否 $\mathbf{v}$ 為 *eigenvector?* 其所對應的 *eigenvalue* 為何?

要如何找到一個 $n\times n$ matrix 的 eigenvector 及其對應的 eigenvalue 呢? 其實一般的找法是先找到 eigenvalue, 然後再找出與其對應的 eigenvector. 首先觀察若 $\lambda\in\mathbb{R}$ 是 A 的 eigenvalue, 表示存在一個非零向量 $\mathbf{v}\in\mathbb{R}^n$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$. 由於 $I_n\mathbf{v}=\mathbf{v}$, 所以看成矩陣的運算 $\lambda\mathbf{v}=(\lambda I_n)\mathbf{v}$. 因此 $A\mathbf{v}=\lambda\mathbf{v}$ 就等同於 $(A-\lambda I_n)\mathbf{v}=\mathbf{O}$. 換言之, λ 是 A 的 eigenvalue 等同於由 $n\times n$ matrix $A-\lambda I_n$ 所對應的 linear system $(A-\lambda I_n)\mathbf{x}=\mathbf{O}$ 有 nontrivial solution $\mathbf{x}=\mathbf{v}$. 由 Theorem 3.5.9, 這又等同於 $A-\lambda I_n$ 不是 invertible, 再由 Theorem 8.2.6(1) 知這也等同於 $\det(A-\lambda I_n)=0$. 總言之, 要找到 A 的 eigenvalue λ 就是要找到 λ 滿足 $\det(A-\lambda I_n)=0$.

要怎樣找到 λ 滿足 $\det(A-\lambda I_n)=0$ 呢? 假設 $A=[a_{ij}]$, 若我們將 t 視為變數, 考慮 $\det(A-tI_n)$. 由於

$$A-tI_n=\begin{bmatrix} a_{11}-t & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22}-t & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}-t\end{bmatrix}$$

利用數學歸納法, 我們可以證明 $\det(A-tI_n)$ 會是一個以 t 為變數的 n 次實係數多項式. 而若 $t=\lambda$ 為此多項式的一實數根, 則 λ 就會滿足 $\det(A-\lambda I_n)=0$, 也就是說 λ 就會是 A 的一個 eigenvalue. 反之, 若 λ 就會是 A 的一個 eigenvalue, 就表示 $t=\lambda$ 會是多項式 $\det(A-tI_n)$ 的一個根. 由此可知多項式 $\det(A-tI_n)$ 可以讓我們完全掌握 A 的 eigenvalue, 我們因而給它一個特別的定義.

Definition 9.1.2. 假設 A 為 $n\times n$ matrix, 考慮以 t 為變數的多項式 $p_A(t)=\det(A-tI_n)$. 我們稱 $p_A(t)$ 為 A 的 *characteristic polynomial* (特徵多項式)..

從上面的討論我們知道 $t=\lambda$ 為 characteristic polynomial $p_A(t)$ 的一個實根若且唯若 λ 為 A 的 eigenvalue. 要注意因為這裡我們談的是實矩陣 A 在 $\mathbb{R}^n$ 的 eigenvectors, 因此若 $t=\tau$ 是 $p_A(t)$ 的一個虛根 (即非實數的根), 此時假設存在 $\mathbf{v}\in\mathbb{R}^n$ 使得 $A\mathbf{v}=\tau\mathbf{v}$. 由於 $A\mathbf{v}\in\mathbb{R}^n$, 但 $\tau\mathbf{v}\notin\mathbb{R}^n$, 所以 $A\mathbf{v}=\tau\mathbf{v}$ 不可能成立. 因此在這個課程裡, 我們考慮 eigenvalue 僅考慮 characteristic polynomial 的實根.

Example 9.1.3. 考慮 $A=\begin{bmatrix}2&0&1\\1&3&1\\1&1&2\end{bmatrix}$, 此時 A 的 characteristic polynomial 為 $p_A(t)=\det(A-tI_3)=\det\begin{bmatrix}2-t&0&1\\1&3-t&1\\1&1&2-t\end{bmatrix}$. 對第一個 row 降階求行列式得

$$p_A(t)=(2-t)\det\begin{bmatrix}3-t&1\\1&2-t\end{bmatrix}+\det\begin{bmatrix}1&3-t\\1&1\end{bmatrix}=(2-t)(t^2-5t+6-1)+(1-(3-t)).$$

化簡可得 $p_A(t)=(2-t)(t^2-5t+4)=(2-t)(t-1)(t-4)$. 由此知 $t=1,2,4$ 為 A 的 characteristic polynomial 的三實根, 也因此得 A 有三個 eigenvalues $1,2,4$.

接下來我們說明當 A 為 $n\times n$ matrix 時, 其 characteristic polynomial $\det(A-tI_n)$ 確實是 t 的多項式. 首先觀察當我們在利用降階求 determinant 時, 其實是一些乘積之和. 利用數學歸納法可得這些乘積是由每一個 column 中的某個元素相乘而得而且它們都不會在同一個 row. 例如當我們計算 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的 characteristic polynomial $\det\begin{bmatrix} a-t & b \\ c & d-t \end{bmatrix}$ 時不難發現會貢獻 t 的最高次項乘積的是 $(a-t)(d-t)$ 而另一個乘積 bc 就僅影響到常數項, 因此其最高次項 t^2 與次高次項 t 的係數就完全由 $(a-t)(d-t)$ 的 t^2 與 t 的係數即 $at^2-(a+d)t$ 所決定. 現考慮 3×3 matrix $A=\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ 的 characteristic polynomial. 利用對第一個 row 降階的方式我們有

$$\det\begin{bmatrix} a-t & b & c \\ d & e-t & f \\ g & h & i-t \end{bmatrix} = (a-t)\det\begin{bmatrix} e-t & f \\ h & i-t \end{bmatrix} - b\det\begin{bmatrix} d & f \\ g & i-t \end{bmatrix} + c\det\begin{bmatrix} d & e-t \\ g & h \end{bmatrix}.$$

從前面 2×2 的情形我們看出 $\det\begin{bmatrix} e-t & f \\ h & i-t \end{bmatrix}$ 的 t^2 與次高次項 t 的係數就完全由 $(e-t)(i-t)$ 的 t^2 與 t 的係數所決定, 因此 $(a-t)(e-t)(i-t)$ 貢獻出 t^3 和 t^2 的係數. 而 $\det\begin{bmatrix} d & f \\ g & i-t \end{bmatrix}$ 和 $\det\begin{bmatrix} d & e-t \\ g & h \end{bmatrix}$ 最多僅有 t 的一次出現, 因此得 $\det(A-tI_3)$ 的 t^3 和 t^2 的係數完全由 $(a-t)(e-t)(i-t)$ 所決定. 也就是說 A 的 chacteristic polynomial $p_A(t)$ 為 3 次多項式且其最高次的兩項為 $(-1)^3t^3+(-1)^2(a+e+i)$. 這裡 a,e,i 為 A 的 diagonal entries, 它們之和 $a+e+i$ 我們稱為 A 的 trace, 用 $\mathrm{tr}(A)$ 來表示. 利用數學歸納法, 我們可得當 $A=[a_{ij}]$ 為 $n\times n$ matrix 時, A 的 characteristic polynomial $p_A(t)=\det(A-tI_n)$ 為 t 的 n 次實係數多項式, 且其最高次的兩項是由 $(a_{11}-t)(a_{22}-t)\cdots(a_{nn}-t)$ 所貢獻因此為 $(-1)^nt^n+(-1)^{n-1}(a_{11}+\cdots+a_{nn})t^{n-1}$. 由於 A 的 diagonal entries 之和 $a_{11}+\cdots+a_{nn}$ 我們定為 $\mathrm{tr}(A)$, 因此有以下之結論.

Proposition 9.1.4. 假設 A 為 $n\times n$ *matrix.* 則 A 的 *characteristic polynomial* 為 t 的 n 次實係數多項式. 其 t^n 項係數為 $(-1)^n$, t^{n-1} 項係數為 $(-1)^{n-1}\mathrm{tr}(A)$ 而常數項係數為 $\det(A)$.

Proof. 令 $p_A(t)=\det(A-tI_n)$, 由前面的討論我們僅剩討論 $p_A(t)$ 的常數項. 由於 $p_A(t)$ 是多項式所以它的常數項是 $p_A(0)=\det(A-0I_n)=\det(A)$. □

Question 9.2. 假設 A 為 $n\times n$ *matrix.* 試問 A 最多會有幾個相異的 *eigenvalues?*

接下來我們介紹一個和 eigenvalue 有關的定義. 若 $\lambda\in\mathbb{R}$ 是 A 的 eigenvalue. 由於 $t=\lambda$ 會是 A 的 characteristic polynomial $p_A(t)=\det(A-tI_n)$ 的一個根. 由因式定理知 $t-\lambda$ 會整除 $p_A(t)$. 若 $(t-\lambda)^m$ 可整除 $p_A(t)$, 但 $(t-\lambda)^{m+1}$ 不能整除 $p_A(t)$, 則我們稱 eigenvalue λ 的 *algebraic multiplicity* (代數重根數) 為 m. 當然了當 $t=\lambda$ 是 $p_A(t)$ 的一個單根, 我們就說 λ 的 algebraic multiplicity 為 1.

Question 9.3. *Identity matrix* I_n 的 *eigenvalue* 有哪些? 其 *algebraic multiplicity* 為何?

最後我們介紹一些和 characteristic polynomial 有關的性質. 一般來說兩個 $n\times n$ matrices 的 characteristic polynomial 可能不相同. 不過在一種特殊情況之下, 它們的

characteristic polynomial 會一樣. 當 A,B 為 $n\times n$ matrices, 若存在 $n\times n$ 的 invertible matrix U, 使得 $B=U^{-1}AU$, 則我們稱 A,B 為 *similar* (關於這個定義的原因我們以後會再詳述). 此時我們可得 A 和 B 的 characteristic polynomial 是相同的.

Proposition 9.1.5. *假設 A,B 為 $n\times n$ matrices 且存在 $n\times n$ 的 invertible matrix U 滿足 $B=U^{-1}AU$. 則 A 和 B 有相同的 characteristic polynomial.*

Proof. 依假設 B 的 characteristic polynomial 為 $\det(B-tI_n)=\det(U^{-1}AU-tI_n)$. 然而依矩陣乘法性質

$$U^{-1}(A-tI_n)U=U^{-1}AU-U^{-1}(tI_n)U=U^{-1}AU-tU^{-1}I_nU=U^{-1}AU-tI_n.$$

因此再由 determinant 的性質 (Theorem 8.2.6) 得

$$\det(B-tI_n)=\det(U^{-1}(A-tI_n)U)=\det(U^{-1})\det(A-tI_n)\det(U)=\det(A-tI_n).$$

得證 A 和 B 有相同的 characteristic polynomial. □

另一個會有相同的 characteristic polynomial 的情況就是 A 和 A^{T} 有相同的 characteristic polynomial.

Proposition 9.1.6. *假設 A 為 $n\times n$ matrix, 則 A 和 A^{T} 有相同的 characteristic polynomial*

Proof. 利用 trnaspose 的性質 $(A-tI_n)^{\mathrm{T}}=A^{\mathrm{T}}-tI_n^{\mathrm{T}}=A^{\mathrm{T}}-tI_n$ (Proposition 3.2.4), 故利用 Theorem 8.2.6 (3), 我們有

$$P_{A^{\mathrm{T}}}(t)=\det(A^{\mathrm{T}}-tI_n)=\det((A-tI_n)^{\mathrm{T}})=\det(A-tI_n)=P_A(t).$$

□

Question 9.4. *試說明 A 和 A^{T} 有相同的 eigenvalues 且對每個 eigenvalue 其在 A 和 A^{T} 的 algebraic multiplicity 也相同.*

9.2. Eigenspace

我們了解了如何找到一個 $n\times n$ matrix 的 eigenvalue 之後, 接下來便是要找出這些 eigenvalue 所對應的 eigenvectors.

假設 A 為 $n\times n$ matrix 且 $\lambda\in\mathbb{R}$ 為 A 的一個 eigenvalue. 由於 $\det(A-\lambda I_n)=0$, 我們知聯立方程組 $(A-\lambda I_n)\mathbf{x}=\mathbf{O}$ 存在非零的 nontrivial solution. 現假設 $\mathbf{v}\in\mathbb{R}^n$ 為非零向量且 $\mathbf{x}=\mathbf{v}$ 為 $(A-\lambda I_n)\mathbf{x}=\mathbf{O}$ 的一組解. 此即表示 $\mathbf{v}$ 滿足 $(A-\lambda I_n)\mathbf{v}=\mathbf{O}$, 亦即 $A\mathbf{v}=\lambda\mathbf{v}$. 故此時 $\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector. 反之, 若 $\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector, 則 $\mathbf{x}=\mathbf{v}$ 必為 $(A-\lambda I_n)\mathbf{x}=\mathbf{O}$ 的一組 nontrivial solution. 因此我們只要掌握 $n\times n$ matrix $A-\lambda I_n$ 的 nullspace (即 $\{\mathbf{v}\in\mathbb{R}^n\mid(A-\lambda I_n)\mathbf{v}=\mathbf{O}\}$) 中的非零向量就會是 A 相對於 λ 的 eigenvector. 由於 nullspace 是 vector space, 因此我們有以下的定義.

Definition 9.2.1. 假設 A 為 $n\times n$ matrix 且 $\lambda\in\mathbb{R}$ 為 A 的一個 eigenvalue. 則 $A-\lambda I_n$ 的 nullspace 稱為 A 對於 eigenvalue λ 的 *eigenspace*. 我們用 $E_A(\lambda)$ 來表示.

要注意對於 λ 的 eigenspace 並不是由以 λ 為 eigenvalue 的 eigenvectors 所組成. 這是因為零向量 $\mathbf{O}$ 不是 eigenvector, 但 vector space 必須包含 $\mathbf{O}$. 所以對於 λ 的 eigenspace 應該是由所有以 λ 為 eigenvalue 的 eigenvectors 和 $\mathbf{O}$ 所組成. 那為什麼要讓它形成 vector space 呢? 因為 vector space 有其方便性, 例如有了 vector space 我們就可以利用 dimension 來知道它的大小. 因此我們定義 $E_A(\lambda)$ 的 dimension 為 eigenvalue λ 的 *geometric multiplicity* (幾何重根數). 要注意 eigenvalue λ 的 algebraic multiplicity 無法讓我們知道 λ 所對應的 eigenvectors 的多寡, 而是 λ 的 geometric multiplicity 可以提供這一個訊息.

