# 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.

# Chapter 1

# 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.

## 1.1. 一次聯立方程組及基本列運算

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

$$a_1x_1+\cdots+a_nx_n=b,$$

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 (system of $m$ linear equations in $n$ variables). 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

$$\begin{aligned} 3x_1-2x_2+9x_4 &= 4 \\ 2x_1+2x_2-4x_4 &= 6 \end{aligned} \tag{1.1}$$

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.

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. 我們將介紹一個有系統的方法來解聯立方程組, 把這三種基本方法看成是對矩陣的運算.

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 (1.1) 中的聯立方程組所對應的 augmented matrix 為

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

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 $[A \mid \mathbf{b}]$ 這一個 matrix. 反之一個 augmented matrix $[A \mid \mathbf{b}]$ 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.

為了方便起見, 我們將上面 (1), (2), (3) 三種 elementary row operation 分別稱為 *type 1*, *type 2* 以及 *type 3* 的 elementary row operation.

**Question 1.1.** 任一 *type* 的 *elementary row operation* 可以用其他兩種 *type* 取代嗎*?(*參見習題*)*

**Example 1.1.1.** 考慮矩陣

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

將 $A$ 的第一、第二兩個 row 互換 (即做一個 type 1 的 elementary row operation), 可得

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

而將 $B$ 的第二個 row 乘上 2 (即做一個 type 2 的 elementary row operation), 可得

$$C = \begin{bmatrix} 2 & 1 & -1 & 3 \\ 2 & 4 & 6 & 8 \\ 4 & 0 & 1 & 2 \end{bmatrix}.$$

而將 $C$ 的第三個 row 乘上 $-3$ 加到第一個 row (即做一個 type 3 的 elementary row operation), 可得

$$D=\begin{bmatrix}-10 & 1 & -4 & -3\\ 2 & 4 & 6 & 8\\ 4 & 0 & 1 & 2\end{bmatrix}.$$

$\sharp$

注意, 若一個矩陣 $P$ 經由一個 elementary row operation 轉換成矩陣 $Q$, 我們也可以對 $Q$ 藉由同樣 type 的 elementary row operation 將之轉換回 $P$. 例如前面 Example 1.1.1 中, 我們可以將 $B$ 的第一、第二兩個 row 互換而得到 $A$. 我們也可將 $C$ 的第二個 row 乘上 1/2 而得到 $B$. 另外我們也可將 $D$ 的第三個 row 乘上 3 加到第一個 row 而轉換回 $C$ (參見習題).

**Question 1.2.** 兩個同樣 *type* 的 *elementary row operation* 交換順序會一樣嗎? 兩個不同 *type* 的 *elementary row operation* 交換順序會一樣嗎?

## 1.2. 解聯立方程組

大家應很容易看出一個 augmented matrix 經過上節所提的 elementary row operation 後所得的 augmented matrix 所對應的聯立方程組就是大家熟悉的加減消去法的三種步驟所得的方程組. 利用加減消去法最常遇到的問題就是 (尤其在處理未知數很多的方程組時), 常常做了幾次後, 混亂到不知道那些式子是消過了以及那些式子還可以再進一步消減. 還有就是, 到底要將方程組的式子消到哪種地步時, 才可以解出方程組. 關於第一個問題, 我們可以理解用矩陣的表示法就可以把這消去的過程記錄下來. 而接下來我們要探討的就是第二個問題, 也就是將矩陣化成某種特定的形式就可以解出方程式來.

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 (未知數) 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 還有其他位置也是 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 所在的位置是 $x_j$, 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*.

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).

很容易就可以發現一個矩陣利用 elementary row operations 化成的 echelon form 的方法有很多種，而所化成的 echelon form 也有可能不同 (例如一個 echelon form 的某一個 row 經由 type 2 的 elementary row operation 作用後仍為 echelon form)。不過以後我們會證明，同一個矩陣不管用哪些 elementary row operations 化成的 echelon form 它們的 pivot variables 皆會相同。也因此一個矩陣化為 echelon form 其 pivot 的個數是固定的, 我們特別有以下的定義.

**Definition 1.2.1.** 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 $\text{rank}(A) = r$ 來表示.

一個矩陣的 rank 是該矩陣很重要的資訊，以後我們會探討它很多的性質。由於它的重要性，我們特別在這裡先介紹它的定義，讓大家先熟悉一下。

**Question 1.3.** *假設矩陣 $A$ 有 $m$ 個 row 以及 $n$ 個 column. 若 $\text{rank}(A) = r$, 試說明 $r \leq \min\{m,n\}$.*

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.

**Question 1.4.** 考慮一個由 $n$ 個 *variables* 的 $m$ 個方程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;

**Question 1.5.** 說明一個由 $n$ 個 *variables* 的 $m$ 個方程式所組成的聯立方程組，當係數矩陣的 *rank* 等於 $m$，此聯立方程組一定有解 (即為 *consistent*). 而當係數矩陣的 *rank* 等於 $n$ 時，此聯立方程組若有解則解唯一。

我們總結, 當 $A$ 是 echelon form 時, 找出 $A\mathbf{x} = \mathbf{b}$ 所有的解的方法. 首先, 當 $A$ 有一個 row 全為 0 但 $\mathbf{b}$ 在該 row 不為 0 時, 我們知該方程組無解, 所以我們僅需討論其他的情況 (即有解的情況). 此時我們先挑出 free variable (即非 pivot variable). 由於 free variable 可以任意取值, 一般來說我們會用一些參數表示之 (注意不同的 free variable 要用不同的參數代號). 接著, 我們由下而上, 從最大編號的 pivot variable 開始, 利用 free variables 的那些參數將它的值寫下來, 再依序寫出所有 pivot variables 的值.

我們看以下幾個解聯立方程組的例子.

**Example 1.2.2.** Solve the linear system

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

此聯立方程組的 augmented matrix 為

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

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

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

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

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

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 1.2.3.** Solve the linear system

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

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 1.2.4.** 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 column vector 且將 $r,s$ 提出. 故將解寫成

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

♯

了解到解聯立方程組的方法及步驟後, 有幾件事必須要說明: (1) 為何經由 elementary row operations 我們可以將一個矩陣化為 echelon form? (2) 為何利用這個 echelon form, 便可得到與原方程組相同的解集合? (3) 為什麼用前面介紹 (pivot variables, free variables) 的方法就可以把係數矩陣是 echelon form 的聯立方程組所有的解找出來? 在下一節, 我們將詳細介紹有關 echelon form 的特性, 然後一一回答這些問題. 不過再次提醒大家務必先熟悉這節介紹解聯立方程組的方法及步驟.

## 1.3. Echelon Form 的性質

這一節中我們將說明前面提到有關 echelon form 的三個問題. 首先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. 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 $4, 5, 6, \ldots$ 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.

我們先看只有一個 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 $b$ 所在位置和第一個 row 的 leading entry $a$ 相同, 我們可將第一個 row 乘以 $-b/a$, 再加到第二個 row 上. 如此一來第二個 row 原本的 leading entry 變為 0, 故其 leading entry 所在位置往右移了, 依定義此時為 echelon form.

接著我們使用數學歸納法的假設, 亦即任何有 $k$ 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. 現假設此時第一個 row 的 leading entry 為 $a$. 接下來我們挑出其他 row 中

leading entry 的位置與第一個 row 的 leading entry 位置一樣的 row. 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 type 2 的 elementary row operation. 事實上在化成 echelon form 的過程確實只需要用到 type 1,3 這兩種 elementary row operations, 至於 type 2 的 elementary row operation 會在以後我們會介紹化為 “reduced” echelon form 的過程是需要的, 留待以後再談.

接下來我們說明為何將 augmented matrix $[A \mid \mathbf{b}]$ 利用 elementary row operations 化成 echelon form $[A' \mid \mathbf{b}']$, 則其對應的聯立方程組 $A'\mathbf{x} = \mathbf{b}'$ 會和原方程組 $A\mathbf{x} = \mathbf{b}$ 有相同的解集合. 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 $A'\mathbf{x} = \mathbf{b}'$. 也就是說 $A\mathbf{x} = \mathbf{b}$ 的解就會是 $A'\mathbf{x} = \mathbf{b}'$ 的解. 不過這不表示它們會有相同的解集合, 我們還要說明 $A'\mathbf{x} = \mathbf{b}'$ 的解也會是 $A\mathbf{x} = \mathbf{b}$ 的解才行. 然而我們前面提及 elementary row operations 是可以還原的. 換句話說 $[A' \mid \mathbf{b}']$ 也可經由 elementary row operations 變換成 $[A \mid \mathbf{b}]$. 所以套用剛才的理由, 我們也知 $A'\mathbf{x} = \mathbf{b}'$ 的解就會是 $A\mathbf{x} = \mathbf{b}$ 的解. 因此得證 $A\mathbf{x} = \mathbf{b}$ 和 $A'\mathbf{x} = \mathbf{b}'$ 會有相同的解集合.

**Example 1.3.1.** 考慮聯立方程組 $\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2 \end{array}\right.$. 如果 $x_1 = c_1, x_2 = c_2$ 是此方程組的一組解，表示 $a_{11}c_1 + a_{12}c_2 = b_1$ 以及 $a_{21}c_1 + a_{22}c_2 = b_2$ 也因此 $x_1 = c_1, x_2 = c_2$ 會是聯立方程組 $\left\{\begin{array}{l} a_{21}x_1 + a_{22}x_2 = b_2 \\ a_{11}x_1 + a_{12}x_2 = b_1 \end{array}\right.$ 的一組解。反之亦然，這說明對此聯立方程組做一個 type 1 的 elementary row operation 並不會影響解集合。

另一方面若給定非 0 實數 $r$, 則 $c_1, c_2$ 也會滿足 $a_{11}c_1 + a_{12}c_2 = b_1$ 以及 $ra_{21}c_1 + ra_{22}c_2 = rb_2$. 也就是說 $x_1 = c_1, x_2 = c_2$ 會是聯立方程組 $\left\{\begin{array}{c} a_{11}x_1 + a_{12}x_2 = b_1 \\ ra_{21}x_1 + ra_{22}x_2 = rb_2 \end{array}\right.$ 的一組解. 這說明了原方程組的解仍會是方程組在做了 type 2 的 elementary row operation 後所得的方程組之解。特別的是，由於 $r \neq 0$, 將新方程組的第二個 row 乘上 $1/r$, 便會得到原方程組。所以同理可知新方程組的解仍會是原方程組的解。這說明對此聯立方程組做一個 type 2 的 elementary row operation 並不會影響解集合。

最後我們看 type 3 的情況。若 $r$ 為給定實數，由於 $c_1,c_2$ 也會滿足 $(ra_{11}+a_{21})c_1+(ra_{12}+a_{22})c_2=(ra_{11}c_1+ra_{12}c_2)+(a_{21}c_1+a_{22}c_2)=rb_1+b_2$. 故知 $x_1=c_1,x_2=c_2$ 也會是聯立方程組 $\begin{cases} a_{11}x_1+a_{12}x_2=b_1 \\ (ra_{11}+a_{21})x_1+(ra_{12}+a_{22})x_2=rb_1+b_2 \end{cases}$ 的一組解。這說明了原方程組的解仍會是方程組在做了 type 3 的 elementary row operation 後所得的方程組之解。由於新方程組也可利用第一個 row 乘上 $-r$ 加到第二個 row 的方式得到原方程組。所以同理可知新方程組的解仍會是原方程組的解。這說明對此聯立方程組做一個 type 3 的 elementary row operation 並不會影響解集合。 $\sharp$

當連立方程組 $A\mathbf{x}=\mathbf{b}$ 和 $A'\mathbf{x}=\mathbf{b}'$ 它們的解集合相同, 這表示兩組方程組是有很特別的關係的. 我們有以下的定義.

**Definition 1.3.2.** 假設 linear systems $A\mathbf{x}=\mathbf{b}$ 和 $A'\mathbf{x}=\mathbf{b}'$ 的解集合相同, 則稱 $A\mathbf{x}=\mathbf{b}$ 和 $A'\mathbf{x}=\mathbf{b}'$ 為 *equivalent linear systems*

從上面的探討我們知道 augmented matrix $[A\,|\,\mathbf{b}]$ 若利用 elementary row operations 化成 $[A'\,|\,\mathbf{b}']$, 則 $A\mathbf{x}=\mathbf{b}$ 和 $A'\mathbf{x}=\mathbf{b}'$ 為 equivalent linear systems.

我們已知要探討聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的解, 僅要考慮 $A$ 為 echelon form 的情形. 接下來我們就是要討論當 $A$ 為 echelon form 時, 聯立方程組 $A\mathbf{x}=\mathbf{b}$ 解的特性. 事實上我們很容易理解利用 1.2 節中所提求解的方法所得的結果皆為方程組的一組解. 這裡要探討的是為何利用 1.2 節中所提求解的方法, 就可得所有的解. 接著我們將說明, 雖然一個矩陣利用 elementary row operations 化為 echelon form 的結果不唯一, 但是它們的 pivot variables 是唯一的.

如果我們得到 1.2 節 (2)(a) 的情形 (即 $A$ 有一個 row 全為 0 但 $\mathbf{b}$ 在該 row 不為 0), 在該節已說明此時方程組無解. 所以我們只要探討有解的情形. 首先回顧一下在 1.2 節所提求解的方法: 首先我們要找到 *free variables*, 也就是是方程組除了 pivot variable 以外的 variable. 接著給這些 free variable 任意的參數值, 然後再利用由下往上代回的方式找到聯立方程組所有的解. 若無 free variable, 就直接由下往上一步一步求值即可.

由於可以忽略 augmented matrix 全為 0 的 row, 所以我們可假設係數矩陣 $A$ 沒有一個 row 全為 0. 因為 $A$ 為 echelon form, 這也表示 $A$ 每一個 row 皆有 leading entry 且為 pivot. 現在我們回答當 $A$ 是 echelon form 時, 1.2 節中所述解聯立方程組 $A\mathbf{x}=\mathbf{b}$ 的方法所求得的解就是所有的解. 也就是說給定 $x_1=c_1,\ldots,x_n=c_n$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解, 我們要說明這組解確實可由 1.2 節所提的方法得到. 為了方便起見我們令 1.2 節所提的方法所得的解所成的集合為 $S$. 我們要說明 $(x_1,\ldots,x_n)=(c_1,\ldots,c_n)$ 確實為 $S$ 中的元素. 現若 $x_n$ 為 pivot variable, 則 $x_n$ 的值是被唯一確定的. 所以 $S$ 的所有解中 $x_n$ 的取值一定也為 $c_n$. 若 $x_n$ 為 free variable, 則因 $S$ 的解中 $x_n$ 可為任意值, 故 $S$ 中一定有一組解其 $x_n$ 的取值為 $c_n$. 也就是說不管 $x_n$ 是否為 pivot variable, $S$ 中必有一組解其 $x_n$ 的取值為 $c_n$. 現若 $x_{n-1}$ 為 pivot variable, 則由此 pivot 所在的 row 所對應的方程式可知 $x_{n-1}$ 的取值會被 $x_n$ 的取值所決定. 今已知 $S$ 中必有一組解其 $x_n$ 的取值為 $c_n$, 故此組解必滿足 $x_{n-1}=c_{n-1},x_n=c_n$; 而若 $x_{n-1}$ 為 free variable, 則因 $S$ 的解中 $x_{n-1}$ 可為任意值且其取值不影響到 $x_n$ 的取值, 故知 $S$ 中必

有一組解其 $x_{n-1}, x_n$ 的取值為 $x_{n-1} = c_{n-1}, x_n = c_n$. 如此一直下去我們知道 $S$ 中必有一組解其 $x_1, \ldots, x_n$ 的取值為 $x_1 = c_1, \ldots, x_n = c_n$.

我們可以利用上面的概念, 推導出當 $A$ 為 echelon form 時, pivot variables 和 free variables 對聯立方程組 $A\mathbf{x} = \mathbf{b}$ 解的影響. 首先看 pivot variable 對聯立方程組的解之影響.

**Lemma 1.3.3.** *假設 $A$ 為一有 $n$ 個 column 的 echelon form 且 $x_1 = c_1, \ldots, x_n = c_n$ 和 $x_1 = d_1, \ldots, x_n = d_n$ 皆為方程組 $A\mathbf{x} = \mathbf{b}$ 的一組解.*

(1) *假設 $x_n$ 為 $A$ 的一個 pivot variable. 則 $c_n = d_n$.*

(2) *假設 $x_k$ 為 $A$ 的一個 pivot variable, 其中 $1 \leq k \leq n-1$. 若 $c_{k+1} = d_{k+1}, \ldots, c_n = d_n$, 則 $c_k = d_k$.*

**Proof.** 假設聯立方程組為

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

其中

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

為 echelon form, 且不失一般性我們假設 $A$ 的每一個 row, $a_{i1}, \ldots, a_{in}$ 皆不全為 0.

(1) 若 $x_n$ 為 $A$ 的一個 pivot variable, 表示 $A$ 的最後一個 row 的 leading entry 所在位置為 $x_n$. 也就是說 $a_{m1} = a_{m2} = \cdots = a_{mn-1} = 0$ 且 $a_{mn} \neq 0$. 這表示此聯立方程組中最後一個式子為 $a_{mn}x_n = b_m$. 故由 $x_1 = c_1, \ldots, x_n = c_n$ 及 $x_1 = d_1, \ldots, x_n = d_n$ 皆為此聯立方程組的一組解知 $x_n = c_n$ 和 $x_n = d_n$ 皆需滿足 $a_{mn}x_n = b_m$, 亦即 $a_{mn}c_n = b_m$ 且 $a_{mn}d_n = b_m$. 故由 $a_{mn} \neq 0$ 得知 $c_n = d_n$.

(2) 若 $x_k$ 為 $A$ 的一個 pivot variable, 表示 $A$ 有一個 row 的 leading entry 所在位置為 $x_k$. 也就是說若此 row 為 $A$ 的第 $i$ 個 row, 則 $a_{i1} = a_{i2} = \cdots = a_{ik-1} = 0$ 且 $a_{ik} \neq 0$. 此 row 所對應的式子為 $a_{ik}x_k + \cdots + a_{in}x_n = b_i$. 故由 $x_1 = c_1, \ldots, x_n = c_n$ 及 $x_1 = d_1, \ldots, x_n = d_n$ 皆為此聯立方程組的一組解知 $x_k = c_k, x_{k+1} = c_{k+1}, \ldots, x_n = c_n$ 和 $x_k = d_k, x_{k+1} = d_{k+1}, \ldots, x_n = d_n$ 皆需滿足 $a_{ik}x_k + \cdots + a_{in}x_n = b_i$. 因此由 $c_{k+1} = d_{k+1}, \ldots, c_n = d_n$ 的假設知

$$a_{ik}c_k = b_i - (a_{ik+1}c_{k+1} + \cdots + a_{in}c_n) = b_i - (a_{ik+1}d_{k+1} + \cdots + a_{in}d_n) = a_{ik}d_k.$$

再由 $a_{ik} \neq 0$ 得知 $c_k = d_k$.

$\square$

相對於 pivot variable 我們知道對於 free variable 我們可以隨意取任何的實數而得到一組解, 所以我們有以下 free variable 對解的影響.

**Lemma 1.3.4.** 假設 $A$ 為一有 $n$ 個 *column* 的 *echelon form* 且沒有一個 *row* 全為 0.

(1) 假設 $x_n$ 為 $A$ 的一個 *free variable*. 則對任意的實數 $r$, 方程組 $A\mathbf{x}=\mathbf{b}$ 皆可找到一組解其 $x_n=r$.

(2) 假設 $x_k$ 為 $A$ 的一個 *free variable*, 其中 $1\leq k\leq n-1$. 若 $x_1=c_1,\dots,x_n=c_n$ 為方程組 $A\mathbf{x}=\mathbf{b}$ 的一組解, 則對任意實數 $r$ 方程組 $A\mathbf{x}=\mathbf{b}$ 皆可找到一組解其 $x_k=r$ 且 $x_{k+1}=c_{k+1},\dots,x_n=c_n$.

**Proof.** 在前面所提的求解過程中我們知道可將 free variable 定為任意的實數, 再一步一步由下往上代回得到一組解. 在這個過程中我們了解到若 $x_i$ 是 free variable, 則它的取值可能會影響到的僅有 $x_j$, 其中 $j<i$ 這樣的變數, 而不會影響到其他變數 $x_l$, 其中 $i<l$ 的取值.

現若 $x_n$ 是 free variable, 這表示我們可以設定 $x_n$ 為任意實數, 再一步一步往上代求得聯立方程組的一組解, 所以對任意的實數 $r$, 方程組 $A\mathbf{x}=\mathbf{b}$ 皆可找到一組解其 $x_n$ 為 $r$.

若 $x_k$ 為 $A$ 的一個 free variable, 其中 $1\leq k\leq n-1$ 且已知 $x_1=c_1,\dots,x_n=c_n$ 為方程組 $A\mathbf{x}=\mathbf{b}$ 的一組解. 換言之, $x_{k+1}=c_{k+1},\dots,x_n=c_n$ 皆滿足方程組 pivot 的位置在 $x_k$ 右方的那些 row 所對應的那些方程式. 由於 $x_k$ 可取任意的實數且不會影響 $x_{k+1},\dots,x_n$ 的取值, 所以我們可令 $x_k=r$ 且 $x_{k+1}=c_{k+1},\dots,x_n=c_n$ 一步一步代回求得聯立方程組的一組解. $\square$

接下來，我們用兩個前面討論過的例子說明一下 Lemma 1.3.3 和 Lemma 1.3.4.

**Example 1.3.5.** 在 Page 6, 我們知道聯立方程組

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

的解為 $x_1=2-t,x_2=1-t,x_3=t,x_4=-1$, 其中 $t$ 為任意實數. 我們可以看出，雖然這組方程組有無窮多解，不過所有的解中 $x_4$ 的值一定是 $-1$. 這就是 Lemma 1.3.3 (1) 中說的因為 $x_4$ 是最後一個 variable 且為 pivot, 所以任兩組解的 $x_4$ 的值都相同。又讓我們隨便選一組解，例如 $x_1=2,x_2=1,x_3=0,x_4=-1$ (即 $t$ 代 0 的情況). 由於 $x_3$ 是 free variable, Lemma 1.3.4 (2) 告訴我們若選定 $x_4=-1$, 我們都可將 $x_3$ 隨便代一個值來得到一組解。要注意的是, $x_2$ 雖為 pivot variable, 不過 Lemma 1.3.3 (2) 並不是告訴我們任選兩組解其 $x_2$ 的值都會一樣。它是說若這兩組解的 $x_3,x_4$ 的值是一致的則它們 $x_2$ 的值也會一樣。例如 $x_1=1,x_2=0,x_3=1,x_4=-1$ 也是一組解，它的 $x_2$ 的值 0 就和前面那組解的 $x_2$ 的值 1 就不同。不過若兩組解它們有一致的 $x_3$ 和 $x_4$, 那麼它們 $x_2$ 的值也一致。這可從前面所給的解集合看出。

在 Page 9, 我們知道聯立方程組

$$\begin{array}{rrrrcr} x_1 & -2x_2 & +x_3 & -x_4 & = & 4 \\ & x_2 & +2x_3 & -x_4 & = & -9 \end{array}$$

的解為 $x_1=-14+3r-5s,x_2=-9+r-2s,x_3=s,x_4=r$, 其中 $s,r$ 為任意實數. 由於最後一個 variables $x_4$ 是 free variable, Lemma 1.3.4 (1) 告訴我們任意給一個實數 $r$ 都可以找到一組解其 $x_4$ 的值為 $r$. 然而 $x_3$ 也是 free variable, 所以 Lemma 1.3.4 (2) 告訴我們再任意選一

個實數 $s$, 我們都可以找到一組解其 $x_4$ 和 $x_3$ 的值分別為 $r$ 和 $s$. 這都可以由解集合看出來。又 $x_2$ 是 pivot variable, 而 Lemma 1.3.3 (2) 告訴我們若有兩組解其 $x_3$ 的值是一樣的且 $x_4$ 的值也一樣，則這兩組解的 $x_2$ 的值也會一樣，因而再由 $x_1$ 是 pivot variable 可知它們 $x_1$ 的值也相同。這也和解集合若 $x_4 = r$, $x_3 = s$, 則可確定 $x_2$ 和 $x_1$ 的值分別為 $-9+r-2s$ 和 $-14+3r-5s$ 相吻合。 $\sharp$

Lemma 1.3.3 和 Lemma 1.3.4 有許多應用. 例如當 $A$ 是 echelon form 時若聯立方程組 $A\mathbf{x} = \mathbf{b}$ 已知有一個解 $x_1 = c_1, \ldots, x_n = c_n$ 且 $x_1, \ldots, x_n$ 每一個都是 pivot variable, 則由 Lemma 1.3.3 知聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的解僅能是 $x_1 = c_1, \ldots, x_n = c_n$. 換句話說此方程組的解唯一. 另一方面, 若聯立方程組 $A\mathbf{x} = \mathbf{b}$ 已知有解且 $x_1, \ldots, x_n$ 中有 free variable, 則由 Lemma 1.3.4 知聯立方程組 $A\mathbf{x} = \mathbf{b}$ 會有無窮多解 (除非我們討論的數系僅有有限多個元素).

當我們給一個矩陣時, 有許多種方法將之化為 echelon form, 而且化成的 echelon form 很可能不一樣. 不過利用 Lemma 1.3.3 和 Lemma 1.3.4 我們可以得到這些 echelon form 雖然可能不一樣, 但他們 pivot 的所在位置都會一致. 由於我們只關心係數矩陣 $A$ 化為 echelon form 後的情形, 所以我們可以考慮 $A\mathbf{x} = \mathbf{0}$ 這一種特殊形式的聯立方程組. 要注意這樣的聯立方程組都會有解, 因為 $x_1 = 0, \ldots, x_n = 0$ 就是一組解. 我們特別稱這樣的聯立方程組為 *homogeneous system.*

**Proposition 1.3.6.** *給定一矩陣 $A$, 若 $A_1, A_2$ 均為 $A$ 利用 elementary row operations 化成的 echelon forms. 則 $A_1$ 和 $A_2$ 的 pivot 個數相同, 事實上他們的 pivot variables 是一致的.*

**Proof.** 我們考慮 $A\mathbf{x} = \mathbf{0}$ 這一組聯立方程組, 其中 $A$ 有 $n$ 個 column (即此方程組有 $n$ 個變數). 因為 $A$ 可利用 elementary row operation 化為 $A_1$ 及 $A_2$, 這表示 augmented matrix $[A \mid \mathbf{0}]$ 可以利用 elementary row operation 化為 $[A_1 \mid \mathbf{0}]$ 及 $[A_2 \mid \mathbf{0}]$. 換句話說聯立方程組 $A_1\mathbf{x} = \mathbf{0}$ 和 $A_2\mathbf{x} = \mathbf{0}$ 皆與聯立方程組 $A\mathbf{x} = \mathbf{0}$ 有同樣的解. 再次強調這些聯立方程組都會有 $x_1 = 0, \ldots, x_n = 0$ 這樣的一組解.

我們要用反證法處理. 假設 $A_1$ 和 $A_2$ 有 pivot variable 不一致, 不失一般性我們就假設對 $A_1$ 來說 $x_i$ 是 pivot variable 但對 $A_2$ 來說 $x_i$ 不是 pivot variable (即 free variable). 假設 $i = n$, 這表示方程組 $A_1\mathbf{x} = \mathbf{0}$ 的解中 $x_n$ 的取值是唯一的 (Lemma 1.3.3), 事實上 $x_n$ 一定為 0; 但 $A_2\mathbf{x} = \mathbf{0}$ 的解中 $x_n$ 的取值卻可以是任意的實數 (Lemma 1.3.4). 這和此二方程組有相同的解相矛盾. 現若 $1 \le i \le n-1$. 利用 $x_1 = 0, \ldots, x_n = 0$ 已是這兩聯立方程組的解, 我們知道方程組 $A_1\mathbf{x} = \mathbf{0}$ 的解中一定找不到一組解其 $x_{i+1}, \ldots, x_n$ 的取值皆為 0 但 $x_i$ 的取值不是 0 (Lemma 1.3.3); 另一方面 Lemma 1.3.4 告訴我們 $A_2\mathbf{x} = \mathbf{0}$ 的解中一定可找到一組解其 $x_{i+1}, \ldots, x_n$ 的取值皆為 0 但 $x_i$ 的取值不是 0 (事實上 $x_i$ 可以是任意實數). 這又和 $A_1\mathbf{x} = \mathbf{0}, A_2\mathbf{x} = \mathbf{0}$ 此二方程組有相同的解相矛盾. 故由反證法知 $A_1$ 和 $A_2$ 的 pivot variables 是一致的. $\square$

在解聯立方程組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 entry 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 entry 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 entry 便是 1 了 (這就是需要 type 2 的 elementary row operation 的地方). 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.

化為 reduced echelon form 後, 我們就可以利用前面由 echelon form 求解的方法求出聯立方程組的解. 由於 reduced echelon form 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 $B'\mathbf{x}=\mathbf{0}$ 為

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

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 $(x_1,x_2,x_3,x_4)=(0,-3t,t,t)=t(0,-3,1,1), t\in\mathbb{R}$.

我們知道每個矩陣皆可經由 elementary row operations 化為 echelon form. 而我們又知每個 echelon form 也可利用 elementary row operations 化為 reduced echelon form. 因此每個矩陣皆可利用 elementary row operations 化為 reduced echelon form. 另外我們也知道將聯立方程組的 augmented matrix 做 elementary row operations 後所對應的聯立方程組會是 equivalent, 所以化為 reduced echelon form 所得的解集合也會和原方程組的解集合相同.

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* 或 *Gaussian elimination*, 而用 reduced echelon form 求解一般稱為 *Gauss-Jordan method*.

**Example 1.3.7.** Example 1.2.2 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 1.2.4 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 $x_1 = -14+3r-5s$. ♯

我們可以套用 Proposition 1.3.6 的證明方法, 證明一個矩陣利用 elementary row operations 化為 reduced echelon form 其結果是唯一的.

**Theorem 1.3.8.** *給定一矩陣 $A$, 若 $A_1,A_2$ 均為 $A$ 利用 elementary row operations 化成的 reduced echelon forms, 則 $A_1 = A_2$.*

**Proof.** 我們考慮 $A\mathbf{x} = \mathbf{0}$ 這一組聯立方程組, 依假設聯立方程組 $A_1\mathbf{x} = \mathbf{0}$ 和 $A_2\mathbf{x} = \mathbf{0}$ 皆與聯立方程組 $A\mathbf{x} = \mathbf{0}$ 有同樣的解.

利用反證法. 假設 $A_1 \neq A_2$ 且假設從下往上, $A_1, A_2$ 第一個發生相異的 row 其 pivot variable 為 $x_k$ (注意由 Proposition 1.3.6, 我們知道 $A_1, A_2$ 的 pivot variables 是一致的). 現假設 $A_1, A_2$ 在此 row 所對應的方程式分別為

$$x_k + a_{k+1}x_{k+1} + \cdots + a_n x_n = 0 \quad \text{與} \quad x_k + b_{k+1}x_{k+1} + \cdots + b_n x_n = 0 \tag{1.2}$$

其中存在 $l$ 滿足 $k+1 \leq l \leq n$ 且 $a_l \neq b_l$. 若 $x_l$ 為 pivot variable, 由於 $A_1, A_2$ 皆為 reduced echelon form, 在這個 row 中, 其他的 pivot variable 對應的係數應為 0, 而導致 $a_l = b_l = 0$ 之矛盾, 故知 $x_l$ 應為 free variable. 我們已知給定一組 free variables 的值, 可以用由下往上代回的方式得到聯立方程組的解. 現考慮除了 $x_l$ 這一個 free variable 代 1, 其他 free variables 代 0 所得的解. 設依此所得 $A_1\mathbf{x} = \mathbf{0}$ 與 $A_2\mathbf{x} = \mathbf{0}$ 的解分別為 $(x_1, \ldots, x_n) = (c_1, \ldots, c_n)$ 與 $(x_1, \ldots, x_n) = (d_1, \ldots, d_n)$. 注意, 若 $x_j$ 為 pivot variable, 其中 $j > k$, 則 $a_j = b_j = 0$, 所以此時 $x_j$ 的取值不會影響到 $x_k$ 的取值, 也就是說式子 (1.2) 帶入這兩組解後分別為

$$c_k + a_l = 0 \quad \text{與} \quad d_k + b_l = 0.$$

又由於依假設 $A_1, A_2$ 在 $x_k$ 為 pivot 這一個 row 以下的每一個 row 都一致, 我們知對所有 $k+1 \leq i \leq n$ 皆有 $c_i = d_i$. 然而這兩組解皆為 $A\mathbf{x} = \mathbf{0}$ 的解且 $x_k$ 為 pivot variable, 故由 Lemma 1.3.3 知 $c_k = d_k$. 可得 $a_l = -c_k = -d_k = b_l$. 此與 $a_l \neq b_l$ 的假設相矛盾, 故知 $A_1 = A_2$. □

利用化成 reduced echelon form 來解 linear system, 雖然步驟較多, 不過仍然有它的好處. 例如因為化成 echelon form 並不唯一, 所以有可能同一組 linear system 因化成 echelon form 求解, 寫下來的解集合的元素的表現 “形式” 會不同 (只是形式不同, 解集合是一樣的). 若化成 reduced echelon form 就不會這樣了, 因為它是唯一的, 所以大家寫下來的解集合的元素表示的 “形式” 是一致的. 另外若要判斷兩個矩陣是否可以用一些 elementary row operations 將其中一個換成另一個, 將這兩個矩陣化成 reduced echelon form 就可以了. 若它們化成 reduced echelon form 是一致的, 那當然表示它們是可以用一些 elementary row operations 將其中一個換成另一個, 而若不一致, 則由唯一性可知它們不可能用一些 elementary row operations 將其中一個換成另一個.