Example 9.2.2. 考慮 $A = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 3 & 1 \\ -1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. 可計算出 A 和 B 有相同的 characteristic polynomial $-(t-1)^2(t-2)$. 也因此 1 和 2 皆為 A 和 B 的 eigenvalues. 而且對於矩陣 A 和 B, eigenvalue 1 的 algebraic multiplicity 皆為 2, 而 eigenvalue 2 的 algebraic multiplicity 皆為 1. 接下來我們分節計算 A 和 B 的 eigenspace.

首先考慮 A 對於 eigenvalue 1 的 eigenspace, 亦即找出 $A - I_3 = \begin{bmatrix} -2 & 4 & 2 \\ -1 & 2 & 1 \\ -1 & 2 & 1 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} 1 & -2 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. 可得 $E_A(1) = \mathrm{Span}(\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 1 的 eigenvector 就是那些由 $\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$ 和 $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ 的 linear combination 所得的 nonzero vector. 例如 $\mathbf{v} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ 就滿足

$$A\mathbf{v} = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \mathbf{v}.$$

由於 $\dim(E_A(1)) = 2$, 我們也得到 A 對於 eigenvalue 1 的 geometric multiplicity 為 2. 至於 A 對於 eigenvalue 2 的 eigenspace, 亦即找出 $A - 2I_3 = \begin{bmatrix} -3 & 4 & 2 \\ -1 & 1 & 1 \\ -1 & 2 & 0 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$. 因此得 $E_A(2) = \mathrm{Span}(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 2 的 eigenvector 就是那些和 $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ 平行的 nonzero vector, 我們也得到 A 對於 eigenvalue 2 的 geometric multiplicity 為 1.

至於 B 對於 eigenvalue 1 的 eigenspace, 即 $B-I_3=\begin{bmatrix}-1&3&1\\-1&2&1\\0&1&0\end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix}1&0&-1\\0&1&0\\0&0&0\end{bmatrix}$. 可得 $E_B(1)=\mathrm{Span}(\begin{bmatrix}1\\0\\1\end{bmatrix})$. 也就是說 B 對於 eigenvalue 為 1 的 eigenvector 就是那些和 $\begin{bmatrix}1\\0\\1\end{bmatrix}$ 平行的 nonzero vector, 我們也得到 B 對於 eigenvalue 1 的 geometric multiplicity 為 1. 至於 B 對於 eigenvalue 2 的 eigenspace, 亦即找出 $B-2I_3=\begin{bmatrix}-2&3&1\\-1&1&1\\0&1&-1\end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix}1&0&-2\\0&1&-1\\0&0&0\end{bmatrix}$. 因此得 $E_A(2)=\mathrm{Span}(\begin{bmatrix}2\\1\\1\end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 2 的 eigenvector 就是那些和 $\begin{bmatrix}2\\1\\1\end{bmatrix}$ 平行的 nonzero vector, 我們也得到 A 對於 eigenvalue 2 的 geometric multiplicity 為 1.

從 Example 9.2.2 我們知道即使兩個矩陣有相同的 characteristic polynomial, 他們的 eigenvectors 有可能有很大的差異, 甚至連 eigenspace 的 dimension 也可能不同. 因此有可能兩個矩陣有相同的 eigenvalues 且相對應的 algebraic multiplicity 也相同, 但其 geometric multiplicity 卻相異.

Question 9.5. *Identity matrix I_n 的 eigenvalue 的 geometric multiplicity 為何?*

由 Proposition 9.1.6 我們知道 A 和 A^{T} 有相同的 characteristic polynomial 所以他們有相同的 eigenvalue 而且這些 eigenvalue 在 A 和 A^{T} 的 algebraic multiplicity 會相同. 這對 geometric multiplicity 也成立, 我們有以下的結果.

Proposition 9.2.3. 假設 A 為 $n\times n$ *matrix* 且 $\lambda\in\mathbb{R}$ 為 A 的一個 *eigenvalue.* 則 λ 對於 A 的 *geometric multiplicity* 與 λ 對於 A^{T} 的 *geometric multiplicity* 相等.

Proof. 我們要說明 $\dim(E_A(\lambda))=\dim(E_{A^{\mathrm{T}}}(\lambda))$, 亦即 $\dim(N(A-\lambda I_n))=\dim(N(A^{\mathrm{T}}-\lambda I_n))$. 由 Theorem 4.4.13 我們知 $\dim(N(A-\lambda I_n))=\mathrm{null}(A-\lambda I_n)=n-\mathrm{rank}(A-\lambda I_n)$, 同理由 $A\in\mathscr{M}_{n\times n}$ 得 $\dim(N(A^{\mathrm{T}}-\lambda I_n))=n-\mathrm{rank}(A^{\mathrm{T}}-\lambda I_n)$. 因為 $A^{\mathrm{T}}-\lambda I_n=(A-\lambda I_n)^{\mathrm{T}}$ 以及 $\mathrm{rank}((A-\lambda I_n)^{\mathrm{T}})=\mathrm{rank}(A-\lambda I_n)$ (Proposition 4.4.14), 得證 $\dim(N(A-\lambda I_n))=\dim(N(A^{\mathrm{T}}-\lambda I_n))$. □

當 $\mathbf{v}$ 為 A 的 eigenvector, 若其 eigenvalue 為 λ, 則和 $\mathbf{v}$ 平行的 nonzero vector 皆為 eigenvalue 為 λ 的 eigenvector. 也因此若 $\mathbf{v},\mathbf{w}$ 為 A 的 eigenvectors 而他們所對應的 eigenvalue 是相異時, $\mathbf{v},\mathbf{w}$ 不可能平行. 也就是說 $\mathbf{v},\mathbf{w}$ 為 linearly independent. 這個結果可推廣到更一般的狀況.

Proposition 9.2.4. 假設 A 為 $n\times n$ *matrix* 且 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 A 的 *eigenvectors*. 若 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 所對應的 *eigenvalues* 皆相異, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 *linearly independent.*

Proof. 我們利用數學歸納法證明. 前面已知 $k=2$ 的情形成立, 接著我們假設有 $k-1$ 個 eigenvectors 的情形也成立. 現考慮 k 個 eigenvectors 的情形. 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 A 的 eigenvectors 且其對應的 eigenvalue 分別為 $\lambda_1,\ldots,\lambda_k$ (亦即 $A\mathbf{v}_i=\lambda_i\mathbf{v}_i$, for $i=1,\ldots,n$). 依歸納法之假設 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$ 為 linearly independent. 現用反證法, 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1},\mathbf{v}_k$ 為 linearly dependent. 依 Lemma 4.2.4, 這表示 $\mathbf{v}_k\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{k-1})$. 也就是說存在 $c_1,\ldots,c_{k-1}\in\mathbb{R}$ 使得

$$\mathbf{v}_k=c_1\mathbf{v}_1+\cdots+c_{k-1}\mathbf{v}_{k-1} \tag{9.1}$$

利用 eigenvector 的定義我們得

$$\lambda_k\mathbf{v}_k=A\mathbf{v}_k=A(c_1\mathbf{v}_1+\cdots+c_{k-1}\mathbf{v}_{k-1})=c_1A\mathbf{v}_1+\cdots+c_{k-1}A\mathbf{v}_{k-1}=c_1\lambda_1\mathbf{v}_1+\cdots c_{k-1}\lambda_{k-1}\mathbf{v}_{k-1}. \tag{9.2}$$

將式子 (9.1) 乘上 λ_k 與式子 (9.2) 相減得

$$c_1(\lambda_k-\lambda_1)\mathbf{v}_1+\cdots+c_{k-1}(\lambda_k-\lambda_{k-1})\mathbf{v}_{k-1}=\mathbf{O}. \tag{9.3}$$

由於 $\mathbf{v}_k\neq\mathbf{O}$, 我們知 $c_1,\ldots,c_{k-1}$ 不全為 0. 而由 eigenvalue 皆相異, 我們知對任意 $i=1,\ldots,k-1$, 皆有 $\lambda_k-\lambda_i\neq 0$. 因此 $c_1(\lambda_k-\lambda_1),\ldots,c_{k-1}(\lambda_k-\lambda_{k-1})$ 為不全為 0 的實數. 換句話說, 式子 (9.3) 告訴我們 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$ 為 linearly dependent, 此與歸納法之假設相矛盾, 得證本定理. □

9.3. Diagonalization

假設 A 為 $n\times n$ matrix, 我們已經知道當 $\mathbf{v}$ 為 A 的 eigenvector 且其對應的 eigenvalue 為 λ, 則我們可以很容易計算 $A^k\mathbf{v}=\lambda^k\mathbf{v}$. 不過若 $\mathbf{v}$ 不是 A 的 eigenvector 怎麼辦? 在一種特殊的情況之下, 我們仍然可以很輕鬆的計算 $A^k\mathbf{v}$ 為何. 事實上, 當 $\mathbf{v}_1,\ldots,\mathbf{v}\in\mathbb{R}^n$ 是 $\mathbb{R}^n$ 的一組 basis 且皆為 A 的 eigenvectors 時, 我們就可以很容易的計算 $A^k\mathbf{v}$. 這是因為此時若 $\mathbf{v}\in\mathbb{R}^n$, 則由 basis 的定義知存在 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 現若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 所對應的 eigenvalues 分別為 $\lambda_1,\ldots,\lambda_n$, 此時

$$A\mathbf{v}=A(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1A\mathbf{v}_1+\cdots+c_nA\mathbf{v}_n=c_1\lambda_1\mathbf{v}_1+\cdots+c_n\lambda_n\mathbf{v}_n.$$

若左邊再乘上 A 得

$$A^2\mathbf{v}=A(A\mathbf{v})=A(c_1\lambda_1\mathbf{v}_1+\cdots+c_n\lambda_n\mathbf{v}_n)=c_1\lambda_1A\mathbf{v}_1+\cdots+c_n\lambda_nA\mathbf{v}_n=c_1\lambda_1^2\mathbf{v}_1+\cdots+c_n\lambda_n^2\mathbf{v}_n.$$

因此利用數學歸納法可得

$$A^k\mathbf{v}=c_1\lambda_1^k\mathbf{v}_1+\cdots+c_n\lambda_n^k\mathbf{v}_n.$$

我們對有這樣特點的 matrix 給了以下的定義.

Definition 9.3.1. 假設 A 為 $n\times n$ matrix. 若存在 $\mathbb{R}^n$ 的一組 basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 A 的 eigenvectors, 則稱 A 為 *diagonalizable* (可對角化).

為甚麼稱為 diagonalizable 呢? 這是因為若 $\mathbf{v}_1,\ldots,\mathbf{v}\in\mathbb{R}^n$ 是 $\mathbb{R}^n$ 的一組 basis 且皆為 A 的 eigenvectors, 又假設它們所對應的 eigenvalues 分別為 $\lambda_1,\ldots,\lambda_n$. 亦即 $A\mathbf{v}_1=\lambda_1\mathbf{v}_1,\ldots,A\mathbf{v}_n=\lambda_n\mathbf{v}_n$. 此時由矩陣乘法的定義我們有

$$A\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}=\begin{bmatrix}| & | & & |\\ A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n\\ | & | & & |\end{bmatrix}=\begin{bmatrix}| & | & & |\\ \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\\ | & | & & |\end{bmatrix}.$$

另一方面若考慮 (i,i)-th entry 為 λ_i 的 $n\times n$ diagonal matrix (即對角線第 i 個位置為 λ_i 而對角線外其餘位置皆為 0), 則我們有

$$\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}D=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \lambda_n\end{bmatrix}=\begin{bmatrix}| & | & & |\\ \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_3\mathbf{v}_n\\ | & | & & |\end{bmatrix}.$$

因此若令 $C=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}$, 則我們有 $AC=CD$. 現又因 C 的 column 之間為 linearly independent 且有 n 個 column, 我們得 C 的 rank 為 n, 因此由 C 為 $n\times n$ matrix 得知 C 為 invertible (參見 Theorem 3.5.2). 因此我們可將 $AC=CD$ 改寫成 $D=C^{-1}AC$. 反之, 若存在一個 $n\times n$ invertible matrix 使得 $C^{-1}AC$ 為 diagonal matrix D, 則因 C 為 $n\times n$ invertible matrix, 所以 C 的 n 個 column vectors 形成 $\mathbb{R}^n$ 的一組 basis. 又因為 $AC=CD$, 由上面矩陣乘法的性質知 C 的 i-th column 就會是 A 以 D 的 (i,i)-th entry 為 eigenvalue 的 eigenvector. 所以 C 的 column vectors 就是 $\mathbb{R}^n$ 的一組 basis 且為 A 的 eigenvectors, 也就是說 A 為 diagonalizable. 前面曾經提過, 形如 $U^{-1}AU$ (其中 U 為 $n\times n$ invertible matrix) 這樣的 matrix 就稱為和 A 為 similar 的 matrix. 因此由這裡的討論, 我們知道 A 為 diagonalizable 就等同於 A 和一個 diagonal matrix 是 similar. 這也就是 diagonalizable 這個名稱的原因.