**Question 1.6.** 假設矩陣 $A, B$ 分別可利用 *elementary row operations* 化成 *reduced echelon form* $A', B'$ 試說明 $A$ 可利用 *elementary row operations* 化成 $B$ 若且唯若 $A' = B'$.

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 variables 和 free variables 對聯立方程組是否有解以及解是否唯一有著重要的關連. 本章中有關聯立方程組的理論對後面線性代數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.

# Matrix

在第一章我們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.

## 2.1. 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 $M_{m\times n}$ 來表示所有係數在 $\mathbb{R}$ 的 $m\times n$ 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. 例如考慮

$$A=\begin{bmatrix}1&0&2&3\\0&1&5&8\\2&1&1&0\end{bmatrix},\tag{2.1}$$

則 $A$ 為一個 $3\times 4$ matrix, 即 $A\in M_{3\times 4}$. 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 (2.1) 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 (2.1) 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 vectors 和 column vectors 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.

**Definition 2.1.1.** 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 $1\leq i\leq m$ 以及 $1\leq j\leq n$ 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 $M_{m\times n}$ 的矩陣才談是否相等, 且兩個 $M_{m\times n}$ 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 $M_{m\times n}$ 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 2.1.2.** 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 $1\leq i\leq m$ 以及 $1\leq j\leq n$ 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 $1\leq i\leq m$ 以及 $1\leq j\leq n$ 皆有 $d_{ij}=ra_{ij}$.

Definition 2.1.2 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

$$rA=\begin{bmatrix}ra_{11}&ra_{12}&\cdots&ra_{1n}\\ra_{21}&ra_{22}&\cdots&ra_{2n}\\\vdots&\vdots&\ddots&\vdots\\ra_{m1}&ra_{m2}&\cdots&ra_{mn}\end{bmatrix}$$

矩陣的加法與係數積的定義可以說是由大家熟悉的 $\mathbb{R}^2,\mathbb{R}^3$ 的向量（甚至 $\mathbb{R}^n$）的加法與係數積延伸而來, 我們有以下這些性質.

**Proposition 2.1.3.** 對於 $M_{m\times n}$ 上的矩陣, 我們有以下的性質:

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

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

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

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

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

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

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

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

Proposition 2.1.3 的證明用到實數 $\mathbb{R}$ 相對應的性質. 例如 (1) 談的是矩陣加法的交換性質, 用到的是 $\mathbb{R}$ 的加法交換性. 事實上若令 $A=[a_{ij}],B=[b_{ij}]$ 以及 $A+B=[c_{ij}],B+A=[d_{ij}]$. 依定義, 我們有 $c_{ij}=a_{ij}+b_{ij}$ 以及 $d_{ij}=b_{ij}+a_{ij}$. 由於實數的加法交換性 $a_{ij}+b_{ij}=b_{ij}+a_{ij}$, 因此得 $c_{ij}=d_{ij}$, 故依矩陣相等的定義，得證 $A+B=B+A$. 其他各項都可用一樣的方法證明, 這裡就不證明了. 不過 (3), (4) 談的是存在性的問題. 這類問題就必須說明存在的是什麼. 例如 $O$ 就是零矩陣，也就是說 $O=[c_{ij}]$ 其中 $c_{ij}=0$, $\forall i\,j$. 而 (4) 的 $A'$ 稱為 $A$ 的加法反元素, 它是跟著 $A$ 而變的. 如果 $A=[a_{ij}]$ 則令 $A'=[a'_{ij}]$ 其中 $a'_{ij}=-a_{ij}$, 就會滿足 $A+A'=O$ 了. 一般來說, 給定 $A$ 我們會將 $A$ 的加法反元素用 $-A$ 來表示.

接著我們定義矩陣間的乘法. 首先回顧當我們要解聯立方程組

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

我們會把它寫成

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

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 2.1.4.** 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

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

注意依此定義, 必需 $A\in M_{m\times n}$ 的 column 的個數 $n$ 等於 $\mathbf{b}\in M_{n\times 1}$ 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

$$A\mathbf{b}=b_1\begin{bmatrix} a_{11}\\ a_{21}\\ \vdots\\ a_{m1}\end{bmatrix}+b_2\begin{bmatrix} a_{12}\\ a_{22}\\ \vdots\\ a_{m2}\end{bmatrix}+\cdots+b_n\begin{bmatrix} a_{1n}\\ a_{2n}\\ \vdots\\ a_{mn}\end{bmatrix}=\begin{bmatrix} b_1a_{11}+b_2a_{12}+\cdots+b_na_{1n}\\ b_1a_{21}+b_2a_{22}+\cdots+b_na_{2n}\\ \vdots\\ b_1a_{m1}+b_2a_{m2}+\cdots+b_na_{mn}\end{bmatrix}. \tag{2.2}$$

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 2.1.4 的矩陣乘法定義

$$\mathbf{a}\,\mathbf{b}=\begin{bmatrix} a_1 & a_2 & \cdots & a_n\end{bmatrix}\begin{bmatrix} b_1\\ b_2\\ \vdots\\ b_n\end{bmatrix}=b_1a_1+b_2a_2+\cdots+b_na_n. \tag{2.3}$$

(注意, 若 $\mathbf{a},\mathbf{b}$ 是 $\mathbb{R}^n$ 中的向量, 則 $\mathbf{ab}$ 就是我們熟悉 $\mathbf{a},\mathbf{b}$ 的內積 $\langle\mathbf{a},\mathbf{b}\rangle$.) 依此看法, 由式子 (2.2), 我們可將 $A\mathbf{b}$ 寫成

$$A\mathbf{b}=\begin{bmatrix} {}_1\mathbf{a}\,\mathbf{b}\\ {}_2\mathbf{a}\,\mathbf{b}\\ \vdots\\ {}_m\mathbf{a}\,\mathbf{b}\end{bmatrix}. \tag{2.4}$$

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.

我們來看一個 $m\times n$ matrix 以及 $\mathbb{R}^n$ 的 column vector 在此乘法的定義之下的基本性質.

**Lemma 2.1.5.** 假設 $A=[a_{ij}]$, $A'=[a'_{ij}]$ 為 $m\times n$ *matrices* 以及 $\mathbf{b}=[b_j]$, $\mathbf{b}'=[b'_j]$ 為 $\mathbb{R}^n$ 中的 *column vectors (*即 $n\times 1$ *matrices)* 以及 $c\in\mathbb{R}$. 我們有以下的性質.

(1) $A(\mathbf{b}+\mathbf{b}')=A\mathbf{b}+A\mathbf{b}'$.

(2) $A(c\mathbf{b})=c(A\mathbf{b})=(cA)\mathbf{b}$.

(3) $(A+A')\mathbf{b}=A\mathbf{b}+A'\mathbf{b}$.

**Proof.** 令 $A$ 的 column vectors 依次為 $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 且 $A'$ 的 column vectors 依次為 $\mathbf{a}'_1,\ldots,\mathbf{a}'_n$.

(1) 依定義

$$A(\mathbf{b}+\mathbf{b}')=\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} b_1+b_1' \\ \vdots \\ b_n+b_n' \end{bmatrix}=(b_1+b_1')\mathbf{a}_1+\cdots+(b_n+b_n')\mathbf{a}_n. \tag{2.5}$$

而 $A\mathbf{b}+A\mathbf{b}'$ 為

$$\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}+\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} b_1' \\ \vdots \\ b_n' \end{bmatrix}=(b_1\mathbf{a}_1+\cdots+b_n\mathbf{a}_n)+(b_1'\mathbf{a}_1+\cdots+b_n'\mathbf{a}_n). \tag{2.6}$$

由矩陣加法和係數積的分配律 (Proposition 2.1.3 的性質 (7),(8)), 我們得證式子 (2.5) 和式子 (2.6) 相等.

(2) 依定義 $c(A\mathbf{b})$ 為

$$c\left(\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}\right)=c(b_1\mathbf{a}_1+b_2\mathbf{a}_2+\cdots+b_n\mathbf{a}_n). \tag{2.7}$$

而 $cA$ 的 column vectors 依次為 $c\mathbf{a}_1,\dots,c\mathbf{a}_n$. 故 $(cA)\mathbf{b}$ 為

$$\begin{bmatrix} | & & | \\ c\mathbf{a}_1 & \cdots & c\mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}=b_1(c\mathbf{a}_1)+\cdots+b_n(c\mathbf{a}_n). \tag{2.8}$$

最後依定義 $c\mathbf{b}$ 為 $\begin{bmatrix} cb_1 \\ \vdots \\ cb_n \end{bmatrix}$, 故 $A(c\mathbf{b})$ 為

$$\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} cb_1 \\ \vdots \\ cb_n \end{bmatrix}=(cb_1)\mathbf{a}_1+\cdots+(cb_n)\mathbf{a}_n. \tag{2.9}$$

由矩陣加法和係數積的結合律 (Proposition 2.1.3 的性質 (6)), 我們得證 (2.7), (2.8), (2.9) 三個式子皆相等.

(3) 依定義 $A+A'$ 的 column vectors 依次為 $\mathbf{a}_1+\mathbf{a}_1',\dots,\mathbf{a}_n+\mathbf{a}_n'$, 所以

$$(A+A')\mathbf{b}=\begin{bmatrix} | & & | \\ \mathbf{a}_1+\mathbf{a}_1' & \cdots & \mathbf{a}_n+\mathbf{a}_n' \\ | & & | \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}=b_1(\mathbf{a}_1+\mathbf{a}_1')+\cdots+b_n(\mathbf{a}_n+\mathbf{a}_n'). \tag{2.10}$$

另一方面, $A\mathbf{b}+A'\mathbf{b}$ 為

$$\begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}+\begin{bmatrix} | & & | \\ \mathbf{a}_1' & \cdots & \mathbf{a}_n' \\ | & & | \end{bmatrix}\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}=(b_1\mathbf{a}_1+\cdots+b_n\mathbf{a}_n)+(b_1\mathbf{a}_1'+\cdots+b_n\mathbf{a}_n'). \tag{2.11}$$

再次由矩陣加法和係數積的分配律, 我們得證式子 (2.10) 和式子 (2.11) 相等. □

Lemma 2.1.5 (1),(2) 告訴我們矩陣對向量的乘法有類似分配律的性質, 這個性質很重要 (以後我們會再提到並稱之為 *linear* 的性質), 我們特別用以下定理表示.

**Proposition 2.1.6.** 假設 $A \in M_{m\times n}$, 且 $\mathbf{b},\mathbf{b}'$ 為 $\mathbb{R}^n$ 中的 *column vectors,* 以及 $c,c' \in \mathbb{R}$. 則

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

**Proof.** 因 $c\mathbf{b},c\mathbf{b}'$ 皆為 $\mathbb{R}^n$ 中的 column vectors, 由 Lemma 2.1.5 (1) 知 $A(c\mathbf{b}+c'\mathbf{b}') = A(c\mathbf{b})+A(c'\mathbf{b}')$. 再由 Lemma 2.1.5 (2) 知 $A(c\mathbf{b}) = c(A\mathbf{b}), A(c'\mathbf{b}') = c'(A\mathbf{b}')$, 故得證本定理. □

**Question 2.1.** 假設 $A \in M_{m\times n}$, 且 $\mathbf{b}_1,\dots,\mathbf{b}_k$ 為 $\mathbb{R}^n$ 中的 *column vectors,* $c_1,\dots,c_k \in \mathbb{R}$. 試利用數學歸納法證明

$$A\big(\sum_{i=1}^{k} c_i\mathbf{b}_i\big) = \sum_{i=1}^{k} c_i(A\mathbf{b}_i).$$

**Question 2.2.** 假設 $A,A' \in M_{m\times n}$, 且 $\mathbf{b}$ 為 $\mathbb{R}^n$ 中的 *column vectors,* 以及 $c,c' \in \mathbb{R}$. 是否 $(cA+c'A')\mathbf{b} = c(A\mathbf{b})+c'(A'\mathbf{b})$ ?

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 $\mathbf{b}_k \in M_{n\times 1}$, $1 \le k \le l$, 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 2.1.7.** 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 $1 \le k \le l$, $C$ 的 $k$-th column $\mathbf{c}_k$ 為

$$\mathbf{c}_k = A\mathbf{b}_k = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{nk} \end{bmatrix} = b_{1k}\mathbf{a}_1 + b_{2k}\mathbf{a}_2 + \cdots + b_{nk}\mathbf{a}_n. \tag{2.12}$$

由此定義, 我們知對於 $1 \le i \le m$, $1 \le k \le l$, $AB$ 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 (2.4) 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

$$c_{ik} = {}_i\mathbf{a}\,\mathbf{b}_k = a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}. \tag{2.13}$$

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 2.1.8.** 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

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

$\sharp$

大部分的書都會用式子 (2.13) 當成矩陣乘法的定義. 我們選用式子 (2.12) 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 (2.13) 利用 entry 來看方便多了. 例如我們有以下的性質.

**Proposition 2.1.9.** 假設 $A,A'\in M_{m\times n}$, $B,B'\in M_{n\times l}$. 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 $B$ 的 column vectors 依次為 $\mathbf{b}_1,\ldots,\mathbf{b}_l$ 且 $B'$ 的 column vectors 依次為 $\mathbf{b}'_1,\ldots,\mathbf{b}'_l$.

(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 $A(\mathbf{b}_k+\mathbf{b}'_k)$. 另一方面, $AB+AB'$ 的 $k$-th column 為 $AB$ 的 $k$-th column $A\mathbf{b}_k$ 加上 $AB'$ 的 $k$-th column $A\mathbf{b}'_k$, 因此 $AB+AB'$ 的 $k$-th column 為 $A\mathbf{b}_k+A\mathbf{b}'_k$. 由 Lemma 2.1.5 (1), 我們得證它們相等.

(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, 即 $(A+A')\mathbf{b}_k$. 另一方面, $AB+A'B$ 的 $k$-th column 為 $AB$ 的 $k$-th column $A\mathbf{b}_k$ 加上 $A'B$ 的 $k$-th column $A'\mathbf{b}_k$, 因此 $AB+A'B$ 的 $k$-th column 為 $A\mathbf{b}_k+A'\mathbf{b}_k$. 因此由 Lemma 2.1.5 (3), 我們得證它們相等. $\square$

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

**Proposition 2.1.10.** 設 $c\in\mathbb{R}$, $A\in M_{m\times n},B\in M_{n\times l}$. 則

$$c(AB)=(cA)B=A(cB).$$

**Proof.** 假設 $B$ 的 column vectors 依次為 $\mathbf{b}_1,\ldots,\mathbf{b}_l$. 首先注意 $c(AB)$, $(cA)B$ 和 $A(cB)$ 皆為 $m\times l$ matrix, 我們僅要證明當 $1\leq k\leq l$ 時, $c(AB)$, $(cA)B$ 和 $A(cB)$ 的 $k$-th column 皆相等.

$c(AB)$ 的 $k$-th column 為 $c$ 乘上 $AB$ 的 $k$-th column, 故為 $c(A\mathbf{b}_k)$. 而 $(cA)B$ 的 $k$-th column 為 $cA$ 右邊乘上 $B$ 的 $k$-th column, 故為 $(cA)\mathbf{b}_k$. 最後由於 $cB$ 的 $k$-th column 為 $c\mathbf{b}_k$, 故 $A(cB)$ 的 $k$-th column 為 $A(c\mathbf{b}_k)$. 因此由 Lemma 2.1.5 (2), 我們得證它們皆相等. □

**Question 2.3.** 假設 $A,A' \in M_{m\times n}$, $B,B' \in M_{n\times l}$ 以及 $c,c' \in \mathbb{R}$. 可證明 $(cA+c'A)B = cAB + c'A'B$ 以及 $A(cB+c'B') = cAB + c'AB'$ 嗎？

由 Proposition 2.1.9 和 Proposition 2.1.10 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, 也可推得乘法具有結合律的性質 (即 $(AB)C = A(BC)$). 這裡要注意 $A$, $B$, $C$ 的階數必須要有限制 $(AB)C$ 和 $A(BC)$ 才會有意義.

**Proposition 2.1.11.** 假設 $A \in M_{m\times n}, B \in M_{n\times l}, C \in M_{l\times k}$, 則 $(AB)C = A(BC)$.

**Proof.** 依定義 $AB \in M_{m\times l}$, 故 $(AB)C \in M_{m\times k}$. 而 $BC \in M_{n\times k}$, 故 $A(BC) \in M_{m\times k}$ 與 $(AB)C$ 同階.

對於 $1 \le j \le k$, 我們要證明 $(AB)C$ 和 $A(BC)$ 的 $j$-th column 相等. 令 $\mathbf{c}_j$ 為 $C$ 的 $j$-th column 依定義 $(AB)C$ 的 $j$-th column 為 $(AB)\mathbf{c}_j$. 至於 $A(BC)$ 的 $j$-th column, 依定義為 $A$ 右邊乘上 $(BC)$ 的 $j$-th column (即 $B\mathbf{c}_j$). 所以我們僅要說明 $(AB)\mathbf{c}_j = A(B\mathbf{c}_j)$, 就可證得結合律.

由於 $\mathbf{c}_j$ 只有一個 column, 為了方便考量, 我們將 $\mathbf{c}_j$ 用單一足碼表達, 即令 $\mathbf{c}_j = \begin{bmatrix} c_1 \\ \vdots \\ c_l \end{bmatrix}$. 現對任意 $i = 1,\ldots,l$, 令 $AB$ 的 $i-th$ column 為 $\mathbf{p}_i$, 則

$$(AB)\mathbf{c}_j = \begin{bmatrix} | & & | \\ \mathbf{p}_1 & \cdots & \mathbf{p}_l \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_l \end{bmatrix} = c_1\mathbf{p}_1 + \cdots + c_l\mathbf{p}_l.$$

然而若 $\mathbf{b}_i$ 為 $B$ 的 $i$-th column, 依定義 $\mathbf{p}_i$ 為 $AB$ 的 $i$-th column 故得 $\mathbf{p}_i = A\mathbf{b}_i$. 因此我們得

$$(AB)\mathbf{c}_j = c_1(A\mathbf{b}_1) + \cdots + c_l(A\mathbf{b}_l).$$

另一方面

$$A(B\mathbf{c}_j) = A(\begin{bmatrix} | & & | \\ \mathbf{b}_1 & \cdots & \mathbf{b}_l \\ | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_l \end{bmatrix}) = A(c_1\mathbf{b}_1 + \cdots + c_l\mathbf{b}_l).$$

注意這裡我們將 $\mathbf{b}_i$ 視為 $\mathbb{R}^n$ 中的 column vector, 故套用 Proposition 2.1.6 (或 Question 2.1) 可得 $A(c_1\mathbf{b}_1 + \cdots + c_l\mathbf{b}_l) = c_1(A\mathbf{b}_1) + \cdots + c_l(A\mathbf{b}_l)$. 所以得證 $(AB)\mathbf{c}_j = A(B\mathbf{c}_j)$. □

有了矩陣乘法的結合律 (Proposition 2.1.11), 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.

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 $A \in M_{2\times 3}$, $B \in M_{3\times 4}$ 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 $A \in M_{2\times 3}, B \in M_{3\times 2}$ 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 2.1.9 和 Proposition 2.1.10 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 $A \in M_{n\times n}$, 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 $\mathbf{e}_i$, 即 $\mathbb{R}^n$ 的 column vector, 其中 $i$-th entry 為 1, 其他 entry 為 0. 事實上 $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ 就是我們熟悉的 $\mathbb{R}^n$ 的 standard basis (標準基底). 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 $A \in M_{m\times n}, B \in M_{n\times l}$ 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 2.4.** 假設 $A \in M_{n\times n}$, 是否 $(A - 2I_n)^2 = A^2 - 4A + 4I_n$ 為對?

**Question 2.5.** 試證明 $I_n$ 是唯一的 $n\times n$ 滿足對任意 $A \in M_{m\times n}$ 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 $A \in M_{m\times n}, B \in M_{n\times l}$ 我們很容易驗證 $rA = A(rI_n), rB = (rI_n)B$.

**Question 2.6.** 試利用 *Proposition 2.1.10* 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 $2\times 2$ 相乘可交換.

## 2.2. Transpose Operation

這一節中我們將介紹矩陣取 transpose (即轉置矩陣) 的概念, 即其相關性質. 最後利用它來探討如何從 row 的角度來看矩陣相乘.

對於一個 $m\times n$ matrix, 簡單來說其轉置矩陣就是將此矩陣的 row 與 column 的腳色互換, 也就是說將 row vectors 依序換成 column vectors. 我們有以下的定義.

**Definition 2.2.1.** 給定 $A\in M_{m\times n}$. 定義 $A^{\mathrm{t}}\in M_{n\times m}$, 其中對於 $1\leq i\leq m$, $A^{\mathrm{t}}$ 的 $i$-th column 就是將 $A$ 的 $i$-th row 寫成 column vector. 我們稱 $A^{\mathrm{t}}$ 為 $A$ 的 *transpose*.

**Example 2.2.2.** 令
$$A=\begin{bmatrix}1&2&3\\-1&-2&-3\end{bmatrix},$$
依定義 $A^{\mathrm{t}}$ 應為 $3\times 2$ matrix. 其中 $A^{\mathrm{t}}$ 的第一個 column 為 $A$ 的第一個 row $\begin{bmatrix}1&2&3\end{bmatrix}$ 寫成 column vector, 即 $\begin{bmatrix}1\\2\\3\end{bmatrix}$. 同理 $A^{\mathrm{t}}$ 的第二個 column 為 $A$ 的第二個 row $\begin{bmatrix}-1&-2&-3\end{bmatrix}$ 寫成 column vector, 即 $\begin{bmatrix}-1\\-2\\-3\end{bmatrix}$. 故得
$$A^{\mathrm{t}}=\begin{bmatrix}1&-1\\2&-2\\3&-3\end{bmatrix}.$$
注意, $A^{\mathrm{t}}$ 的 1-st, 2-nd 和 3-rd row 也恰為 $A$ 的 1-st, 2-nd 和 3-rd column 寫成 row 而得.

由上面的例子我們看到, 當 $A\in M_{m\times n}$, 對於 $1\leq j\leq n$, $A^{\mathrm{t}}$ 的 $j$-th row 就是將 $A$ 的 $j$-th column 寫成 row vector. 事實上若將 $A$ 寫成 $A=[a_{ij}]$. 對於 $1\leq i\leq m$, $A^{\mathrm{t}}$ 的 $i$-th column 就是將 $A$ 的 $i$-th row $\begin{bmatrix}a_{i1}&\cdots&a_{in}\end{bmatrix}$ 寫成 column vector $\begin{bmatrix}a_{i1}\\\vdots\\a_{im}\end{bmatrix}$. 因此 $A^{\mathrm{t}}$ 的 $(1,i)$-th entry 就是 $A$ 的 $(i,1)$-th entry $a_{i1}$, 而 $A^{\mathrm{t}}$ 的 $(2,i)$-th entry 就是 $A$ 的 $(i,2)$-th entry $a_{i1}$. 依此類推我們可以得到對於 $1\leq j\leq n$, $A^{\mathrm{t}}$ 的 $(j,i)$-th entry 就是 $A$ 的 $(i,j)$-th entry $a_{ij}$. 也就是說若我們將 $A^{\mathrm{t}}$ 寫成 $A^{\mathrm{t}}=[a'_{sk}]$, 則 $1\leq s\leq n$, $1\leq k\leq m$, 且 $a'_{sk}=a_{ks}$. 因此對於 $1\leq j\leq n$, $A^{\mathrm{t}}$ 的 $j$-th row $\begin{bmatrix}a'_{j1}&\cdots&a'_{jm}\end{bmatrix}$ 即為 $\begin{bmatrix}a_{1j}&\cdots&a_{mj}\end{bmatrix}$, 恰為 $A$ 的 $j$-th column 寫成 row vector. 我們將以上的討論寫成以下的結論.

**Lemma 2.2.3.** 假設 $A=[a_{ij}]\in M_{m\times n}$ 且 $A^{\mathrm{t}}=[a'_{sk}]\in M_{n\times m}$. 則對於 $1\leq s\leq n$, $1\leq k\leq m$, $a'_{sk}=a_{ks}$ 且 $A^{\mathrm{t}}$ 的 *$s$-th row* 就是將 $A$ 的 *$s$-th column* 寫成 *row vector.*

由 Lemma 2.2.3, 以後要談論 $A$ 和 $A^{\mathrm{t}}$ 間的關係, 我們可以用 row 換成 column, column 換成 row 以及 $(i,j)$-th entry 換成 $(j,i)$-th entry 三種看法處理. 現在我們來看矩陣取 transpose 的基本性質.

**Proposition 2.2.4.** 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 $A^{\mathrm{t}}$ 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 $C^{\mathrm{t}}$ 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 依定義知就是 $A^{\mathrm{t}}$ 的 $i$-th row 寫成 column vector, 而 $A^{\mathrm{t}}$ 的 $i$-th row 依 Lemma 2.2.3 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 $A^{\mathrm{t}}$ 和 $B^{\mathrm{t}}$ 的 $i$-th column 之和. 依 transpose 定義知它就是 $A$ 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. 我們將這些矩陣分別用 $A = [a_{ij}]$, $A^{\mathrm{t}} = [a'_{ji}]$, $C = [c_{sk}]$, $C^{\mathrm{t}} = [c'_{ks}]$ 表示. 對於 $1 \leq t \leq l$, $1 \leq i \leq m$, $(AC)^{\mathrm{t}}$ 的 $(k,i)$-th entry 為 $AC$ 的 $(i,k)$-th entry, 由式子 (2.13) 知應為

$$a_{i1}c_{1k} + a_{i2}c_{2k} + \cdots + a_{in}c_{nk}.$$

另一方面, $C^{\mathrm{t}}A^{\mathrm{t}}$ 的 $(k,i)$-th entry 為

$$c'_{k1}a'_{1i} + c'_{k2}a'_{2i} \cdots + c'_{kn}a'_{ni}.$$

利用 Lemma 2.2.3 知此即為

$$c_{1k}a_{i1} + c_{2k}a_{i2} \cdots + c_{nk}a_{in}.$$

故得證 $(AC)^{\mathrm{t}} = C^{\mathrm{t}}A^{\mathrm{t}}$. □

**Question 2.7.** 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 2.2.5.** 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 2.2.4, 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 $BB^{\mathrm{t}}$ 為 symmetric matrix. 同理可得 $B^{\mathrm{t}}B$ 亦為 symmetric matrix. □

利用 Proposition 2.2.4, 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 $A\in M_{1\times m}, B\in M_{m\times n}$, 令

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

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

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

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 2.2.4 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

$$\begin{bmatrix} a_1 & \cdots & a_m \end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ldots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix}=a_1\begin{bmatrix} b_{11} & \cdots & b_{1n} \end{bmatrix}+\cdots+a_m\begin{bmatrix} b_{m1} & \cdots & b_{mn} \end{bmatrix} \tag{2.14}$$

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 $B^{\mathrm{t}}$ 右邊乘上 $A^{\mathrm{t}}$ 的 $i$-th column. 然而 $A^{\mathrm{t}}$ 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 2.2.4 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

$$AB=\begin{bmatrix} - & {}_1\mathbf{a} & - \\ - & {}_2\mathbf{a} & - \\ & \vdots & \\ - & {}_m\mathbf{a} & - \end{bmatrix}B=\begin{bmatrix} - & {}_1\mathbf{a}B & - \\ - & {}_2\mathbf{a}B & - \\ & \vdots & \\ - & {}_m\mathbf{a}B & - \end{bmatrix}. \tag{2.15}$$

結合式子 (2.14), 我們有以下之結果.

**Proposition 2.2.6.** 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

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

## 2.3. Elementary Matrix

前面提過, 我們將聯立方乘式用矩陣 $A\mathbf{x}=\mathbf{b}$ 來表示, 是想利用矩陣的乘法來處理聯立方程式. 事實上 elementary row operation 已可以看成是矩陣的乘法運算. 首先考慮 $n\times n$ 的單位矩陣 $I_n$ (即 $I_n$ 的對角線位置皆為 1, 其他位置為 0). 若用 $i$-th row 和 $j$-th row 交換的 type 1 elementary row operation 將 $I_n$ 轉換成矩陣 $E_1$, 可得

$$E_1=\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}. \tag{2.16}$$

同樣的若使用 type 2 elementary row operation 將 $I_n$ 的 $i$-th row 乘上非零實數 $r$ 轉換成矩陣 $E_2$, 可得

$$E_2=\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & r & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}. \tag{2.17}$$

最後若使用 type 3 elementary row operation 將 $I_n$ 的 $i$-th row 乘上實數 $r$ 加到 $I_m$ 的 $j$-th row 所得的矩陣為 $E_3$, 可得

$$E_3=\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & r & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}. \tag{2.18}$$

這樣的矩陣我們稱之為 *elementary matrix*. 而我們分別稱 $E_1,E_2,E_3$ 為 *type 1*, *type 2* 以及 *type 3* 的 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. 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 2.2.6, $I_mA$ 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 type 1 elementary row operation 將 $I_m$ 轉換成矩陣 $E$, 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 type 1 elementary row operation 變換所得的矩陣.

同樣的若將 $I_m$ 的 $i$-th row 乘上非零實數 $r$ 所得的 type 2 elementary matrix 為 $E$, 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 type 2 elementary row operation 變換所得的矩陣.

最後若將 $I_m$ 的 $i$-th row 乘上實數 $r$ 加到 $I_m$ 的 $j$-th row 所得的 type 3 elementary matrix 為 $E$, 則因 $E$ 的 $j$-th row 的 $i$-th entry 為 $r$, $j$-th entry 為 1. 故由 Proposition 2.2.6, $EA$ 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 type 3 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相對應的 elementary matrix. (2.16), (2.17), (2.18) 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 2.8.** 試找出那些同階的 *elementary matrices* 其相乘是不可以交換的.

**Example 2.3.1.** 考慮 Example 1.1.1 的情形. $A$ 是 $3 \times 4$ matrix 且 $B$ 是將 $A$ 的第一個 row 和第二個 row 交換所得. 考慮將 $I_3$ 的第一個 row 和第二個 row 交換所得的 elementary

matrix
$$E_1=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}.$$
可得
$$E_1A=\begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&2&3&4\\2&1&-1&3\\4&0&1&2\end{bmatrix}=\begin{bmatrix}2&1&-1&3\\1&2&3&4\\4&0&1&2\end{bmatrix}=B.$$
同樣的 $C$ 是由 $B$ 的第二個 row 乘上 2 所得, 所以我們考慮將 $I_3$ 的第二個 row 乘上 2 所得的 elementary matrix
$$E_2=\begin{bmatrix}1&0&0\\0&2&0\\0&0&1\end{bmatrix}.$$
可得
$$E_2B=\begin{bmatrix}1&0&0\\0&2&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&1&-1&3\\1&2&3&4\\4&0&1&2\end{bmatrix}=\begin{bmatrix}2&1&-1&3\\2&4&6&8\\4&0&1&2\end{bmatrix}=C.$$
最後 $D$ 是將 $C$ 的第三個 row 乘上 $-3$ 加到第一個 row, 所以我們考慮將 $I_3$ 的第三個 row 乘上 $-3$ 加到第一個 row 所得的 elementary matrix
$$E_3=\begin{bmatrix}1&0&-3\\0&1&0\\0&0&1\end{bmatrix}.$$
可得
$$E_3C=\begin{bmatrix}1&0&-3\\0&1&0\\0&0&1\end{bmatrix}\begin{bmatrix}2&1&-1&3\\2&4&6&8\\4&0&1&2\end{bmatrix}=\begin{bmatrix}-10&1&-4&-3\\2&4&6&8\\4&0&1&2\end{bmatrix}=D.$$

上面所提這種將一個矩陣做 elementary row operation 可視為將此矩陣左邊乘上其對應的 elementary matrix 的看法, 將來對我們探討矩陣的性質是很有幫助的. 這種看法的互換, 也時能讓我們得到有趣的結果. 例如前面提過一個矩陣經由一個 elementary row operation 轉變成另一個矩陣後, 我們可以再用相同 type 的 elementary row operation 將其轉換回原來的矩陣. 這個事實用 elementary matrices 的角度來看, 可以有以下的看法:

(1) 設 $E_1$ 是將 $I_m$ 的 $i$-th row 和 $j$-th row 互換的 type 1 elementary matrix. 我們將 $E_1$ 的 $i$-th row 和 $j$-th row 再互換就可轉換回 identity matrix $I_m$. 所以我們有 $E_1E_1=I_m$.

(2) 設 $E_2$ 是將 $I_m$ 的 $i$-th row 乘上非零實數 $r$ 的 type 2 elementary matrix. 我們將 $E_2$ 的 $i$-th row 乘上 $r^{-1}$ 就可轉換回 $I_m$. 所以若令 $E_2'$ 為將 $I_m$ 的 $i$-th row 乘上 $r^{-1}$ 的 type 2 elementary matrix, 我們有 $E_2'E_2=I_m$. 同理可得 $E_2E_2'=I_m$.

(3) 設 $E_3$ 是將 $I_m$ 的 $i$-th row 乘上實數 $r$ 加到 $j$-th row 所得的矩陣的 type 3 elementary matrix. 我們將 $E_3$ 的 $i$-th row 乘上 $-r$ 再加到 $j$-th row 就可轉換回 $I_m$. 所以若令 $E_3'$ 為將 $I_m$ 的 $i$-th row 乘上 $-r$ 的 type 3 elementary matrix, 我們有 $E_3'E_3=I_m$. 同理可得 $E_3E_3'=I_m$.

我們知道當一個 $m\times m$ 的矩陣 $A$ 若可找到矩陣 $B$ 使得 $BA=AB=I_m$, 則稱 $A$ 為一個 *invertible matrix* (可逆矩陣), 且 $B$ 為 $A$ 的 *inverse* (反矩陣). 從上面的探討我們有以下之結論.

**Proposition 2.3.2.** *假設 $E$ 是一個 elementary matrix, 則 $E$ 為 invertible 且 $E$ 的 inverse 是和 $E$ 相同 type 的 elementary matrix.*

既然有所謂的 elementary row operations 當然也會有 elementary column operations. 它的概念只是將 row operation 對 row 的動作改為對 column 的動作. 我們將一個矩陣的 $i$-th column 和 $j$-th column 對調, 這一個動作及稱為 type 1 的 elementary column operation. 若將矩陣的 $i$-th column 上的數皆乘上非零實數 $r$, 則稱 type 2 的 elementary column operation. 至於 type 3 的 elementary column operation 就是把矩陣的 $i$-th column 乘上 $r$ 後加到其 $j$-th column. 由於 column operations 並未用在解聯立方乘組的問題, 所以這裡我們僅約略介紹其相關的概念, 不再像前面依樣詳述. 事實上 column operations 的概念和 row operations 是相呼應的, 大家可以用前面探討的方式檢驗.

將 identity matrix $I_m$ 做 elementary column operation 後會得到甚麼樣的矩陣呢? 結果也會是前面提到的 elementary matrix (這也是 elementary matrix 沒有區分 row 和 column 的原因). 例如將 $I_m$ 的 $i$-th column 和 $j$-th column 互換所得的矩陣就是將 $I_m$ 的 $i$-th row 和 $j$-th row 互換的 type 1 elementary matrix. 而將 $I_m$ 的 $i$-th column 乘上非零實數 $r$ 的矩陣, 就是將 $I_m$ 的 $i$-th row 乘上 $r$ 的 type 2 elementary matrix. 不過要注意, 將 $I_m$ 的 $i$-th column 乘上實數 $r$ 加到 $j$-column 所得的矩陣不是將 $I_m$ 的 $i$-th row 乘上實數 $r$ 加到 $j$-th row 所得的 elementary matrix, 而是將 $I_m$ 的 $j$-th row 乘上實數 $r$ 加到 $i$-th row 所得的矩陣的 type 3 elementary matrix. 這一部分請務必檢驗, 就能了解其中原因.

既然一個 elementary matrix 同時可對應到 elementary row operation 也可對應到 elementary column operation, 那要如何區分呢? 別忘了, 矩陣的乘法是沒有交換性的. 前面我們知道, 當一個 elementary matrix $E$ 乘在一個矩陣 $A$ 的左邊時, 所得的矩陣 $EA$ 會是對 $A$ 做 $E$ 所對應的 elementary row operation. 而若將 $E$ 乘在矩陣 $B$ 的右邊, 則所得的矩陣 $BE$ 會是對 $B$ 做 $E$ 所對應的 elementary column operation. 為了方便起見, 我們綜合成以下的結論.

**Theorem 2.3.3.** *假設 $A$ 是一個 $m\times n$ matrix. 若 $E$ 是對 $I_m$ 做 elementary row operation 所得的 elementary matrix, 則 $EA$ 就會是對 $A$ 作相對應的 elementary row operation 所得的矩陣. 若 $E'$ 是對 $I_n$ 做 elementary column operation 所得的 elementary matrix, 則 $AE'$ 就會是對 $A$ 作相對應的 elementary column operation 所得的矩陣.*

**Example 2.3.4.** 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

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

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

這裡我們再說明一下, 當 $A$ 是一個 $m \times n$ matrix, 因為 $A$ 有 $m$ 個 row, 所以乘在左邊的 elementary matrix (對應到 elementary row operation) 必須是一個 $m$ 階方陣. 同樣的, 因為 $A$ 有 $n$ 個 column, 所以乘在右邊的 elementary matrix (對應到 elementary column operation) 必須是一個 $n$ 階方陣.

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 $A$ 被轉換為 $E_1A$. 現若對 $[A|I_m]$ 做相同的 elementary row operation 的話, 所得的結果會是 $E_1[A|I_m]$. 然而此時原先 $A$ 的部分會變成 $E_1A$, 而 $I_m$ 的部分經同樣的 elementary row operation, 所以 $I_m$ 這部分會是 $E_1I_m$. 因此我們知

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

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 2.3.5.** *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 2.3.6.** 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 $3\times 4$ matrix, 所以第一個 elementary row operation 所對應的 elementary matrix $E_1$ 就是將 $3\times 3$ 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

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

**Question 2.9.** 假設 $A$ 為 $m\times n$ *matrix.* 若要記錄對 $A$ 所做的 *column operations* 所對應的 *elementary matrices* 的乘積，增廣矩陣應如何擺放？

## 2.4. Matrix 和 System of Linear Equations 的連結

我們曾經利用 elementary row operations 將增廣矩陣化為 echelon form 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.

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 $x_1 = c_1, x_2 = c_2, \ldots, x_n = c_n$, 為此聯立方程組的一組解, 我們便會用

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

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 $A\mathbf{c} = \mathbf{b}$.

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

**Lemma 2.4.1.** *假設 $A \in M_{m\times n}$ 且 $\mathbf{b} \in \mathbb{R}^m$. 則 $A\mathbf{x} = \mathbf{b}$ 有解若且唯若 $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$, 其中 $\mathbf{a}_1, \ldots, \mathbf{a}_n$ 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 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) \subseteq \mathbb{R}^m$. 然而由 Lemma 2.4.1 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 $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. 故知此時 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \mathbb{R}^m$. 反之, 若 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \mathbb{R}^m$, 表示對任意 $\mathbf{b} \in \mathbb{R}^m$ 皆有 $\mathbf{b} \in \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$. 同樣由 Lemma 2.4.1 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 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \mathbb{R}^m$ 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 $\mathbf{e}_1,\ldots,\mathbf{e}_m\in\mathbb{R}^m$ 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 $i=1,\ldots,m$, 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 $i=1,\ldots,m$ 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.

我們曾探討過, 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 1.2 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 2.4.2.** *假設 $A\in M_{m\times n}$, 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 2.4.2, 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 2.4.2 並不適用 (不過 Lemma 2.4.1 是適用的).

我們曾提及, 當 $A\in M_{m\times n}$, 將 $A$ 化為 echelon form 後其 pivot 的個數不可能多於 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 2.4.2 中的情況不可能發生. 我們有以下的結論.

**Corollary 2.4.3.** *假設 $A \in M_{m\times n}$, 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 2.4.2 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 2.4.2 知不會存在 $n \times m$ matrix $C$ 使得 $AC = I_m$. □

**Question 2.10.** *假設 $A \in M_{m\times n}$, 其中 $m < n$. 是否存在 $n \times m$ matrix $C$ 使得 $CA = I_n$?*

前面提過 Theorem 2.4.2 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 2.4.3 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.

**2.4.2. 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 $A \in M_{m\times n}$ 以及 $\mathbf{b} \in \mathbb{R}^m$. 如果 $A\mathbf{x} = \mathbf{b}$ 有解, 則 $A\mathbf{x} = \mathbf{b}$ 的解和 $A\mathbf{x} = \mathbf{0}$ 的解 (這裡 $\mathbf{0}$ 是 $\mathbb{R}^m$ 的零向量) 息息相關, 我們有以下之定理.

**Lemma 2.4.4.** *給定 $A \in M_{m\times n}$ 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 $A\mathbf{x} = \mathbf{0}$ 的一組解.*

(2) *若 $\mathbf{x} = \mathbf{u} \in \mathbb{R}^n$ 為 $A\mathbf{x} = \mathbf{0}$ 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

$$A(\mathbf{c}' - \mathbf{c}) = A\mathbf{c}' - A\mathbf{c} = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

因此 $\mathbf{x} = \mathbf{c}' - \mathbf{c}$ 會是 $A\mathbf{x} = \mathbf{0}$ 的一組解.

(2) 若 $\mathbf{x} = \mathbf{u} \in \mathbb{R}^n$ 為 $A\mathbf{x} = \mathbf{0}$ 的一組解, 則

$$A(\mathbf{c} + \mathbf{u}) = A\mathbf{c} + A\mathbf{u} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

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

Lemma 2.4.4 告訴我們若已知 $\mathbf{x} = \mathbf{c}$ 為 $A\mathbf{x} = \mathbf{b}$ 的一組解, 且知道 $A\mathbf{x} = \mathbf{0}$ 所有的解, 就能利用 $\mathbf{c}$ 以及 $A\mathbf{x} = \mathbf{0}$ 所有的解得到 $A\mathbf{x} = \mathbf{b}$ 所有的解. 所以了解 $A\mathbf{x} = \mathbf{0}$ 所有的解是很重要的課題 (以後我們會深入探討). 回顧一下 $A\mathbf{x} = \mathbf{0}$ 這樣的 linear system, 我們稱之為 homogeneous linear system. Homogeneous linear system 一定有解, 事實上當 $A \in M_{m\times n}$ 時, $x_1 = 0, \ldots, x_n = 0$ 就是 $A\mathbf{x} = \mathbf{0}$ 的一組解. 這組解 $\mathbf{x} = \mathbf{0} \in \mathbb{R}^n$ 因為不需任何計算就能得到, 我們稱之為 $A\mathbf{x} = \mathbf{0}$ 的 trivial solution. 注意 trivial solution $\mathbf{x} = \mathbf{0}$ 這裡的 $\mathbf{0}$ 是 $\mathbb{R}^n$ 的零向量, 而 $A\mathbf{x} = \mathbf{0}$ 這裡的 $\mathbf{0}$ 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 $A\mathbf{x} = \mathbf{0}$ 除了 trivial solution 外還有其他的 solution (即解不唯一), 我們稱這些不為 $\mathbf{0}$ 的 solution 為 nontrivial solution.

從 Lemma 2.4.4 我們知, 若 $A\mathbf{x}=\mathbf{0}$ 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 2.4.5.** 假設 $A\in M_{m\times n}$. 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* $A\mathbf{x}=\mathbf{0}$ 沒有 *nontrivial solution.*

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

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

**Proof.** (1) $\Rightarrow$ (2): 利用反證法, 假設 $\mathbf{x}=\mathbf{u}$ 為 $A\mathbf{x}=\mathbf{0}$ 的一組 nontrivial solution 而 $\mathbf{x}=\mathbf{c}$ 為 $A\mathbf{x}=\mathbf{b}$ 的一組解, 則由 Lemma 2.4.4 知 $\mathbf{x}=\mathbf{c}+\mathbf{u}\neq\mathbf{c}$ 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 $A\mathbf{x}=\mathbf{0}$ 沒有 nontrivial solution.

(2) $\Rightarrow$ (3): $A\mathbf{x}=\mathbf{0}$ 沒有 nontrivial solution, 表示 $A$ 化成 echelon form 後沒有 free variables. 也就是說所有的 variables 皆為 pivot variables. 因此 pivot 的個數就是未知數的個數 $n$, 故得 $\operatorname{rank}(A)=n$.

(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 $A'=\begin{bmatrix} I_n \\ \mathbf{0}\end{bmatrix}$, 即 $A'$ 的前 $n$ 個 row 就是 $I_n$. 由 Lemma 2.3.5, 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 (2.15)), 我們有

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

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 $B\in M_{n\times m}$ 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 2.4.5, 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 $A\mathbf{x} = \mathbf{0}$ 有 nontrivial solution, 則 $A\mathbf{x} = \mathbf{b}$ 要不然無解, 要不然會有無窮多解.

**Question 2.11.** *假設 $A \in M_{m\times n}$, $\mathbf{b}, \mathbf{b}' \in \mathbb{R}^m$. 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 $A \in M_{m\times n}$, 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 2.4.5 中的情況不可能發生. 我們有以下的結論.

**Corollary 2.4.6.** *假設 $A \in M_{m\times n}$, 其中 $m < n$. 若 $\mathbf{b} \in \mathbb{R}^m$ 且聯立方程組 $A\mathbf{x} = \mathbf{b}$ 有解, 則解不唯一 (即必有兩個以上的解). 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 2.4.5 知 homogeneous linear system $A\mathbf{x} = \mathbf{0}$ 有 nontrivial solution. 亦即在 $\mathbb{R}^n$ 存在非零向量 $\mathbf{c}$ 使得 $\mathbf{x} = \mathbf{c}$ 為 $A\mathbf{x} = \mathbf{0}$ 之一組解. 所以 Lemma 2.4.4 告訴我們, 若 $A\mathbf{x} = \mathbf{b}$ 有解, 則解不唯一.

另一方面 Theorem 2.4.5 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 2.4.5 也和 Theorem 2.4.2 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 2.4.6 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 $A\mathbf{x} = \mathbf{b}$ 不可能有唯一解.

## 2.5. 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 $B$ 滿足 $BA$ 以及 $AB$ 為 identity. 不過當 $A \in M_{m\times n}$ 且 $m \neq n$ 時, 由 Corollary 2.4.3 以及 Corollary 2.4.6, 我們知道不可能存在 $B$ 同時滿足 $BA$ 和 $AB$ 皆為 identity matrix (因為 $\operatorname{rank}(A)$ 不可能同時為 $m$ 和 $n$). 所以我們僅對 $m = n$, 即 $A$ 為 square matrix 時有以下的定義.

**Definition 2.5.1.** 假設 $A \in M_{n\times n}$ 為 $n$ 階 square matrix, 若存在 $B \in M_{n\times n}$ 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 2.5.2.** 假設 $A \in M_{n\times n}$ 為 $n$ 階 *square matrix.* 則下列是等價的.

(1) $A$ 為 *invertible matrix.*

(2) 存在 $B \in M_{n\times n}$ 使得 $BA = I_n$.

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

(4) 存在 $C \in M_{n\times n}$ 使得 $AC = I_n$.

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

由 $A$ 為 $n \times n$ matrix 以及 Theorem 2.4.5 知存在 $B \in M_{n\times n}$ 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 2.4.2 知存在 $C \in M_{n\times n}$ 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 $B, C \in M_{n\times n}$ 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. □

當一個 $n \times n$ matrix 的 rank 為 $n$ 時, 有的書為了強調這個 rank 和階數相等的特殊情況, 特別稱之為 *nonsingular matrix.* 所以由 Theorem 2.5.2 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. 以後我們一律採用 invertible 和 non-invertible 這樣的說法, 而不用 nonsingular 和 singular 這樣的說法.

由 Theorem 2.5.2 的證明我們知若 $A \in M_{n\times n}$ 且存在 $B, C \in M_{n\times n}$ 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 $B, B' \in M_{n\times n}$ 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 2.5.3.** 假設 $A \in M_{n\times n}$ 且 $B, B' \in M_{n\times n}$ 滿足 $BA = I_n$ 以及 $B'A = I_n$. 則 $B = B'$.

**Proof.** 由 Theorem 2.5.2 我們知 $A$ 為 invertible 且由其證明知 $BA = AB = I_n$ 以及 $B'A = AB' = I_n$. 故

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

□

由 Corollary 2.5.3, 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 2.5.4.** 假設 $A\in M_{n\times n}$ 為 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 2.5.5.** 假設 $A,B\in M_{n\times n}$. 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 $A^{\mathrm{t}}$ 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 2.5.2, 我們要說一個 $n\times n$ matrix 為 invertible, 只要找到 $B\in M_{n\times n}$ 使得 $BA=I_n$ 或 $AB=I_n$ 且此時由唯一性 (Corollary 2.5.3) 知 $B=A^{-1}$.

(1) 依定義 $A^{-1}$ 亦為 $n\times n$ matrix 故 Theorem 2.5.2 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 $A^{\mathrm{t}}$ 亦為 $n\times n$ matrix 故 Theorem 2.5.2 適用. 由 $A^{-1}A=I_n$ 利用 Proposition 2.2.4 得
$$I_n=(A^{-1}A)^{\mathrm{t}}=A^{\mathrm{t}}(A^{-1})^{\mathrm{t}}$$
故知 $A^{\mathrm{t}}$ 為 invertible 且 $(A^{\mathrm{t}})^{-1}=(A^{-1})^{\mathrm{t}}$. 反之若 $A^{\mathrm{t}}$ 為 invertible, 由前知 $(A^{\mathrm{t}})^{\mathrm{t}}$ 為 invertible, 故由利用 Proposition 2.2.4 $(A^{\mathrm{t}})^{\mathrm{t}}=A$ 得證 $A$ 為 invertible.

(3) 依定義 $AB$ 為 $n\times n$ matrix 故 Theorem 2.5.2 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 2.5.2 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 2.5.5 (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 2.4.2 中我們知道有可能 $A\in M_{m\times n},C\in M_{n\times m}$ 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 2.12.** 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 2.5.2 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 2.3.5 我們知存在 $E\in M_{n\times n}$ 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 2.5.6.** 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

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

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

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

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

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

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 $A \in M_{n\times n}$ 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 2.5.5 (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 2.5.7.** $A$ 為 *invertible matrix* 若且唯若 $A$ 為一些 *elementary matrices* 的乘積.

**Example 2.5.8.** 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

$$\left[\begin{array}{ccc|ccc} 0 & 2 & 0 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 1 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 1 \end{array}\right], E_1 = E_1^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

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

$$A^{-1} = E_4E_3E_2E_1 = \begin{bmatrix} \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, A = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1} = \begin{bmatrix} 0 & 2 & 0 \\ 1 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix}.$$

最後讓我們回到解聯立方程組的問題. 怎樣的 $A \in M_{m\times n}$ 會使得對任意 $\mathbf{b} \in \mathbb{R}^m$, 聯立方程組 $A\mathbf{x} = \mathbf{b}$ 皆有解且解唯一呢? 由 Theorem 2.4.2 和 Theorem 2.4.5 知此時 $\mathrm{rank}(A) = m$ 且 $\mathrm{rank}(A) = n$, 即 $m = n$. 也就是說 $A$ 必須是 $n \times n$ 且 $\mathrm{rank}(A) = n$. 因此由 Theorem 2.5.2 知 $A$ 為 $n \times n$ invertible matrix. 事實上我們有以下的等價關係. 由於它們直接套用 Theorem 2.4.2 和 Theorem 2.4.5 就可推得, 我們就不再證明了.

**Theorem 2.5.9.** 假設 $A \in M_{n\times n}$, 令 $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^n$ 為 $A$ 的 *column vectors.* 則下列是等價的.

(1) $A$ 為 *invertible matrix.*

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

(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 $A\mathbf{x} = \mathbf{0}$ 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 $A \in M_{n\times n}$ 為 invertible matrix, 則對任意 $\mathbf{b} \in \mathbb{R}^n$, 我們可以利用 $A$ 的反矩陣 $A^{-1}$ 得到聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的唯一解. 事實上若令 $\mathbf{x} = A^{-1}\mathbf{b}$, 此時 $A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = \mathbf{b}$. 又由 Theorem 2.5.9 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 2.5.10.** 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 2.5.6 中的 $4 \times 4$ matrix $A$ 且 $\mathbf{b} = [b_1\ b_2\ b_3\ b_4]^{\mathrm{t}}$. 因 $A$ 為 invertible, 故由 Theorem 2.5.9 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

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

*Chapter 3*

# Vector Spaces

在這一章中, 我們利用大家熟悉的坐標平面中的向量, 將之推廣到所謂的 vector space (向量空間) 這一種有特定代數結構的系統, 是線性代數中主要的探討對象.

## 3.1. 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.

在平面中的向量我們可以用幾何的方式規定向量的加法及其倍數關係. 相信大家對這種定法已相當熟悉, 在這裡我們不再重複. 我們可以將平面坐標化, 這就是所謂的坐標平面. 這種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 $a,b$ 皆為實數).

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 $b=d$; 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 3.1.1.** 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

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

這裡我們要強調, Definition 3.1.1 中所定義的加法及係數積, 和矩陣的加法及係數積是一致的. 基於符號的方便性, 當我們要用符號來表示一個向量時, 會用 $\mathbf{u},\mathbf{v}$ 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 $\mathbf{v}=(a,b)$.

一般來說有了定義之後, 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, 我們可已先依定義推導出一些性質, 利用這些性質簡化處理程序, 再處理更進一步的問題. 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 3.1.2.** 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 $\mathbf{0}\in\mathbb{R}^2$ 滿足對任意 $\mathbf{u}\in\mathbb{R}^2$ 皆有 $\mathbf{0}+\mathbf{u}=\mathbf{u}$.

(4) 對任意 $\mathbf{u}\in\mathbb{R}^2$ 皆可找到 $\mathbf{u}'\in\mathbb{R}^2$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{0}$.

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

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

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

通常一個定理敘述完就要證明, 不過這幾項的證明都僅是一般制式的代數操作, 相信大家都很熟悉, 這裡就不再證明了. 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. 在證明時會發現這個性質會成立主要是實數加法有交換性. 不過數學上是存在許多“抽象”的數系它的運算是不能交換的. 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. 因為向量的加法是定義兩個向量的加法, 所以兩個以上的向量相加結合律就顯得重要了. 有了結合律, 我們就不必擔心哪兩個向量先加. 結合律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 $\mathbf{u}+\mathbf{u}'=\mathbf{0}$. 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) 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 (6) 的性質兩邊乘上 $1/2$, 得

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

(6),(7),(8) 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 $r$, 而 $(rs)\mathbf{u}$ 是表示先將 $r,s$ 乘在一起得 $rs$ 再乘上 $\mathbf{u}$. 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.

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. 這些性質也讓坐標平面上向量的系統享有許多豐富的性質.

**Question 3.1.** *利用 $\mathbb{R}^2$ 向量加法的定義, 試證明以下性質:*

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

(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 $\mathbf{u}+\mathbf{u}'=\mathbf{0}$.

坐標平面上向量的運算也可推廣到坐標空間, 即 $\mathbb{R}^3$. 同樣的概念也可推廣到更一般的 $\mathbb{R}^n$, $n\in\mathbb{N}$. 這些系統中的運算都享有 Proposition 3.1.2 的 8 項性質, 而這些性質讓這些系統有著豐富的性質. 所以接下來我們將專注於有這 8 項性質的系統, 稱之為 vector space (向量空間).

## 3.2. Vector Space 的定義及其基本性質

我們曾經提過像 $\mathbb{R}^2$ 這樣, 裡面任意兩個向量相加仍在 $\mathbb{R}^2$ 中且向量乘上任意的實數後也仍在 $\mathbb{R}^2$, 而向量的運算又符合 Proposition 3.1.2 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 相關性質.

給定一非空集合 $V$, 我們說 $V$ 中有加法運算 (addition) $+$, 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). 至於係數積 (scaler multiplication) 我們要注意的是, 可以乘在向量上的數所在的數系必須像實數一樣有加法與乘法, 且加法, 乘法都是有交換律及結合律, 還有加法乘法之間要有分配律. 更重要的是這個數系裡非 0 的元素都有乘法反元素. 這樣的數系我們稱之為 *field* (體). 例如實數 $\mathbb{R}$ 和有理數 $\mathbb{Q}$ 在我們一般熟悉的加法, 乘法運算下都是 field, 但是整數 $\mathbb{Z}$ 就不是 field, 因為除了 $\pm 1$ 以外其他的非 0 整數在 $\mathbb{Z}$ 中就無法找到乘法反元素. 由於以後我們談的向量空間, 向量前所乘的係數所在的數系只要是 field, 則我們所要探討的性質都會成立. 所以係數積我們都不會強調是哪一個 field, 而用 $\mathbb{F}$ 來表示. 不過由於我們給的例子大多是係數積為 $\mathbb{R}$ 的情況, 所以若對 field 的概念覺得陌生, 不妨就用 $\mathbb{F}=\mathbb{R}$ 的情況來思考即可. 現若 $\mathbb{F}$ 是一個 field, 我們說 $\mathbb{F}$ 對 $V$ 有係數積表示對任意 $c\in\mathbb{F}$ 以及 $\mathbf{v}\in V$, 皆有 $c$ 對 $\mathbf{v}$ 所得係數積 $c\mathbf{v}$ 仍然在 $V$ 中 (此為係數積封閉性). 當一個集合 $V$ 上有加法運算, 且 field $\mathbb{F}$ 對其有係數積, 則我們可以探討其是否為 vector space, 也就是說探討它是否符合以下之定義.

**Definition 3.2.1.** 假設非空集合 $V$ 中有加法運算 $+$, 以及 field $\mathbb{F}$ 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 over* $\mathbb{F}$.

(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 $\mathbf{0}\in V$ 滿足對任意 $\mathbf{u}\in V$ 皆有 $\mathbf{0}+\mathbf{u}=\mathbf{u}$.

(4) 對任意 $\mathbf{u}\in V$ 皆可找到 $\mathbf{u}'\in V$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{0}$.

(5) 對任意 $\mathbf{u}\in V$, 皆有 $1\mathbf{u}=\mathbf{u}$.

(6) 對任意 $r,s\in\mathbb{F}$ 以及 $\mathbf{u}\in V$, 皆有 $r(s\mathbf{u})=(rs)\mathbf{u}$.

(7) 對任意 $r\in\mathbb{F}$ 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 $r,s\in\mathbb{F}$ 以及 $\mathbf{u}\in V$, 皆有 $(r+s)\mathbf{u}=r\mathbf{u}+s\mathbf{u}$.

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 over $\mathbb{F}$ 時就隱含其中有加法運算且直接用 $+$ 表示, 同時也隱含 $\mathbb{F}$ 是一個 field 且 $V$ 中有 $\mathbb{F}$ 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. 尤其要注意, 我們必須明確說是 over 哪一個 field 的 vector space (以後我們會看到例子, 同樣的集合看成 over 不同的 field 的 vector space 影響會很大).

接下來我們看一些有關 vector space 的例子.

**Example 3.2.2.** (A) 考慮 $S$ 為所有次數等於 2 的有理係數多項式所成的集合. 對 $S$ 中兩多項式 $f(x)=ax^2+bx+c$, $g(x)=a'x^2+b'x+c'$ 我們定義

$$f(x)+g(x)=(a+a')x^2+(b+b')x+(c+c').$$

$S$ 在這加法定義下並無封閉性. 例如 $f(x)=x^2+2x+1\in S$ 且 $g(x)=-x^2\in S$, 但 $f(x)+g(x)=2x+1\not\in S$. 所以在此加法下 $S$ 不是 vector space. 現考慮 $P_2(\mathbb{Q})$ 為次數小於等於 2 的有理係數多項式所成的集合. 利用剛才的加法定義, 我們可得這個加法對 $P_2(\mathbb{Q})$ 有封閉性. 另外若對任意實數 $r\in\mathbb{R}$, 我們定義 $r$ 對 $f(x)=ax^2+bx+c\in P_2(\mathbb{Q})$ 的係數積為 $r\cdot f(x)=(ra)x^2+(rb)x+(rc)$. 在此定義之下實數對 $P_2(\mathbb{Q})$ 的係數積並無封閉性, 例如 $\sqrt{2}$ 乘上 $P_2(\mathbb{Q})$ 的元素 $x^2+x+1$ 會是 $\sqrt{2}x^2+\sqrt{2}x+\sqrt{2}$ 就不再是 $P_2(\mathbb{Q})$ 的元素, 所以在此定義之下 $P_2(\mathbb{Q})$ 也不是 over $\mathbb{R}$ 的 vector space. 不過若同樣的定義考慮有理數 $\mathbb{Q}$ 對 $P_2(\mathbb{Q})$ 的係數積, 則會符合封閉性. 所以我們可以考慮 $P_2(\mathbb{Q})$ 是否是 vector space over $\mathbb{Q}$. 事實上我們很容易驗證此時的加法與係數積會符合 vector space 的 8 個性質, 所以在此定義之下 $P_2(\mathbb{Q})$ 確實是 vector space over $\mathbb{Q}$. 同樣的若 $n$ 是正整數, 若令 $P_n(\mathbb{F})$ 為次數小於 $n$ 且係數在 $\mathbb{F}$ 的多項式所成的集合, 利用上述加法及係數積的定義, 我們可得 $P_n(\mathbb{F})$ 是一個 over $\mathbb{F}$ 的 vector space.

(B) 對任意的 field $\mathbb{F}$, 考慮 $P(\mathbb{F})$ 為所有以 $\mathbb{F}$ 的元素為係數的多項式所成的集合. 利用如 (A) 中定義多項式的加法與係數積, 我們可以證明 $P(\mathbb{F})$ 為 vector space. 首先若 $f(x)=a_nx^n+\cdots+a_1x+a_0, g(x)=b_mx^m+\cdots+b_1x+b_0\in P(\mathbb{F})$, 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 $r\in\mathbb{F}$, 係數積 $rf(x)$ 的定義為 $rf(x)=\sum_{i=0}^{n}(ra_i)x^i$. 利用這個定義以及 $\mathbb{F}$ 是 field 的假設, 我們知道在此定義之下加法和係數積確為 $P(\mathbb{F})$ 中的運算 (有封閉性). 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 $f(x)=\sum_{i=0}^{n}a_ix^i$, $g(x)=\sum_{i=0}^{n}b_ix^i\in P(\mathbb{F})$ 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 $f(x)=\sum_{i=0}^{n}a_ix^i$, $g(x)=\sum_{i=0}^{n}b_ix^i, h(x)=\sum_{i=0}^{n}c_ix^i\in P(\mathbb{F})$ 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 $g(x)=0=\sum_{i=0}^{n}b_ix^i\in P(\mathbb{F})$, 其中 $b_i=0$, $\forall i=0,1,\dots,n$. 此時對任意 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{F})$, 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 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{F})$, 我們考慮 $h(x)=\sum_{i=0}^{n}(-a_i)x^i\in P(\mathbb{F})$, 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 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{F})$, 皆有

$$1f(x)=\sum_{i=0}^{n}(1a_i)x^i=\sum_{i=0}^{n}a_ix^i=f(x).$$

(6) 對任意 $r,s\in\mathbb{R}$ 以及 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{F})$, 我們有

$$r(sf(x))=r(\sum_{i=0}^{n}(sa_i)x^i)=\sum_{i=0}^{n}(r(sa_i))x^i=\sum_{i=0}^{n}((rs)a_i)x^i=(rs)\sum_{i=0}^{n}a_ix^i=(rs)f(x).$$

(7) 對任意 $r,s\in\mathbb{R}$ 以及 $f(x)=\sum_{i=0}^{n}a_ix^i\in P(\mathbb{F})$, 皆有

$$(r+s)f(x)=\sum_{i=0}^{n}((r+s)a_i)x^i=\sum_{i=0}^{n}(ra_i+sa_i)x^i=\sum_{i=0}^{n}(ra_i)x^i+\sum_{i=0}^{n}(sa_i)x^i=rf(x)+sf(x).$$

(8) 對任意 $r\in\mathbb{R}$ 以及 $f(x)=\sum_{i=0}^{n}a_ix^i$, $g(x)=\sum_{i=0}^{n}b_ix^i\in P(\mathbb{F})$ 皆有

$$r(f(x)+g(x))=r(\sum_{i=0}^{n}(a_i+b_i)x^i)=\sum_{i=0}^{n}(r(a_i+b_i))x^i=\sum_{i=0}^{n}(ra_i+rb_i)x^i=rf(x)+rg(x).$$

因為 $P(\mathbb{F})$ 的加法與 $\mathbb{F}$ 的係數積符合 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 $P(\mathbb{F})$ 是一個 vector space over $\mathbb{F}$. 例如所有有理係數多項式所成的集合 $P(\mathbb{Q})$ 就是一個 over $\mathbb{Q}$ 的 vector space, 而實係數多項式所成的集合 $P(\mathbb{R})$ 就是一個 vector space over $\mathbb{R}$. 有趣的是 $P(\mathbb{R})$ 也是一個 vector space over $\mathbb{Q}$, 大家想想看為什麼.