要如何知道一個 $n\times n$ matrix A 是否為 diagonalizable 呢? 從其定義, 我們知道它必須要有夠多的 eigenvectors. 以下我們要看一種特殊的情況可以確保 A 有夠多的 eigenvectors 從而得到 A 為 diagonalizable. 首先要有夠多的 eigenvectors 就表示要有夠多的 eigenvalues, 所以我們假設 A 的 characteristic polynomial 可以在 $\mathbb{R}$ 中完全分解. 也就是存在 $\lambda_1,\ldots,\lambda_k\in\mathbb{R}$ 皆相異且滿足 $p_A(t)=(-1)^n(t-\lambda_1)^{m_1}\cdots(t-\lambda_k)^{m_k}$. 依定義對於 $i=1,\ldots,k$, m_i 就是 λ_i 的 algebraic multiplicity 而且因 $p_A(t)$ 的次數為 n, 我們有 $m_1+\cdots+m_k=n$. 等一下我們會證明對於每個 eigenvalue, 其 geometric multiplicity 會小於等於其 algebraic multiplicity. 所以這裡 A 的 eigenvectors 要夠多, 最好的狀況就是每一個 eigenvalue 其 geometric multiplicity 等於其 algebraic multiplicity. 所以這裡我們假設對於 $i=1,\ldots,k$, λ_i 的 geometric multiplicity 等於其 algebraic multiplicity, 亦即 $\dim(E_A(\lambda_i))=m_i$. 此時我們令 $\mathbf{v}_{i,1},\ldots,\mathbf{v}_{i,m_i}$ 為 $E_A(\lambda_i)$ 的一組 basis. 將這 k 組 vectors 收集在一起後, 我們要說明它們 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly independent. 此時由於它們在 $\mathbb{R}^n$ 中且共有 $m_1+\cdots+m_k=n$ 個向量, 因此由 Corollary 4.3.5, 知它們是 $\mathbb{R}^n$ 中的一組 basis. 再加上它們皆為 A 的 eigenvectors, 所以可知此時 A 為 diagonalizable.

現假設 $\mathbf{v}_{1,1},\dots,\mathbf{v}_{1,m_1},\dots,\mathbf{v}_{k,1},\dots,\mathbf{v}_{k,m_k}$ 是 linearly dependent. 亦即存在不全為 0 的實數 $c_{1,1},\dots,c_{1,m_1},\dots, c_{k,1},\dots,c_{k,m_k}$ 使得

$$c_{1,1}\mathbf{v}_{1,1}+\cdots+c_{1,m_1}\mathbf{v}_{1,m_1}+\cdots+c_{k,1}\mathbf{v}_{k,1}+\cdots+c_{k,m_k}\mathbf{v}_{k,m_k}=\mathbf{O}.$$

此時對任意 $i\in\{1,\dots,k\}$, 我們令 $\mathbf{w}_i=c_{i,1}\mathbf{v}_{i,1}+\cdots+c_{i,m_i}\mathbf{v}_{i,m_i}$. 因此由於 $\mathbf{v}_{i,1},\dots,\mathbf{v}_{i,m_i}$ 為 linearly independent, 如果 $c_{i,1},\dots,c_{i,m_i}$ 不全為 0, 可得 $\mathbf{w}_i\neq\mathbf{O}$. 但由於 $w_i\in E_A(\lambda_i)$, 故此時 $\mathbf{w}_i$ 為 eigenvalue 為 λ_i 的 eigenvector. 也就是說, 若存在某些 $c_{i,j}\neq 0$, 則對於那些 i, $\mathbf{w}_i$ 會是 eigenvalue 為 λ_i 的 eigenvectors 滿足 $\mathbf{w}_1+\cdots+\mathbf{w}_k=\mathbf{O}$. 此與 Proposition 9.2.4 所述, 不同 eigenvalue 的 eigenvectors 之間是 linearly independent 的結果相矛盾, 故得證 $\mathbf{v}_{1,1},\dots,\mathbf{v}_{1,m_1},\dots,\mathbf{v}_{k,1},\dots,\mathbf{v}_{k,m_k}$ 是 linearly independent. 我們因此證得了當 A 的 characteristic polynomial 可以在 $\mathbb{R}$ 中完全分解且 A 的每一個 eigenvalue 的 geometric multiplicity 等於其 algebraic multiplicity, 則 A 為 diagonalizable.

其實反過來也是對的, 也就是說若 A 為 diagonalizable, 則 A 的 characteristic polynomial 可以在 $\mathbb{R}$ 中完全分解而且 A 的每一個 eigenvalue 的 geometric multiplicity 等於其 algebraic multiplicity. 不過在證明之前我們先證明前面提過的一般來說一個 eigenvalue 的 geometric multiplicity 會小於等於其 algebraic multiplicity.

Proposition 9.3.2. *假設 A 為 $n\times n$ matrix. 若 λ 為 A 的一個 eigenvalue 且其 geometric multiplicity 為 d 以及 algebraic multiplicity 為 m, 則 $d\leq m$.*

Proof. 依假設 $\dim(E_A(\lambda))=d$, 故令 $\mathbf{v}_1,\dots,\mathbf{v}_d$ 為 $E_A(\lambda)$ 的一組 basis. 由於 $\mathbf{v}_1,\dots,\mathbf{v}_d$ 為 linearly independent, 我們可以將之拓展成 $\mathbb{R}^n$ 中的一組 basis $\mathbf{v}_1,\dots,\mathbf{v}_d,\mathbf{v}_{d+1},\dots,\mathbf{v}_n$. 令 C 為 i-th column 為 $\mathbf{v}_i$ 的 $n\times n$ invertible matrix. 此時利用矩陣乘法可得 $AC=CE$ 其中 $E=\left[\begin{array}{c|c}\lambda I_d & M_1\\ \hline \mathbf{O} & M_2\end{array}\right]$. 對於 1-st column 展開, 利用數學歸納法我們可以證明 $\det(E-tI_n)=(\lambda-t)^d\det(M_2-tI_{n-d})$. 換言之, E 的 characteristic polynomial 可以被 $(t-\lambda)^d$ 所整除. 然而 A 和 E 為 similar (因為 $E=C^{-1}AC$), 所以它們有相同的 characteristic polynomial (參見 Proposition 9.1.5), 因此得 $(t-\lambda)^d$ 可整除 $p_A(t)$. 然而 λ 的 algebraic multiplicity 為 m, 表示 m 為 $t-\lambda$ 可以整除 $p_A(t)$ 的最高次數, 因此得證 $d\leq m$. □

現假設 $n\times n$ matrix A 是 diagonalizable. 依定義令 $\mathbf{v}_{1,1},\dots,\mathbf{v}_{1,d_1},\dots,\mathbf{v}_{k,1},\dots,\mathbf{v}_{k,d_k}$ 是 $\mathbb{R}^n$ 的一組 basis, 且對任意 $i\in\{1,\dots,k\}$, $\mathbf{v}_{i,1},\dots,\mathbf{v}_{i,d_i}$ 為 A 以 λ_i 為 eigenvalue 的 eigenvector, 其中 $\lambda_1,\dots,\lambda_k$ 皆相異. 由於 $\mathbf{v}_{i,1},\dots,\mathbf{v}_{i,d_i}\in E_A(\lambda_i)$ 且為 linearly independent, 我們知 λ_i 的 geometric multiplicity $\dim(E_A(\lambda_i))\geq d_i$. 現又假設每個 λ_i 的 algebraic multiplicity 為 m_i, 由 Proposition 9.3.2 我們有

$$m_i\geq\dim(E_A(\lambda_i))\geq d_i,\forall i=1,\dots,k. \tag{9.4}$$

由於 $m_1+\cdots+m_k$ 表示 A 的 characteristic polynomial $p_A(t)$ 的實根個數 (含重根), 其值會小於等於 $p_A(t)$ 的次數 n. 而 $m_1+\cdots+m_k$ 表示 $\mathbb{R}^n$ 的 dimension, 即 n. 因此將式子 (9.4) 中 $i=1,\dots,k$ 加起來可得

$$n\geq m_1+\cdots+m_k\geq\dim(E_A(\lambda_i))+\cdots+\dim(E_A(\lambda_k))\geq d_1+\cdots+d_k=n.$$

因此得知上式中 “$\geq$” 應為 “$=$” (否則有一項為不等會造成 $n>n$ 之矛盾). 也就是說 $n=m_1+\cdots+m_k$ (這表示 $p_A(t)$ 可以在實數中完全分解) 以及 $m_i=\dim(E_A(\lambda_i)),\forall i=1,\ldots,k$ (這表示每個 eigenvalue 的 geometric multiplicity 等於其 algebraic multiplicity). 綜合以上的討論我們有以下的結論.

Theorem 9.3.3. 假設 A 為 $n\times n$ *matrix.* 以下敘述是等價的.

(1) $\mathbb{R}^n$ 中存在一組 *basis* 是由 A 的 *eigenvectors* 所組成.

(2) 存在一個 $n\times n$ *invertible matrix* C 使得 $C^{-1}AC$ 為 *diagonal matrix.*

(3) A 的 *characteristic polynomial* 可在實數中完全分解且 A 的每個 *eigenvalue* 的 *geometric multiplicity* 等於其 *algebraic multiplicity.*

依照 diagonalizable matrix 的定義, 我們可以將 Theorem 9.3.3 中任一項當成檢驗矩陣是否為 diagonalizable 的方法, 其中以 (3) 是最常見的.

Question 9.6. 假設 A 為 $n\times n$ *matrix.* 試說明 A 為 *diagonalizable* 若且唯若 A^{T} 為 *diagonalizable.*

Example 9.3.4. 我們考慮 Example 9.2.2 中的矩陣 $A=\begin{bmatrix}-1&4&2\\-1&3&1\\-1&2&2\end{bmatrix}, B=\begin{bmatrix}0&3&1\\-1&3&1\\0&1&1\end{bmatrix}$. 前面已知它們有相同的 characteristic polynomial $-(t-1)^2(t-2)$. 也因此 A,B 的 eigenvalue 1 其 algebraic multiplicity 皆為 2, 而 eigenvalue 2 的 algebraic multiplicity 皆為 1. 不過在 Example 9.2.2 中我們知道 B 在 eigenvalue 1 的 geometric multiplicity 為 1, 所以 B 不是 diagonalizable matrix. 反之, A 在 eigenvalue 1 和 eigenvalue 2 的 geometric multiplicity 皆等於其 algebraic multiplicity, 所以 A 為 diagonalizable matrix. 我們看如何將 A 對角化.

由於 A 對於 eigenvalue 為 1 和 2 的 eigenspace 分別為 $E_A(1)=\mathrm{Span}(\begin{bmatrix}2\\1\\0\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix})$ 和 $E_A(2)=\mathrm{Span}(\begin{bmatrix}2\\1\\1\end{bmatrix})$, 可得 $\begin{bmatrix}2\\1\\0\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix},\begin{bmatrix}2\\1\\1\end{bmatrix}$ 就是一組由 A 的 eigenvectors 所形成的 $\mathbb{R}^3$ 的 basis. 因此若令 $C=\begin{bmatrix}2&1&2\\1&0&1\\0&1&1\end{bmatrix}$ 以及 $D=\begin{bmatrix}1&0&0\\0&1&0\\0&0&2\end{bmatrix}$, 則

$$AC=\begin{bmatrix}-1&4&2\\-1&3&1\\-1&2&2\end{bmatrix}\begin{bmatrix}2&1&2\\1&0&1\\0&1&1\end{bmatrix}=\begin{bmatrix}2&1&4\\1&0&2\\0&1&2\end{bmatrix}=\begin{bmatrix}2&1&2\\1&0&1\\0&1&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\0&0&2\end{bmatrix}=CD.$$

再由 C 為 invertible, 得 $C^{-1}AC=D$.

最後我們要強調的是, 由於 A 的 eigenvalue λ 的 geometric multiplicity 必大於 0 (因對應 λ 的 eigenvector 必存在) 且其值必小於等於其 algebraic multiplicity (Proposition 9.3.2). 因此若 λ 是 A 的 characteristic polynomial 的單根 (即 λ 的 algebraic multiplicity 為 1), 其 geometric multiplicity 一定等於其 algebraic multiplicity (皆為 1). 所以在檢查一個矩陣

是否為 diagonalizable 時, 對於 algebraic multiplicity 為 1 的 eigenvalue 我們就不必檢查其 geometric multiplicity 了. 舉例來說, 若 A 的 characteristic polynomial 可在 $\mathbb{R}$ 中完全分解且其根皆為單根 (無重根), 則 A 一定為 diagonalizable. 另外還有一種矩陣不必檢查就知道一定是 diagonalizable, 就是 symmetric matrix. 以後我們將會證明所有的 symmetric matrix 皆為 diagonalizable.

9.4. The Spectral Theorem

在這一節中我們要探討 symmetric matrix. 我們將證明 symmetric matrix 皆為 diagonalizable, 更重要的是它們都是所謂的 *orthogonal diagonalizable*. 這個結果在數學和物理方面都有很重要的應用, 不過我們不會深入探討它的應用, 而著重於說明如何將 symmetric matrix 對角化.

首先我們來看 2×2 symmetric matrix 的情形. 假設 $A=\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, 其中 $b\neq 0$ (因為若 $b=0$, 此時 A 已為 diagonal matrix 不必對角化). 此時 A 的 characteristic polynomial 為 $P_A(t)=t^2-(a+c)t+(ac-b^2)$. 由於 $p_A(t)$ 的判別式 $(a+c)^2-4(ac-b^2)=(a-c)^2+4b^2>0$, 我們得 $P_A(t)=0$ 有兩相異實根 λ_1,λ_2. 也就是說 λ_1,λ_2 為 A 的兩相異 eigenvealue, 故知 A 為 diagonalizable. 事實上若令 $\mathbf{v}_1=\begin{bmatrix} b \\ \lambda_1-a \end{bmatrix}$, 我們有

$$A\mathbf{v}_1=\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} b \\ \lambda_1-a \end{bmatrix}=\begin{bmatrix} \lambda_1 b \\ b^2+\lambda_1 c-ac \end{bmatrix}=\lambda_1\begin{bmatrix} b \\ \lambda_1-a \end{bmatrix}=\lambda_1\mathbf{v}_1.$$

注意這裡我們用到了 $\lambda_1^2-(a+c)\lambda_1+(ac-b^2)=0$. 由於 $b\neq 0$, 我們知 $\mathbf{v}_1\neq\mathbf{O}$, 故 $\mathbf{v}_1$ 是 A 的 eigenvector 其 eigenvalue 為 λ_1. 同理令 $\mathbf{v}_2=\begin{bmatrix} b \\ \lambda_2-a \end{bmatrix}$, 我們可得 $\mathbf{v}_2$ 為 A 的 eigenvector 其 eigenvalue 為 λ_2. 重要的是, 我們有 $\langle\mathbf{v}_1,\mathbf{v}_2\rangle=b^2+\lambda_1\lambda_2-a(\lambda_1+\lambda_2)+a^2$. 利用根與係數關係, 即 $\lambda_1\lambda_2=ac-b^2$ 以及 $\lambda_1+\lambda_2=a+c$, 我們得 $\langle\mathbf{v}_1,\mathbf{v}_2\rangle=0$. 也就是說 $\mathbf{v}_1,\mathbf{v}_2$ 這組 $\mathbb{R}^2$ 的 basis 不只是由 A 的 eigenvectors 所組成, 而且它們倆倆互相垂直. 這種比一般 diagonalizable 更強的條件我們便稱之為 orthogonal diagonalizable. 其正式的定義如下.

Definition 9.4.1. 假設 $A\in\mathscr{M}_{n\times n}$, 若存在一組 $\mathbb{R}^n$ 的 orthogonal basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 A 的 eigenvectors, 則稱 A 為 *orthogonal diagonalizable*.

當然了, 在 Definition 9.4.1 中若令 $\mathbf{u}_i=\frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$ 則 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 為 $\mathbb{R}^n$ 的一組 orthonormal basis 且皆為 A 的 eigenvectors. 所以 A 為 orthogonal diagonalizable 也等同於 $\mathbb{R}^n$ 中有一組 orthonormal basis 是由 A 的 eigenvector 所組成. 此時若 $\mathbf{u}_i$ 所對應的 eigenvalue 為 λ_i 且令 $Q=\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix}$ 則可得 $AQ=QD$ 其中 D 為 (i,i)-th entry 為 λ_i 的 diagonal matrix, 也就是說我們可以將 A 對角化成 $Q^{-1}AQ=D$. 一般由 eigenvectors 所形成的 basis 都可以達到這個對角化的目的, 為何特別考慮 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 為 orthonormal basis 的情形呢? 這是因為當 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 為 $\mathbb{R}^n$ 的 orthonormal basis 時, 我們會有 $Q^{\mathrm{T}}Q=I_n$, 也因此由 inverse matrix 的唯一性, 我們知 $Q^{\mathrm{T}}=Q^{-1}$. 也就是說當 Q 的 column vectors 為 $\mathbb{R}^n$ 的 orthonormal basis

時, 我們可以馬上得知 $Q^{-1}=Q^{\mathrm{T}}$. 就因為這個特性, 當一個 $n\times n$ matrix 其 column vectors 是由 $\mathbb{R}^n$ 的 orthonormal basis 所組成時, 我們特別稱之為 *orthogonal matrix* (注意不是稱為 orthonormal matrix). 也因此我們可以將 A 對角化成 $Q^{\mathrm{T}}AQ=D$, 故稱 A 為 orthogonal diagonalizable.

Question 9.7. 假設 $Q\in\mathscr{M}_{n\times n}$, 是否 $Q^{-1}=Q^{\mathrm{T}}$ 即表示 Q 為 *orthogonal matrix?*

反之, 若存在 Q 為 $n\times n$ orthogonal matrix 以及 $D=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \end{bmatrix}$ 為 $n\times n$ diagonal matrix 使得 $Q^{\mathrm{T}}AQ=D$. 此時由 $AQ=QD$, 知 Q 的 i-th column 為 A 的 eigenvalue 為 λ_i 的 eigenvector, 也因此由 Q 的 column vectors 形成 $\mathbb{R}^n$ 的 orthonormal basis, 我們有以下之結果.

Proposition 9.4.2. 假設 $A\in\mathscr{M}_{n\times n}$. 則 A 為 *orthogonal diagonalizable* 若且唯若存在 $n\times n$ 的 *orthogonal matrix* Q 使得 $Q^{\mathrm{T}}AQ$ 為 *diagonal matrix.*

利用 Proposition 9.4.2, 我們知當 A 為 orthogonal diagonalizable 時存在 $Q,D\in\mathscr{M}_{n\times n}$ 其中 Q 為 orthogonal matrix, D 為 diagonal matrix 使得 $A=QDQ^{\mathrm{T}}$. 此時 $A^{\mathrm{T}}=(QDQ^{\mathrm{T}})^{\mathrm{T}}=(Q^{\mathrm{T}})^{\mathrm{T}}D^{\mathrm{T}}Q^{\mathrm{T}}$. 由於 $(Q^{\mathrm{T}})^{\mathrm{T}}=Q$ 且 $D^{\mathrm{T}}=D$ (因為 D 為 diagonal matrix), 我們得 $A^{\mathrm{T}}=QDQ^{\mathrm{T}}=A$, 亦即 A 為 symmetric. 得證了以下結果.

Corollary 9.4.3. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *orthogonal diagonalizable,* 則 A 為 *symmetric matrix.*

所謂 Spectral Theorem 指的就是 Corollary 9.4.3 的反向也是對的. 也就是說我們要證明當 A 為 symmetric 時, A 必為 orthogonal diagonalizable. 首先我們需要知道 symmetric matrix 和內積之間的關係.

Lemma 9.4.4. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *symmetric,* 則對於任意 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ 皆有 $\langle A\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{v},A\mathbf{w}\rangle$.

Proof. 回顧一下, 若將內積寫成矩陣乘法的形式, 對於任意 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ 我們有 $\langle\mathbf{v},\mathbf{w}\rangle=\mathbf{v}^{\mathrm{T}}\mathbf{w}$ (注意此處 $\mathbf{v},\mathbf{w}$ 皆視為 $n\times 1$ matrix). 因此得

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

最後由 $A^{\mathrm{T}}=A$ 之假設得證 $\langle A\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{v},A\mathbf{w}\rangle$. □

一個 $n\times n$ matrix 是否為 diagonalizable 第一個要檢查的條件就是其 characteristic polynomial 須在實數中完全分解. 接下來我們便是要說明一個 symmetric matrix 其 characteristic polynomial 確實可以在實數中完全分解.

Lemma 9.4.5. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *symmetric,* 則 A 的 *characteristic polynomial* $p_A(t)$ 的根皆為實根.

Proof. 假設 $\lambda = a + b\imath$ (此處 $\imath$ 為虛數滿足 $\imath^2 = -1$) 為 $p_A(t)$ 的一個虛根, 即 $a, b \in \mathbb{R}$ 且 $b \neq 0$. 接下來我們要考慮複數矩陣, 極其 entry 為複數的矩陣. 要注意複述矩陣的運算以及行列式和實數矩陣有相同的規則. 所以依 $a + b\imath$ 為 $p_A(t)$ 的一根, 矩陣 $A - (a + b\imath)I_n$ 的行列式值為 0. 現將矩陣 $A - (a + b\imath)I_n$ 和矩陣 $A - (a - b\imath)I_n$ 相乘得

$$(A - (a + b\imath)I_n)(A - (a - b\imath)I_n) = A^2 - 2aA + (a^2 + b^2)I_n.$$

注意由於 $a, b \in \mathbb{R}$ 以及 A 為實數矩陣, 所以 $A^2 - 2aA + (a^2 + b^2)I_n$ 亦為實數矩陣. 另外由於 $\det(A - (a + b\imath)I_n) = 0$, 故有

$$\det(A^2 - 2aA + (a^2 + b^2)I_n) = \det(A - (a + b\imath)I_n)\det(A - (a - b\imath)I_n) = 0.$$

也就是說 $A^2 - 2aA + (a^2 + b^2)I_n$ 為 singular, 亦即存在 $\mathbf{v} \in \mathbb{R}^n$ 且 $\mathbf{v} \neq 0$ 使得

$$(A^2 - 2aA + (a^2 + b^2)I_n)\mathbf{v} = A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v} = \mathbf{O}.$$

然而

$$\langle A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v}, \mathbf{v}\rangle = \langle A^2\mathbf{v}, \mathbf{v}\rangle - 2a\langle A\mathbf{v}, \mathbf{v}\rangle + a^2\langle \mathbf{v}, \mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle.$$

又利用 A 為 symmetric, Lemma 9.4.4 告訴我們 $\langle A^2\mathbf{v}, \mathbf{v}\rangle = \langle A(A\mathbf{v}), \mathbf{v}\rangle = \langle A\mathbf{v}, A\mathbf{v}\rangle$, 故得

$$\langle A\mathbf{v} - a\mathbf{v}, A\mathbf{v} - a\mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle = \langle A^2\mathbf{v}, \mathbf{v}\rangle - 2a\langle A\mathbf{v}, \mathbf{v}\rangle + a^2\langle \mathbf{v}, \mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle,$$

亦即

$$\|A\mathbf{v} - a\mathbf{v}\|^2 + b^2\|\mathbf{v}\|^2 = \langle A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{O}, \mathbf{v}\rangle = 0.$$

因為 $\|A\mathbf{v} - a\mathbf{v}\| \geq 0$, $\|\mathbf{v}\| > 0$, 我們得 $b = 0$. 此與當初假設 $b \neq 0$ 相矛盾, 故知 $p_A(t) = 0$ 沒有虛根, 即所有的根都是實根. □

知道一個 symmetric matrix 的 characteristic polynomial 的根皆為實根, 我們便可以證明 symmetric matrix 皆為 orthogonal diagonalizable. 這裡我們要用數學歸納法, 也就是因為已證得 2×2 symmetric matrix 皆為 orthogonal diagonalizable. 現假設 $(n-1) \times (n-1)$ symmetric matrix 皆為 orthogonal diagonalizable. 我們要利用此證明當 A 為 $n \times n$ symmetric matrix 時亦為 orthogonal diagonalizable. 首先由 Lemma 9.4.5 知存在實數 λ 為 A 的一個 eigenvalue. 令 $\mathbf{u}_1$ 為 A 對於 λ 的 eigenvector 且 $\|\mathbf{u}_1\| = 1$. 利用 Gram-Schmidt process, 我們可以將 $\mathbf{u}_1$ 拓展成 $\mathbb{R}^n$ 的一組 orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. 現考慮 orthogonal matrix $Q = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix}$, 對於 $j = 1, \ldots, n$ 若 $A\mathbf{u}_j = c_{1\,j}\mathbf{u}_1 + \cdots + c_{n\,j}\mathbf{u}_n$, 則依舉陣乘法定應我們有 $AQ = QC$, 其中 $C = [c_{i\,j}]$. 因 Q 為 orthogonal matrix, 我們得 $C = Q^{-1}AQ = Q^{\mathrm{T}}AQ$. 因此再由 A 為 symmetric 得 $C^{\mathrm{T}} = Q^{\mathrm{T}}AQ = C$, 亦即 C 亦為 symmetric. 另一方面依假設 $A\mathbf{u}_1 = \lambda\mathbf{u}_1$, 我們知 C 的 1-st column 為 $\begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, 故由 C 為 symmetric 知 C 的 1-st row 為

$[\lambda \quad 0 \quad \cdots \quad 0]$. 也就是說 C 可以寫成以下的形式

$$C = \left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right].$$

由於 C 為 symmetric, 這裡 B 是 $(n-1)\times(n-1)$ symmetric matrix. 依歸納假設, 我們知 B 為 orthogonal diagonalizable, 亦即存在 $\mathbf{w}_1,\ldots,\mathbf{w}_{n-1}$ 為 $\mathbb{R}^{n-1}$ 的一組 orthonomal basis 且為 B 的 eigenvectors. 此時令 $R = \begin{bmatrix} | & | & & | \\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_{n-1} \\ | & | & & | \end{bmatrix}$, 我們得 R 為 $(n-1)\times(n-1)$ orthogonal matrix 且存在 $(n-1)\times(n-1)$ digonal matrix D 滿足 $R^{\mathrm{T}}BR = D$. 現在令 $P = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & R & \\ 0 & & & \end{array}\right]$. 依矩陣乘法, 我們有

$$P^{\mathrm{T}}CP = \left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & R^{\mathrm{T}}BR & \\ 0 & & & \end{array}\right] = \left[\begin{array}{c|ccc} \lambda & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & D & \\ 0 & & & \end{array}\right].$$

也就是說 $P^{\mathrm{T}}CP$ 為 diagonal matrix, 也因此得 $(QP)^{\mathrm{T}}A(QP) = P^{\mathrm{T}}(Q^{\mathrm{T}}AQ)P = P^{\mathrm{T}}CP$ 為 diagonal matrix. 注意由於 Q,P 皆為 orthogonal matrix, $(QP)^{\mathrm{T}}(QP) = P^{\mathrm{T}}(Q^{\mathrm{T}}Q)P = P^{\mathrm{T}}P = I_n$, 也就是說 QP 亦為 orthogonal matrix. 因此由 Proposition 9.4.2, 得 A 為 orthogonal diagonalizable, 也因此證明了 Spectral Theorem.

Theorem 9.4.6 (Spectral Theorem)**.** 假設 A 為 $n\times n$ *symmetric matrix,* 則 A 為 *orthogonal diagonalizable.*

接下來我們來探討, 給定一個 $n\times n$ symmetric matrix A, 如何找到 orthogonal matrix Q 使得 $Q^{\mathrm{T}}AQ$ 為 diagonal matrix. 當然了, 我們可以如 Theorem 9.4.6 的證明, 利用數學歸納法一步一步地將 Q 找到. 不過這要重複做好幾次的 Gram-Schmidt process, 頗為複雜. 利用以下的 Proposition, 我們可以將步驟簡化許多.

Proposition 9.4.7. 假設 A 為 $n\times n$ *symmetric matrix.* 若 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ 為 A 的 *eigenvectors* 且其對應的 *eigenvalue* 為相異實數, 則 $\langle\mathbf{v},\mathbf{w}\rangle = 0$.