(C) 給定任意 $n \in \mathbb{N}$, 以及一個 field $\mathbb{F}$. 我們令 $\mathbb{F}^n = \{(a_1,\ldots,a_n) : a_i \in \mathbb{F}\}$. 我們沿用 $\mathbb{R}^2$ 中向量的加法及係數積來定義 $\mathbb{F}^n$ 中向量的加法以及係數積. 令 $\mathbf{u} = (a_1,\ldots,a_n), \mathbf{v} = (b_1,\ldots,b_n) \in \mathbb{F}^n$ 以及 $r \in \mathbb{F}$. 我們定義

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

依此定義我們很容易驗證此加法和係數積運算在 $\mathbb{F}^n$ 是封閉的, 而且符合 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 $\mathbb{F}^n$ 是一個 vector space over $\mathbb{F}$. 同樣的考慮所有 entry 皆為 $\mathbb{F}$ 中元素的 $m \times n$ 矩陣所成的集合 $M_{m,n}(\mathbb{F})$ 利用一般矩陣的加法及係數積, 我們也可利用之前談論矩陣加法性質 (Proposition 2.1.3) 一樣的方法驗證 $M_{m,n}(\mathbb{F})$ 是一個 vector space over $\mathbb{F}$.

(D) 給定一非空集合 $S$ 以及 field $\mathbb{F}$, 我們令 $F(S,\mathbb{F})$ 表示所有定義域是 $S$ 且對應域是 $\mathbb{F}$ 的函數所成的集合. 現若 $f,g \in F(S,\mathbb{F})$, 表示對任意 $s \in S$, $f(s),g(s)$ 都會是 $\mathbb{F}$ 中的元素. 所以我們定義 $f+g$ 為定義域是 $S$ 且對應域是 $\mathbb{F}$ 的函數, 其定義為 $f+g$ 在任意 $s \in S$ 的取值為 $f(s)+g(s)$. 亦即 $(f+g)(s) = f(s)+g(s), \forall s \in S$. 對任意 $c \in \mathbb{F}$ 我們也定義 $cf$ 為定義域是 $S$ 且對應域是 $\mathbb{F}$ 的函數, 其定義為 $cf$ 在任意 $s \in S$ 的取值為 $c \cdot f(s)$. 亦即 $(cf)(s) = c \cdot f(s), \forall s \in S$. 由於在此定義之下 $f+g$ 和 $cf$ 仍然在 $F(S,\mathbb{F})$ 中, 所以此加法和係數積運算在 $F(S,\mathbb{F})$ 是封閉的. 我們可以驗證這兩種運算也符合 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 $F(S,\mathbb{F})$ 是一個 vector space over $\mathbb{F}$.

**Question 3.2.** 依照 *Example 3.2.2 (D)* 的定義, 試說明 $F(\mathbb{R},\mathbb{R})$, $F(\mathbb{R},\mathbb{Q})$, $F(\mathbb{Q},\mathbb{Q})$ 以及 $F(\mathbb{Q},\mathbb{R})$ 中那些是 *vector space over* $\mathbb{R}$? 哪些是 *vector space over* $\mathbb{Q}$?

或許很多同學會疑惑, 為何在上一節和 Example 3.2.2中這 8 個性質要證明, 而一開始卻是定義不必證明呢? 要回答這個問題就要回歸到整個過程的演變. 上一節中我們定義了坐標平面向量加法及係數積, 然後驗證它們符合 Proposition 3.1.2 這 8 個性質. 而後由這些性質得到許多運算上很方便且豐富的性質. 事實上這些豐富的性質成立的原因, 主因並不是這些平面向量的加法和係數積是如何定義的, 而是由於它們符合 Proposition 3.1.2 這 8 個性質. 注意到這一點後, 我們專注於符合這 8 個性質的系統稱之為向量空間. 希望以後探討向量空間的問題, 可以不必用到它們真正的運算僅利用這 8 個性質就能得到向量空間所有的性質. 所以當我們想推導向量空間的性質時, 我們就可以直接套用定義中這 8 性質去推導, 這樣推導出來的結果便適用於所有的向量空間. 反過來說, 當我們遇到一個系統有加法, 有係數積, 只要我們能利用該系統的運算 "證明" 它符合這基本的 8 個性質, 那它就是向量空間. 因而所有向量空間的性質它都會符合了, 而不必再用該系統的運算一一去推導.

例如, 在 Question 3.1 中我們利用 $\mathbb{R}^2$ 加法的定義說明 vector space 性質 (3) 中 $\mathbf{0}$ 是唯一的而 (4) 中若給定 $\mathbf{u}$, 則反向量 $\mathbf{u}'$ 也是唯一的. 這兩個唯一性, 事實上不需要知道向量的加法如何定義, 直接用 vector space 的性質就可以證明, 也就是說這個結果對一般的 vector space 都成立. 首先我們看以下之定理.

**Proposition 3.2.3.** 假設 $V$ 為 *vector space,* 且 $\mathbf{u},\mathbf{v},\mathbf{w} \in V$. 若 $\mathbf{u}+\mathbf{w} = \mathbf{v}+\mathbf{w}$, 則 $\mathbf{u} = \mathbf{v}$.

**Proof.** 利用 vector space 的性質 (4), 我們知道存在 $\mathbf{w}' \in V$ 滿足 $\mathbf{w}+\mathbf{w}' = \mathbf{0}$. 然而由 $\mathbf{u}+\mathbf{w} = \mathbf{v}+\mathbf{w}$ 知 $(\mathbf{u}+\mathbf{w})+\mathbf{w}' = (\mathbf{v}+\mathbf{w})+\mathbf{w}'$. 左式由性質 (2),(3) 可得 $(\mathbf{u}+\mathbf{w})+\mathbf{w}' = \mathbf{u}+(\mathbf{w}+\mathbf{w}') = \mathbf{u}+\mathbf{0} = \mathbf{u}$. 同理右式可得 $(\mathbf{v}+\mathbf{w})+\mathbf{w}' = \mathbf{v}+(\mathbf{w}+\mathbf{w}') = \mathbf{v}+\mathbf{0} = \mathbf{v}$, 得證 $\mathbf{u} = \mathbf{v}$. □

Proposition 3.2.3 告訴我們以後在 vector space 處理向量加法問題時, 可以自然地像處理實數一樣使用消去法. 利用這個結果, 我們可以用來處理上述零向量以及反向量的唯一性.

**Corollary 3.2.4.** 假設 $V$ 為 *vector space,* 則在 $V$ 中存在唯一的向量 $\mathbf{0}$ 滿足對任意 $\mathbf{u} \in V$ 皆有 $\mathbf{0}+\mathbf{u} = \mathbf{u}$. 另外, 給定 $\mathbf{u} \in V$, 存在唯一的 $\mathbf{u}' \in V$ 滿足 $\mathbf{u}+\mathbf{u}' = \mathbf{0}$.

**Proof.** 首先證明 $\mathbf{0}$ 是唯一的. 假設 $\mathbf{0}'$ 也滿足 (3) 的性質, 即對任意 $\mathbf{u} \in V$ 皆有 $\mathbf{0}'+\mathbf{u} = \mathbf{u}$. 因此任取 $\mathbf{u} \in V$, 我們有 $\mathbf{0}+\mathbf{u} = \mathbf{u} = \mathbf{0}'+\mathbf{u}$, 故由 Proposition 3.2.3 知 $\mathbf{0} = \mathbf{0}'$, 得證唯一性.

另一方面, 給定 $\mathbf{u} \in V$, 若 $\mathbf{u}', \mathbf{u}''$ 皆滿足 (4) 的性質, 即 $\mathbf{u}+\mathbf{u}' = \mathbf{u}+\mathbf{u}'' = \mathbf{0}$. 故由 Proposition 3.2.3 知 $\mathbf{u}'' = \mathbf{u}'$, 得證唯一性. □

既然 $\mathbf{0}$ 是唯一的, 以後就用 $\mathbf{0}$ 這個專屬的符號來表示 $V$ 中唯一符合 $\mathbf{0}+\mathbf{u} = \mathbf{u}, \forall \mathbf{u} \in V$ 的這個元素, 且稱之為 $V$ 的 *additive identity* 或依慣例稱之為 *zero vector.* 又給定 $\mathbf{u} \in V$, 存在唯一的 $\mathbf{u}'$ 使得 $\mathbf{u}+\mathbf{u}' = \mathbf{0}$, 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.* 而 $\mathbf{u}$ 和 $\mathbf{v}$ 的 additive inverse $-\mathbf{v}$ 相加, 即 $\mathbf{u}+(-\mathbf{v})$, 我們就用 $\mathbf{u}-\mathbf{v}$ 來表示.

**Question 3.3.** 假設 $V$ 為 *vector space.* 試證明對任意 $\mathbf{v} \in V$ 皆有 $-(-\mathbf{v}) = \mathbf{v}$.

接著我們要再強調的是, 雖然在性質 (3) 中提到 $\mathbf{0}$ 是必須滿足對所有 $\mathbf{u} \in V$ 皆有 $\mathbf{0}+\mathbf{u} = \mathbf{u}$ 才可以 (事實上 Proposition 3.2.3 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 $\mathbf{w}$ 是零向量. 不過當我們確定 $V$ 是 vector space 之後, 就可利用 Corollary 3.2.4, 知道只要有一個 $\mathbf{u} \in V$, 會使得 $\mathbf{w}+\mathbf{u} = \mathbf{u}$, 就可以認定 $\mathbf{w} = \mathbf{0}$ 了. 利用這一個唯一性, 我們可以推得許多有關於 $\mathbf{0}$ 的性質. 為了方便起見, 我們列出以下的結果.

**Proposition 3.2.5.** 假設 $V$ 為 *vector space over* $\mathbb{F}$, 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 $\mathbf{w} = \mathbf{0}$.

(2) 對任意 $\mathbf{v} \in V$ 皆有 $0\mathbf{v} = \mathbf{0}$.

(3) 對任意 $r \in \mathbb{F}$ 皆有 $r\mathbf{0} = \mathbf{0}$.

(4) 對任意 $r \in \mathbb{F}, \mathbf{v} \in V$ 皆有 $(-1)(r\mathbf{v}) = -(r\mathbf{v}) = r(-\mathbf{v})$.

**Proof.** (1) 由於 $\mathbf{w}$ 和 $\mathbf{0}$ 皆滿足 $\mathbf{w}+\mathbf{u} = \mathbf{u} = \mathbf{0}+\mathbf{u}$, 故利用 Proposition 3.2.3 推得 $\mathbf{w} = \mathbf{0}$.

(2) 要證明 $0\mathbf{v} = \mathbf{0}$, 利用 (1), 我們只要檢查是否存在 $\mathbf{u} \in V$ 使得 $0\mathbf{v}+\mathbf{u} = \mathbf{u}$. 事實上, 若考慮 $\mathbf{u} = \mathbf{v}$ 的情形, 此時因 $\mathbf{v} = 1\mathbf{v}$, 故 $0\mathbf{v}+\mathbf{v} = 0\mathbf{v}+1\mathbf{v} = (0+1)\mathbf{v} = 1\mathbf{v} = \mathbf{v}$. 故得證 $0\mathbf{v} = \mathbf{0}$.

(3) 同理, 利用 (1), 我們考慮 $\mathbf{u} = r\mathbf{0}$ 的情況, 此時 $r\mathbf{0}+\mathbf{u} = r\mathbf{0}+r\mathbf{0} = r(\mathbf{0}+\mathbf{0}) = r\mathbf{0} = \mathbf{u}$, 得證 $r\mathbf{0} = \mathbf{0}$.

(4) 這裡要證明的是 $(-1)(r\mathbf{v})$ 和 $r(-\mathbf{v})$ 都是 $r\mathbf{v}$ 的 additive inverse (反向量). 由 Corollary 3.2.4 我們知到只要驗證它們是否加上 $r\mathbf{v}$ 都會是 $\mathbf{0}$ 即可. 然而由 (2) 我們有

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

由 (3) 我們有

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

因此得證 $(-1)(r\mathbf{v}) = -(r\mathbf{v}) = r(-\mathbf{v})$. □

由 Proposition 3.2.5 我們知當 $r = 0$ 或 $\mathbf{v} = \mathbf{0}$ 時會有 $r\mathbf{v} = \mathbf{0}$, 但若 $r \neq 0$ 且 $\mathbf{v} \neq \mathbf{0}$, 是否有可能 $r\mathbf{v} = \mathbf{0}$ 呢? 答案是不可能. 這是因為若 $r\mathbf{v} = \mathbf{0}$, 首先由 $r \neq 0$ 且 $\mathbb{F}$ 是一個 field, 我們知道存在 $r' \in \mathbb{F}$ 滿足 $r'r = 1$. 因此可以考慮 $r'$ 乘上 $r\mathbf{v}$ 由 Proposition 3.2.5 (3) 得到 $r'(r\mathbf{v}) = r'\mathbf{0} = \mathbf{0}$. 然而由 vector space 運算性質 (5), (6) 我們有 $r'(r\mathbf{v}) = (rr')\mathbf{v} = 1\mathbf{v} = \mathbf{v}$. 也就是說 $\mathbf{v} = \mathbf{0}$, 此與假設 $\mathbf{b} \neq \mathbf{0}$ 相矛盾, 故知 $r\mathbf{v}$ 絕對不會是 $\mathbf{0}$.

**Question 3.4.** 假設 $V$ 為 *vector space over* $\mathbb{F}$.

(1) 已知 $\mathbf{v} \in V$ 且 $\mathbf{v} \neq \mathbf{0}$. 試證明若 $r, s \in \mathbb{F}$ 且 $r\mathbf{v} = s\mathbf{v}$, 則 $r = s$.

(2) 已知 $r \in \mathbb{F}$ 且 $r \neq 0$. 試證明若 $\mathbf{u}, \mathbf{v} \in V$ 且 $r\mathbf{u} = r\mathbf{v}$, 則 $\mathbf{u} = \mathbf{v}$.

利用以上結果證明若 $V$ 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 $\mathbf{u} = \frac{1}{2}(\mathbf{w}-\mathbf{v})$.

## 3.3. Subspaces

在這一節, 我們介紹 subspace 的概念, 簡單地說, 對於一個 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 3.3.1.** 假設 $V$ 為 vector space over $\mathbb{F}$ 且 $W$ 為 $V$ 的 nonempty subset. 若對 $W$ 的元素利用原先 $V$ 的加法及 $\mathbb{F}$ 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. 以下的定理告訴我們要辨認是否為 subspace, 只要檢查封閉性即可.

**Proposition 3.3.2.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $W$ 為 $V$ 的非空子集合. 則 $W$ 為 $V$ 的 *subspace* 若且唯若對任意 $\mathbf{u},\mathbf{v}\in W$ 且 $r\in\mathbb{F}$ 皆有 $\mathbf{u}+\mathbf{v}\in W$ 以及 $r\mathbf{u}\in W$.

**Proof.** 首先假設 $W$ 為 $V$ 的 subspace. 由於 vector space 的首要條件就是加法與係數積的封閉性, 因此依 subspace 的定義 $W$ 在 $V$ 的加法以及係數積之下應有封閉性, 亦即對任意 $\mathbf{u},\mathbf{v}\in W$ 且 $r\in\mathbb{F}$ 皆有 $\mathbf{u}+\mathbf{v}\in W$ 以及 $r\mathbf{u}\in W$.

反之, 若 $W$ 在 $V$ 的加法以及係數積之下應有封閉性, 則依定義若此加法及係數積符合 vector space 的 8 項性質, 則 $W$ 就是 over $\mathbb{F}$ 的 vector space, 因此依定義就是 $V$ 的 subspace. 然而這 8 項性質中除了 (3), (4) 兩項外, 其餘了性質由於在 $V$ 中的元素皆成立, 所以當然限制在 $W$ 上依然成立. 因此我們僅要驗證 (3), (4) 兩項即可.

性質 (3) 要求的是在 $W$ 中存在一元素 $\mathbf{w}\in W$ 滿足對任意 $\mathbf{u}\in W$ 皆有 $\mathbf{w}+\mathbf{u}=\mathbf{u}$. 然而由於這些元素皆在 $V$ 中, 而 Proposition 3.2.5 (1) 告訴我們此 $\mathbf{w}$ 就是 $V$ 的 zero vector $\mathbf{0}$. 所以我們要檢查的是 $\mathbf{0}\in W$. 現因 $W$ 不是空集合, 所以必定存在 $\mathbf{u}\in W$, 此時因 $0\in\mathbb{F}$ 且由封閉性 $0\mathbf{u}\in W$, 因此由 Proposition 3.2.5 (2) 得證 $\mathbf{0}=0\mathbf{u}\in W$.

性質 (4) 要求的是對任意 $W$ 中的元素 $\mathbf{u}\in W$ 皆存在 $\mathbf{u}'\in W$ 滿足 $\mathbf{u}+\mathbf{u}'=\mathbf{0}$. 由於 $W\subseteq V$, $\mathbf{u}$ 亦在 $V$ 中, 故由 additive inverse 的唯一性 (Proposition 3.2.4知 $\mathbf{u}'=-\mathbf{u}$ 再由 Proposition 3.2.5 (4) 知 $-\mathbf{u}=(-1)\mathbf{u}$. 因此由 $-1\in\mathbb{F}$ 以及係數積的封閉性知 $-\mathbf{u}\in W$. □

注意在這個證明裡, 我們僅利用係數積的封閉性證明 (3), (4) 成立, 不過在驗證 subspace 時一定還要驗證加法的封閉性, 否則無加法封閉性根本沒資格成為 vector space.

一個 vector space $V$ 中有兩個 trivial subspace, 即 $V$ 和 $\{\mathbf{0}\}$. 其中 $\{\mathbf{0}\}$ 稱為 *zero subspace* of $V$, 以後我們用 $\mathbf{O}$ 來表示. 例外要注意 subspace 不能是空集合, 又因為 $\mathbf{0}$ 一定在其中, 所以以後檢查 $V$ 中的子集合是否為 subspace, 我們可以先檢查 $\mathbf{0}$ 是否在其中. 一來可以知道它是不是空集合, 而且若 $\mathbf{0}$ 不在其中就可以斷定它不是 subspace, 真是一舉兩得啊! 以下我們寫下一個檢查是否為 subspace 更簡明的方法.

**Corollary 3.3.3.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $W$ 為 $V$ 的子集合. 則 $W$ 為 $V$ 的 *subspace* 若且唯若 $\mathbf{0}\in W$ 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 $\mathbf{u},\mathbf{v}\in W$, $r\in\mathbb{F}$ 由係數積的封閉性得 $r\mathbf{v}\in W$ 再由加法封閉性得 $\mathbf{u}+r\mathbf{v}\in W$. 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 $0\mathbf{w}=\mathbf{0}$ (Proposition 3.2.5(2)). 故得證 $\mathbf{0}=0\mathbf{w}\in W$.

$(\Leftarrow)$ : 由 $\mathbf{0}\in W$, 我們知 $W$ 為 $V$ 的非空子集合. 故由 Proposition 3.3.2, 我們僅要證明 $W$ 有加法和係數積的封閉性. 我們要用對任意 $\mathbf{u},\mathbf{v}\in W$, $r\in\mathbb{F}$ 皆有 $\mathbf{u}+r\mathbf{v}\in W$ 這個假設證明封閉性. 因 $1\in\mathbb{F}$, 故對任意 $\mathbf{u},\mathbf{v}\in W$ 考慮 $r=1$ 的情形可得 $\mathbf{u}+\mathbf{v}\in W$, 證得加法封閉性. 又若 $\mathbf{v}\in W$ 以及 $r\in\mathbb{F}$, 因為已知 $\mathbf{0}\in W$, 故考慮 $\mathbf{u}=\mathbf{0}$ 的情形得證 $r\mathbf{v}=\mathbf{0}+r\mathbf{y}\in W$. □

由 Corollary 3.3.2, 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) $\mathbf{0} \in W$

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

是否成立即可. 我們看以下的例子.

**Example 3.3.4.** (A) 考慮 $M_{m\times n}(\mathbb{F})$, 即所有 entries 在 $\mathbb{F}$ 中的 $m \times n$ matrices 所成的 vector space. 所謂 $M_{m\times n}(\mathbb{F})$ 的 *upper triangular* matrix 表示當 $i > j$ 時該矩陣第 $(i, j)$-th entry 為 0. 我們要證明 $M_{m\times n}(\mathbb{F})$ 中所有的 upper triangular matrix 所成的集合是 $M_{m\times n}(\mathbb{F})$ 的 subspace. 首先觀察 $m \times n$ 階零矩陣, 由於其任意 entry 皆為 0, 當然 $(i, j)$-th entry 當 $i > j$ 時亦為 0, 因此零矩陣是 upper triangular. 現考慮 $A, B \in M_{m\times n}(\mathbb{F})$ 皆為 upper triangular, 設 $a_{ij}$, $b_{ij}$ 分別表示 $A, B$ 的 $(i, j)$-th entry. 對任意 $r \in \mathbb{F}$, 我們有 $A + rB$ 的 $(i, j)$-th entry 為 $a_{ij} + rb_{ij}$. 現當 $i > j$ 時 $a_{ij} = b_{ij} = 0$, 故得 $a_{ij} + rb_{ij} = 0$. 證得 $A + rB$ 亦為 upper triangular. 因此得證 $M_{m\times n}(\mathbb{F})$ 中所有的 upper triangular matrix 所成的集合是 $M_{m\times n}(\mathbb{F})$ 的 subspace.

(B) 考慮 $M_{n\times n}(\mathbb{F})$, 即所有 entries 在 $\mathbb{F}$ 中的 $n \times n$ 方陣所成的 vector space. 我們想知道 $M_{n\times n}(\mathbb{F})$ 中所有的 symmetric matrices (對稱矩陣) 所成的集合是否為 $M_{n\times n}(\mathbb{F})$ 的 subspace. 首先回顧, 對一個 $m \times n$ matrix $A$, 我們定義 $A$ 的 *transpose* 為一個 $n \times m$ matrix, 記為 $A^{\mathrm{t}}$, 滿足對任意 $1 \leq i \leq n$, $1 \leq j \leq m$, $A^{\mathrm{t}}$ 的 $(i, j)$-th entry 為 $A$ 的 $(j, i)$-th entry. 利用矩陣加法及係數積運算, 我們很容易驗證對任意 $m \times n$ 矩陣 $A, B$ 以及 $r \in \mathbb{F}$ 皆滿足 $(A + rB)^{\mathrm{t}} = A^{\mathrm{t}} + (rB)^{\mathrm{t}} = A + rB$. 現回到對稱矩陣的定義, 對於 $A \in M_{n\times n}(\mathbb{F})$ 我們稱 $A$ 為 symmetric matrix, 表示 $A^{\mathrm{t}} = A$. 很明顯的 $n \times n$ 階零矩陣 $\mathbf{0}$ 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 $r \in \mathbb{R}$, 利用 $A^{\mathrm{t}} = A, B^{\mathrm{t}} = B$, 我們有 $(A + rB)^{\mathrm{t}} = A^{\mathrm{t}} + (rB)^{\mathrm{t}} = A + rB = A + rB$. 亦即 $A + rB$ 亦為 symmetric matrix, 得證 $M_{n\times n}(\mathbb{F})$ 中所有的 symmetric matrices 所成的集合為 $M_{n\times n}(\mathbb{F})$ 的 subspace.

$M_{n\times n}(\mathbb{F})$ 中所有的 invertible matrices (可逆矩陣) 所成的集合是否為 $M_{n\times n}(\mathbb{F})$ 的 subspace 呢? 答案是否定的. 很明顯的零矩陣 $\mathbf{0}$ 就不是 invertible, 所以由 $\mathbf{0}$ 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 $M_{n\times n}(\mathbb{F})$ 的 subspace. 其實即使我們考慮 invertible matrices 所成的集合與 $\{\mathbf{0}\}$ 的聯集, 仍不會是 $M_{n\times n}(\mathbb{F})$ 的 subspace. 因為即使此時 $\mathbf{0}$ 在其中, 但仍有可能兩個 invertible matrices 相加後就不是 invertible. 例如在 $2 \times 2$ 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.

(C) 考慮 $P(\mathbb{F})$, 即所有以 $\mathbb{F}$ 的元素為係數的多項式所成的 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 $P_n(\mathbb{F})$ 是 $P(\mathbb{F})$ 的 subspace. 首先我們可以將 $P_n(\mathbb{F})$ 寫成 $P_n(\mathbb{F}) = \{\sum_{i=0}^{n} a_i x^i \mid a_i \in \mathbb{F}\}$. 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 $P_n(\mathbb{F})$ 為 $P(\mathbb{F})$ 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 $P(\mathbb{F})$ 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.

(D) 對一非空集合 $S$ 以及 field $\mathbb{F}$, 考慮 $F(S,\mathbb{F})$ 為所有從 $S$ 映射到 $\mathbb{F}$ 的函數所成的 vector space. 現假設 $T$ 是 $S$ 的一個非空子集合, 考慮 $N_T=\{f\in F(S,\mathbb{F}): f(t)=0, \forall t\in T\}$, 亦即 $N_T$ 為 $S$ 到 $\mathbb{F}$ 的函數, 但將 $T$ 中的元素都映射到 0. 我們要說明 $N_T$ 是 $F(S,\mathbb{F})$ 的 subspace. 首先 $F(S,\mathbb{F})$ 中的 zero vector $\mathbf{0}$ 就是零函數, 也就是把 $S$ 中的元素都映射到 0 的函數. 現由於 $T\subseteq S$, 所以此零函數當然把 $T$ 中的元素都映射到 0. 得證 $\mathbf{0}\in N_T$. 現若 $f,g\in N_T$ 且 $r\in\mathbb{F}$. 依定義 $(f+rg)(s)=f(s)+rg(s)$, $\forall s\in S$, 故由 $f,g\in N_T$ 的假設知對任意 $t\in T$, $(f+rg)(t)=f(t)+rg(t)=0+0=0$, 得證 $f+rg\in N_T$, 也因此證明了 $N_T$ 是 $F(S,\mathbb{F})$ 的 subspace.

**Question 3.5.** 在 $M_{m\times n}(\mathbb{F})$ 這個 *vector space* 中, 考慮在固定的特定位置 (例如對角線位置) 為 0 的矩陣所成的集合, (例如對角線位置皆為 0 的矩陣所成的集合). 試問這樣的集合是否為 $M_{m\times n}(\mathbb{F})$ 的 *subspace?* 又若考慮在固定的特定位置皆不是 0 的矩陣所成的集合, 是否為 $M_{m\times n}(\mathbb{F})$ 的 *subspace?*

Subspace 是 vector space 中特殊的子集合, 所以我們當然希望能利用已知的 subspace “製造” 出新的 subspace. 這裡我們介紹兩種常見的方法.

**Proposition 3.3.5.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $W_1,W_2$ 為 $V$ 的 *subspace,* 則 $W_1\cap W_2$ 亦為 $V$ 的 *subspace.*

**Proof.** 首先 $W_1,W_2$ 為 subspace, 故 $\mathbf{0}\in W_1$ 且 $\mathbf{0}\in W_2$, 故得 $\mathbf{0}\in W_1\cap W_2$. 現若 $\mathbf{u},\mathbf{v}\in W_1\cap W_2$ 且 $r\in\mathbb{F}$, 由於 $\mathbf{u},\mathbf{v}$ 皆屬於 $W_1$ 且 $W_1$ 是 subspace, 故知 $\mathbf{u}+r\mathbf{v}\in W_1$, 同理可得 $\mathbf{u}+r\mathbf{v}\in W_2$, 故得證 $\mathbf{u}+r\mathbf{v}\in W_1\cap W_2$. □

**Question 3.6.** 證明任意多個 $V$ 的 *subspace* 的交集依然是 $V$ 的 *subspace.* (注意不要用數學歸納法, 數學歸納法僅能證明有限多個的情況)

雖然兩個 subspaces 的交集仍為 subspace, 但他們的聯集就未必是 subspace 了. 當然了, 當 $W_1,W_2$ 為 $V$ 的 subspace, 若 $W_1\subseteq W_2$, 則 $W_1\cup W_2=W_2$ 當然就是 $V$ 的 subspace. 同樣的, 若 $W_2\subseteq W_1$, 則 $W_1\cup W_2=W_1$ 當然也是 $V$ 的 subspace. 下一個定理就是告訴我們, 除了這兩個明顯的情況外, 兩個 subspaces 的聯集不會是 subspace.

**Proposition 3.3.6.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $W_1,W_2$ 為 $V$ 的 *subspace.* 若 $W_1\not\subseteq W_2$ 且 $W_2\not\subseteq W_1$, 則 $W_1\cup W_2$ 不是 $V$ 的 *subspace.*

**Proof.** 依假設 $W_1\not\subseteq W_2$ 表示存在一個元素在 $W_1$ 中但不在 $W_2$, 我們假設 $\mathbf{w}_1\in W_1$ 但 $\mathbf{w}_1\not\in W_2$, 同理由 $W_2\not\subseteq W_1$, 我們假設 $\mathbf{w}_2\in W_2$ 但 $\mathbf{w}_2\not\in W_1$. 當然了, 依定義我們有 $\mathbf{w}_1,\mathbf{w}_2\in W_1\cup W_2$, 我們要利用 $\mathbf{w}_1,\mathbf{w}_2$ 這兩個 $W_1\cup W_2$ 中的元素說明 $W_1\cup W_2$ 在加法之下沒有封閉性, 因此得證 $W_1\cup W_2$ 不是 $V$ 的 subspace.

我們用反證法，假設 $\mathbf{w}_1+\mathbf{w}_2\in W_1\cup W_2$. 這表示 $\mathbf{w}_1+\mathbf{w}_2\in W_1$ 或 $\mathbf{w}_1+\mathbf{w}_2\in W_2$. 若 $\mathbf{w}_1+\mathbf{w}_2\in W_1$, 由 $\mathbf{w}_1\in W_1$ 以及 $W_1$ 為 vector space, 得 $\mathbf{w}_2=(\mathbf{w}_1+\mathbf{w}_2)-\mathbf{w}_1\in W_1$. 此與 $\mathbf{w}_2\not\in W_2$ 相矛盾. 同理若 $\mathbf{w}_1+\mathbf{w}_2\in W_2$, 我們也會推得 $\mathbf{w}_1\in W_2$ 的矛盾. 因此知 $\mathbf{w}_1+\mathbf{w}_2\not\in W_1\cup W_2$, 得證 $W_1\cup W_2$ 不是 $V$ 的 subspace. □

從 Proposition 3.3.6 的證明中, 我們發現 $W_1\cup W_2$ 不是 vector space 最主要的原因是沒有加法封閉性 (它有係數積的封閉性), 我們可以考慮以下的集合故意收集加法所得的元素讓它有加法封閉性.

**Definition 3.3.7.** 假設 $V$ 為 vector space $W_1,W_2$ 為其 subspace, 定義集合

$$W_1+W_2=\{\mathbf{w}_1+\mathbf{w}_2\mid \mathbf{w}_1\in W_1,\mathbf{w}_2\in W_2\}$$

並稱之為 the *sum* of $W_1$ and $W_2$.

對任意 $\mathbf{w}_1\in W_1$, 由於 $\mathbf{w}_1=\mathbf{w}_1+\mathbf{0}$, 且 $\mathbf{0}\in W_2$ (因 $W_2$ 為 subspace), 我們有 $\mathbf{w}_1\in W_1+W_2$, 亦即 $W_1\subseteq W_1+W_2$. 同理可得 $W_2\subseteq W_1+W_2$. 以下定理告訴我們 $W_1+W_2$ 也會是 $V$ 的 subspace, 事實上它是包含 $W_1$ 和 $W_2$ 最小的 subspace.

**Proposition 3.3.8.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $W_1,W_2$ 為 $V$ 的 *subspace,* 則 $W_1+W_2$ 也是 $V$ 的 *subspace.* 特別若 $W$ 是 $V$ 的 *subspace* 且滿足 $W_1\subseteq W$ 以及 $W_2\subseteq W$, 則 $W_1+W_2\subseteq W$.