Proof. 假設 $\mathbf{v},\mathbf{w}$ 所對應的 eigenvalue 分別為 λ,λ'. 也就是說 $A\mathbf{v} = \lambda\mathbf{v}, A\mathbf{w} = \lambda'\mathbf{w}$. 考慮 $\langle A\mathbf{v},\mathbf{w}\rangle = \langle\lambda\mathbf{v},\mathbf{w}\rangle = \lambda\langle\mathbf{v},\mathbf{w}\rangle$. 同理我們有 $\langle\mathbf{v},A\mathbf{w}\rangle = \lambda'\langle\mathbf{v},\mathbf{w}\rangle$. 然而 Lemma 9.4.4 告訴我們 $\langle A\mathbf{v},\mathbf{w}\rangle = \langle\mathbf{v},A\mathbf{w}\rangle$, 故得 $(\lambda-\lambda')\langle\mathbf{v},\mathbf{w}\rangle = 0$. 因此由題設 $\lambda\neq\lambda'$ 推得 $\langle\mathbf{v},\mathbf{w}\rangle = 0$. □

當 A 為 $n\times n$ symmetric matrix, 我們簡單說明一下如何找到一組 A 的 eigenvectors 形成 $\mathbb{R}^n$ 的 orthonormal basis. 首先我們列出 A 的所有相異的 eigenvalues $\lambda_1,\ldots,\lambda_k$, 然後求出它們所對應的 eigenspace $E_A(\lambda_1),\ldots,E_A(\lambda_k)$. 若我們將每個 $E_A(\lambda_i)$ 的 basis 放在一起, 由於

A 為 diagonalizable 它們會是 $\mathbb{R}^n$ 的一組 basis. 雖然 Proposition 9.4.7 告訴我們, 當 $\lambda_i \neq \lambda_j$ 時, $E_A(\lambda_i)$ 和 $E_A(\lambda_j)$ 之間的向量是相互垂直的. 不過 $E_A(\lambda_i)$ 本身的那一組 basis 之間的向量未必倆倆相互垂直. 所以我們必須利用 Gram-Schmidt process 分別找到 A 每個 eigenspace $E_A(\lambda_i)$ 的一組 orthonormal basis. 再將這些 eigenspace 的 basis 放在一起它們自然兩兩互相垂直, 也因此它們就是由 A 的 eigenvectors 所組成的 $\mathbb{R}^n$ 的一組 orthonormal basis. 我們看以下的例子.

Example 9.4.8. (1) 考慮 symmetric matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. 我們有 A 的 characteristic polynomial 為 $p_A(t) = -(t+1)(t-1)(t-2)$. 所以 A 有三個相異的 eigenvalues, $-1,1,2$. 知道 A 必能對角化, 而且由 Proposition 9.4.7 知它們所對應的 eigenvector 會倆倆互相垂直. 事實上我們可求出對應到 $-1,1,2$ 的 eigenvector 分別為

$$\mathbf{v}_1 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

很容易檢查它們確實倆倆互相垂直. 此時對於 $i = 1,2,3$ 令 $\mathbf{u}_i = \frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$, 我們得

$$\mathbf{u}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

為 $\mathbb{R}^3$ 的一組 orthonromal basis. 故可將 A 對角化成

$$\begin{bmatrix} -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(2) 考慮 symmetric matrix $B = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}$. 我們有 B 的 characteristic polynomial 為 $p_B(t) = -t(t-9)^2$. 所以 B eigenvalues 為 $0,9$. 知道 B 必能對角化, 我們知 $\dim(E_B(0)) = 1$, $\dim(E_B(9)) = 9$. 事實上 $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ 為 $E_B(0) = N(B)$ 的 basis, 而 $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$ 為 $E_B(9)$ 的 basis. 而且由 Proposition 9.4.7 知 $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = 0$, 事實上很容易檢查它們確實成立. 不過 $\langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 1 \neq 0$, 我們必須利用 Gram-Schmidt process 將 $\mathbf{v}_2, \mathbf{v}_3$ 換成 $E_B(9)$ 的一組 orthogonal basis. 令 $\mathbf{w}_2 = \mathbf{v}_2$ 且

$$\mathbf{w}_3 = \mathbf{v}_3 - \mathrm{Proj}_{\mathbf{w}_2}\mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}.$$

此時令 $\mathbf{u}_1=\frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1$, $\mathbf{u}_2=\frac{1}{\|\mathbf{w}_2\|}\mathbf{w}_2$, $\mathbf{u}_3=\frac{1}{\|\mathbf{w}_3\|}\mathbf{w}_3$ 我們得

$$\mathbf{u}_1=\frac{1}{\sqrt{3}}\begin{bmatrix}2\\2\\1\end{bmatrix},\quad \mathbf{u}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix},\quad \mathbf{u}_3=\frac{1}{3\sqrt{2}}\begin{bmatrix}-1\\-1\\4\end{bmatrix}$$

為 $\mathbb{R}^3$ 的一組 orthonromal basis. 故可將 B 對角化成

$$\begin{bmatrix}\frac{2}{3}&\frac{2}{3}&\frac{1}{3}\\-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}&0\\-\frac{1}{3\sqrt{2}}&-\frac{1}{3\sqrt{2}}&\frac{4}{3\sqrt{2}}\end{bmatrix}\begin{bmatrix}5&-4&-2\\-4&5&-2\\-2&-2&8\end{bmatrix}\begin{bmatrix}\frac{2}{3}&-\frac{1}{\sqrt{2}}&-\frac{1}{3\sqrt{2}}\\\frac{2}{3}&\frac{1}{\sqrt{2}}&-\frac{1}{3\sqrt{2}}\\\frac{1}{3}&0&\frac{4}{3\sqrt{2}}\end{bmatrix}=\begin{bmatrix}0&0&0\\0&9&0\\0&0&9\end{bmatrix}.$$

9.5. Application: Conics and Quadric Surfaces

我們將利用 symmetric matrix 是 orthogonal diagonalizable 的特性將坐標平面上的二次曲線以及坐標空間上的二次曲面的方程式化成標準式, 以方便我們判別它們是哪一類的圖形.

一般來說我們是利用平移和旋轉的方法將二次曲線和二次曲面的方程式化成標準式. 其中旋轉的部分牽涉到對角化, 我們首先利用 quadratic form 來談對角化的問題. 所謂 n 個變數的 quadratic form 指的就是形如

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j$$

這樣的二次式. 例如 $x^2+3xy-y^2$, $3x^2+y^2-z^2+5xy+xz+3yz$ 就是分別是兩個變數和三個變數的 quadratic form. 令 $\mathbf{x}=\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}$, 所有 n 個變數的 quadratic form 都可以用矩陣表示成 $\mathbf{x}^{\mathrm{T}}A\mathbf{x}$ 的形式, 其中 A 為 $n\times n$ symmetric matrix. 例如兩個變數的 quadratic form $ax_1^2+bx_1x_2+cx_2^2$ 就可以寫成

$$ax_1^2+bx_1x_2+cx_2^2=\begin{bmatrix}x_1&x_2\end{bmatrix}\begin{bmatrix}a&b/2\\b/2&c\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}.$$

而三個變數的 quadratic form $ax_1^2+bx_2^2+cx_3^2+rx_1x_2+sx_1x_3+tx_2x_3$ 就可以寫成

$$ax_1^2+bx_2^2+cx_3^2+rx_1x_2+sx_1x_3+tx_2x_3=\begin{bmatrix}x_1&x_2&x_3\end{bmatrix}\begin{bmatrix}a&r/2&s/2\\r/2&b&t/2\\s/2&t/2&c\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}.$$

將 quadratic form 寫成這樣的矩陣表示的好處是因為 A 是 symmetric, 故存在 orthogonal matrix Q 使得 $Q^{\mathrm{T}}AQ$ 為 diagonal matrix $\begin{bmatrix}\lambda_1&&\\&\ddots&\\&&\lambda_n\end{bmatrix}$. 因此如果我們將變數 $\mathbf{x}=\begin{bmatrix}x_1\\\vdots\\x_n\end{bmatrix}$ 變換成 $\mathbf{t}=\begin{bmatrix}t_1\\\vdots\\t_n\end{bmatrix}$ 其中 $\mathbf{t}=Q^{\mathrm{T}}\mathbf{x}$ (注意因 $Q^{\mathrm{T}}=Q^{-1}$, 這等同於令 $\mathbf{x}=Q\mathbf{t}$), 則

$$\mathbf{x}^{\mathrm{T}}A\mathbf{x}=(Q\mathbf{t})^{\mathrm{T}}A(Q\mathbf{t})=\mathbf{t}^{\mathrm{T}}(Q^{\mathrm{T}}AQ)\mathbf{t}=\begin{bmatrix}t_1&\cdots&t_n\end{bmatrix}\begin{bmatrix}\lambda_1&&\\&\ddots&\\&&\lambda_n\end{bmatrix}\begin{bmatrix}t_1\\\vdots\\t_n\end{bmatrix}=\lambda_1t_1^2+\cdots+\lambda_nt_n^2.$$

也就是說, 我們可以藉由變換變數將一個 quadratic form 變成只有平方項. 我們看以下的例子.

Example 9.5.1. 考慮 quadratic form $x_1^2+4x_1x_2-2x_2^2$. 我們先寫下其矩陣形式

$$x_1^2+4x_1x_2-2x_2^2=\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

由於 $\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ 為 symmetric matrix, 故為 orthogonal diagonalizable, 事實上我們有

$$\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}=\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}.$$

因此若令 $\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}=\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 則

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} t_1 & t_2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}=2t_1^2-3t_2^2.$$

對於 quadratic form $x_2^2+x_3^2+2x_1x_2+2x_1x_3$, 其矩陣形式為

$$x_2^2+x_3^2+2x_1x_2+2x_1x_3=\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

我們曾在 Example 9.4.8 計算過 $Q^{\mathrm{T}}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}Q=\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ 其中 Q 為 orthogonal matrix $\begin{bmatrix} -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}$. 因此若令 $\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}=\begin{bmatrix} -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ 則

$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}=-t_1^2+t_2^2+2t_3^2.$$

現在我們回到二次曲線的情況. 對於坐標平面上的二次曲線其一般的通式為 $ax^2+bxy+cy^2+dx+ey+f=0$. 我們可以將此式表為矩陣形式, 即

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} d & e \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+f=0. \tag{9.5}$$

假設 symmetric matrix $A=\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$ 可對角化成 $Q^{\mathrm{T}}AQ=\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. 此時考慮變換變數 $\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}=Q^{\mathrm{T}}\begin{bmatrix} x \\ y \end{bmatrix}$ (也就是 $\begin{bmatrix} x \\ y \end{bmatrix}=Q\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$), 則式子 (9.5) 可寫成

$$\begin{bmatrix} \overline{x} & \overline{y} \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}+\begin{bmatrix} d & e \end{bmatrix}Q\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}+f=0.$$

寫回方程式的樣子就是

$$\lambda_1\overline{x}^2+\lambda_2\overline{y}^2+d'\overline{x}+e'\overline{y}+f=0, \tag{9.6}$$

其中 $\begin{bmatrix} d' & e' \end{bmatrix}=\begin{bmatrix} d & e \end{bmatrix}Q$.

首先我們考慮 λ_1, λ_2 皆不為 0 的情形, 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+\lambda_2(\bar{y}-k)^2=f'.$$

我們分成下面幾種情形討論.

(A) λ_1, λ_2 同號:

(1) f' 與 λ_1, λ_2 同號: 此時圖形為 *ellipse* (橢圓). 注意當 $\lambda_1=\lambda_2$ 時會是圓, 不過這裡我們將之視為橢圓的一種.

(2) $f'=0$: 此時很容易看出圖形為 $(\bar{x},\bar{y})=(h,k)$ 這一點.

(3) f' 與 λ_1, λ_2 異號: 此時很容易看出圖形為空集合.

(B) λ_1, λ_2 異號:

(1) $f'\neq 0$: 此時圖形為 *hyperbola* (雙曲線).

(2) $f'=0$: 此時圖形為兩相交直線.

Example 9.5.2. 考慮二次曲線方程式 $2xy+\sqrt{2}x+\sqrt{2}y=1$. 此方程式可用矩陣表示成

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=1.$$

由於

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

考慮變數變換 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$, 我們得

$$\begin{bmatrix} \bar{x} & \bar{y} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}+\begin{bmatrix} \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}=1.$$

因此此曲線用新的變數其方程式為 $\bar{x}^2-\bar{y}^2+2\bar{x}=1$. 利用配方法得 $(\bar{x}+1)^2-\bar{y}^2=2$, 故其圖形為雙曲線.

同理若原方程式為 $2xy+\sqrt{2}x+\sqrt{2}y=-1$, 變換變數後的方程式為 $(\bar{x}+1)^2-\bar{y}^2=0$ 其圖形便會是兩相交直線 $\bar{x}+\bar{y}+1=0$ 和 $\bar{x}-\bar{y}+1=0$.

另一種情況是 λ_1, λ_2 其中有一個為 0. 注意 λ_1, λ_2 不可能同時為 0, 否則會是一次方程式. 不失一般性, 我們假設 $\lambda_1\neq 0, \lambda_2=0$ 的情形. 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+e'\bar{y}=f'.$$

我們分成下面幾種情形討論.

(C) λ_1, λ_2 其中有一個為 0 (不失一般性假設 $\lambda_1\neq 0, \lambda_2=0$):

(1) $e'\neq 0$: 此時圖形為 *parabola* (拋物線).

(2) $e'=0$ 且 λ_1, f' 同號: 此時圖形為兩平行直線 (與直線 $\bar{x}=0$ 平行).

(3) $e'=0$ 且 $f'=0$: 此時圖形為一直線 $\bar{x}=h$.

(4) $e'=0$ 且 λ_1, f' 異號: 此時圖形為空集合.

Example 9.5.3. 考慮二次曲線方程式 $x^2-2xy+y^2+4\sqrt{2}x=4$. 此方程式可用矩陣表示成

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} 4\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=4.$$

由於

$$\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}=\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

考慮變數變換 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$, 我們得

$$\begin{bmatrix} \bar{x} & \bar{y} \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}+\begin{bmatrix} 4\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}=4.$$

因此此曲線用新的變數其方程式為 $2\bar{x}^2+4\bar{x}+4\bar{y}=4$. 利用配方法得 $2(\bar{x}+1)^2+4\bar{y}=6$, 故其圖形為拋物線.