**Proof.** 首先因 $\mathbf{0}\in W_1$ 且 $\mathbf{0}\in W_2$, 故由 $\mathbf{0}=\mathbf{0}+\mathbf{0}$ 可得 $\mathbf{0}\in W_1+W_2$. 現若 $\mathbf{u},\mathbf{v}\in W_1+W_2$ 且 $r\in\mathbb{F}$, 此時由 $\mathbf{u}\in W_1+W_2$, 知存在 $\mathbf{u}_1\in W_1$ 以及 $\mathbf{u}_2\in W_2$ 使得 $\mathbf{u}=\mathbf{u}_1+\mathbf{u}_2$, 同理存在 $\mathbf{v}_1\in W_1$ 以及 $\mathbf{v}_2\in W_2$ 使得 $\mathbf{v}=\mathbf{v}_1+\mathbf{v}_2$. 因此 $\mathbf{u}+r\mathbf{v}=(\mathbf{u}_1+\mathbf{u}_2)+r(\mathbf{v}_1+\mathbf{v}_2)=(\mathbf{u}_1+r\mathbf{v}_1)+(\mathbf{u}_2+r\mathbf{v}_2)$. 然而 $W_1$ 是 subspace, 由 $\mathbf{u}_1,\mathbf{v}_1\in W_1$ 以及 $r\in\mathbb{F}$ 知 $\mathbf{u}_1+r\mathbf{v}_1\in W_1$. 同理知 $\mathbf{u}_2+r\mathbf{v}_2\in W_2$, 故得 $W_1+W_2$ 是 $V$ 的 subspace.

現對任意 $\mathbf{u}\in W_1+W_2$, 因為存在 $\mathbf{w}_1\in W_1,\mathbf{w}_2\in W_2$ 滿足 $\mathbf{u}=\mathbf{w}_1+\mathbf{w}_2$, 故由 $W_1\subseteq W$ 以及 $W_2\subseteq W$ 知 $\mathbf{w}_1,\mathbf{w}_2\in W$, 因此由 $W$ 是 subspace 知 $\mathbf{u}=\mathbf{w}_1+\mathbf{w}_2\in W$, 得證 $W_1+W_2\subseteq W$. □

**Question 3.7.** 假設 $n\geq 3$, $W_1,W_2,\ldots,W_n$ 皆為 *vector space* $V$ 的 *subspaces.* 學習 *Definition 3.3.7* 的定義方法, 你覺得要如何定義 $W_1+W_2+\cdots+W_n$ 才能讓它為包含 $W_1,W_2,\ldots,W_n$ 最小的 *subspace* 呢?

## 3.4. Linear Combination and Span of Vectors

在這一節中, 我們將介紹線性組合的概念.

當 $V$ 是 vector space over $\mathbb{F}$, 要如何得到 $V$ 的 subspace 呢? 我們可以在 $V$ 中先找到一個 $\mathbf{v}\in V$ 然後找包含 $\mathbf{v}$ 的集合使其為包含 $\mathbf{v}$ 最小的 subspace. 首先這個集合必須包含所有 $\mathbb{F}$ 中的元素與 $\mathbf{v}$ 的係數積, 如此方可保證係數積的封閉性. 所以我們考慮集合 $\{r\mathbf{v}\mid r\in\mathbb{F}\}$. 這個集合不只對 $\mathbb{F}$ 的係數積有封閉性, 而且有加法的封閉性, 事實上它就是包含 $\mathbf{v}$ 最小的 subspace 了. 我們用 $\mathrm{Span}(\mathbf{v})$ 來表示它, 意即 $\mathbf{v}$ 所 *span* (展成) 的向量空間. 我們來驗證 $\mathrm{Span}(\mathbf{v})$ 確為 $V$ 的 subspace. 首先由於 $\mathbf{0}=0\mathbf{v}$, 所以的確有 $\mathbf{0}\in\mathrm{Span}(\mathbf{v})$.

接著, 若 $\mathbf{u},\mathbf{w}\in\mathrm{Span}(V)$, 表示存在 $s,t\in\mathbb{F}$ 滿足 $\mathbf{u}=s\mathbf{v}$ 且 $\mathbf{w}=t\mathbf{v}$, 因此對任意 $r\in\mathbb{F}$, 我們有 $\mathbf{u}+r\mathbf{w}=(s\mathbf{v})+r(t\mathbf{v})=(s+rt)\mathbf{v}$. 由於 $s+rt\in\mathbb{F}$, 我們有 $(s+rt)\mathbf{v}\in\mathrm{Span}(\mathbf{v})$, 亦即 $\mathbf{u}+r\mathbf{w}\in\mathrm{Span}(\mathbf{v})$. 得證 $\mathrm{Span}(\mathbf{v})$ 為 $V$ 的 subspace.

現若 $\mathbf{u},\mathbf{v}\in V$, 既然 $\mathrm{Span}(\mathbf{u}),\mathrm{Span}(\mathbf{v})$ 為 $V$ 的 subspace, 由上一節 subspace 的 sum 的概念, 我們知 $\mathrm{Span}(\mathbf{u})+\mathrm{Span}(\mathbf{v})$ 亦為 $V$ 的 subspace. 依定義

$$\mathrm{Span}(\mathbf{u})+\mathrm{Span}(\mathbf{v})=\{r\mathbf{u}+s\mathbf{v}\mid r,s\in\mathbb{F}\}.$$

由於它是包含 $\mathbf{u},\mathbf{v}$ 最小的 subspace, 我們視之為由 $\mathbf{u},\mathbf{v}$ 所展成的 subspace, 故一般用 $\mathrm{Span}(\mathbf{u},\mathbf{v})$ 來表示. 而 $\mathrm{Span}(\mathbf{u},\mathbf{v})$ 中的元素, $r\mathbf{u}+s\mathbf{v}$ 就稱之為 $\mathbf{u},\mathbf{v}$ 的 *linear combination* (線性組合). 這個概念可以推廣到一般有限多個向量的情況, 我們有以下的定義.

**Definition 3.4.1.** 假設 $V$ 為 vector space over $\mathbb{F}$, 且 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$. 對於任意 $c_1,\dots,c_n\in\mathbb{F}$, 我們稱 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 為 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 的 *linear combination.* 所有 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 的 linear combination 所成的集合, 我們用 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 來表示, 亦即

$$\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_n)=\{\sum_{i=1}^{n}c_i\mathbf{v}_i\mid c_1,\dots,c_n\in\mathbb{F}\}.$$

我們可以直接驗證 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 會是 $V$ 的 subspace (或是利用 sum 的概念, 參見 Question 3.7). 事實上它是包含 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 最小的 subspace. 這是因為若 $W$ 是 $V$ 的 subspace 且 $\mathbf{v}_1,\dots,\mathbf{v}_n\in W$, 則由 $W$ 的加法與係數積的封閉性得 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_n)\subseteq W$.

其實我們不只談論有限多個 $V$ 中的向量所展成的 subspace, 我們也可談論 $V$ 中任意的非空子集合所展成的 subspace. 不過這裡要注意的是, 我們每次只能處理有限多個向量的加法, 所以線性組合也僅能是有限多個向量的線性組合. 因此當 $V$ 中的子集合 $S$ 有無窮多個元素時, $S$ 的 span 是由 $S$ 中有限多個向量的線性組合所組成的. 我們有以下的定義.

**Definition 3.4.2.** 假設 $V$ 為 vector space over $\mathbb{F}$ 且 $S$ 是 $V$ 的非空子集. 則定義

$$\mathrm{Span}(S)=\{\sum_{i=1}^{n}c_i\mathbf{v}_i\mid n\in\mathbb{N},\mathbf{v}_1,\dots,\mathbf{v}_n\in S,c_1,\dots,c_n\in\mathbb{F}\}.$$

要提醒大家注意, 在 Definition 3.4.1 中 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 的定義, 由於涉及 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 這 $n$ 個給定的向量, 所以在此我們用集合表示法說明 $\mathrm{Span}(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 的元素時不必提及 $n$ 和 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是什麼. 然而在 Definition 3.4.2 中當我們用集合表示法說明 $\mathrm{Span}(S)$ 的元素時, 我們是在 $S$ 中任選 $n$ 個元素 $\mathbf{v}_1,\dots,\mathbf{v}_n$, 這 $n$ 和 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是變動的, 所以必須寫下表示 $n$ 是任意可能的正整數, 而 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是 $S$ 中任意可能的向量.

事實上 $\mathrm{Span}(S)$ 也會是 $V$ 的 subspace. 首先 $S$ 不是空集合, 所以存在 $\mathbf{v}\in S$, 此時考慮 $0\mathbf{v}=\mathbf{0}$, 由 $\mathrm{Span}(S)$ 的定義 (取 $n=1$, $\mathbf{v}_1=\mathbf{v}$, $c_1=0$), 知 $\mathbf{0}\in\mathrm{Span}(S)$. 現若 $\mathbf{u},\mathbf{v}\in\mathrm{Span}(S)$ 且 $r\in\mathbb{F}$, 則由於 $\mathbf{u}=c_1\mathbf{u}_1+\cdots+c_n\mathbf{u}_n$, 其中 $n\in\mathbb{N}$, $\mathbf{u}_1,\dots,\mathbf{u}_n\in S$ 且 $c_1,\dots,c_n\in\mathbb{F}$ 以及 $\mathbf{v}=c'_1\mathbf{v}_1+\cdots+c'_m\mathbf{v}_m$, 其中 $m\in\mathbb{N}$, $\mathbf{v}_1,\dots,\mathbf{v}_m\in S$ 且 $c'_1,\dots,c'_m\in\mathbb{F}$, 我們有

$$\mathbf{u}+r\mathbf{v}=c_1\mathbf{u}_1+\cdots+c_n\mathbf{u}_n+rc'_1\mathbf{v}_1+\cdots+rc'_m\mathbf{v}_m,$$

仍符合 $\mathrm{Span}(S)$ 中元素之定義, 故知 $\mathbf{u}+r\mathbf{v}\in\mathrm{Span}(S)$ 得證 $\mathrm{Span}(S)$ 為 $V$ 的 subspace.

在我們寫下一些向量的線性組合時, 前面乘的係數若是 0, 通常我們的省略不寫. 例如 $2\mathbf{v}_1+3\mathbf{v}_2+0\mathbf{v}_3$ 是 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 的一個線性組合, 不過我們通常寫成 $2\mathbf{v}_1+3\mathbf{v}_2$. 基於這個原因再加上我們希望任何集合的 span 皆為 vector space, 因此當 $S$ 是空集合 (用 $\emptyset$ 表示), 我們定義 $\mathrm{Span}(S)=\mathrm{Span}(\emptyset)=\{\mathbf{0}\}$ 這一個 zero subspace.

給定一個 vector space $V$, 若能找到一個 subset $S$ 使得 $\mathrm{Span}(S)=V$, 這當然很好, 表示我們可以用較小的集合 $S$ 就能描述 $V$. 特別的, 若 $S$ 是有限個元素的集合那就更好, 表示僅需有限多個元素就能 "掌握" $V$ 中所有的元素.

**Definition 3.4.3.** 假設 $V$ 為 vector space over $\mathbb{F}$ 且 $S\subseteq V$. 若 $\mathrm{Span}(S)=V$, 則稱 $S$ 為 $V$ 的一組 *spanning set*. 此時我們說 $S$ generates (或 spans $V$). 特別的, 若能找到 finite set (即僅有有限個元素的集合) $S$ 滿足 $\mathrm{Span}(S)=V$, 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. 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 3.4.4.** 我們討論前面提的一些 vector space 哪些是 finitely generated vector space.

(A) $M_{m\times n}(\mathbb{F})$ 是 finitely generated. 因為考慮 $E_{ij}\in M_{m\times n}(\mathbb{F})$, 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 $E_{ij}$ 其中 $1\leq i\leq m$, $1\leq j\leq n$ 的 linear combination. 所以 $\{E_{ij}\mid 1\leq i\leq m, 1\leq j\leq n\}$ 是 $M_{m\times n}(\mathbb{F})$ 的 spanning set, 也因此 $M_{m\times n}(\mathbb{F})$ 是 finitely generated vector space.

(B) $P(\mathbb{F})$ 不是 finitely generated vector space. 這是因為如果 $\{f_1(x),\ldots,f_n(x)\}$ 為 $P(\mathbb{F})$ 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 $P(\mathbb{F})$ 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 $P_n(\mathbb{F})$ 就是 finitely generated vector space. 很容易看出 $\{x^n,\ldots,x,1\}$ 就是 $P_n(\mathbb{F})$ 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.

談到 $\mathrm{Span}(S)$, 我們自然會需要探討那些元素會是 $\mathrm{Span}(S)$ 的元素. 我們看以下的例子.

**Example 3.4.5.** 在 $P_3(\mathbb{R})$ 中考慮 $\mathbf{u}=x^3-2x^2-5x-3$ and $\mathbf{v}=3x^3-5x^2-4x-9$. 我們要檢查 $2x^3-2x^2+12x-6$ 和 $3x^3-2x^2+7x+8$ 是否屬於 $\mathrm{Span}(\mathbf{u},\mathbf{v})$. 首先檢查是否存在 $a,b\in\mathbb{R}$ 使得 $2x^3-2x^2+12x-6=a\mathbf{u}+b\mathbf{v}=a(x^3-2x^2-5x-3)+b(3x^3-5x^2-4x-9)$ 比較係數後我

們發現 $a,b$ 需滿足聯立方程組

$$\begin{cases} a+3b &= 2 \\ -2a-5b &= -2 \\ -5a-4b &= 12 \\ -3a-9b &= -6 \end{cases}$$

利用上一章解聯立方程組的方法, 我們解得 $a=-4, b=2$, 因此得到 $2x^3-2x^2+12x-6 \in \mathrm{Span}(\mathbf{u},\mathbf{v})$. 同樣的我們要檢查是否存在 $a,b \in \mathbb{R}$ 使得 $3x^3-2x^2+7x+8 = a\mathbf{u}+b\mathbf{v} = a(x^3-2x^2-5x-3)+b(3x^3-5x^2-4x-9)$ 比較係數後我們發現 $a,b$ 需滿足聯立方程組

$$\begin{cases} a+3b &= 3 \\ -2a-5b &= -2 \\ -5a-4b &= 7 \\ -3a-9b &= 8 \end{cases}$$

結果發現此聯立方程組無解, 因此知 $3x^3-2x^2+7x+8 \notin \mathrm{Span}(\mathbf{u},\mathbf{v})$.

我們可以討論更一般的情形, 即探討何時 $f(x)=c_1x^3+c_2x^2+c_3x+c_4$ 會屬於 $\mathrm{Span}(\mathbf{u},\mathbf{v})$. 依定義, 這表示存在 $a,b \in \mathbb{R}$ 使得 $c_1x^3+c_2x^2+c_3x+c_4 = a(x^3-2x^2-5x-3)+b(3x^3-5x^2-4x-9)$ 比較係數後我們發現 $a,b$ 需滿足聯立方程組

$$\begin{cases} a+3b &= c_1 \\ -2a-5b &= c_2 \\ -5a-4b &= c_3 \\ -3a-9b &= c_4 \end{cases}$$

利用 elementary row operation 我們將 augmented matrix

$$\left[\begin{array}{rr|l} 1 & 3 & c_1 \\ -2 & -5 & c_2 \\ -5 & -4 & c_3 \\ -3 & -9 & c_4 \end{array}\right].$$

轉換得 reduced echelon form

$$\left[\begin{array}{cc|r} 1 & 0 & -5c_1-3c_2 \\ 0 & 1 & 2c_1+c_2 \\ 0 & 0 & -17c_1-11c_2+c_3 \\ 0 & 0 & 3c_1+c_4 \end{array}\right].$$

這告訴我們此聯立方程組有解若且唯若 $-17c_1-11c_2+c_3=0$ 且 $3c_1+c_4=0$ 且有解時其解為 $a=-5c_1-3c_2, b=2c_1+c_2$. 也就是說當 $-17c_1-11c_2+c_3=0$ 且 $3c_1+c_4=0$ 時, 多項式 $f(x)=c_1x^3+c_2x^2+c_3x+c_4$ 會屬於 $\mathrm{Span}(\mathbf{u},\mathbf{v})$ 且此時 $f(x)=(-5c_1-3c_2)\mathbf{u}+(2c_1+c_2)\mathbf{v}$.

在這一節最後, 我們列下一些有關 $\mathrm{Span}(S)$ 的性質.

**Lemma 3.4.6.** *假設 $V$ 為 vector space over $\mathbb{F}$ 且 $S \subseteq V$, 則 $\mathrm{Span}(S)$ 是 $V$ 中包含 $S$ 最小的 subspace. 換句話說, 若 $W$ 是 $V$ 的 subspace 且 $S \subseteq W$, 則 $\mathrm{Span}(S) \subseteq W$.*

**Proof.** 依定義 $S \subseteq \mathrm{Span}(S)$ 且我們已知 $\mathrm{Span}(S)$ 是 $V$ 的 subspace. 現假設 $W$ 是 $V$ 的 subspace 且 $S \subseteq W$, 我們要說明 $\mathrm{Span}(S) \subseteq W$. 對任意 $\mathbf{v} \in \mathrm{Span}(S)$, 我們知存在 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in S$ 以及 $c_1,\ldots,c_n \in \mathbb{F}$ 使得 $\mathbf{v}=\sum_{i=1}^n c_i\mathbf{v}_i$. 現因 $S \subseteq W$, 我們有 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in W$, 故由 $W$ 是 subspace, 得證 $\mathbf{v}=\sum_{i=1}^n c_i\mathbf{v}_i \in W$. $\square$

利用 Lemma 3.4.6, 我們馬上可知若 $S_1 \subseteq S_2$, 則 $\mathrm{Span}(S_1) \subseteq \mathrm{Span}(S_2)$. 這當然可直接用定義來證明, 不過利用 Lemma 3.4.6, 我們就可以直接套用, 而省去許多繁瑣的論證. 這是因為 $\mathrm{Span}(S_2)$ 是 $V$ 的 subspace 又 $S_1 \subseteq S_2 \subseteq \mathrm{Span}(S_2)$, 故套用 Lemma 3.4.6 (考慮 $S = S_1$, $W = \mathrm{Span}(S_2)$ 的情形) 得證 $\mathrm{Span}(S_1) \subseteq \mathrm{Span}(S_2)$. 利用這個概念我們也可以很快的處理 Span 對兩集合交集及聯集的影響.

當 $S_1, S_2$ 是 $V$ 的 subsets, 我們可以考慮 $\mathrm{Span}(S_1 \cap S_2)$ 和 $\mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$ 的關係. 由於 $S_1 \cap S_2 \subseteq S_1$, 我們有 $\mathrm{Span}(S_1 \cap S_2) \subseteq \mathrm{Span}(S_1)$. 同理 $\mathrm{Span}(S_1 \cap S_2) \subseteq \mathrm{Span}(S_2)$. 由於 $\mathrm{Span}(S_1 \cap S_2)$ 同時包含於 $\mathrm{Span}(S_1)$ 和 $\mathrm{Span}(S_2)$ 可推得 $\mathrm{Span}(S_1 \cap S_2) \subseteq \mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$. 不過反向 $\mathrm{Span}(S_1 \cap S_2) \supseteq \mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$ 就不一定成立, 主要原因是取集合的交集遠比 Span 後再取交集小的多. 例如在 $\mathbb{R}^2$ 上考慮 $S_1 = \{(1,1)\}$, $S_2 = \{(2,2)\}$. 我們有 $S_1 \cap S_2 = \emptyset$ 所以 $\mathrm{Span}(S_1 \cap S_2) = \mathrm{Span}(\emptyset) = \{\mathbf{0}\}$; 不過 $\mathrm{Span}(S_1) = \{(r,r) \mid r \in \mathbb{R}\} = \mathrm{Span}(S_2)$, 所以 $\mathrm{Span}(S_1 \cap S_2) = \{\mathbf{0}\} \subsetneq \mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$.

接下來我們看聯集的情況. 因為 $S_1 \subseteq S_1 \cup S_2$ 且 $S_2 \subseteq S_1 \cup S_2$, 我們有 $\mathrm{Span}(S_1) \subseteq \mathrm{Span}(S_1 \cup S_2)$ 以及 $\mathrm{Span}(S_2) \subseteq \mathrm{Span}(S_1 \cup S_2)$, 所以當然 $\mathrm{Span}(S_1) \cup \mathrm{Span}(S_2) \subseteq \mathrm{Span}(S_1 \cup S_2)$. 不過 $\mathrm{Span}(S_1) \cup \mathrm{Span}(S_2)$ 比起 $\mathrm{Span}(S_1 \cup S_2)$ 太小了, 事實上我們知道 $\mathrm{Span}(S_1) \cup \mathrm{Span}(S_2)$ 在大多數的情況甚至不是 subspace (Proposition 3.3.6). 不過 Proposition 3.3.8 告訴我們 $\mathrm{Span}(S_1) + \mathrm{Span}(S_2)$ 是包含 $\mathrm{Span}(S_1)$ 和 $\mathrm{Span}(S_2)$ 最小的 subspace, 因此由 $\mathrm{Span}(S_1 \cup S_2)$ 包含 $\mathrm{Span}(S_1)$ 和 $\mathrm{Span}(S_2)$ 且是 subspace 得 $\mathrm{Span}(S_1) + \mathrm{Span}(S_2) \subseteq \mathrm{Span}(S_1 \cup S_2)$. 另一方面 $\mathrm{Span}(S_1 \cup S_2)$ 是包含 $S_1 \cup S_2$ 最小的 subspace, 然而 $S_1 \subseteq \mathrm{Span}(S_1) + \mathrm{Span}(S_2)$ 且 $S_2 \subseteq \mathrm{Span}(S_1) + \mathrm{Span}(S_2)$, 再加上 $\mathrm{Span}(S_1) + \mathrm{Span}(S_2)$ 是 subspace, 所以我們也有 $\mathrm{Span}(S_1 \cup S_2) \subseteq \mathrm{Span}(S_1) + \mathrm{Span}(S_2)$. 因此證得 $\mathrm{Span}(S_1 \cup S_2) = \mathrm{Span}(S_1) + \mathrm{Span}(S_2)$. 我們總結以上的結果如下.

**Proposition 3.4.7.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $S_1, S_2$ 皆為 $V$ 的 *subsets.*

(1) 若 $S_1 \subseteq S_2$, 則 $\mathrm{Span}(S_1) \subseteq \mathrm{Span}(S_2)$.

(2) $\mathrm{Span}(S_1 \cap S_2) \subseteq \mathrm{Span}(S_1) \cap \mathrm{Span}(S_2)$.

(3) $\mathrm{Span}(S_1 \cup S_2) = \mathrm{Span}(S_1) + \mathrm{Span}(S_2)$.

## 3.5. Linearly Dependence and Independence

Spanning set 的概念是處理 linear combination 的存在性, 也就是了解一個向量可不可以寫成一些特定向量的線性組合; 而 linear independence 的概念便是處理 linear combination 的唯一性, 也就是說寫成線性組合的方法是否唯一. 這一節中便是探討 linearly independent 的概念.

考慮 over $\mathbb{F}$ 的 vector space $V$. 若 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$, 我們知道 $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ 為 $V$ 中包含 $\mathbf{v}_1, \mathbf{v}_2$ 以及 $\mathbf{v}_3$ 最小的 subspace. 這個 subspace 很容易掌握, 因為每個元素都是 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 的線性組合, 我們只要了解 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 就可以瞭解這個 subspace. 當然了, 若能用較少的元素就能掌握整個 subspace 就更好, 所以我們自然會問這裡 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 有沒有多餘, 不必要的呢? 這是有可能的, 例如若 $\mathbf{v}_3 \in \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$, 則 $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$. 所以此時僅要了

解 $\mathbf{v}_1,\mathbf{v}_2$ 就夠了. 當 $\mathbf{v}_3\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$ 表示存在 $r_1,r_2\in\mathbb{F}$ 使得 $\mathbf{v}_3=r_1\mathbf{v}_1+r_2\mathbf{v}_2$, 此時表示 $\mathbf{v}_3$ 和 $\mathbf{v}_1,\mathbf{v}_2$, 是有關係的, 我們稱 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 是 *linearly dependent* (線性相依或線性相關). 不過要注意 $\mathbf{v}_3\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$ 並不表示 $\mathbf{v}_1\in\mathrm{Span}(\mathbf{v}_2,\mathbf{v}_3)$. 我們看以下的例子.

**Example 3.5.1.** 在 $\mathbb{R}^2$ 中考慮 $\mathbf{v}_1=(1,0)$, $\mathbf{v}_2=(1,1)$, $\mathbf{v}_3=(2,2)$. 我們有 $\mathbf{v}_3\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$, 因為 $\mathbf{v}_3=(2,2)=0(1,0)+2(1,1)=0\mathbf{v}_1+2\mathbf{v}_2\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$. 不過 $\mathbf{v}_1\not\in\mathrm{Span}(\mathbf{v}_2,\mathbf{v}_3)$, 因為 $(1,0)$ 無法寫成 $(1,1)$ 和 $(2,2)$ 的線性組合.

因此當我們要檢查 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 之間是否有線性關係, 不能只檢查是否 $\mathbf{v}_3\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$ 還要檢查是否 $\mathbf{v}_1\in\mathrm{Span}(\mathbf{v}_2,\mathbf{v}_3)$ 以及 $\mathbf{v}_2\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_3)$. 不過分開檢查這三種情況有點麻煩. 我們再深入看一下這三種情況代表甚麼. 當 $\mathbf{v}_1\in\mathrm{Span}(\mathbf{v}_2,\mathbf{v}_3)$, 表示存在 $r,s\in\mathbb{F}$ 使得 $\mathbf{v}_1=r\mathbf{v}_2+s\mathbf{v}_3$, 也就是說 $1\mathbf{v}_1+(-r)\mathbf{v}_2+(-s)\mathbf{v}_3=\mathbf{0}$. 同理, $\mathbf{v}_2\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_3)$ 和 $\mathbf{v}_3\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$ 分別表示存在 $r',s'\in\mathbb{F}$ 和 $r'',s''\in\mathbb{F}$ 分別使得 $\mathbf{v}_2=r'\mathbf{v}_1+s'\mathbf{v}_3$ 和 $\mathbf{v}_3=r''\mathbf{v}_1+s''\mathbf{v}_2$. 也就是說當 (1) $\mathbf{v}_1\in\mathrm{Span}(\mathbf{v}_2,\mathbf{v}_3)$, (2) $\mathbf{v}_2\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_3)$ 或 (3) $\mathbf{v}_3\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$ 其中有一個發生時, 我們會有相對應的

$$\begin{array}{rcrcrcll}
1\mathbf{v}_1 & + & (-r)\mathbf{v}_2 & + & (-s)\mathbf{v}_3 & = & \mathbf{0} & \qquad (1)\\
(-r')\mathbf{v}_1 & + & 1\mathbf{v}_2 & + & (-s')\mathbf{v}_3 & = & \mathbf{0} & \qquad (2)\\
(-r'')\mathbf{v}_1 & + & (-s'')\mathbf{v}_2 & + & 1\mathbf{v}_3 & = & \mathbf{0} & \qquad (3)
\end{array}$$

這三種可能的情況發生, 其中 $r,s,r',s',r'',s''\in\mathbb{F}$. 不過不管 (1), (2), (3) 哪一種情況發生, 總有一個 $\mathbf{v}_i$ 其前面的係數是 1, 不為 0. 也就是說我們可找到不全為 0 的 $c_1,c_2,c_3$ 使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$. 反之若能找到 $c_1,c_2,c_3$ 不全為 0 使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$, 我們就可將前面係數 $c_i$ 不等於 0 的 $\mathbf{v}_i$ 寫成另兩個向量的線性組合. 例如若 $c_1\neq 0$, 則可得 $\mathbf{v}_1=(-c_1'c_2)\mathbf{v}_2+(-c_1'c_3)\mathbf{v}_3$, 其中 $c_1'$ 為 $c_1$ 的乘法反元素 (因 $c_1\neq 0$). 由此可知, 存在 $c_1,c_2,c_3$ 不全為 0 使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$ 和前面所提 (1), (2), (3) 三種情況是等價的, 因此我們用此方法來定義 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 linearly dependent.

讓我們把以上概念推廣到任意多個向量的情況. 我們稱一組向量為 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

$$r_1\mathbf{v}_1+\cdots+r_{n-1}\mathbf{v}_{n-1}+(-1)\mathbf{v}_i+r_{n+1}\mathbf{v}_{n+1}+\cdots+r_n\mathbf{v}_n=\mathbf{0}.$$

也就是說我們找到一組不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$. 反之, 若存在一組不全為 0 的實數 $c_1,\dots,c_n$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$. 我們假設 $c_i\neq 0$, 此時令 $c_i'$ 為 $c_i$ 的乘法反元素 (即 $c_ic_i'=1$), 可得

$$\mathbf{v}_i=(-c_1c_i')\mathbf{v}_1+\cdots+(-c_{i-1}c_i')\mathbf{v}_{i-1}+(-c_{i+1}c_i')\mathbf{v}_{i+1}+\cdots+(-c_nc_i')\mathbf{v}_n,$$

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 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$ 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 3.5.2.** 假設 $V$ 是一個 vector space over $\mathbb{F}$, 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$, 若存在一組不全為 0 的 $c_1,\ldots,c_n \in \mathbb{F}$ 使得

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

則稱 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly dependent* (線性相依或線性相關). 反之, 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 不是 linearly dependent, 則稱為 *linearly independent* (線性獨立).

這個定義可以推廣到 $V$ 的任意子集合. 若 $S \subseteq V$, 且 $S$ 中存在相異 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in S$ 使得 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly dependent, 則稱集合 $S$ 為 linearly dependent; 反之, 若 $S$ 不是 linearly dependent, 表示任意的一組相異向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in S$ 皆為 linearly independent, 則稱 $S$ 為 linearly independent. 在此定義之下空集合 $\emptyset$ 是 linearly independent, 因為我們無法在 $\emptyset$ 中找到任何 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是 linearly dependent. 另一方面, 若 $\mathbf{0} \in S$, 則 $S$ 一定是 linearly dependent. 因為 $\mathbf{0}$ 本身一個元素就是 linearly dependent. 原因是 $1\mathbf{0}=\mathbf{0}$, 所以我們找到一個係數不為 0 的 $\mathbf{0}$ 的線性組合. 因此依定義 $S$ 為 linearly dependent.

一般來說, 除了上述 $S=\emptyset$ 和 $\mathbf{0}\in S$ 這兩種情況外, 要說明 $S$ 是否為 linearly independent 並不是馬上能看出來的. 通常當我們要證明一組向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent, 我們有以下兩個方法: 第一個方法就是先設 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$, 再證明此時 $c_1,\ldots,c_n$ 必全為 0. 第二種方法, 就是所謂的反證法, 亦即先假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly dependent (也就是說假設存在不全為 0 的 $c_1,\ldots,c_n \in \mathbb{F}$ 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$), 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 3.5.3.** 假設 $V$ 為 vector space over $\mathbb{F}$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$ 為 linearly independent. 這表示 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 之間沒有線性關係. 因此可以理解若我們在 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 中移除 $\mathbf{v}_n$, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_{n-1}$ 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 $c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}=\mathbf{0}$ 推得 $c_1,\ldots,c_{n-1}$ 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 $c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}=\mathbf{0}$. 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

$$c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}+c_n\mathbf{v}_n=c_1\mathbf{v}_1+\cdots+c_{n-1}\mathbf{v}_{n-1}=\mathbf{0}.$$

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.

**Question 3.8.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $S \subseteq S' \subseteq V$. 試說明以下的對錯.

(1) 若 $S$ 為 *linearly independent,* 則 $S'$ 為 *linearly independent.*

(2) 若 $S$ 為 *linearly dependent,* 則 $S'$ 為 *linearly dependent.*

(3) 若 $S'$ 為 *linearly independent,* 則 $S$ 為 *linearly independent.*

(4) 若 $S'$ 為 *linearly dependent,* 則 $S$ 為 *linearly dependent.*

在 Example 3.5.3 中, 我們知道當 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$ 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 3.5.4.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{v}_{n+1} \in V$. 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 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n+c_{n+1}\mathbf{v}_{n+1}=\mathbf{0}$. 我們知此時 $c_{n+1}$ 必為 0, 否則由 $c_{n+1}\neq 0$ 知存在 $c'_{n+1} \in \mathbb{F}$ 使得 $c_{n+1}c'_{n+1}=1$ 此時

$$\mathbf{v}_{n+1}=(-c_1c'_{n+1})\mathbf{v}_1+\cdots+(-c_nc'_{n+1})\mathbf{v}_n,$$

會得到 $\mathbf{v}_{n+1} \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 之矛盾. 因此由 $c_{n+1}=0$, 得 $c_1,\ldots,c_n$ 不全為 0 且使得

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

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. $\square$

我們常常看到一個 vector space 中, 若一個集合中向量的個數太多時, 就不會是 linearly independent 了. 例如在 $\mathbb{R}^2$ 中任意 3 個向量 $\mathbf{v}_1=(a_1,b_1),\mathbf{v}_2=(a_2,b_2),\mathbf{v}_3=(a_3,b_3)$ 就一定會 linearly dependent. 這事實的原因便是我們要找到 $c_1,c_2,c_3 \in \mathbb{R}$ 使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$, 就等同於解

$$x_1\mathbf{v}_1+x_2\mathbf{v}_2+x_3\mathbf{v}_3=x_1(a_1,b_1)+x_2(a_2,b_2)+x_3(a_3,b_3)=(0,0),$$

亦即解聯立方程組

$$\begin{cases} a_1x_1+a_2x_2+a_3x_3 &= 0 \\ b_1x_1+b_2x_2+b_3x_3 &= 0 \end{cases}$$

這是有三個未知數 $x_1,x_2,x_3$ 但僅有兩個方程式的齊次聯立方程組, 我們知道一定有無窮多解, 也就是說存在 $c_1,c_2,c_3 \in \mathbb{R}$ 不全為 0 使得 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$. 因此 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 一定是 linearly dependent. 我們可以利用這個概念證明以下著個重要的定理.

**Lemma 3.5.5.** 設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$. 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.** 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 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m=\mathbf{0}$, 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

$$(c_1a_{1,1}+\cdots+c_ma_{1,m})\mathbf{v}_1+\cdots+(c_1a_{i,1}+\cdots+c_ma_{i,m})\mathbf{v}_i+\cdots+(c_1a_{m,1}+\cdots+c_ma_{m,m})\mathbf{v}_m. \quad (3.1)$$

因此若我們能找到 $c_1,\ldots,c_m\in\mathbb{F}$ 使得式子 (3.1) 中每個 $\mathbf{v}_i$ 的係數等於 0, 便可得到 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m=\mathbf{0}$. 因此我們只要找到聯立方程組

$$\begin{cases} a_{1,1}x_1+\cdots+a_{1,m}x_m &=0\\ \qquad\vdots & \\ a_{i,1}x_1+\cdots+a_{i,m}x_m &=0\\ \qquad\vdots & \\ a_{n,1}x_1+\cdots+a_{n,m}x_m &=0\end{cases}$$

的一組解 $x_1=c_1,\ldots,x_m=c_m$, 就可以使得 $c_1\mathbf{w}_1+\cdots+c_m\mathbf{w}_m=\mathbf{0}$. 然而這個 homogeneous linear system 的方程式個數 $n$ 少於未知數個數 $m$. 也就是說將它對應的矩陣化為 echelon form 時, 其 pivot 的個數 (小於等於 $n$) 必少於 variables 的個數 $m$, 也就是存在著 free variables, 因此由此方程組有 $x_1=\cdots=x_n=0$ 這一組解知此方程組有其他解 (參考 Lemma 1.3.4), 即存在不全為 0 的 $c_1,\ldots,c_m\in\mathbb{F}$ 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. □

**Question 3.9.** 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly independent.* 若 $\mathbf{w}_1,\ldots,\mathbf{w}_k\in\operatorname{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 且 $k<n$, 試證明 $\operatorname{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)\neq\operatorname{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$.

假設 $W$ 為 $V$ 的 subspace, 且 $\mathbf{w}_1,\ldots,\mathbf{w}_n\in W$ 為 linearly independent. 如果 $\mathbf{w}_1,\ldots,\mathbf{w}_n$ 不是 $W$ 的 spanning vectors (即 $\operatorname{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)\subsetneq W$), 則我們可以在 $W$ 中選取 $\mathbf{w}_{n+1}$ 滿足 $\mathbf{w}_{n+1}\not\in\operatorname{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_n)$. 此時 Lemma 3.5.4 告訴我們 $\mathbf{w}_1,\ldots,\mathbf{w}_n,\mathbf{w}_{n+1}$ 仍保持 linearly independent. 利用這個概念我們可以回答 finitely generated vector space 的 subspace 也是 finitely generated.

**Proposition 3.5.6.** 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 $\operatorname{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)=V$. 由於 $\{\mathbf{0}\}=\operatorname{Span}(\mathbf{0})$ 為 finitely generated, 我們僅需要考慮 $W\neq\{\mathbf{0}\}$ 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 $\mathbf{w}_1\neq\mathbf{0}$. 由於 $W$ 不是 finitely generated, 我們知 $\operatorname{Span}(\mathbf{w}_1)\neq W$, 亦即存在 $\mathbf{w}_2\in W$ 且 $\mathbf{w}_2\not\in\operatorname{Span}(\mathbf{w}_1)$. 由 Lemma 3.5.5 知 $\mathbf{w}_1,\mathbf{w}_2$ 為 linearly independent. 同理, 因 $W$ 不是 finitely generated, 我們知 $\operatorname{Span}(\mathbf{w}_1,\mathbf{w}_2)\neq W$ 亦即存在 $\mathbf{w}_3\in W$ 且 $\mathbf{w}_3\not\in\operatorname{Span}(\mathbf{w}_1,\mathbf{w}_2)$. 由 Lemma 3.5.4 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 $\operatorname{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)\neq W$, 存在 $\mathbf{w}_{k+1}\in W$ 且 $\mathbf{w}_{k+1}\not\in\operatorname{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)$. 因此再由 Lemma 3.5.4 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 3.5.5 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 3.10.** 利用在 $P(\mathbb{F})$ 中對於任意 $n \in \mathbb{N}$, $x^n, x^{n-1},\ldots,x,1$ 為 *linearly independent*, 證明 $P(\mathbb{F})$ 不是 *finitely generated vector space.*

前面提過若 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$, 要知道 $\mathbf{v} \in V$ 是否屬於 $\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 這是一個存在性的問題, 也就是問是否存在 $c_1,\ldots,c_n \in \mathbb{F}$ 使得 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. 但若已知 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 是否會存在另一組 $c'_1,\ldots,c'_n \in \mathbb{F}$ 使得 $\mathbf{v} = c'_1\mathbf{v}_1 + \cdots + c'_n\mathbf{v}_n$ 呢? 這個唯一性的問題就會和 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是否為 linearly independent 有關了. 我們有以下的結果.

**Proposition 3.5.7.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in V$.

(1) 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly dependent,* 則對任意 $\mathbf{v} \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 將 $\mathbf{v}$ 寫成 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 *linear combination* 的寫法不唯一. 也就是說存在 $c_1,\ldots,c_n \in \mathbb{F}$ 以及 $c'_1,\ldots,c'_n \in \mathbb{F}$ 其中某個 $c_i \neq c'_i$ 使得 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = c'_1\mathbf{v}_1 + \cdots + c'_n\mathbf{v}_n$.

(2) 若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 *linearly independent,* 則對任意 $\mathbf{v} \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 將 $\mathbf{v}$ 寫成 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 *linear combination* 的寫法唯一. 也就是說若 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ 則對任意 $c'_1,\ldots,c'_n \in \mathbb{F}$ 其中某個 $c_i \neq c'_i$, 皆有 $\mathbf{v} \neq c'_1\mathbf{v}_1 + \cdots + c'_n\mathbf{v}_n$.

**Proof.** (1) 因為 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly dependent, 故存在 $d_1,\ldots,d_n \in \mathbb{F}$ 且某個 $d_i \neq 0$ 使得 $d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n = \mathbf{0}$. 現若 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 考慮 $c'_j = c_j + d_i \in \mathbb{F}$, for $j = 1,\ldots,n$. 此時因 $d_i \neq 0$, 故知 $c_i \neq c'_i$, 但我們仍有

$$c'_1\mathbf{v}_1 + \cdots + c'_n\mathbf{v}_n = (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_n + d_n)\mathbf{v}_n =$$
$$(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) + (d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n) = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n + \mathbf{0} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

(2) 現假設 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent, 對任意 $\mathbf{v} \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 假設存在 $c_1,\ldots,c_n \in \mathbb{F}$ 以及 $c'_1,\ldots,c'_n \in \mathbb{F}$ 滿足 $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ 以及 $\mathbf{v} = c'_1\mathbf{v}_1 + \cdots + c'_n\mathbf{v}_n$, 此時 $(c_1 - c'_1)\mathbf{v}_1 + \cdots + (c_n - c'_n)\mathbf{v}_n = \mathbf{0}$, 故由 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent 知 $c_1 - c'_1 = \cdots = c_n - c'_n = 0$, 即 $c_1 = c'_1,\ldots,c_n = c'_n$, 得證 $\mathbf{v}$ 寫成 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的 linear combination 的寫法唯一. □

## 3.6. Basis and Dimension

對於一個 vector space $V$, 我們希望能找到一個集合 $S$ 使得 $V = \mathrm{Span}(S)$. 這樣我們要了解 $V$ 便只要了解 $S$ 即可. 當然了 $S$ 越大越容易展成 $V$, 但是我們又希望其越小越好, 這樣就可以用小一點的集合來了解 $V$. 如何找到這樣大小合宜的 $S$ 便是 basis 的概念.

假設 $V$ 是 over $\mathbb{F}$ 的 vector space 且 $V \neq \{\mathbf{0}\}$. 我們可以在 $V$ 中任取非零向量 $\mathbf{v}_1$, 考慮 $\mathrm{Span}(\mathbf{v}_1)$. 若 $\mathrm{Span}(\mathbf{v}_1) = V$, 則我們找到集合 $S_1 = \{\mathbf{v}_1\}$ 使得 $\mathrm{Span}(S_1) = V$, 且由於 $\mathbf{v}_1 \neq \mathbf{0}$, 依定義 $S_1$ 是 linearly independent. 若 $\mathrm{Span}(\mathbf{v}_1) \subsetneq V$, 表示存在 $\mathbf{v}_2 \in V$ 且 $\mathbf{v}_2 \notin \mathrm{Span}(\mathbf{v}_1)$, 此時考慮 $S_2 = \{\mathbf{v}_1, \mathbf{v}_2\}$. 由於 $\mathbf{v}_2 \notin \mathrm{Span}(\mathbf{v}_1)$, 我們知 $\mathbf{v}_1, \mathbf{v}_2$ 是 linearly independent, 故若

$\mathrm{Span}(S_2) = V$, 我們找到了 $S_2$ 是 $V$ 的 spanning set 且 $S_2$ 為 linearly independent. 這樣一直下去假設我們找到 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 $V$ 的 spanning set 且為 linearly independent. 此時有了 spanning set 的存在性以及 linearly independent 的唯一性, 則 $V$ 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 3.6.1.** 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*.

首先要注意 spanning set 未必是 linearly independent. 例如在 $\mathbb{R}^2$ 中 $\{\mathbf{v}_1 = (1,0), \mathbf{v}_2 = (0,1), \mathbf{v}_3 = (1,1)\}$ 是 $\mathbb{R}^2$ 的 spanning set, 但不是 linearly independent (因 $\mathbf{v}_3 = \mathbf{v}_1 + \mathbf{v}_2$). 同樣的 linearly independent 的元素未必形成 spanning set. 例如在 $\mathbb{R}^3$ 中 $\mathbf{w}_1 = (1,0,0), \mathbf{w}_2 = (0,1,0)$ 就是 linearly independent, 但 $\mathrm{Span}(\mathbf{w}_1, \mathbf{w}_2) \neq \mathbb{R}^3$. 因此要說明 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 $V$ 的一組 basis, 必須說明 $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_n) = V$ 以及 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly independent, 缺一不可. 當然了, 若 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 $V$ 的 spanning set 或 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 linearly independent 其中只要任一項不滿足, 則知 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 不是 $V$ 的 basis. 例如前面舉的例子 $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ 就不是 $\mathbb{R}^2$ 的 basis, 而 $\mathbf{w}_1, \mathbf{w}_2$ 也不是 $\mathbb{R}^3$ 的 basis. 接下來我們看幾個常見的 vector space 的 basis.

**Example 3.6.2.** 假設 $\mathbb{F}$ 是一個 field, 我們考慮以下常見 over $\mathbb{F}$ 的 vector spaces.

(A) 在 $\mathbb{F}^n$ 中考慮 $\mathbf{e}_1 = (1,0,\ldots,0), \mathbf{e}_2 = (0,1,0,\ldots,o), \ldots, \mathbf{e}_n = (0,0,\ldots,o,1)$ (即 $\mathbf{e}_i$ 為 $i$-th entry 為 1, 其他 entry 為 0 的向量), 由於對任意 $(c_1, \ldots, c_n) \in \mathbb{F}^n$, 我們有 $(c_1, \ldots, c_n) = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n$, 知 $\mathrm{Span}(\mathbf{e}_1, \ldots, be_n) = \mathbb{F}^n$. 又若 $c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n = \mathbf{0}$, 表示 $(c_1, \ldots, c_n) = (0, \ldots, 0)$, 亦即 $c_1 = \cdots = c_n = 0$, 故知 $\mathbf{e}_1, \ldots, be_n$ 為 linearly independent. 因此 $\mathbf{e}_1, \ldots, \mathbf{e}_n$ 為 $\mathbb{F}^n$ 的一組 basis. 這一組 basis 是 $\mathbb{F}^n$ 中最直接且最常用的 basis 所以我們又稱之為 $\mathbb{F}^n$ 的 *standard basis*.

(B) 在 $M_{m\times n}(\mathbb{F})$ 中, 考慮 $E_{ij}$, 為 $(i,j)$-th entry 為 1, 其他 entry 為 0 的 $m \times n$ matrix. 則利用和 $\mathbb{F}^n$ 上類似的證法, 我們可以知 $\{E_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ 是 $M_{m\times n}(\mathbb{F})$ 的 spanning set 且為 linearly independent, 故為 $M_{m\times n}(\mathbb{F})$ 的一組 basis.

(C) 在 $P_n(\mathbb{F})$ 中 $\{1, x, x^2, \ldots, x^n\}$ 可展成 $P_n(\mathbb{F})$ 且為 linearly independent, 所以是 $P_n(\mathbb{F})$ 的 basis. 這組 basis 也稱為 $P_n(\mathbb{F})$ 的 *standard basis*.

在 Definition 3.6.1 中 $V$ 可由有限多個元素所展成, 所以此時的 $V$ 依定義是 finitely generated vector space (Definition 3.4.3), 不過我們提過一般的 vector space 未必會是 finitely generated, 所以對於一般的 vector space, 我們有以下 basis 的定義.

**Definition 3.6.3.** 假設 $V$ 為 vector space 且 $S \subseteq V$. 若 $S$ 為 $V$ 的 spanning set 且為 linearly independent, 則稱 $S$ 為 $V$ 的一組 *basis*.

由於我們已定義 $\mathrm{Span}(\emptyset) = \{\mathbf{0}\}$ 且 $\emptyset$ 為 linearly independent, 所以依照 Definition 3.6.3, 我們說空集合 $\emptyset$ 是 zero vector space $\{\mathbf{0}\}$ 的 basis. 另外在 Example 3.4.4 中我們知道 $P(\mathbb{F})$ 不是 finitely generated. 然而很容易看出 $\{1, x, x^2, \ldots\}$ 是 $P(\mathbb{F})$ 的 spanning set 且為 linearly independent, 所以 $\{1, x, x^2, \ldots\}$ 是 $P(\mathbb{F})$ 的 basis.

我們碰到的第一個問題便是 basis 的存在性問題. 也就是說, 對於一般的 vector space 是不是都會有 basis. 接下來我們要說明非零的 finitely generated vector space 都會有 basis. 其實這對於不是 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. 在本講義中我們僅探討 finitely generated vector space. 既然 $V$ 為 finitely generated, 我們假設 $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為其 spanning set. 此時若 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly independent, 則 $S$ 就會是 $V$ 的一組 basis. 但如果不幸的 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 linearly dependent, 這表示存在某個 $\mathbf{v}_i$ 可寫成其他向量的線性組合，為了方便起見就設此向量為 $\mathbf{v}_n$ 好了, 也就是說 $\mathbf{v}_n \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$. 此時集合 $S_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}\}$ 雖然元素變少了但仍為 $V$ 的 spanning set. 這是因為 $V = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}, \mathbf{v}_n) = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1}) + \mathrm{Span}(\mathbf{v}_n) = \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_{n-1})$. 現若 $S_1$ 為 independent, 則 $S_1$ 便會是 $V$ 的 basis. 若不是 independent, 我們又可如法炮製，找到更少元素的 $S_2$ 滿足仍為 $V$ 的 spanning set. 因為 $S$ 的元素僅有有限多個，這樣的程序一直下去，最終一定會停止。也就是最後會找到 $S$ 的一個子集合，仍為 $V$ 的 spanning set 且為 linearly independent, 也因此找到 $V$ 的一組 basis. 這是我們找到 basis 的基本方法，不過“一直下去”這樣的說法不太好，最好用數學歸納法解釋，以下我們便用數學歸納法證明。

**Proposition 3.6.4.** 假設 $V \neq \{\mathbf{0}\}$ 為 *vector space over* $\mathbb{F}$, 且 $S \subseteq V$ 為一個 *finite set* 滿足 $V = \mathrm{Span}(S)$. 則存在 $S' \subseteq S$ 為 $V$ 的一組 *basis*. 也就是說, 若 $V \neq \{\mathbf{0}\}$ 為 *finitely generated vector space over* $\mathbb{F}$, 則存在 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 為 $V$ 的一組 *basis*.

**Proof.** 我們對 $S$ 的元素個數 $n$ 做數學歸納法. 假設 $n = 1$, 即 $V = \mathrm{Span}(\mathbf{v}_1)$, 因 $V \neq \{\mathbf{0}\}$, 知 $\mathbf{v}_1 \neq \mathbf{0}$. 故由 $\mathbf{v}_1$ 本身是 linearly independent, 得證 $S = \{\mathbf{v}_1\}$ 是 $V$ 的 basis. 假設 $n = k$ 時成立, 亦即對所有有 $k$ 元素的集合 $S$, 若 $V = \mathrm{Span}(S)$, 則存在 $S' \subseteq S$ 為 $V$ 的 basis. 我們要證明當 $n = k+1$ 時亦成立. 現假設 $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}\}$ 且 $\mathrm{Span}(S) = V$, 若 $S$ 為 linearly independent, 則依定義令 $S' = S$, 即為 $V$ 的一組 basis. 而若 $S$ 不是 linearly independent, 則不失一般性, 我們假設 $\mathbf{v}_{k+1} \in \mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. 此時令 $\tilde{S} = S \setminus \{\mathbf{v}_{k+1}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$, 因 $V = \mathrm{Span}(\tilde{S})$ 且 $\tilde{S}$ 的元素個數為 $k$, 故由歸納假設知存在 $S' \subseteq \tilde{S} \subseteq S$ 為 $V$ 的一組 basis. 得證本定理. □

Proposition 3.6.4 說的是當 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 $V$ 的 spanning set, 則我們可以在 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 中捨去一些元素使其為 linearly independent 且仍為 $V$ 的 spanning set, 故可成為 $V$ 的一組 basis. 反之, 若 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 $V$ 的一組 linearly independent set, 則我們可以加入一些元素擴大 $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, 使其成為 $V$ 的 spanning set 且仍保持 linearly independent, 故可成為 $V$ 的一組 basis. 原因是若 $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ 為 $V$ 的 spanning set, 則 $S$ 自然是 $V$ 的一組 basis. 但如果不是 spanning set, 表示存在 $\mathbf{v}_{n+1} \in V$ 但 $\mathbf{v}_{n+1} \notin \mathrm{Span}(S)$. 此時考慮 $S_1 = S \cup \{\mathbf{v}_{n+1}\} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}\}$, 由 Lemma 3.5.4, $S_1$ 仍保持 linearly independent. 然而此時 $S_1$ 的元素變多了且其所展成的空間也變大了。如果 $\mathrm{Span}(S_1) = V$, 則我們得 $S_1$ 為 $V$ 的一組 basis. 若 $S_1$ 仍不是 spanning set, 則我們可以如法炮製，找到更多元素的 $S_2$ 使其仍保持 linearly independent. 因為 linear independent 的集合其元素個數不可能大於 $V$ 的

spanning set 的個數 (Lemma 3.5.5), 這樣的程序一直下去，最終一定會停止。也就是最後會找到一個包含 $S$ 的一個集合，仍為 linearly independent 且為 $V$ 的 spanning set, 也因此找到 $V$ 的一組 basis. 這是我們另一個找到 basis 的方法，不過“一直下去”這樣的說法不太好，所以以下我們依然便用數學歸納法證明。

**Proposition 3.6.5.** 假設 $V \neq \{\mathbf{0}\}$ 為 *vector space over* $\mathbb{F}$, 且 $V = \mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_m)$. 若 $S = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\} \subseteq V$ 為 *linearly independent,* 則存在 $S' \subseteq \{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 使得 $S \cup S'$ 為 $V$ 的 *basis.*

**Proof.** 我們依然對 $S$ 的元素個數做歸納法, 不過這次是做反向的歸納法. 注意因 $V = \mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_m)$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent, 故由 Lemma 3.5.5 知 $n \leq m$, 所以我們可以假設 $n = m-t$, 其中 $0 \leq t \leq m-1$. 我們要對 $t$ 做 mathematical induction. 當 $t = 0$ 時表示 $m = n$, 此時我們要說明 $S = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 為 $V$ 的 spanning set. 這是因為若 $\mathrm{Span}(S) \neq V$, 表示存在 $\mathbf{w} \in V$ 且 $\mathbf{w} \notin \mathrm{Span}(S)$, 故由 Lemma 3.5.4 知 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n,\mathbf{w}\}$ 為 linearly independent, 但此時 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n.\mathbf{w}\}$ 有 $n+1 = m+1$ 個元素, 多於 $\mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_m)$ 的 $m$ 個元素, 與 Lemma 3.5.5 相矛盾, 故知 $S$ 為 $V$ 的 spanning set. 再利用已知 $S$ 為 linearly independent, 故令 $S' = \emptyset$ 得證 $S = S \cup S'$ 為 $V$ 的 basis. 現假設 $t = k$ 時成立, 即若 $S = \{\mathbf{v}_1,\ldots,\mathbf{v}_{m-k}\}$ 為 linearly independent, 則存在 $S' \subseteq \{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 使得 $S \cup S'$ 為 $V$ 的 basis. 現考慮 $t = k+1$ 的情形, 即 $S = \{\mathbf{v}_1,\ldots,\mathbf{v}_{m-k-1}\}$ 為 linearly independent. 首先考慮 $S$ 是否為 $V$ 的 spanning set. 若 $V = \mathrm{Span}(S)$, 則依定義知 $S$ 為 $V$ 的一組 basis, 故取 $S' = \emptyset$, 則 $S' \subseteq \{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$, 且 $S \cup S' = S$ 為 $V$ 的 basis. 而若 $\mathrm{Span}(S) \neq V$, 表示 $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 中必有一元素 $\mathbf{u}_i \notin \mathrm{Span}(S)$, 否則若 $\mathbf{u}_1,\ldots,\mathbf{u}_m$ 皆屬於 $\mathrm{Span}(S)$, 會造成 $V = \mathrm{Span}(\mathbf{u}_1,\ldots,\mathbf{u}_m) \subseteq \mathrm{Span}(S)$ 之矛盾. 現考慮 $\tilde{S} = S \cup \{\mathbf{u}_i\}$, 因 $\mathbf{u}_i \notin \mathrm{Span}(S)$ 由 Lemma 3.5.4 知 $\tilde{S}$ 為 linearly independent, 再由 $\tilde{S}$ 的元素個數為 $m-k$, 故由歸納假設知存在 $\tilde{S}' \subseteq \{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 使得 $\tilde{S} \cup \tilde{S}'$ 為 $V$ 的 basis. 故令 $S' = \{\mathbf{u}_i\} \cup \tilde{S}'$, 我們有 $S' \subseteq \{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 且 $S \cup S' = S \cup \{\mathbf{u}_i\} \cup \tilde{S}' = \tilde{S} \cup \tilde{S}'$ 為 $V$ 的 basis. □

我們已經知道 finitely generated vector space 的 basis 是存在的, 不過它並不唯一. 例如在 $\mathbb{R}^2$ 中除了 standard basis $\{(1,0),(0,1)\}$ 外, 我們很容易看出 $\{(1,1),(0,1)\}$ 也可以是 $\mathbb{R}^2$ 的 basis. 不過 basis 雖然不唯一, 不過在 finitely generated vector space 中組成 basis 的元素個數是固定的. 我們有以下的定理.

**Theorem 3.6.6.** 假設 $V$ 為 *vector space over* $\mathbb{F}$, 且 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 和 $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 皆為 $V$ 的 *basis,* 則 $n = m$.

**Proof.** 我們用反證法, 假設 $n \neq m$, 不失一般性我們就假設 $m > n$. 因 $V = \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 故 $\mathbf{u}_i \in \mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, $\forall i = 1,\ldots,m$. 因此由 Lemma 3.5.5 知 $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 是 linearly dependent. 此與 $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ 是 basis 的假設相矛盾, 故得證 $m = n$. □

Theorem 3.6.6 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 3.6.7.** 假設 $V$ 是一個 finitely generated vector space over $\mathbb{F}$. 組成 $V$ 的一組 basis 的元素個數稱為 $V$ over $\mathbb{F}$ 的 *dimension* (維度), 用 $\dim_{\mathbb{F}}(V)$ 來表示.

由於組成 finitely generated vector space 的一組 basis 的元素個數是有限的, 所以以後我們稱 finitely generated vector space 為 *finite dimensional vector space.*

**Example 3.6.8.** 我們探討在 Example 3.6.2 中的 finite dimensional vector space 的維度為多少.

(A) 考慮 $\mathbb{F}^n$ 中的 standard basis $\{\mathbf{e}_i \mid 1 \leq i \leq n\}$, 因為共有 $n$ 個元素所以 $\dim_{\mathbb{F}}(\mathbb{F}^n) = n$.

(B) 我們知道 $\{E_{ij} \in \mathscr{M}_{m\times n} \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ 是 $\mathscr{M}_{m\times n}(\mathbb{F})$ 的一組 basis. 因此 $\dim \mathbb{F}(\mathscr{M}_{m\times n}) = m \times n$.

(C) 我們知道 $\{x^n, \ldots, x, 1\}$ 是 $P_n(\mathbb{F})$ 的 spanning set 且為 linearly independent. 故知 $\{x^n, \ldots, x, 1\}$ 為 $P_n(\mathbb{F})$ 的一組 basis, 因此 $\dim_{\mathbb{F}}(P_n(\mathbb{F})) = n+1$.

大家或許注意到, 我們在表示維度的 dim 符號的下標特別標上 $\mathbb{F}$, 即 $\dim_{\mathbb{F}}$. 這個原因是強調我們將 vector space 看成 over $\mathbb{F}$ 的 vector space 所得的 dimension. 我們曾經提過, 同樣的集合我們有可能看成 over 不同的 field 的 vector space. 在此情況之下它們的 basis 也就會不同, 也因此我們要標示用哪一個 field. 我們看以下的例子.

**Example 3.6.9.** 我們用 $\mathbb{C}$ 表示 complex numbers (複數) 所成的 field, 而用 $\mathbb{R}$ 表示 real numbers (實數) 所成的 field. 現考慮集合 $\mathbb{C}^2 = \{(z_1, z_2) \mid z_1, z_2 \in \mathbb{C}\}$. 很容易檢查用一般的加法及係數積, $\mathbb{C}^2$ 是 vector space over $\mathbb{C}$, 也會是 vector space over $\mathbb{R}$. 首先我們知道 $\{(1,0),(0,1)\}$ 是 $\mathbb{C}^2$ 看成 over $\mathbb{C}$ 的 vector space 的 basis (Example 3.6.2 (A) $n=2$, $\mathbb{F} = \mathbb{C}$ 的情況), 所以我們得 $\dim_{\mathbb{C}}(\mathbb{C}^2) = 2$. 不過 $\{(1,0),(0,1)\}$ 在 over $\mathbb{R}$ 之下就不是 basis 了. 很容易看出來任何 $(1,0),(0,1)$ over $\mathbb{R}$ 的 linear combination 都無法表示 $(i,0)$ 這一個 $\mathbb{C}^2$ 的元素 (這裡 $i$ 是滿足 $i^2 = -1$ 的純虛數). 不過 $\{(1,0),(i,0),(0,1),(0,i)\}$ 就是 $\mathbb{C}^2$ over $\mathbb{R}$ 的 spanning set. 這是因為任意 $\mathbb{C}^2$ 的元素都可以寫成 $(a+bi, c+di)$ 其中 $a,b,c,d \in \mathbb{R}$, 所以可得 $(a+bi, c+di) = a(1,0)+b(i,0)+c(0,1)+d(0,i)$. 又 $\{(1,0),(i,0),(0,1),(0,i)\}$ over $\mathbb{R}$ 是 linearly independent. 這是因為若 $a,b,c,d$ 是實數滿足 $a(1,0)+b(i,0)+c(0,1)+d(0,i) = (0,0)$, 即表示 $a+bi = 0$ 且 $c+di = 0$, 故得 $a=b=c=d=0$. 由此知 $\{(1,0),(i,0),(0,1),(0,i)\}$ 是 $\mathbb{C}^2$ over $\mathbb{R}$ 的 basis, 所以我們有 $\dim_{\mathbb{R}}(\mathbb{C}^2) = 4$.

由 Example 3.6.9, 我們知道要說明一個 vector space 的 dimension 為何, 一定要說明其 over 的 field 是甚麼. 不過一般情形, 當我們很明確知道 over 的 field 是甚麼而沒有如 Example 3.6.9 這種模稜兩可的情形, 我們便會省略直接用 $\dim(V)$ 來表示其 dimension.

對於 finite dimensional vector space 有關於 dimension 的性質, 我們匯集如下. 再次強調, 由於這裡我們只考慮 over 一個固定的 field $\mathbb{F}$, 所以我們僅用 $\dim(V)$ 來表示其 dimension.

**Proposition 3.6.10.** 假設 $V$ 為 *finite dimensional vector space over* $\mathbb{F}$.

(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 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 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) 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.*

(4) 若 $W$ 為 $V$ 的 *subspace,* 則 $\dim(W)\leq\dim(V)$. 特別的, 若 $\dim(W)=\dim(V)$ 則 $W=V$.

**Proof.** 為了方便起見, 我們令 $S=\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$.

(1) 依假設 $V=\mathrm{Span}(S)$, 故利用 Proposition 3.6.4 知存在 $S'\subseteq S$ 為 $V$ 的 basis. 也就是說 $S'$ 的元素個數就是 $V$ 的 dimension. 然而 $S'$ 是 $S$ 的 subset, 所以其元素個數小於等於 $S$ 的元素個數 $n$, 故得證 $\dim(V)\leq n$. 現若 $S$ 為 linearly dependent, 即表示存在 $\mathbf{v}_i$ 可寫成 $S$ 中其他元素的線性組合, 因此考慮 $\tilde{S}=S\setminus\{\mathbf{v}_i\}$, 我們仍有 $\mathrm{Span}(\tilde{S})=V$. 此時 $\tilde{S}$ 的元素個數為 $n-1$, 所以再套用前面所證的可得 $\dim(V)\leq n-1<n$.

(2) 依假設 $S$ 是 linearly independent, 故利用 Proposition 3.6.5 知存在某個有限集合 $S'$ 使得 $S\cup S'$ 為 $V$ 的 basis. 也就是說 $S\cup S'$ 的元素個數就是 $V$ 的 dimension. 然而 $S\subseteq S\cup S'$, 所以 $S\cup S'$ 的元素個數大於等於 $S$ 的元素個數 $n$, 故得證 $\dim(V)\geq n$. 現若 $S$ 不是 $V$ 的 spanning set, 表示存在 $\mathbf{w}\in V$ 且 $\mathbf{w}\not\in\mathrm{Span}(S)$, 因此考慮 $\tilde{S}=S\cup\{\mathbf{w}\}$, 我們仍有 $\tilde{S}$ 為 linearly independent (Lemma 3.5.4). 此時 $\tilde{S}$ 的元素個數為 $n+1$, 所以再套用前面所證的可得 $\dim(V)\geq n+1>n$.

(3) 我們要證明 (a) 可推得 (b), (b) 可推得 (c) 以及 (c) 可推得 (a). 因此知 (a),(b),(c) 是等價的.

**(a) $\Rightarrow$ (b):** 假設 $S$ 是 $V$ 的 basis, 當然 $S$ 是 $V$ 的 spanning set. 又由於 $S$ 的元素個數為 $n$, 依定義 $\dim(V)=n$.

**(b) $\Rightarrow$ (c):** 由於 $S$ 是 $V$ 的 spanning set, 由前面 (1) 的結果, 若 $S$ 是 linearly dependent, 則 $\dim(V)<n$. 此與 $\dim(V)=n$ 假設相矛盾, 故推得 $S$ 是 linearly independent.