總而言之, 我們可以從二次曲線的 quadratic form 部分得到其 eigenvalue λ_1,λ_2, 然後由 λ_1,λ_2 的正負號判斷此二次曲線應歸類於哪一種曲線. 若 λ_1,λ_2 同號, 則為橢圓類; 而 λ_1,λ_2 異號, 則為雙曲線類; 而若 λ_1,λ_2 有一個為 0, 則為拋物線類. 不過最後我們還是得經由配方法求得其一次項與常數項, 這樣才能確認此曲線是否為 degenerated (退化) 的情形 (即直線, 點或空集合).

Question 9.8. *假設二次曲線方程式的 quadratic form 的部分可表成* $\begin{bmatrix} x & y \end{bmatrix}A\begin{bmatrix} x \\ y \end{bmatrix}$*, 中* A *為* 2×2 *symmetric matrix. 試問是否可由* $\det(A)$ *來判斷此曲線是橢圓類, 雙曲線類還是拋物線類 (不考慮退化情形)?*

對於坐標空間的二次曲面我們也是用同樣方法處理. 首先寫成矩陣的形式

$$\begin{bmatrix} x & y & z \end{bmatrix}A\begin{bmatrix} x \\ y \\ z \end{bmatrix}+\begin{bmatrix} c & d & e \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}+f=0,$$

其中 A 為 3×3 symmetric matrix. 再將 A 對角化然後變換變數成

$$\lambda_1\bar{x}^2+\lambda_2\bar{y}^2+\lambda_3\bar{z}^2+c'\bar{x}+d'\bar{y}+d'\bar{z}+f=0. \tag{9.7}$$

二次曲面的分類頗為複雜, 大家不必記下這些分類. 不過為了完整性, 我們還是列出這些分類供同學參考. 由於此處無法利用圖形來解釋, 建議有興趣的同學參考課本上的圖形.

首先我們考慮 $\lambda_1,\lambda_2,\lambda_3$ 皆不為 0 的情形, 此時可以利用配方法將式子 (9.7) 改寫成

$$\lambda_1(\bar{x}-h)^2+\lambda_2(\bar{y}-k)^2+\lambda_3(\bar{z}-l)^2=f'.$$

我們分成下面幾種情形討論.

(A) $\lambda_1,\lambda_2,\lambda_3$ 同號:

(1) f' 與 $\lambda_1,\lambda_2,\lambda_3$ 同號: 此時曲面為有界的, 且與 $\bar{x}=h$, $\bar{y}=k$ 和 $\bar{z}=l$ 三個平面所交的圖形為橢圓. 曲面有點像橄欖球表面一樣, 我們稱之為 *ellipsoid*. 注意當 $\lambda_1=\lambda_2=\lambda_3$ 時會是球面, 不過這裡我們將之視為 ellipsoid 的一種.

(2) $f'=0$: 此時很容易看出圖形為 $(\bar{x},\bar{y},\bar{z})=(h,k,l)$ 這一點.

(3) f' 與 $\lambda_1,\lambda_2,\lambda_3$ 異號: 此時很容易看出圖形為空集合.

(B) $\lambda_1,\lambda_2,\lambda_3$ 異號 (不失一般性假設 λ_1,λ_2 同號):

(1) f' 與 λ_1,λ_2 同號: 此時曲面與 $\bar{z}=l$ 所交的圖形為橢圓, 而分別和 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為雙曲線. 因為曲面整體上只有一片, 我們稱之為 *hyperboloid* of one sheet.

(2) f' 與 λ_1,λ_2 異號: 此時曲面與平面 $\bar{z}=l$ 不相交, 不過若將平面往上或往下移動夠多的話會交出橢圓. 此曲面分別和 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為雙曲線. 因為曲面整體上有兩片, 我們稱之為 *hyperboloid* of two sheets.

(3) $f'=0$: 此時曲面與平面 $\bar{z}=l$ 交於一點, 不過若將平面往上或往下移的話會交出橢圓. 此區面分別和 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為兩相交直線. 圖形有點像甜筒, 我們稱之為 *elliptic cone*.

另一種情況是 $\lambda_1,\lambda_2,\lambda_3$ 其中有一個為 0. 注意 $\lambda_1,\lambda_2,\lambda_3$ 不可能皆為 0, 否則會是一次方程式. 不失一般性, 我們假設 $\lambda_1\neq 0$. 我們又可分成下面幾種情形討論.

(C) λ_2,λ_3 僅有一個為 0 (不失一般性假設 $\lambda_2\neq 0$): 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+\lambda_2(\bar{y}-k)^2+e'\bar{z}=f'.$$

(1) $e'\neq 0$ 且 λ_1,λ_2 同號: 此曲面會完全在平面 $e'\bar{z}=f'$ 之上方或下方, 不過若將平面往上或往下移動會交出橢圓. 而此曲面分別與 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為凹向一致的拋物線. 我們稱之為 *elliptic paraboloid*.

(2) $e'\neq 0$ 且 λ_1,λ_2 異號: 此曲面與平面 $e'\bar{z}=f'$ 交於兩相交直線, 不過若將平面往上或往下移動會交出雙曲線. 此曲面分別與 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為凹向相反的拋物線. 我們稱之為 *elliptic paraboloid*. 此曲面上的一點 $(\bar{x},\bar{y},\bar{z})=(h,k,f'/e')$ 就是所謂的 saddle point (鞍點).

(3) $e'=0$ 且 λ_1,λ_2,f' 同號: 此時曲面與任何的水平平面 $\bar{z}=s$ 所交的圖形為橢圓. 圖形像橢圓柱面, 稱為 *elliptic cylinder*.

(4) $e'=0$ 且 λ_1,λ_2 異號又 $f'\neq 0$: 此時曲面與任何的水平平面 $\bar{z}=s$ 所交的圖形為雙曲線. 圖形像雙曲柱面, 稱為 *hyperbolic cylinder*.

(5) $e'=0$ 且 λ_1,λ_2 同號但與 f' 異號: 此時是空集合.

(6) $e'=0$ 且 λ_1,λ_2 同號又 $f'=0$: 此時曲面與任何的水平平面 $\bar{z}=s$ 僅交於一點. 圖形為一鉛直線.

(7) $e'=0$ 且 λ_1,λ_2 異號又 $f'=0$: 此時曲面與任何的水平平面 $\bar{z}=s$ 交於兩相交直線. 圖形為兩相交平面.

(D) λ_2,λ_3 皆為 0: 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+d'\bar{y}+e'\bar{z}=f'.$$

(1) d', e' 不全為 0: 此時令 $r = \sqrt{(d')^2 + (e')^2}$ 利用變換變數

$$\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & d'/r & -e'/r \\ 0 & e'/r & d'/r \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

我們又可將上式改寫成

$$\lambda_1(t_1 - h)^2 + rt_2 = f'.$$

此曲面與任何的水平平面 $t_3 = s$ 所交的圖形為拋物線. 圖形像拋物柱面, 稱為 *parabolic cylinder.*

(2) $d' = e' = 0$ 且 f' 與 λ_1 同號: 此時圖形為兩平行平面 (與 $\bar{x} = 0$ 平行).

(3) $d' = e' = 0$ 且 $f' = 0$: 此時圖形為平面 $\bar{x} = h$.

(4) $d' = e' = 0$ 且 f' 與 λ_1 異號: 此時為空集合.

Example 9.5.4. 考慮坐標空間中曲面 $5x^2 + 5x^2 + 8z^2 - 8xy - 4xz - 4yz + 2x + 2y + z = 9$. 寫成矩陣形式為

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 9.$$

由於

$$\begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{-1}{3\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{4}{3\sqrt{2}} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3} \\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(請參考 Example 9.4.8 (2)). 考慮變數變換 $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3} \\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix}$, 我們得

$$\begin{bmatrix} \bar{x} & \bar{y} & \bar{z} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} + \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3} \\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3} \\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix} = 9.$$

因此此曲面用新的變數其方程式為 $9\bar{x}^2 + 9\bar{y}^2 + 3\bar{z} = 9$, 為前面列出的 (C)(1) 這個情形, 故知其為 elliptic paraboloid.

Question 9.9. 空間中曲面 $5x^2 + 5x^2 + 8z^2 - 8xy - 4xz - 4yz + 2x + 2y + z = 0$ 會是怎樣的圖形?

9.6. Application: Markov Processes

當 $A \in \mathscr{M}_{n\times n}$ 為 diagonalizable 時, 對於 $k \in \mathbb{N}$ 我們可以利用對角舉陣很容易求出 A^k. 進而對於任意 $\mathbf{v} \in \mathbb{R}^n$, 推算出 $A^k\mathbf{v}$. 其實還有一種情況 (即使不是 diagonalizable), 當 k 很大時我們也能 "估計" $A^k\mathbf{v}$ 大約為何. 這就是本節要探討的課題.

首先我們看 $A \in \mathscr{M}_{n\times n}$ 為 diagonalizable 的情形. 此時由於存在 diagonal matrix $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ 以及 invertible matrix $P \in \mathscr{M}_{n\times n}$ 使得 $D = P^{-1}AP$, 因此 $A = PDP^{-1}$. 依此我們可以推得

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1} = P\begin{bmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{bmatrix}P^{-1},$$

然後用數學歸納法推得

$$A^k = P\begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix}P^{-1}.$$

Example 9.6.1. 我們利用 Fibonacci sequence $0,1,1,2,3,5,8,13,\ldots$, 來說明如何利用對角化. Fibonacci sequence 是一組滿足 $F_{k+1} = F_k + F_{k-1}$ 的遞迴數列, 其中 $F_0 = 0, F_1 = 1$. 我們令 $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ 且對任意 $k \geq 1$ 令 $\mathbf{v}_k = \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}$, 則

$$A\mathbf{v}_k = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix} = \begin{bmatrix} F_k + F_{k-1} \\ F_k \end{bmatrix} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \mathbf{v}_{k+1}.$$

因此我們有 $\mathbf{v}_{k+1} = A^k\mathbf{v}_1$. 也就是說對於任意 $k \geq 1$, 我們只要能算出 $A^k\mathbf{v}_1$, 就能求出 F_{k+1} 為何. 然而 A 的 characteristic polynomial 為 $P_A(t) = t^2 - t - 1$, 故得 A 的 eigenvalues 為 $\lambda_1 = (1-\sqrt{5})/2$, $\lambda_2 = (1+\sqrt{5})/2$. 因 A 是 2×2 matrix, 所以由 A 兩個相異的 eigenvalues 得 A 為 diagonalizable. 事實上 A 對於 λ_1, λ_2 的 eigenvector 分別為 $\mathbf{v}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$, $\lambda_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$. 因此若令 $P = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}$, 我們有

$$P = \begin{bmatrix} \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}, \quad P^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 & \frac{1+\sqrt{5}}{2} \\ 1 & \frac{-1+\sqrt{5}}{2} \end{bmatrix}.$$

因此將 A 對角化得 $A = P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}P^{-1}$, 也因此求出對任意 $k \in \mathbb{N}$,

$$A^k = P\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}P^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}\begin{bmatrix} -1 & \lambda_2 \\ 1 & -\lambda_1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} \lambda_2^{k+1} - \lambda_1^{k+1} & \lambda_2^k - \lambda_1^k \\ \lambda_2^k - \lambda_1^k & \lambda_2^{k-1} - \lambda_1^{k-1} \end{bmatrix}.$$

所以由 $\mathbf{v}_1 = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 我們得

$$\mathbf{v}_{k+1} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = A^k\mathbf{v}_1 = A^k\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} \lambda_2^{k+1} - \lambda_1^{k+1} \\ \lambda_2^k - \lambda_1^k \end{bmatrix},$$

故得

$$F_{k+1} = \frac{1}{5}(\lambda_2^{k+1} - \lambda_1^{k+1}) = \frac{1}{5}\left((\frac{1+\sqrt{5}}{2})^{k+1} - (\frac{1-\sqrt{5}}{2})^{k+1}\right).$$

接下來我們要探討的是, 有時即使 A 不是 diagonalizable, 但我們仍能估計 $A^k\mathbf{v}$. 這裡要探討的情況是所謂 *Markov Peocesses*, 是機率統計上的課題. 由於我們僅專注於線性代數的部分, 在這裡就不多談它的由來和例子, 直接切入主題.

Definition 9.6.2. 對於一 $\mathbb{R}^n$ 上的 vector $\mathbf{v}=\begin{bmatrix} c_1\\ \vdots\\ c_n\end{bmatrix}$, 若 $c_1+\cdots+c_n=1$ 且對於所有 $i=1,\dots,n$, 皆有 $c_i\geq 0$, 則稱 $\mathbf{v}$ 為 *probability vector*. 若 $A\in\mathscr{M}_{n\times n}$ 且其每一個 column vector 皆為 probability vector, 則稱 A 為 *stochastic matrix*. 另外一個 stochastic matrix A 若存在 $r\in\mathbb{N}$ 使得 A^r 的每個 entry 皆為正實數, 則稱 A 為 *regular*.

Example 9.6.3. $A=\begin{bmatrix} 1/2 & 1\\ 1/2 & 0\end{bmatrix}$ 和 $I_2=\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix}$ 皆為 stochastic matrix. 而且 A 為 regular, 因為 $A^2=\begin{bmatrix} 3/4 & 1/2\\ 1/4 & 1/2\end{bmatrix}$, 每個 entry 皆為正. 然而 I_2 不是 regular, 因為對於任意 $r\in\mathbb{N}$ 皆有 $I_2^n=I_2$ (除了對角線, 其他位置的 entry 皆為 0).

接下來我們看幾個有關 stochastic matrix 的性質.

Lemma 9.6.4. *假設 $A\in\mathscr{M}_{n\times n}$ 為 stochastic matrix 且 $\mathbf{v}\in\mathbb{R}^n$ 為 probability vector. 則 $A\mathbf{v}$ 亦為 probability vector. 另外若 A 的每一個 entry 皆為正實數, 則 $A\mathbf{v}$ 的每個 entry 亦皆為正實數.*

Proof. 令 $A=\begin{bmatrix} | & & |\\ \mathbf{a}_1 & \cdots & \mathbf{a}_n\\ | & & |\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix} c_1\\ \vdots\\ c_n\end{bmatrix}$, 則 $A\mathbf{v}=c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$. 因此 $A\mathbf{v}$ 所有 entries 之和就是 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$ 所有 entries 之和. 這等同於個別算出每個 $c_i\mathbf{a}_i$ 的所有 entries 之和再全部加起來. 然而因 $\mathbf{a}_i$ 為 probability vector, $c_i\mathbf{a}_i$ 的所有 entries 之和為 c_i, 所以 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$ 所有 entries 之和為 $c_1+\cdots+c_n=1$. 又因為 $c_1,\dots,c_n$ 以及 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 中的每個 entry 皆為非負實數, 所以 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$ 的每個 entry 皆為非負實數. 得證 $A\mathbf{v}$ 為 probability vector.