**(c) $\Rightarrow$ (a):** 由於 $S$ 是 linearly independent, 由前面 (2) 的結果, 若 $S$ 不是 $V$ 的 spanning set, 則 $\dim(V)>n$. 此與 $\dim(V)=n$ 假設相矛盾, 故推得 $S$ 是 $V$ 的 spanning set. 因此得證 $S$ 是 $V$ 的 basis.

(4) 因 $W$ 是 $V$ 的 subspace, 故由 Proposition 3.5.6 知 $W$ 亦為 finite dimensional vector space, 我們假設 $S=\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 為 $W$ 的 basis. 由於 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 且為 linearly independent, 故由 (2) 的結果知 $\dim(W)=n\leq\dim(V)$. 而若 $\dim(V)=n$, 則由 $S$ 是 linearly independent 利用 (3)((c) $\Rightarrow$ (a)) 知 $S$ 也是 $V$ 的 basis, 故得證 $W=V$. □

強調一下, Proposition 3.6.10 告訴我們知道 $V$ 的 dimension 的好處. 若我們知道 $\dim(V)$ 恰好是 $n$, 則 (3) 告訴我們當要檢查 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 是否為 $V$ 的一組 basis 時, 則僅要檢查它們是否為 spanning set 或 linearly independent 其中一項就可. 所以我們只要選擇檢查哪一個較好處理即可. 另外若已知 $W$ 為 $V$ 的 subspace, 要檢查 $W$ 是否為 $V$, 我們不必再像以前檢查是否每個 $V$ 中的元素都在 $W$, 而只要算出 $\dim(W)$ 是否等於 $n$ 即可.

**Example 3.6.11.** 很容易看出在 $P(\mathbb{R})$ 中 $x^n,x^{n-1},\ldots,x,1$ 為 linearly independent. 這是因為如果 $c_n,\ldots,c_1,c_0\in\mathbb{R}$ 使得 $c_nx^n+\cdots+c_1x+c_01$ 為零多項式, 則依零多項式的定義, $c_n,\ldots,c_1,c_0$ 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.

我們知道了 $p_1(x),p_2(x),p_3(x)\in P_2(\mathbb{R})$ 為 linearly independent, 事實上 $p_1(x),p_2(x),p_3(x)$ 會是 $P_2(\mathbb{R})$ 的 spanning set. 不過要證明這一點, 需用到次數小於 3 的實係數非零多項式不會有 3 個相異實根這個事實, 說明起來有點麻煩. 不過由於我們已知 $\dim(P_2(\mathbb{R}))=3$, 故由 Proposition 3.6.10 (3) 知 $p_1(x),p_2(x),p_3(x)$ 為 $P_2(\mathbb{R})$ 的一組 basis. 也就是說任何實係數的次數小於 3 的多項式 $f(x)$ 都可以都可以找到唯一的一組 $c_1,c_2,c_3\in\mathbb{R}$ 使得 $f(x)=c_1p_1(x)+c_2p_2(x)+c_3p_3(x)$. 事實上代入 $x=a,b,c$, 我們知道 $c_1=f(a)$, $c_2=f(b)$, $c_3=f(c)$ 就是這唯一的一組.

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.

找到一個 over $\mathbb{F}$ 的向量空間 $V$ 之一組 basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 的好處是, 由 $V=\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 我們知對任意 $\mathbf{v}\in V$, 皆可找到 $c_1,\ldots,c_n\in\mathbb{F}$ 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 又因 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 linearly independent, 我們知這些 $c_1,\ldots,c_n$ 是唯一的 (Corollary 3.5.7). 因此當我們固定 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 這一組 basis, 對任意 $V$ 中一個向量 $\mathbf{v}$, 若 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 我們可以用坐標的

方式來表示它, 即 $(c_1,\cdots,c_n)$. 這樣我們就給了 $V$ 中的向量和 $\mathbb{F}^n$ 中的向量之間一個一對一的對應關係. 換言之, 我們可以將 $V$ 這種抽象的向量空間視為 $\mathbb{F}^n$ 這種具體的向量空間. 這種坐標化的概念, 在之後是非常重要的.

## 3.7. Column Space and Null space

我們將介紹一個矩陣的 column space, row space 以及 nullspace 並探討如何找到它們的 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 null space 的重要性, 我們將之正式定義如下:

**Definition 3.7.1.** 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 $\mathrm{Col}(A)$ 來表示 $A$ 的 column space.

(2) 我們稱 homogeneous linear system $A\mathbf{x}=\mathbf{0}$ 所有解所成的集合為 $A$ 的 *null space* 且用 $N(A)$ 表示 $A$ 的 null space. 即 $N(A)=\{\mathbf{u}\in\mathbb{R}^n \mid A\mathbf{u}=\mathbf{0}\}$.

要注意當 $A\in M_{m\times n}$, 則 $A$ 的 column space $\mathrm{Col}(A)$ 會是 $\mathbb{R}^m$ 的 subspace, 而 $A$ 的 null space $N(A)$ 會是 $\mathbb{R}^n$ 的 subspace (請自行檢驗). 利用 Lemma 2.4.1 以及 Theorem 2.4.5 我們馬上有以下的結果.

**Proposition 3.7.2.** 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 $\mathbf{b}\in\mathrm{Col}(A)$.

(2) 假設 $A\mathbf{x}=\mathbf{b}$ 有解則其解唯一若且唯若 $N(A)=\{\mathbf{0}\}$.

接下來我們就是要找到一個矩陣的 column space 以及 null space 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 3.7.3.** 考慮 $\mathbb{R}^3$ 中

$$\mathbf{v}_1=\begin{bmatrix}0\\3\\0\\1\end{bmatrix},\mathbf{v}_2=\begin{bmatrix}2\\-5\\0\\0\end{bmatrix},\mathbf{v}_3=\begin{bmatrix}0\\7\\-1\\0\end{bmatrix}.$$

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 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$. 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 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$, 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 $c_1\mathbf{v}_1+c_2\mathbf{v}_2+c_3\mathbf{v}_3=\mathbf{0}$, 故知 $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ 為 linearly independent.

從 Example 3.7.3 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 3.7.3 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 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$, 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 null space $N(A)$ 就是 homogeneous linear system $A\mathbf{x}=\mathbf{0}$ 的所有解所成的集合. 由於我們已經知道如何找到 $A\mathbf{x}=\mathbf{0}$ 的解, 所以我們就從如何找 null space 的 basis 開始.

回顧我們找 $A\mathbf{x}=\mathbf{0}$ 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 $A'\mathbf{x}=\mathbf{0}$ 的解集合就是 $A\mathbf{x}=\mathbf{0}$ 的解集合, 也就是說 $A$ 和 $A'$ 有相同的 null space. 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 null space 的一組 spanning vectors. 我們證明了 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 就是 $A$ 的 null space 的一組 basis, 也因此得知 $A$ 的 null space 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 3.7.4.** *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 null space 的 dimension 為 $n-r$. 假設 $A'\mathbf{x}=\mathbf{0}$ 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 null space 的一組 basis.*

由於一個矩陣的 null space 不會因為其化為 echelon form 的不同而改變, 而且 null space 的 dimension 是固定的, 所以 Proposition 3.7.4 也說明了“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”, 也就是這個矩陣的 rank.

**Example 3.7.5.** 考慮 $A$ 的 null space, 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 null space 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 null space 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 null space 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 3.11.** *試將 Example 3.7.5 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 $\mathrm{Col}(A)$ 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. 我們看以下的例子.

**Example 3.7.6.** 考慮 Example 3.7.5 中的 $4\times 6$ 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.

$$\begin{array}{ccccccccc} 2x_1 & +x_2 & +x_3 & & & & = & b_1 \\ x_1 & & & +x_4 & & & = & b_2 \\ x_1 & +x_2 & +x_3 & & +x_5 & +2x_6 & = & b_3 \\ x_1 & +2x_2 & +2x_3 & -2x_4 & +x_5 & +2x_6 & = & b_4 \end{array}$$

考慮 augmented matrix $[A\,|\,\mathbf{b}]$, 利用 Example 3.7.5 相同的 elementary row operations 我們得

$$\left[\begin{array}{cccccc|c} 2 & 1 & 1 & 0 & 0 & 0 & b_1 \\ 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 1 & 1 & 1 & 0 & 1 & 2 & b_3 \\ 1 & 2 & 2 & -2 & 1 & 2 & b_4 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & -2 & 0 & 0 & b_1-2b_2 \\ 0 & 1 & 1 & -1 & 1 & 2 & b_3-b_2 \\ 0 & 2 & 2 & -3 & 1 & 2 & b_4-b_2 \end{array}\right] \rightsquigarrow$$

$$\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & -2 & 0 & 0 & b_1-2b_2 \\ 0 & 0 & 0 & 1 & 1 & 2 & b_3+b_2-b_1 \\ 0 & 0 & 0 & 1 & 1 & 2 & b_4+3b_2-2b_1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 1 & 0 & 0 & b_2 \\ 0 & 1 & 1 & -2 & 0 & 0 & b_1-2b_2 \\ 0 & 0 & 0 & 1 & 1 & 2 & b_3+b_2-b_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & b_4-b_3+2b_2-b_1 \end{array}\right].$$

由解聯立方程組的方法 (即 1.2 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 null space. 由於 $x_1$ 為 pivot variable, $x_2,x_3,x_4$ 為 free variable. 利用前面求 null space 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 null space 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 3.7.6 找 column space 所用的方法缺點就是還要再求另一個矩陣的 null space 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 $A\mathbf{x}=\mathbf{0}$ 和 $A'\mathbf{x}=\mathbf{0}$ 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 $A'\mathbf{x}=\mathbf{0}$ 的一組解, 表示 $c_1\mathbf{a}'_1+\cdots+c_n\mathbf{a}'_n=\mathbf{0}$, 此時由於 $x_1=c_1,\ldots,x_n=c_n$ 亦為 $A\mathbf{x}=\mathbf{0}$ 的一組解故我們亦有 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{0}$. 同理若 $c_1,\cdots,c_n\in\mathbb{R}$ 使得 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{0}$, 我們亦會有 $c_1\mathbf{a}'_1+\cdots+c_n\mathbf{a}'_n=\mathbf{0}$. 這告訴我們存在不全為 0 的 $c_i$ 使得 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\mathbf{0}$ 若且唯若存在不全為 0 的 $c_i$ 使得 $c_1\mathbf{a}'_1+\cdots+c_n\mathbf{a}'_n=\mathbf{0}$. 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 3.7.7.** 考慮 Example 3.7.5 中的 $4\times6$ matrix $A$, 且利用 elementary row operation 將之化為 reduced echelon form $A'$. 也就是將 Example 3.7.5 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

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

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 $4\times3$ 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, $\mathbf{a}_3, \mathbf{a}_5, \mathbf{a}_6 \in \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4)$. 故知 $A$ 的 column space 為

$$\mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4, \mathbf{a}_5, \mathbf{a}_6) = \mathrm{Span}(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4).$$

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 3.7.7 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) $A'$ 的 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 $A'$ 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 $x_{i_1}, \ldots, x_{i_r}$, 則由於 $A'$ 的 pivot 所在的 column vectors $\mathbf{a}'_{i_1}, \ldots, \mathbf{a}'_{i_r}$ 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 $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}$ 亦為 linearly independent. 同理, 由於 $A'$ 的其他 column vectors $\mathbf{a}'_j$ 皆符合 $\mathbf{a}'_j \in \mathrm{Span}(\mathbf{a}'_{i_1}, \ldots, \mathbf{a}'_{i_r})$, 我們得 $A$ 的其他 column vectors $\mathbf{a}_j$ 也符合 $\mathbf{a}_j \in \mathrm{Span}(\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r})$. 因此得 $\mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \mathrm{Span}(\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r})$. 我們證得了 $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}$ 為 $A$ 的 column space 的 spanning vectors 且為 linearly independent, 故 $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}$ 為 $A$ 的 column space 的一組 basis. 我們有以下的定理.

**Proposition 3.7.8.** *假設 $A \in M_{m\times n}$ 且 $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ 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 $x_{i_1}, \ldots, x_{i_r}$, 則 $\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_r}$ 為 $A$ 的 column space 的一組 basis.*

相對於矩陣的 column space, 我們也可考慮矩陣的 row space. 我們有以下的定義.

**Definition 3.7.9.** 假設 $A = \begin{bmatrix} — & {}_1\mathbf{a} & — \\ & \vdots & \\ — & {}_m\mathbf{a} & — \end{bmatrix}$ 為以 $\mathbb{R}^n$ 中的向量 ${}_1\mathbf{a}_1,\ldots,{}_m\mathbf{a}$ 為 row vectors 的 $m\times n$ matrix. 則 $A$ 的 row space 為 $\mathrm{Span}({}_1\mathbf{a},\ldots,{}_m\mathbf{a})$, 且用 $\mathrm{Row}(A)$ 來表示.

如何求 $A$ 的 row space 的 basis 呢? 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 3.7.10.** 考慮 Example 3.7.5 中的 $4\times 6$ matrix $A$, 且利用 elementary row operation 將之化為 reduced echelon form $A'$ (參見 Example 3.7.7), 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 3.7.5 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 3.7.7 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 3.7.10 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, 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 Proposition 3.6.10 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 3.7.11.** 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 3.7.12.** 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 3.7.10 中的矩陣 $A$. 利用 Example 3.7.10 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 4-th entry 的地方 (這些地方就是 reduced echelon form 的 pivot 的位置) 需有 $c_1=1,c_2=-2,c_4=3$. 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 3.7.12 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 $6\times4$ matrix, 而考慮成 row vectors, 所得的矩陣為 $4\times6$ 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.

我們再次強調，elementary row operations 會保持 column vectors 之間的線性關係 (但一般不會保持 column space)；不過 elementary row operations 會保持 row space (但一

般不會保持 row vectors 之間的線性關係). 因此給定一個矩陣 $A$, 利用 elementary row operations 將 $A$ 化為 echelon form $A$. 此時 $\mathrm{Col}(A)$ 未必等於 $\mathrm{Col}(A')$, 但由於 $A'$ 中 pivot 所在的 column vectors 皆為 linearly independent 且能展成 $\mathrm{Col}(A')$, 所以回到原先的矩陣 $A$, 其對應的 column vectors 就會形成 $\mathrm{Col}(A)$ 的一組 basis. 而由於 $\mathrm{Row}(A') = \mathrm{Row}(A)$, 而且很明顯的 $A'$ 的 pivot 所在的 row vectors 會是 $\mathrm{Row}(A')$ 的一組 basis, 所以也自然是 $\mathrm{Row}(A)$ 的一組 basis. 不過要注意，這些 $A'$ 的 pivot 的位置對 $\mathrm{Row}(A)$ 是沒有意義的，它並未擔保其對應到 $A$ 的 row vectors 會是 $\mathrm{Row}(A)$ 的一組 basis.

從以上的看法（或從 Proposition 3.7.8 和 Proposition 3.7.11），我們知道組成 $A$ 的 column space 和 row space 的 basis 的向量個數皆為 pivot 的個數, 即 $A$ 的 rank. 所以 $A$ 的 column space 和 row space 的 dimension 都是 $A$ 的 rank, 至於 $A$ 的 null space 的維度, 我們也給予一特殊的名稱, 即以下的定義.

**Definition 3.7.13.** 假設 $A$ 為 $m \times n$ matrix. $A$ 的 null space 的 dimension 稱為 $A$ 的 *nullity*, 記為 $\mathrm{nullity}(A)$, 亦即 $\mathrm{nullity}(A) = \dim(N(A))$.

依此定義, 由 Proposition 3.7.8 和 Proposition 3.7.11 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 3.7.4 我們知道 $\mathrm{nullity}(A)$ 就是 homogeneous linear system $A\mathbf{x} = \mathbf{0}$ 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 3.7.14** (Dimension Theorem)**.** *假設 $A$ 為 $m \times n$ matrix. 則*

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

**Question 3.12.** *假設 $A$ 為 $n \times n$ invertible matrix. 試求 $\mathrm{rank}(A)$ 以及 $\mathrm{nullity}(A)$.*

**Question 3.13.** *假設 $A \in M_{m\times n}$.*

(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 $\mathrm{nullity}(A)$.*

(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 $\mathrm{nullity}(A)$.*

Proposition 3.7.8 告訴我們 $A$ 的 column space 的維度就是 $A$ 的 rank, 亦即 $\dim(\mathrm{Col}(A)) = \mathrm{rank}(A)$, 而 Proposition 3.7.11 告訴我們 $\dim(\mathrm{Row}(A)) = \mathrm{rank}(A)$, 因此得 $\dim(\mathrm{Col}(A)) = \dim(\mathrm{Row}(A))$. 也就是說一個矩陣的 column space 和 row space 有相同的維度. 現考慮矩陣 $A$ 的 transpose $A^{\mathrm{t}}$. 由於 $A$ 的 column space 就是 $A^{\mathrm{t}}$ 的 row space (且 $A$ 的 row space 就是 $A^{\mathrm{t}}$ 的 column space), 所以我們有 $\mathrm{rank}(A) = \dim(\mathrm{Col}(A)) = \dim(\mathrm{Row}(A^{\mathrm{t}})) = \mathrm{rank}(A^{\mathrm{t}})$. 得證以下的性質.

**Proposition 3.7.15.** *假設 $A \in M_{m\times n}$. 則*

$$\dim(\mathrm{Col}(A)) = \dim(\mathrm{Row}(A)) = \mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{t}}).$$

*亦即利用 elementary row operations 將 $A$ 以及 $A^{\mathrm{t}}$ 化為 echelon form 後, 它們的 pivot 個數相同.*

**Question 3.14.** 假設 $A$ 為 $m\times n$ *matrix.* 試證明 $\text{nullity}(A^{\mathrm{t}})=m-\text{rank}(A)$.

注意, 對於 Proposition 3.7.15 的證明, 若想要直接證明利用 elementary row operations 將 $A$ 以及 $A^{\mathrm{t}}$ 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. 最後我們利用類似的概念來處理兩個矩陣相乘後 rank 的變化關係.

**Proposition 3.7.16.** 假設 $A\in M_{m\times n}$, $B\in M_{n\times l}$. 則

(1) $\text{Col}(AB)\subseteq\text{Col}(A)$ 且 $\text{rank}(AB)\leq\text{rank}(A)$.

(2) $\text{Row}(AB)\subseteq\text{Row}(B)$ 且 $\text{rank}(AB)\leq\text{rank}(B)$.

(3) 若 $E\in M_{n\times n}$ 為 *invertible,* 則 $\text{Col}(AE)=\text{Col}(A)$ 且 $\text{rank}(AE)=\text{rank}(A)$.

(4) 若 $H\in M_{m\times m}$ 為 *invertible,* 則 $\text{Row}(HA)=\text{Row}(A)$ 且 $\text{rank}(HA)=\text{rank}(A)$.

**Proof.** 令 $A$ 和 $B$ 的 column vectors 依序為 $\mathbf{a}_1,\dots,\mathbf{a}_n$ 和 $\mathbf{b}_1,\dots,\mathbf{b}_l$ 且令 $A$ 和 $B$ 的 row vectors 依序為 ${}_1\mathbf{a},\dots,{}_m\mathbf{a}$ 和 ${}_1\mathbf{b},\dots,{}_n\mathbf{b}$.

(1) 依定義 $\text{Col}(AB)=\text{Span}(A\mathbf{b}_1,\dots,A\mathbf{b}_l)$, 而對任意 $\mathbf{b}=\begin{bmatrix}b_1\\ \vdots\\ b_n\end{bmatrix}$, 我們有

$$A\mathbf{b}=\begin{bmatrix}| & & |\\ \mathbf{a}_1 & \cdots & \mathbf{a}_n\\ | & & |\end{bmatrix}\begin{bmatrix}b_1\\ \vdots\\ b_n\end{bmatrix}=b_1\mathbf{a}_1+\cdots+b_n\mathbf{a}_n\in\text{Span}(\mathbf{a}_1,\dots,\mathbf{a}_n)=\text{Col}(A).$$

因此 $A\mathbf{b}_i\in\text{Col}(A)$, $\forall 1\leq i\leq l$. 因此得 $\text{Col}(AB)=\text{Span}(A\mathbf{b}_1,\dots,A\mathbf{b}_l)\subseteq\text{Col}(A)$. 換言之 $\text{Col}(AB)$ 為 $\text{Col}(A)$ 的 subspace, 故由 Proposition 3.6.10 (4) 知 $\text{rank}(AB)=\dim(\text{Col}(AB))\leq\dim(\text{Col}(A))=\text{rank}(A)$.

(2) 依定義 $\text{Row}(AB)=\text{Span}({}_1\mathbf{a}B,\dots,{}_m\mathbf{a}B)$, 而對任意 $\mathbf{a}=\begin{bmatrix}a_1 & \cdots & a_n\end{bmatrix}$, 我們有

$$\mathbf{a}B=\begin{bmatrix}a_1 & \cdots & a_n\end{bmatrix}\begin{bmatrix}- & {}_1\mathbf{b} & -\\ & \vdots & \\ - & {}_n\mathbf{b} & -\end{bmatrix}=a_1({}_1\mathbf{b})+\cdots+a_n({}_n\mathbf{b})\in\text{Span}({}_1\mathbf{b},\dots,{}_n\mathbf{b})=\text{Row}(B).$$

因此 ${}_i\mathbf{a}B\in\text{Row}(B)$, $\forall 1\leq i\leq m$. 因此得 $\text{Row}(AB)=\text{Span}({}_1\mathbf{a}B,\dots,{}_m\mathbf{a}B)\subseteq\text{Row}(B)$. 換言之 $\text{Row}(AB)$ 為 $\text{Row}(B)$ 的 subspace, 故由 Proposition 3.6.10 (4) 知 $\text{rank}(AB)=\dim(\text{Row}(AB))\leq\dim(\text{Row}(B))=\text{rank}(B)$.

(3) 利用前面 (1) 的結果我們知 $\text{Col}(AE)\subseteq\text{Col}(A)$. 因 $E$ 為 invertible, 考慮 $(AE)E^{-1}=A(EE^{-1})=A$. 再利用 (1) 知 $\text{Col}(A)=\text{Col}((AE)E^{-1})\subseteq\text{Col}(AE)$. 因此得證 $\text{Col}(AE)=\text{Col}(A)$ 且取 dimension 得 $\text{rank}(AE)=\dim(\text{Col}(AE))=\dim(\text{Col}(A))=\text{rank}(A)$.

(4) 利用前面 (2) 的結果我們知 $\mathrm{Row}(HA) \subseteq \mathrm{Row}(A)$. 因 $H$ 為 invertible, 考慮 $H^{-1}(HA) = (H^{-1}H)A = A$. 再利用 (2) 知 $\mathrm{Row}(A) = \mathrm{Row}(H^{-1}(HA)) \subseteq \mathrm{Row}(HA)$. 因此得證 $\mathrm{Row}(HA) = \mathrm{Row}(A)$ 且取 dimension 得 $\mathrm{rank}(HA) = \dim(\mathrm{Row}(HA)) = \dim(\mathrm{Row}(A)) = \mathrm{rank}(A)$. □

注意, 在 Proposition 3.7.16 (3) 中我們知道當 $E$ 為 invertible 時 $\mathrm{rank}(AE) = \mathrm{rank}(A)$, 也因此知 $\dim(\mathrm{Row}(AE)) = \mathrm{rank}(AE) = \mathrm{rank}(A) = \dim(\mathrm{Row}(A))$. 不過這並不代表 $\mathrm{Row}(AE) = \mathrm{Row}(A)$. 這是因為 $\mathrm{Row}(AE)$ 和 $\mathrm{Row}(A)$ 未必有包含關係, 因此即使 $\dim(\mathrm{Row}(AE)) = \dim(\mathrm{Row}(A))$, 也無法推得 $\mathrm{Row}(AE) = \mathrm{Row}(A)$. 同理在 Proposition 3.7.16 (4) 中我們知道當 $H$ 為 invertible 時 $\dim(\mathrm{Col}(HA)) = \dim(\mathrm{Col}(A))$, 但 $\mathrm{Col}(HA)$ 也未必等於 $\mathrm{Col}(A)$.

**Question 3.15.** 試找到例子 $A, E \in M_{3\times 3}(\mathbb{R})$ 其中 $E$ 為 *invertible* 使得 $\mathrm{Row}(AE) \neq \mathrm{Row}(A)$, $\mathrm{Col}(EA) \neq \mathrm{Col}(A)$. *(Hint: 考慮 $E$ 為 elementary matrix.)*

**Question 3.16.** 試用取轉置矩陣的方法利用 *Proposition 3.7.16 (1)* 證明 *(2)*, 且利用 *(3)* 證明 *(4)*.

**Question 3.17.** 試利用 *Proposition 2.5.7*, 即 *invertible matrix* 皆可寫成 *elementary matrices* 的乘積, 證明 *Proposition 3.7.16 (3), (4)*.

*Chapter 4*

# Inner Product Space

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

## 4.1. Dot Product

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

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

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

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

向量的內積和向量的運算有一定的關係, 以下就是它們之間的關係

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

□

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

利用 Proposition 4.1.5, 我們可以得到所謂的三角不等式.

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

**Proof.** 由 Lemma 4.1.4 以及 Proposition 4.1.5, 我們有

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

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

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

利用內積我們可以知道坐標平面或空間中向量之間的一些幾何關係. 例如若兩非零向量 $\mathbf{u},\mathbf{v}$ 的夾角為 $\theta$, 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 $\theta$, 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 $\theta$ 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 $\theta$ 的存在性就關係到 $\mathbb{R}^n$ 中兩個非零向量 $\mathbf{u},\mathbf{v}$ 是否會滿足

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

然而 Proposition 4.1.5 告訴我們這是一定對的, 所以這裡 $\theta$ 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 $\theta$ 要滿足 $0 \leq \theta \leq \pi$, 如此才能確保所得的夾角會是唯一的. 也就是說用這種方法定義兩非零向量的夾角是沒有問題的, 我們就稱這樣的定義是 well-defined.

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

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

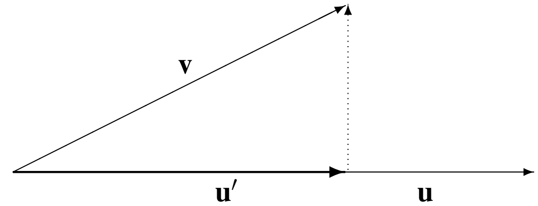
得知 $\theta = 120^\circ$.

利用夾角的定義我們進而定義出何謂「垂直」.

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

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

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



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

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

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

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

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

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

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

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

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

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

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

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

又我們有

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

其中

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

## 4.2. Inner Product

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

## 4.3. Projection and Gram-Schmidt Process

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

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

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

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

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

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

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

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

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

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

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

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

□

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

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

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

現在我們可以證明 projection 的唯一性.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

最後和前面同樣的理由, 我們有 $\mathrm{Span}(\mathbf{w}_1,\mathbf{w}_2,\mathbf{w}_3)=\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3)$. 這樣一直下去, 我們可以得到 $W_k=\mathrm{Span}(\mathbf{w}_1,\ldots,\mathbf{w}_k)=\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_k)$ 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 $\mathbf{v}_{k+1}\not\in W_k$. 故令

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

則得 $\mathbf{w}_{k+1}\neq\mathbf{0}$ 且 $\langle\mathbf{w}_{k+1},\mathbf{w}_i\rangle=0,\forall i=1,\ldots,k$. 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 vector space 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. 因此如果一開始已給定 $W$ 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 4.3.9** (Gram-Schmidt Process)**.** 假設 $V$ 為 *inner product space, W* 為 $V$ 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

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

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

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

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

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

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

雖然 Theorem 4.3.9 的敘述是找到 $V$ 的 subspace $W$ 的 orthogonal basis, 不過因 $W$ 是 $V$ 中任意的 subspace, 所以當 $W$ 為 $V$ 時, 我們也就找到 $V$ 的 orthogonal basis 了. 所以對任意 finite dimensional inner product space, Gram-Schmidt process 都能幫我們找到 orthogonal basis. 注意這裡需要有限維的假設, 因為整個過程我們是一個一個置換這些向量, 所以有限多個向量才可全部置換完成. Theorem 4.3.9 的敘述牽涉到 $V$ 的 subspace $W$ 主要用意是, 我們不只可找到 $W$ 的 orthogonal basis, 也可繼續這個 process, 而將 $W$ 的 orthogonal basis 擴大成 $V$ 的 orthogonal basis. 這是因為若 $W\neq V$, 當找到 $\mathbf{w}_1,\dots,\mathbf{w}_n$ 為 $W$ 的 orthogonal basis 後我們可以繼續考慮 $\mathbf{v}_{n+1}\in V$ 但 $\mathbf{v}_{n+1}\not\in W$, 然後利用 Gram-Schmidt process 得到 $\mathbf{w}_{n+1}=\mathbf{v}_{n+1}-\mathrm{Proj}_W(\mathbf{v}_{n+1})\in W^\perp$. 這樣一直下去直到得到 $V$ 的一組 orthogonal basis 為止. 也因此我們得到以下之結果.

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

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

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

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

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

最後得

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

□

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

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

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

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

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

## 4.4. 矩陣運算和內積的連結

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

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

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

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

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

結合式子 (4.2), (4.3) 我們可得以下性質.

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

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

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

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

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

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

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

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

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

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

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

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

這與 Dimension Theorem 相吻合.

Lemma 4.4.1 還有許多的應用. 我們有以下的定理.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

很容易檢查, 我們確有 $A=QR$.

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

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

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

## 4.5. 聯立方程組和內積的連結

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

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

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

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

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

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

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

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

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

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

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

注意當 $\mathrm{rank}(A)<n$ 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 $A\mathbf{x}=\mathrm{Proj}_{\mathrm{Col}(A)}(\mathbf{b})$ 的解, 故由 $A$ 為 $m\times n$ matrix 以及 Theorem 2.4.5 知其解不唯一 (會有無窮多解). 有關於解不唯一的情形, 我們等一下再談.

由 Proposition 4.5.1, 我們特別有以下的定義.

**Definition 4.5.2.** 假設 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$ 且 $\mathbf{b}\in\mathbb{R}^m$. 考慮聯立方程組 $A\mathbf{x}=\mathbf{b}$. 我們稱聯立方程組 $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 為 $A\mathbf{x}=\mathbf{b}$ 的 *normal equations.*

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

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

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

解得 $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ 為其解. 我們也可由 $\operatorname{rank}(A)=2$, 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 $\operatorname{Col}(A)$ 的 projection (參見 Example 4.4.8).

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

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

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

給定 $m,c\in\mathbb{R}$, 對於 $i=1,\dots,n$, 令 $y_i'=mx_i+c$. 也就是說 $(x_i,y_i')$ 會在直線 $y=mx+c$ 上. 為方便起見令 $\mathbf{b}' = \begin{bmatrix} y_1' \\ \vdots \\ y_n' \end{bmatrix}$, 此時 $\mathbf{b}' = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}\begin{bmatrix} m \\ c \end{bmatrix} \in \operatorname{Col}(A)$. 又令 $\varepsilon_i = y_i' - y_i$, 此即將直線 $y=mx+c$ 代 $x_i$ 所得的 $y_i'$ 與實際資料中的 $y_i$ 之誤差. 我們知道代這些 $x_i$ 後所得的誤差越小越好, 不過由於 $\varepsilon_i$ 會有正有負, 怕正負抵銷了影響誤差的判定, 所以我們取平方, 也就是說希望 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 的值越小越好. 然而依定義 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 就是 $\mathbf{b}'$ 和 $\mathbf{b}$ 距離的平方, 即 $\|\mathbf{b}-\mathbf{b}'\|^2$. 另一方面, 我們利用 $A\mathbf{x}=\mathbf{b}$ 求出其 normal equation $A^{\mathrm{t}}A\mathbf{x}=A^{\mathrm{t}}\mathbf{b}$ 的解 $\mathbf{x} = \begin{bmatrix} m \\ c \end{bmatrix}$, 這裡的 $m,c$ 就是讓直線 $y=mx+c$ 所估計得的 $\mathbf{b}'$ 會是 $\mathbf{b}$ 在 $\operatorname{Col}(A)$ 的投影, 也就是說會讓 $\|\mathbf{b}-\mathbf{b}'\|$ 的值最小. 所以符合我們希望誤差 $\varepsilon_1^2+\cdots+\varepsilon_n^2$ 最小的要求. 也因此這樣所得的直

線, 我們稱之為 *least squares line* (最適合直線, 或最小平方直線), 在統計學中稱之為 *line of regression* (迴歸直線).