另外若 A 的每一個 entry 皆為正實數, 即 $\mathbf{a}_1,\dots,\mathbf{a}_n$ 的每一個 entry 皆為正實數, 此時由於 $c_1,\dots,c_n$ 為非負實數, 故有 $A\mathbf{v}=c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$ 的每個 entry 皆大於等於 $c_i\mathbf{a}_i$ 所相對應的 entry. 因 $c_1,\dots,c_n$ 不全為 0, 故若 $c_i>0$, 則 $c_i\mathbf{a}_i$ 的每個 entry 皆為正實數, 因此得證 $A\mathbf{v}$ 的每個 entry 亦皆為正實數. $\square$

現若 $A=\begin{bmatrix} | & & |\\ \mathbf{a}_1 & \cdots & \mathbf{a}_n\\ | & & |\end{bmatrix}$ 為 stochastic matrix, 則依矩陣乘法定義 A^2 的 i-th column 為 $A\mathbf{a}_i$, 故由 Lemma 9.6.4 知, A^2 的每個 column 皆為 probability vector, 亦即 A^2 亦為 stochastic matrix. 同理對任意 $k\geq 2$, A^k 的 i-th column 為 $A^{k-1}\mathbf{a}_i$, 因此利用數學歸納法以及 Lemma 9.6.4, 我們得證 A^k 亦為 stochastic matrix. 同樣的利用數學歸納法以及 Lemma 9.6.4, 我們可以證明若 A^r 的每一個 entry 皆為正實數, 則對於所有 $k\in\mathbb{N}$, $A^{r+k}=A^{r+k-1}A$ 的每個 entry 亦皆為正實數. 因此有以下的定理 (證明從略).

Proposition 9.6.5. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix*, 則對所有 $k \in \mathbb{N}$, A^k 亦為 *stochastic matrix*. 又若 A 為 *regular* 且 A^r 的每個 *entry* 皆為正實數, 則對所有 $k \in \mathbb{N}$, A^{r+k} 的每個 *entry* 亦皆為正實數.

接下來我們要談論 stochastic matrix 的 eigenvalues 以及 eigenvectors. 不像前面的情況, 由於我門探討的是一般的 stochastic matrix 而不是具體的矩陣, 所以我們無法從它的 characteristic polynomial 來處理. 這裡我們需要特定的技巧, 首先我們從轉置矩陣出發.

Lemma 9.6.6. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix*. 則 1 為 A^{T} 的一個 *eigenvalue* 且 $\mathbf{v} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ 為其 *eigenvector*. 另外若 A 的每個 *entry* 皆為正實數, 則對於 A^{T}, 其 *eigenvalue* 1 的 *geometric multiplicity* 為 1.

Proof. 由於 A 為 stochastic matrix, A 每一個 column vector $\mathbf{a}_i$ 皆為 probability vector, 亦即 $\langle \mathbf{a}_i, \mathbf{v} \rangle = 1$. 因此我們有 $A^{\mathrm{T}}\mathbf{v} = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{v} \rangle \\ \vdots \\ \langle \mathbf{a}_n, \mathbf{v} \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{v}$. 得證 $\mathbf{v}$ 為 A^{T} 的 eigenvector 且其 eigenvalue 為 1.

現假設 A 的每個 entry 皆為正實數且 $\mathbf{w} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 為 eigenvalue 為 1 的 eigenvector. 注意 $\mathbf{w} \neq \mathbf{O}$, 因此不失一般性, 我們可假設 $c_1, \ldots, c_n$ 的最大值不為 0 (因為若最大值為 0, 表示每個 $c_i \leq 0$, 故此時考慮 $-\mathbf{w}$, 其仍為 A^{T} 的一個 eigenvalue 為 1 的 eigenvector 且此時 $-\mathbf{w}$ 每個 entry 的最大值為正實數). 假設 c_j 為 $c_1, \ldots, c_n$ 的最大值. 考慮 $A\mathbf{w}$ 的 j-th entry, 依定義其值為 $\langle \mathbf{a}_j, \mathbf{w} \rangle = a_{j1}c_1 + \cdots + a_{ji}c_i + \cdots + a_{jn}c_n$. 因為 $a_{j1}, \ldots, a_{jn}$ 皆為正實數且 $c_j > 0$ 為 $c_1, \ldots, c_n$ 的最大值, 我們有

$$a_{j1}c_1 + \cdots + a_{ji}c_i + \cdots + a_{jn}c_n \leq a_{j1}c_j + \cdots + a_{ji}c_j + \cdots + a_{jn}c_j = (a_{j1} + \cdots + a_{jn})c_j = c_j. \quad (9.8)$$

由於依假設 $A^{\mathrm{T}}\mathbf{w} = \mathbf{w}$, 所以 A^{T} 的 j-th entry 應為 c_j, 也就是說式子 (9.8) 中的小於等於的符號應為等號, 也因此證得了 $c_1 = \cdots = c_j = \cdots = c_n = r$. 這說明了 $\mathbf{w} = r\mathbf{v}$, 亦即所有 A^{T} 的 eigenvalue 為 1 的 eigenvector 皆在 $\mathrm{Span}(\mathbf{v})$ 中. 因此得證 A^{T} 其 eigenvalue 1 的 geometric multiplicity 為 1 $\square$

回顧 Proposition 9.1.6 和 Proposition 9.2.3 告訴我們 A 和 A^{T} 有相同的 eigenvalues 而且每個 eigenvalue 對於 A 和 A^{T} 的 geometric multiplicity 相同. 因此我們有以下的結果.

Proposition 9.6.7. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix*. 則 1 為 A 的一個 *eigenvalue*. 另外若 A 為 *regular*, 則對於 A, 其 *eigenvalue* 1 的 *geometric multiplicity* 為 1.

Proof. 因 A 為 stochastic matrix, 由 Lemma 9.6.6 知 1 為 A^{T} 的一個 eigenvalue. 故由 Proposition 9.1.6 知 1 亦為 A 的一個 eigenvalue. 另外, 若 A 為 regular 且假設 $r \in \mathbb{N}$ 使得 A^r 的每個 entry 皆為正實數, 則由 Lemma 9.6.6 知 $(A^r)^{\mathrm{T}}$ 的 eigenvalue 1 其 geometric multiplicity 為 1. 也因此由 Proposition 9.2.3 知 A^r 的 eigenvalue 1, 其 geometric multiplicity 亦為 1, 也就是說 $\dim(E_{A^r}(1))=1$. 現對於任意 $\mathbf{v}\in E_A(1)$, 由於 $A\mathbf{v}=\mathbf{v}$, 我們得 $A^r\mathbf{v}=\mathbf{v}$, 亦即 $\mathbf{v}\in E_{A^r}(1)$. 因此得 $E_A(1)\subseteq E_{A^r}(1)$, 所以 $\dim(E_A(1))\leq\dim(E_{A^r}(1))=1$. 然而前面已知 1 為 A 的一個 eigenvalue, 因此 $\dim(E_A(1))>0$. 得證 $\dim(E_A(1))=1$, 亦即 1 對於 A 的 geometric multiplicity 為 1. □

現在我們來探討一個 stochastic matrix $A\in\mathscr{M}_{n\times n}$ 其 eigenvector 有何特性. 對任意 $\mathbf{x}\in\mathbb{R}^n$ 假設 $\mathbf{x}=\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix}$ 且 $A\mathbf{x}=\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$, 對於 $i=1,\ldots,n$ 我們有 $y_i=\sum_{j=1}^n a_{ij}x_j$ 且因為對於任意 $i,j\in\{1,\ldots,n\}$ 皆有 $a_{ij}\geq 0$, 我們得

$$\begin{array}{rcl}|y_1|=|a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n| & \leq & a_{11}|x_1|+a_{12}|x_2|+\cdots+a_{1n}|x_n|\\ & \vdots & \\ |y_i|=|a_{i1}x_1+a_{i2}x_2+\cdots+a_{in}x_n| & \leq & a_{i1}|x_1|+a_{i2}|x_2|+\cdots+a_{in}|x_n|\\ & \vdots & \\ |y_n|=|a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n| & \leq & a_{n1}|x_1|+a_{n2}|x_2|+\cdots+a_{nn}|x_n|\end{array}\tag{9.9}$$

將式子 (9.9) 由上往下加起來且將右式同樣的 $|x_j|$ 加在一起, 由於 A 為 stochastic matrix, 對於 $j=1,\ldots,n$, 我們有 $a_{1j}+\cdots+a_{ij}+\cdots+a_{nj}=1$, 故得

$$|y_1|+|y_2|+\cdots+|y_n|\leq|x_1|+|x_2|+\cdots+|x_n|.\tag{9.10}$$

現若 $\mathbf{x}$ 是 A 的一個 eigenvalue 為 λ 的 eigenvector, 我們有 $A\mathbf{x}=\lambda\mathbf{x}$, 也就是 $y_i=\lambda x_i$, $\forall i=1,\ldots,n$. 代入式子 (9.10) 得 $|\lambda|(|x_1|+|x_2|+\cdots+|x_n|)\leq|x_1|+|x_2|+\cdots+|x_n|$. 由於 $x_1,\ldots,x_n$ 不全為 0 得證 $|\lambda|\leq 1$. 也就是說當 A 為 stochastic matrix, 它的 eigenvalue λ 必須滿足 $|\lambda|\leq 1$.

我們先考慮 $\lambda\neq 1$ 的情形. 此時對於 $i=1,\ldots,n$ 我們有 $\sum_{j=1}^n a_{ij}x_j=y_i=\lambda x_i$. 將此 n 個等式全部加起來且等式左邊同樣的 x_i 項合併, 由於 A 為 stochastic matrix 我們得 $x_1+\cdots+x_n=\lambda(x_1+\cdots+x_n)$. 又由於 $\lambda\neq 1$ 得證 $x_1+\cdots+x_n=0$.

現在我們回到探討 A 為 regular 的情形. 首先考慮 eigenvalue 為 -1 的情形. 假設 $\mathbf{w}\in\mathbb{R}^n$ 為 A 的 eigenvalue 為 -1 的 eigenvector. 此時由於 $A\mathbf{w}=-\mathbf{w}$, 我們得 $A^k\mathbf{w}=(-1)^k\mathbf{w}$. 也就是說當 $k>0$ 為偶數時 $\mathbf{w}$ 就會是 A^k 的 eigenvalue 為 1 的 eigenvector. 又 A 必有 eigenvalue 為 1 的 eigenvector $\mathbf{v}$, 且當 A 為 regular 時我們知必存在 $k>0$ 為偶數使得 A^k 的每個 entry 皆為正實數 (Proposition 9.6.5), 故由 Proposition 9.6.7 的證明我們知此時 $E_A(1)=E_{A^k}(1)=\mathrm{Span}(\mathbf{v})$. 然而 $\mathbf{w}\in E_{A^k}(1)=\mathrm{Span}(\mathbf{v})$, 此與 $\mathbf{v},\mathbf{w}$ 為 linearly independent (因 $\mathbf{v},\mathbf{w}$ 分別為 A 對應到 $1,-1$ 的 eigenvector, 由 Proposition 9.2.4 知它們為 linearly independent) 相矛盾. 故知當 A 為 regular stochastic matrix 時 -1 不可能會是 A 的 eigenvalue.

最後我們考慮 A 為 regular 且 eigenvalue 為 1 的情形. 要注意由 Proposition 9.6.7 的證明我們知道, 若 A^r 的每個 entry 皆為正實數, 則 eigenspace $E_A(1)=E_{A^r}(1)$. 所以我們只要考慮 A 為 stochastic matrix 且 $A=(a_{ij})$ 的每個 entry a_{ij} 皆為正實數的情形即可. 假設 $\mathbf{x}$ 為 A 的 eigenvalue 為 1 的 eigenvector. 此時由於 $A\mathbf{x}=\mathbf{x}$, 我們有 $y_i=x_i$, $\forall i=1,\ldots,n$, 因此得 $|y_1|+|y_2|+\cdots+|y_n|=|x_1|+|x_2|+\cdots+|x_n|$. 也就是說式子 (9.10) 的等式必須成立. 然而式子 (9.10) 的等式成立若且唯若式子 (9.9) 的每一項等式皆成立. 而又對於任意 $i,j\in\{1,\ldots,n\}$ 皆有 $a_{ij}>0$ 故式子 (9.9) 的每一項等式皆成立若且唯若 $x_1,\ldots,x_n$ 皆同時大於等於 0 或同時小於等於 0. 由於 $x_1,\ldots,x_n$ 不全為 0, 我們有 $x_1+\cdots+x_n\neq 0$, 故令 $\mathbf{v}=\frac{1}{x_1+\cdots+x_n}\mathbf{x}$, 我們得 $\mathbf{v}$ 為 probability vector 且為 A 的 eigenvalue 為 1 的 eigenvector. 因為 $\dim(E_A(1))=1$, 我們知道 $\mathbf{v}$ 會是 $\mathbb{R}^n$ 中唯一同時符合這兩個要求的 vector. 綜合以上的討論, 我們有以下的結論.

Proposition 9.6.8. *假設 $A\in\mathscr{M}_{n\times n}$ 為 stochastic matrix, 則 A 的任一 eigenvalue λ 皆需滿足 $|\lambda|\leq 1$. 若 $\mathbf{w}=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}\in\mathbb{R}^n$ 為 A 的 eigenvector 且其 eigenvalue 不等於 1, 則 $c_1+\cdots+c_n=0$.*

另外若 A 為 regular, 則 -1 不可能會是 A 的 eigenvalue 且在 $\mathbb{R}^n$ 中存在唯一的 probability vector 會是 A 的 eigenvalue 為 1 的 eigenvector.

Question 9.10. *假設 A 為 regular stochastic matrix 且 $\mathbf{v}$ 為 A 的 eigenvalue 為 1 的 eigenvector. 試證明 $\mathbf{v}$ 每一個 entry 皆同時為正或同時為負 (即沒有一個 entry 會是 0).*

現在我們假設 $A\in\mathscr{M}_{n\times n}$ 為 regular stochastic matrix 且為 diagonalizable. 令 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 $\mathbb{R}^n$ 的一組 basis 且為 A 的 eigenvectors. 假設 $\mathbf{v}_i$ 所對應的 eigenvalue 為 λ_i, 其中 $\lambda_1=1$. 由 Proposition 9.6.8 我們可設 $\mathbf{v}_1$ 為 probability vector, 又對於 $i=2,\ldots,n$, $|\lambda_i|<1$ 故若 $\mathbf{v}_i=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$ 我們有 $c_1+\cdots+c_n=0$. 現令 $P=\begin{bmatrix}| & & |\\ \mathbf{v}_1 & \cdots & \mathbf{v}_n\\ | & & |\end{bmatrix}$, 我們得 $A=P\begin{bmatrix}1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}P^{-1}$.
由於矩陣的乘法僅牽涉到每個 entry 之間的加法與乘法, 所以當取極限時可以分別取極限再乘在一起. 因此我們有

$$\lim_{k\to\infty}A^k=\lim_{k\to\infty}P\begin{bmatrix}1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}^k P^{-1}=\lim_{k\to\infty}P\begin{bmatrix}1 & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k\end{bmatrix}P^{-1}=P\begin{bmatrix}1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{bmatrix}P^{-1}.$$

注意由於 $\mathbf{v}_1$ 的每個 entry 之和為 1, 而對於 $i=2,\ldots,n$, $\mathbf{v}_i$ 每一個 entry 之和為 0, 利用 $P^{-1}P=I_n$, 我們很容易看出 P^{-1} 的 1-st row 的每一個 entry 皆為 1. 故得

$$P\begin{bmatrix}1 & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{bmatrix}P^{-1}=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}\begin{bmatrix} & 1\ \cdots\ 1 & \\ - & \mathbf{O} & - \\ & \vdots & \\ - & \mathbf{O} & -\end{bmatrix}=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_1 & \cdots & \mathbf{v}_1\\ | & | & & |\end{bmatrix}.$$

也就是說當 A 為 regular stochastic matrix 且為 diagonalizable, 則當 k 越來越大時 A^k 會趨近一個每個 column 皆為 $\mathbf{v}_1$ 的矩陣, 其中 $\mathbf{v}_1$ 為 $\mathbb{R}^n$ 中唯一的 probability vector 滿足 $A\mathbf{v}_1=\mathbf{v}_1$. 這個結果其實對於一般的 regular stochastic matrix (不需 diagonalizable 之假設) 皆成立. 它的證明並沒有用到太多線性代數的技巧, 而著重於數列極限的估計. 不過為了完整性, 我們仍簡略的證明一下. 同學們可以忽略它的證明, 不過我們希望大家仍能充分了解此定理的敘述以及其衍伸的結果 (Corollary 9.6.10).

Theorem 9.6.9. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *regular stochastic matrix* 且 $\mathbf{v}\in\mathbb{R}^n$ 為唯一的 *probability vector* 滿足 $A\mathbf{v}=\mathbf{v}$. 則

$$\lim_{k\to\infty}A^k=\begin{bmatrix}| & | & & |\\ \mathbf{v} & \mathbf{v} & \cdots & \mathbf{v}\\ | & | & & |\end{bmatrix}.$$

Proof. 當 $n=1$ 時 $A=[1]$ 是唯一的 stochastic matrix, 此時本定理自然成立, 所以我們考慮 $n\geq 2$ 的情形. 此時我們要證明 $\lim_{k\to\infty}A^k$ 的每一個 column 是相同的 vector. 也就是說我們要先證明存在 $\mathbf{w}\in\mathbb{R}^n$ 使得 $\lim_{k\to\infty}A^k=\begin{bmatrix}| & | & & |\\ \mathbf{w} & \mathbf{w} & \cdots & \mathbf{w}\\ | & | & & |\end{bmatrix}$. 若證明了這部分, 則因 $A\mathbf{v}=\mathbf{v}$, 我們有 $A^k\mathbf{v}=\mathbf{v}$, $\forall k\in\mathbb{N}$, 故得 $\lim_{k\to\infty}(A^k\mathbf{v})=\mathbf{v}$. 然而若 $\mathbf{v}=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$, 因 $\mathbf{v}$ 為 probability vector, 我們有 $c_1+\cdots+c_n=1$. 因此利用 $\lim_{k\to\infty}A^k=\begin{bmatrix}| & | & & |\\ \mathbf{w} & \mathbf{w} & \cdots & \mathbf{w}\\ | & | & & |\end{bmatrix}$. 得

$$\lim_{k\to\infty}(A^k\mathbf{v})=(\lim_{k\to\infty}A^k)\mathbf{v}=\begin{bmatrix}| & | & & |\\ \mathbf{w} & \mathbf{w} & \cdots & \mathbf{w}\\ | & | & & |\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix}=c_1\mathbf{w}+c_2\mathbf{w}+\cdots+c_n\mathbf{w}=\mathbf{w}.$$

得證 $\mathbf{w}=\mathbf{v}$. 要如何證明 $\lim_{k\to\infty}A^k$ 的每一個 column 是相同的 vector 呢? 給定 $i\in\{1,\ldots,n\}$, 我們考慮所有 A^k 的 i-th row. 若令 M_k, m_k 分別為 A^k 的 i-th row 的 entry 中的最大值與最小值, 我們要證明 $\lim_{k\to\infty}(M_k-m_k)=0$, 因此得證 $\lim_{k\to\infty}A^k$ 的 i-th row 的每個 entry 皆為同一個數. 因為這是對所有 $i=1,\ldots,n$ 皆成立, 也因此得證 $\lim_{k\to\infty}A^k$ 的每一個 column 皆為相同的 vector.

我們令 a_{ij} 為 A 的 (i,j)-th entry 且當 $k>1$ 時令 $a_{ij}^{(k)}$ 為 A^k 的 (i,j)-th entry. 依矩陣乘法的定義以及 $A^{k+1}=A^kA$, 我們有 $a_{ij}^{(k+1)}=\sum_{q=1}^n a_{iq}^{(k)}a_{qj}$. 假設 A^k 的 i-th row 的最大值和最小值分別發生在 A^k 的 (i,s)-th entry 以及 (i,t)-th entry (即 $M_k=a_{is}^{(k)}$, $m_k=a_{it}^{(k)}$), 又假設 A^k 的 i-th row 的最大值和最小值分別發生在 A^{k+1} 的 (i,j)-th entry 以及 (i,l)-th entry (即 $M_{k+1}=a_{ij}^{(k+1)}$, $m_{k+1}=a_{il}^{(k+1)}$). 此時我們有

$$M_{k+1}=a_{ij}^{(k+1)}=a_{it}^{(k)}a_{tj}+\sum_{q\neq t}a_{iq}^{(k)}a_{qj}\leq m_ka_{tj}+M_k\sum_{q\neq t}a_{qj}=m_ka_{tj}+M_k(1-a_{tj}). \tag{9.11}$$

$$m_{k+1}=a_{il}^{(k+1)}=a_{is}^{(k)}a_{sl}+\sum_{q\neq s}a_{iq}^{(k)}a_{ql}\geq M_ka_{sl}+m_k\sum_{q\neq s}a_{ql}=M_ka_{sl}+m_k(1-a_{sl}). \tag{9.12}$$

注意式子 (9.11), (9.12) 用到了 A 為 stochastic matrix 之假設, 即 $\sum_{q=1}^{n} a_{qj} = \sum_{q=1}^{n} a_{ql} = 1$. 現在將式子 (9.11) 減去式子 (9.12) 得

$$M_{k+1} - m_{k+1} \leq M_k(1 - a_{tj} - a_{sl}) + m_k(a_{tj} - (1 - a_{sl})) = (M_k - m_k)(1 - a_{tj} - a_{sl}). \tag{9.13}$$

由於 $a_{tj} \geq 0, a_{sl} \geq 0$, 我們有 $1 - a_{tj} - a_{sl} \leq 1$, 故由式子 (9.13) 得 $M_{k+1} - m_{k+1} \leq M_k - m_k$. 也就是說數列 $(M_k - m_k)_{k=1}^{\infty}$ 是一個遞減的數列.

如何證明 $(M_k - m_k)_{k=1}^{\infty}$ 這個遞減的數列其極限為 0 呢? 我們只要找到它的一個 subsequence 趨近於 0 即可. 這裡就需要用到 A 為 regular 的假設. 也就是說存在 $r \in \mathbb{N}$ 使得 A^r 的每個 entry 皆為正實數. 為了方便起見, 我們令 $B = A^r$ 且令 b_{ij} 為 B 的 (i,j)-th entry 且當 $k > 1$ 時令 $b_{ij}^{(k)}$ 為 $B^k = A^{rk}$ 的 (i,j)-th entry. 注意依假設我們有 $b_{ij} > 0$ 且 $b_{ij}^{(k)} > 0$. 又給定 $i \in \{1, \ldots, n\}$, 我們有 M_{rk}, m_{rk} 分別為 $B^k = A^{rk}$ 的 i-th row 的 entry 中的最大值與最小值. 因此用相同的方法, 套用式子 (9.11), (9.12) 在 B^k 和 B^{k+1} 的情形, 可以推得式子 (9.13) 在 B^k 與 B^{k+1} 相對應的結果

$$M_{r(k+1)} - m_{r(k+1)} \leq M_{rk}(1 - b_{tj} - b_{sl}) + m_{rk}(b_{tj} - (1 - b_{sl})) = (M_{rk} - m_{rk})(1 - b_{tj} - b_{sl}). \tag{9.14}$$

要注意式子 (9.14) 與式子 (9.13) 最大的不同處在於 $b_{tj} > 0$ 且 $b_{sl} > 0$. 現令 β 為 B 的所有 entry 中的最小值. 依 β 的定義, 我們有 $b_{tj} \geq \beta, b_{sl} \geq \beta$, 故得 $1 - b_{tj} - b_{sl} \leq 1 - 2\beta$. 因此由式子 (9.13) 得

$$M_{r(k+1)} - m_{r(k+1)} \leq (M_{rk} - m_{rk})(1 - 2\beta). \tag{9.15}$$

因 β 所在的那個 column 所有的 entry 之和等於 1 且大於等於 $n\beta$, 故得 $0 < \beta \leq \frac{1}{n} \leq \frac{1}{2}$ (注意我們有 $n \geq 2$ 之假設). 所以 $0 \leq 1 - 2\beta < 1$, 利用式子 (9.15) 以及數學歸納法得

$$M_{r(k+1)} - m_{r(k+1)} \leq (M_r - m_r)(1 - 2\beta)^k,$$

且因此得證 $(M_{rk} - m_{rk})_{k=1}^{\infty}$ 這個 subsequence 滿足 $\lim_{k\to\infty}(M_{rk} - m_{rk}) = 0$, 故得原 sequence $(M_k - m_k)_{k=1}^{\infty}$ 滿足 $\lim_{k\to\infty}(M_k - m_k) = 0$. 得證本定理. □

再次重申 Theorem 9.6.9 對於一般的 stochastic matrix 未必會成立, 須加上 regular 的條件才一定成立. 這個定理的最重要的結果就是在 Markov process 中不管初始的 probability vector 為何, 經過 regular stochastic matrix 多次的作用之下, 最後會趨近於一個穩定的狀態. 我們將此結果敘述如下.

Corollary 9.6.10. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *regular stochastic matrix* 且 $\mathbf{v} \in \mathbb{R}^n$ 為唯一的 *probability vector* 滿足 $A\mathbf{v} = \mathbf{v}$. 則對於任意 $\mathbb{R}^n$ 的 *probability vector* $\mathbf{w}$ 皆有

$$\lim_{k\to\infty}(A^k\mathbf{w}) = \mathbf{v}.$$

Proof. 令 $\mathbf{w} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. 依 $\mathbf{w}$ 為 probability vector 之假設, 我們知 $c_1 + \cdots + c_n = 1$. 故由 Theorem 9.6.9 知

$$\lim_{k\to\infty}(A^k\mathbf{w}) = (\lim_{k\to\infty} A^k)\mathbf{w} = \begin{bmatrix} | & | & & | \\ \mathbf{v} & \mathbf{v} & \cdots & \mathbf{v} \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = c_1\mathbf{v} + c_2\mathbf{v} + \cdots + c_n\mathbf{v} = \mathbf{v}.$$

□

Question 9.11. *Corollary 9.6.10 中* $\mathbf{w}$ *需要假設是 probability vector 才會成立嗎?*

9.7. 結論

從 linear operator 的觀念來看 eigenvector 就是經過 linear operator 運算後仍保持平行的向量. 所以若有一組 basis 是由 eigenvectors 所組成, 則此 linear operator 利用這組 ordered basis 所得的 matrix representation 就會是 diagonal matrix. 這種情況之下便會讓我們很容易掌握這個 linear operator (特別是對其多次重複合成的情況之下). 因此要了解一個 square matrix 是否可以對角化, 是一個重要的課題. 利用 eigenvalue 的 algebraic multiplicity 以及 geometric multiplicity, 可以幫助我們判別一個 square matrix 是否為 diagonalizable. 有一種特別的情形是不必計算就可以知道可以對角化, 就是 symmetric matrix. 事實上 symmetric matrix 不只是 diagonalizable, 其實它還是 orthogonal diagonalizable. 也就是說 symmetric matrix 都會存在一組由 eigenvectors 所組成的 orthogonal basis. 這就是所謂的 Spectral Theorem, 它在數學和物理上有許多的應用. 我們特別利用它, 可將坐標平面和坐標空間的二次方程式, 利用坐標變換 (旋轉和平移) 方法將之變換成標準式, 以便判斷其圖形.