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

原先要解的方程組是

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

其矩陣表示法為 $\begin{bmatrix} -1 & 1\\ 1 & 1\\ 2 & 1\end{bmatrix}\begin{bmatrix} a\\ b\end{bmatrix}=\begin{bmatrix}0\\1\\3\end{bmatrix}$. 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

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

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

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

接著我們探討幾個和 least squares line 有關的性質.

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

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

(2) $x_1y_1+\cdots+x_ny_n=x_1y_1'+\cdots+x_ny_n'$.

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

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

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

得證 (1), (2). 又令 $\overline{y'}=(y_1'+\cdots+y_n')/n$, 由 (1) 知 $\overline{y}=\overline{y'}$. 又由於對所有 $i=1,\ldots,n$, 皆有 $y_i'=mx_i+c$, 故知 $\overline{y'}=m\overline{x}+c$. 得證 (3), 即 $\overline{y}=m\overline{x}+c$. □

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

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

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

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

如何找到 $A\mathbf{x} = \mathbf{b}$ 的 minimal solution 呢? 假設 $\mathbf{v}, \mathbf{w}$ 皆為 $A\mathbf{x} = \mathbf{b}$ 的一組解, 則由 $A\mathbf{v} = \mathbf{b}$ 以及 $A\mathbf{w} = \mathbf{b}$ 可得 $A(\mathbf{v} - \mathbf{w}) = \mathbf{0}$. 因此若令 $\mathbf{u} = \mathbf{v} - \mathbf{w}$, 則得 $\mathbf{u} \in N(A)$ 且 $\mathbf{w} = \mathbf{v} - \mathbf{u}$. 反之, 當我們已找到 $\mathbf{v}$ 為 $A\mathbf{x} = \mathbf{b}$ 的一組解時, 對任意 $\mathbf{u} \in N(A)$, 皆有 $A(\mathbf{v} - \mathbf{u}) = A\mathbf{v} - A\mathbf{u} = \mathbf{b} - \mathbf{0} = \mathbf{b}$, 亦即 $\mathbf{v} - \mathbf{u}$ 也是 $A\mathbf{x} = \mathbf{b}$ 的一組解. 因此聯立方程組 $A\mathbf{x} = \mathbf{b}$ 的解集合可以寫成 $\{\mathbf{v} - \mathbf{u} \mid \mathbf{u} \in N(A)\}$, 其中 $\mathbf{v}$ 為 $A\mathbf{x} = \mathbf{b}$ 的一組解. 也就是說當我們已找到 $A\mathbf{x} = \mathbf{b}$ 的一組解 $\mathbf{v}$, 我們要找到 $\mathbf{u} \in N(A)$ 使得 $\|\mathbf{v} - \mathbf{u}\| \leq \|\mathbf{v} - \mathbf{u}'\|$, $\forall \mathbf{u}' \in N(A)$, 此時 $\mathbf{v} - \mathbf{u}$ 就會是 $A\mathbf{x} = \mathbf{b}$ 的 minimal solution. 如何在 $N(A)$ 中找到 $\mathbf{u}$ 呢? 由於 $N(A)$ 是 $\mathbb{R}^n$ 的 subspace, Proposition 4.3.15 告訴我們 $\mathbf{u}$ 就是 $\mathbf{v}$ 在 $N(A)$ 的 orthogonal projection. 也因此可得 $\mathbf{v} - \mathbf{u} \in N(A)^{\perp}$. 再由 Theorem 4.4.2, 知 $N(A)^{\perp} = \mathrm{Col}(A^{\mathrm{t}})$, 故知此 minimal solution $\mathbf{v} - \mathbf{u} \in \mathrm{Col}(A^{\mathrm{t}})$. 這表示 $\mathbf{v} - \mathbf{u}$ 會在 $n \times m$ 矩陣 $A^{\mathrm{t}}$ 的 column space, 因此存在 $\mathbf{x}' \in \mathbb{R}^m$ 使得 $\mathbf{v} - \mathbf{u} = A^{\mathrm{t}}\mathbf{x}'$. 此時由 $A(\mathbf{v} - \mathbf{u}) = A\mathbf{v} - A\mathbf{u} = \mathbf{b}$, 得 $(AA^{\mathrm{t}})\mathbf{x}' = A(A^{\mathrm{t}}\mathbf{x}') = A(\mathbf{v} - \mathbf{u}) = \mathbf{b}$. 所以我們只要解聯立方程組 $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$, 找出的解 $\mathbf{x}' \in \mathbb{R}^m$ 由於會使得 $A^{\mathrm{t}}\mathbf{x}' \in \mathrm{Col}(A^{\mathrm{t}})$ 且符合方程組 $A(A^{\mathrm{t}}\mathbf{x}') = \mathbf{b}$, 因此令 $\mathbf{x} = A^{\mathrm{t}}\mathbf{x}'$ 就會是 linear system $A\mathbf{x} = \mathbf{b}$ 的 minimal solution. 現在問題是 linear system $(AA^{\mathrm{t}})\mathbf{x}' = \mathbf{b}$ 會有解嗎?

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

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

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

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

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

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

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

**Example 4.5.9.** 我們用一個簡單的例子說明 Proposition 4.5.7 以及 Corollary 4.5.8. 考慮 $A=\begin{bmatrix}1&0&1\\1&1&2\\0&1&1\end{bmatrix}$, $\mathbf{b}=\begin{bmatrix}2\\2\\0\end{bmatrix}$. 很容易看出此 linear system $A\mathbf{x}=\mathbf{b}$ 有一組解 $\mathbf{v}_1=\begin{bmatrix}2\\0\\0\end{bmatrix}$ 由於 $\mathrm{rank}(A)=2$, 我們知到此方程組有無窮多解. 事實上此 linear system $A\mathbf{x}=\mathbf{b}$ 有另一組解 $\mathbf{v}_2=\begin{bmatrix}1\\-1\\1\end{bmatrix}$. 我們有 $\|\mathbf{v}_2\|=\sqrt{3}<\|\mathbf{v}_1\|=2$, 還有沒有長度更小的解呢? 我們要找到其 minimal solution. 由於 $AA^{\mathrm{t}}=\begin{bmatrix}2&3&1\\3&6&3\\1&3&2\end{bmatrix}$, 我們解 $(AA^{\mathrm{t}})\mathbf{x}'=\mathbf{b}$, 即 $\begin{bmatrix}2&3&1\\3&6&3\\1&3&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}2\\2\\0\end{bmatrix}$ 得 $\mathbf{x}'=\begin{bmatrix}2\\-2/3\\0\end{bmatrix}$ 為其一組解, 故知 $\mathbf{x}=A^{\mathrm{t}}\mathbf{x}'=\begin{bmatrix}1&1&0\\0&1&1\\1&2&1\end{bmatrix}\begin{bmatrix}2\\-2/3\\0\end{bmatrix}=\begin{bmatrix}4/3\\-2/3\\2/3\end{bmatrix}$ 為原方程組 $A\mathbf{x}=\mathbf{b}$ 的 minimal solution. 事實上 $N(A)=\mathrm{Span}(\begin{bmatrix}1\\1\\-1\end{bmatrix})$. 所以 $A\mathbf{x}=\mathbf{b}$ 所有的解為 $\mathbf{v}+r\mathbf{u}$ 的形式,

其中 $\mathbf{v} = \begin{bmatrix} 4/3 \\ -2/3 \\ 2/3 \end{bmatrix}, \mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$, $r \in \mathbb{R}$. 由於 $\langle \mathbf{v}, \mathbf{u} \rangle = 0$, 我們有 $\|\mathbf{v} + r\mathbf{u}\|^2 = \langle \mathbf{v} + r\mathbf{u}, \mathbf{v} + r\mathbf{u} \rangle = \|\mathbf{v}\|^2 + r^2\|\mathbf{u}\|^2 = (24/9) + 3r^2$. 知 $\mathbf{v}$ 確為 linear system $A\mathbf{x} = \mathbf{b}$ 之 minimal solution.

# 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.

## 5.1. Signed Area in $\mathbb{R}^2$ and Properties of Determinant Function

我們都知道當 $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathsf{M}_{2\times 2}(\mathbb{R})$, 令 $\mathbf{u} = (a,b)$, $\mathbf{v} = (c,d)$. 此時 $\det(A) = ad - bc$ 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 (也就是說將 $\mathbf{u}$ 逆時鐘旋轉某個小於 $180^\circ$ 的角度後會與 $\mathbf{v}$ 平行). 反之, $\det(A)$ 為負的表示 $\mathbf{v}$ 在 $\mathbf{u}$ 的順時針方向. 因此 $\det(A)$ 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*.

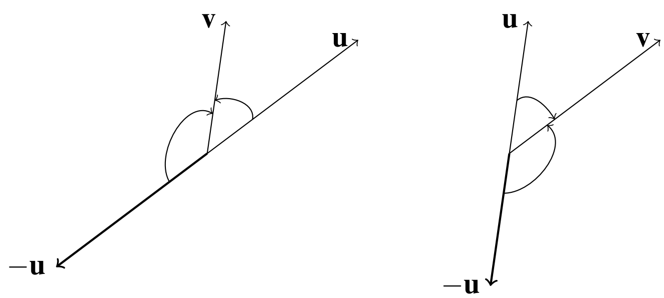
我們希望將此推廣到 $n \times n$ matrix. 也就是說當 $A \in \mathsf{M}_{n\times n}(\mathbb{R})$, 令 $\mathbf{v}_1,\ldots,\mathbf{v}_n \in \mathbb{R}^n$ 依序為 $A$ 的 row vectors. 我們希望能定義 $\det(A)$ 使其值為 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 在 $\mathbb{R}^n$ 所張成的“平行多面體”的 *signed volume*. 也就是說, 希望 $\det(A)$ 的絕對值表示 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 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 $n \geq 4$ 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 $\mathsf{M}_{n\times n}(\mathbb{R})$ 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 $2\times 2$ 的 determinant $\det\begin{bmatrix} a & b\\ c & d\end{bmatrix}=ad-bc=-(bc-ad)=-\det\begin{bmatrix} c & d\\ a & b\end{bmatrix}$ 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 $A\in \mathsf{M}_{n\times n}(\mathbb{R})$, 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 $2\times 2$ 的 determinant

$$\det\begin{bmatrix} -a & -b\\ c & d\end{bmatrix}=\det\begin{bmatrix} a & b\\ -c & -d\end{bmatrix}=-ad+bc=-(ad-bc)=-\det\begin{bmatrix} a & b\\ c & d\end{bmatrix}$$

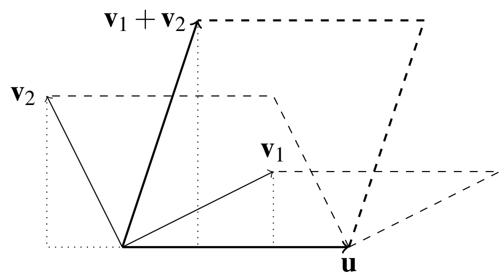
也符合這個性質. 因此我們認為 $\mathbb{R}^n$ 中 $n$ 個向量 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 若將其中一個 $\mathbf{v}_i$ 改為 $-\mathbf{v}_i$ 就會改變方向性. 也就是說當 $A\in \mathsf{M}_{n\times n}(\mathbb{R})$, 若將 $A$ 的某個 row 乘上 $-1$ 所得的矩陣為 $A'$, 則我們要求 $\det(A')=-\det(A)$.

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 $2\times 2$ 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 $A\in \mathsf{M}_{n\times n}(\mathbb{R})$ 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 $2\times 2$ 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 (2) 這個性質稱為 determinant 的 *alternating* 性質; 而 (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 5.1.** 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.

## 5.2. 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. 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, 這樣一直到將原本的 3-rd row 換到 6-th 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:

$$\begin{bmatrix} \vdots \\ -\mathbf{v}_3- \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_3- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ -\mathbf{v}_6- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix}$$

接著我們從下而上的將 5-th 和 4-th row 交換 (即原本 6-rd 已到達 4-th row 的位置), 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 $5-3=2$ 次的相鄰兩個 row 互換的動作圖示如下:

$$\begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_6- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_4- \\ -\mathbf{v}_6- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix} \to \begin{bmatrix} \vdots \\ -\mathbf{v}_6- \\ -\mathbf{v}_4- \\ -\mathbf{v}_5- \\ -\mathbf{v}_3- \\ \vdots \end{bmatrix}$$

在一般的情況, 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 $j$-th row. 注意此時原本 $i+1$-th row 到 $j$-th row 其實都只是往上移一個 row, 而我們共做了 $j-i$ 次的相鄰兩個 row 互換的動作. 接下來從 $j-1$-th row 開始, 先和 $j-2$-th row 交換 (此時原本的 $j$-th row 已換到 $j-2$-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 $j-1$-th row 和 $j-2$-th row 交換一直到 $i+1$-th row 和 $i$-th row 交換共做了 $(j-1)-i$ 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. 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 5.2.1.** *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 5.2.1, 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 5.2.1 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 5.2.1, 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 5.2.2.** *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 5.2.1 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 5.2.3.** 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 5.2.2 知 $\det(B)=0$. 得知 $\det(A')=\det(A)$, 因此我們有以下的性質.

**Lemma 5.2.4.** 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 5.2.4, 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 5.2.5.** 假設 $A$ 為 $n\times n$ *matrix.* 若 $E$ 為 *elementary matrix,* 則
$$\det(EA)=\det(E)\det(A).$$

Theorem 5.2.5 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 5.2.5 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
$$\det(E_k\cdots E_1A)=\det(E_k)\cdots\det(E_1)\det(A). \tag{5.1}$$
利用這些 elementary matrices 的 determinant, 我們有以下關於 determinant 的重要性質.

**Theorem 5.2.6.** 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 $\mathrm{rank}(A)<n$ (參見 Theorem 2.5.2). 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 5.2.3 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 (5.1) 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 2.5.7). 故存在 elementary matrices $E_1\ldots E_k$ 使得 $A=E_k\cdots E_1$. 因此由式子 (5.1) 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 2.5.5(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 (5.1) 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 $A^{\mathrm{t}}$ 也不是 invertible (參見 Proposition 2.5.5(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 2.2.4). 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, $E^{\mathrm{t}}$ 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 5.2.6 (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 5.2.5 也可改寫成以下的形式.

**Corollary 5.2.7.** 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 5.2.8.** 最多僅有唯一的函數 $\det:\mathsf{M}_{n\times n}(\mathbb{R})\to\mathbb{R}$ 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 $\det:\mathsf{M}_{n\times n}(\mathbb{R})\to\mathbb{R}$ 和 $\det':\mathsf{M}_{n\times n}(\mathbb{R})\to\mathbb{R}$ 皆滿足這四項規則. 考慮 $n\times n$ matrix $A$, 若 $A$ 不是 invertible, 則由 Theorem 5.2.6 (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 5.2.6 (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 $A\in\mathsf{M}_{n\times n}(\mathbb{R})$, 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 5.2.9.** 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 1.3), 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 $\det(A)=a_{11}\cdots a_{nn}\det(A')=a_{11}\cdots a_{nn}$ □

**Question 5.2.** 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 5.2.10.** 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 5.2.9 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

$$\det\begin{bmatrix} 0 & 2 & -1 & 1 \\ 1 & 2 & 0 & 2 \\ 1 & 4 & -2 & 6 \\ 2 & 6 & -1 & 8 \end{bmatrix} = -\det\begin{bmatrix} 1 & 2 & 0 & 2 \\ 0 & 2 & -1 & 1 \\ 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 3 \end{bmatrix} = 6.$$

## 5.3. Determinant of $3\times 3$ Matrix

我們利用上一節有關 determinant 的性質, 寫出 $3\times 3$ matrix 的 determinant 可能的形式, 從而證明 $3\times 3$ 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 *signed volume*.

在 Theorem 5.2.6 (3) 中, 我們知道 $\det(A^t)=\det(A)$. 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 $3\times 3$ 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 $\mathrm{M}_{3\times 3}$ 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 $3\times 3$ 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 $2\times 2$ 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

$$\begin{aligned}\det\begin{bmatrix} a_{11}+rb_{11} & a_{12}+rb_{12} & a_{13}+rb_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{bmatrix} &= (a_{11}+rb_{11})\det\begin{bmatrix} a_{22} & a_{23}\\ a_{32} & a_{33}\end{bmatrix} -\\ a_{21}\det\begin{bmatrix} a_{12}+rb_{12} & a_{13}+rb_{13}\\ a_{32} & a_{33}\end{bmatrix} &+ a_{31}\det\begin{bmatrix} a_{12}+rb_{12} & a_{13}+rb_{13}\\ a_{22} & a_{23}\end{bmatrix}.\end{aligned}\tag{5.2}$$

而 $2\times 2$ 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 (5.2) 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 $3\times 3$ 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 $2\times 2$ matrix 的 determinant 存在來證明 $3\times 3$ 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 5.3.1.** 假設 $A=[a_{ij}]$ 為 $3\times 3$ matrix. 將 $A$ 的 $i$-th row 和 $j$-th column 除去所得的 $2\times 2$ 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 5.2.8), 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 $\det(A)=\det(A^t)$, 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 $2\times 2$ 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 $\det(A)=\det(A^t)=a_{11}a'_{11}+a_{12}a'_{12}+a_{13}a'_{13}$, 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 5.3.2.** 假設 $A=[a_{ij}]$ 為 $3\times 3$ *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 $3\times 3$ 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 signed 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 $3\times 3$ 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 signed 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 5.1 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 $\mathbf{v}\times\mathbf{u}$ 是不相等的, 除非 $\mathbf{u}\times\mathbf{v}=\mathbf{0}$. 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 $\mathbf{0}$ 呢? 依定義 $\mathbf{u}\times\mathbf{v}=\mathbf{0}$ 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

$$\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v})=c_1\det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3\end{bmatrix}+c_2\det\begin{bmatrix} a_3 & a_1 \\ b_3 & b_1\end{bmatrix}+c_3\det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2\end{bmatrix} \tag{5.3}$$

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 (5.3) 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

$$\mathbf{w}\cdot(\mathbf{u}\times\mathbf{v}) = (\mathbf{u}\times\mathbf{v})\cdot\mathbf{w} = \det\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} = \det\begin{bmatrix} -\mathbf{u}- \\ -\mathbf{v}- \\ -\mathbf{w}- \end{bmatrix}. \tag{5.4}$$

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 signed 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 5.2.2). 因此由式子 (5.4) 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 signed 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 signed 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 5.3.3.** 給定 $\mathbf{u},\mathbf{v}\in\mathbb{R}^3$. 則 $\mathbf{u}\times\mathbf{w}\neq\mathbf{0}$ 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 *signed volume.*

## 5.4. Existence of the Determinant Function

在上一節中, 我們利用降階的方法以及 $2\times2$ matrix 的 determinant 的存在性建構了 $3\times3$ matrix 的 determinant, 因而得到其存在性. 接著我們可利用 $3\times3$ matrix 的 determinant 存在性得到 $4\times4$ 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 5.3.1 的定義推廣到一般的情形.

**Definition 5.4.1.** 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 5.2.8 的唯一性, 我們有以下的結論.

**Theorem 5.4.2.** 存在唯一的函數 $\det:\mathsf{M}_{n\times n}(\mathbb{R})\to\mathbb{R}$ 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 $3\times 3$ 的情形相同, 由於 $\det(A^{t})=\det(A)$, 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 5.4.3.** 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 5.3.** 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 5.4.4.** 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 $\det(A)=\det(B)=2\det(C)=2\det(D)=4$.

## 5.5. 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

$$AC_k=\begin{bmatrix}| & & |\\ \mathbf{a}_1 & \cdots & \mathbf{a}_n\\ | & & |\end{bmatrix}\begin{bmatrix}| & & c_1 & & |\\ \mathbf{e}_1 & \cdots & \vdots & \cdots & \mathbf{e}_n\\ | & & c_n & & |\end{bmatrix}=\begin{bmatrix}| & & | & & |\\ \mathbf{a}_1 & \cdots & c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n & \cdots & \mathbf{a}_n\\ | & & | & & |\end{bmatrix}. \tag{5.5}$$

也就是說當 $j\neq k$ 時, $AC_k$ 的 $j$-th column 為 $\mathbf{a}_j$, 而 $AC_k$ 的 $k$-th column 為 $c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n$. 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

$$c_1\mathbf{a}_1+\cdots+c_n\mathbf{a}_n=\begin{bmatrix}| & & |\\ \mathbf{a}_1 & \cdots & \mathbf{a}_n\\ | & & |\end{bmatrix}\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}=\begin{bmatrix}b_1\\ \vdots\\ b_n\end{bmatrix}=\mathbf{b}. \tag{5.6}$$

因此若令 $B_k$ 表示將 $A$ 的 $k$-th column 用 $\mathbf{b}$ 取代的 $n\times n$ matrix, 即

$$B_k=\begin{bmatrix}| & & | & & |\\ \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n\\ | & & | & & |\end{bmatrix},$$

則結合式子 (5.5) (5.6), 我們有 $AC_k=B_k$. 因此由 determinant 的乘法性質 (Theorem 5.2.6 (2)), 得 $\det(A)\det(C_k)=\det(B_k)$. 然而對 $C_k$ 的 $k$-th row 展開, 我們有 $\det(C_k)=(-1)^{k+k}c_k\det(I_{n-1})=c_k$. 因此得證以下之定理.

**Theorem 5.5.1.** *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 5.5.1 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 5.5.1 我們可以將此組解具體寫出. 這就是所謂的 *Cramer's Rule.*

**Corollary 5.5.2** (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 5.5.1 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 5.5.3.** 考慮 Example 5.4.4 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 5.5.1 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 5.5.1 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 5.5.4.** 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 5.5.4 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 5.5.5.** *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. $\square$

另外一個 Cramer's rule 的應用就是幫我們找到 invertible matrix 的 inverse. 假設 $A \in \mathsf{M}_{n \times n}(\mathbb{R})$ 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 $\det(A(j,i)) = (-1)^{j+i}\det(A_{ji}) = a'_{ji}$. 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 5.5.6.** 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 5.5.7.** 假設 $A$ 為 $n\times n$ *invertible matrix.* 則

$$A^{-1}=\frac{1}{\det(A)}\operatorname{adj}(A).$$

**Example 5.5.8.** 考慮 Example 5.4.4 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 5.5.3 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 5.5.3 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 $\det(A)=\pm1$, 則由 Proposition 5.5.7 我們有以下的結果.

**Corollary 5.5.9.** 假設 $A$ 為 $n\times n$ *matrix* 其中 $A$ 的每一個 *entry* 皆為整數. 若 $\det(A)=\pm1$, 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 5.5.7 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 5.2.6 (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 5.5.10.** 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 $A^{\mathrm{t}}$ 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 $A^{\mathrm{t}}$ 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 $A^{\mathrm{t}}$ 的 $(i,j)$ cofactor 就是 $A$ 的 $(j,i)$ cofactor. 因此將 $A$ 的 cofactor matrix $A'$ 取轉置就是 $A^{\mathrm{t}}$ 的 cofactor matrix. 換言之, 若我們先將 $A$ 轉置得 $A^{\mathrm{t}}$ 再取 $A^{\mathrm{t}}$ 的 cofactor matrix 也會是 $\operatorname{adj}(A)$. 也因此若將 $A^{\mathrm{t}}$ 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 $\operatorname{adj}(A)$, 因此得 $\operatorname{adj}(A)=(\operatorname{adj}(A^{\mathrm{t}}))^{\mathrm{t}}$. 再取轉置得證以下的結果.

**Proposition 5.5.11.** 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 5.5.7 我們知 $\operatorname{adj}(A)=\det(A)A^{-1}$. 因此由矩陣乘法與係數積關係 (Proposition 2.1.10) 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. 首先考慮一般的 $n\times n$ 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

$$\sum_{k=1}^{n}a_{ik}a'_{jk}=\sum_{k=1}^{n}a_{ik}(-1)^{j+k}\det(A_{jk}).\tag{5.7}$$

當 $i=j$ 時, 式子 (5.7) 恰為對 $A$ 的 $i$-th row 展開所得的 $\det(A)$. 因此得 $A(\operatorname{adj}(A))$ 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 (5.7) 是一致的. 也就是說當 $i \neq j$ 時, $A(\operatorname{adj}(A))$ 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, $A(\operatorname{adj}(A))$ 的 $(i,j)$-th entry 等於 0, 也因此證得了 $A(\operatorname{adj}(A))=\det(A)I_n$. 利用類似的方法考慮 column 的展開, 我們也可證得 $(\operatorname{adj}(A))A=\det(A)I_n$. 不過我們這裡想用前面有關 adjoint 和 transpose 的關係 (Proposition 5.5.11) 證明以下的定理.

**Theorem 5.5.12.** 假設 $A$ 為 $n \times n$ *matrix,* 則

$$A(\operatorname{adj}(A))=(\operatorname{adj}(A))A=\det(A)I_n.$$

**Proof.** 前面已證得 $A(\operatorname{adj}(A))=\det(A)I_n$. 因為這是對所有 $n \times n$ matrix 皆成立, 故將 $A$ 以 $A^{\mathrm{t}}$ 代換得 $A^{\mathrm{t}}(\operatorname{adj}(A^{\mathrm{t}}))=\det(A^{\mathrm{t}})I_n$. 由 Proposition 5.5.11 以及 $\det(A^{\mathrm{t}})=\det(A)$, 我們得 $A^{\mathrm{t}}(\operatorname{adj}(A)^{\mathrm{t}})=\det(A)I_n$. 將此等式兩邊取 transpose 得證 $(\operatorname{adj}(A))A=\det(A)I_n$. □

再次強調, Proposition 5.5.7 的證明是利用 Cramer's rule 因此需用到 $\det(A) \neq 0$ 之假設. 而 Theorem 5.5.12 是對一般 $n \times n$ matrix 皆成立的. 然而我們可以利用 Theorem 5.5.12 推得 Proposition 5.5.7, 所以可以說 Theorem 5.5.12 是 Proposition 5.5.7 的推廣.

**Example 5.5.13.** 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

$$\operatorname{adj}(A)=\begin{bmatrix} \det\begin{bmatrix} b_2 & b_3 \\ c_2 & c_3 \end{bmatrix} & -\det\begin{bmatrix} a_2 & a_3 \\ c_2 & c_3 \end{bmatrix} & \det\begin{bmatrix} a_2 & a_3 \\ b_2 & b_3 \end{bmatrix} \\ -\det\begin{bmatrix} b_1 & b_3 \\ c_1 & c_3 \end{bmatrix} & \det\begin{bmatrix} a_1 & a_3 \\ c_1 & c_3 \end{bmatrix} & -\det\begin{bmatrix} a_1 & a_3 \\ b_1 & b_3 \end{bmatrix} \\ \det\begin{bmatrix} b_1 & b_2 \\ c_1 & c_2 \end{bmatrix} & -\det\begin{bmatrix} a_1 & a_2 \\ c_1 & c_2 \end{bmatrix} & \det\begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \end{bmatrix}.$$

依矩陣乘法定義 $A(\operatorname{adj}(A))$ 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 5.2.9 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 5.5.14.** *假設 $A$ 為 $n \times n$ matrix, 則*

$$\det(\mathrm{adj}(A)) = \det(A)^{n-1}.$$

## 5.6. 結論

Determinant 是 square matrix 上特有的一種函數. 它是從 $\mathsf{M}_{n\times n}(\mathbb{R})$ 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 6

# Linear Transformations

在本章中我們介紹在 vector spaces 之間重要的函數, 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.

## 6.1. 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.

**6.1.1. Function.** 給定 vector spaces $V$, $W$. 若有一個從 $V$ 的向量對應到 $V$ 的向量的對應關係, 即對任意 $\mathbf{v} \in V$, 此對應會將 $\mathbf{v}$ 對應到 $W$ 中一個向量 $\mathbf{w}$. 若對每一個 $\mathbf{v} \in V$, 所對應到的 $\mathbf{w}$ 我們用 $T(\mathbf{v})$ 來表示, 我們就用 $T : V \to W$, $T(\mathbf{v}) = \mathbf{w} \in W$, $\forall \mathbf{v} \in V$ 來表示這一個對應關係, 而稱 $T$ 是一個從 $V$ 到 $W$ 的 *function* (函數). 這裡 $V$ 稱為 $T$ 的 *domain* (定義域), 而 $W$ 就稱為 $T$ 的 *codomain* (對應域). 注意依函數的定義, 若 $T : V \to W$ 是一個函數, 則對任意 $\mathbf{v} \in V$, $T(\mathbf{v})$ 一定要是 $W$ 中的一個確定的向量. 也就是說 $T(\mathbf{v}) \in W$, 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 $T : V \to W$ 是一個函數, 是容許在 $V$ 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 $V$ 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 $T : V \to W$ 是一個函數, 是容許在 $W$ 中有向量 $\mathbf{w}$, 沒有任何 $V$ 的向量會對應到 $\mathbf{w}$ (即不存在 $\mathbf{v} \in V$ 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 $W$ 中任意向量 $\mathbf{w}$, 皆可找到 $\mathbf{v}\in V$ 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).

**6.1.2. Linear Transformation.** 要了解 vector spaces, 若僅是考慮一般的函數, 並無法利用向量之間的運算, 幫助我們了解這些 vector spaces 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 6.1.1.** 假設 $V,W$ 皆為 vector space over $\mathbb{F}$ 且 $T:\mathbb{R}^n\to\mathbb{R}^m$ 為一個函數, 若 $T$ 滿足對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in V$ 以及 $c_1,\ldots,c_k\in\mathbb{F}$ 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 $V$ 中向量的線性組合, 而 $c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)$ 是 $W$ 中向量的線性組合, 要區分清楚. 尤其在 $V\neq W$ 時要特別注意. 特別是當 $\mathbf{0}\in V$ 是 $V$ 的 zero vector 時, 依 linear transformation 的定義, 我們有 $T(\mathbf{0})=T(\mathbf{0}+\mathbf{0})=T(\mathbf{0})+T(\mathbf{0})$. 此時兩邊加上 $T(\mathbf{0})$ 的加法反元素, 得 $T(\mathbf{0})$ 應為 $W$ 中的 zero vector. 也就是說一個 linear transformation $T:V\to W$, 會將 $V$ 中的零向量映射到 $W$ 中的零向量. 雖然當 $V\neq W$ 時, $V$ 和 $W$ 的零向量有可能是不同的, 不過一般我們都用 $\mathbf{0}$ 來表示, 而不區分它. 所以我們仍用 $T(\mathbf{0})=\mathbf{0}$ 來表示 linear transformation 會將 $V$ 中的零向量映射到 $W$ 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 6.1.2.** 假設 $T:V\to W$ 為一個 *linear transformation.* 則 $T$ 會將 $V$ 中的零向量映射到 $W$ 中的零向量, 亦即 $T(\mathbf{0})=\mathbf{0}$.

再次提醒, 這裡兩個 $\mathbf{0}$ 哪一個是 $V$ 的零向量, 哪一個是 $W$ 的零向量, 一定要區分清楚.

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 Corollary 3.3.3), 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 6.1.3.** 假設 $V,W$ 皆為 *vector space over* $\mathbb{F}$ 且 $T:V\to W$ 為一個函數, 則 $T$ 為 *linear transformation* 若且唯若對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{F}$ 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 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{F}$ 皆有 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$.

$(\Leftarrow)$: 我們要利用對任意 $\mathbf{u},\mathbf{v}\in V$, $r\in\mathbb{F}$ 皆有 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$ 這個性質來證明對任意 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in V$ 以及 $c_1,\ldots,c_k\in\mathbb{F}$ 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 $\mathbf{u}=\mathbf{0}$, $\mathbf{v}=\mathbf{v}_1$, $r=c_1$, 依 Lemma 6.1.2, 我們有 $T(c_1\mathbf{v}_1)=T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})=\mathbf{0}+rT(\mathbf{v})=c_1T(\mathbf{v}_1)$. 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 $\mathbf{v}_1,\ldots,\mathbf{v}_k\in V$ 以及 $c_1,\ldots,c_k\in\mathbb{F}$ 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 $\mathbf{v}_1,\ldots,\mathbf{v}_k,\mathbf{v}_{k+1}\in V$ 以及 $c_1,\ldots,c_k,c_{k+1}\in\mathbb{F}$ 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

$$
\begin{aligned}
&T(c_1\mathbf{v}_1+\cdots+c_k\mathbf{v}_k+c_{k+1}\mathbf{v}_{k+1})=\\
&\qquad T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})=c_1T(\mathbf{v}_1)+\cdots+c_kT(\mathbf{v}_k)+c_{k+1}T(\mathbf{v}_{k+1}).
\end{aligned}
$$

故由數學歸納法知 $T$ 為 linear transformation. □

**Example 6.1.4.** (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 $T(\mathbf{0})=\begin{bmatrix}1\\0\end{bmatrix}\neq\mathbf{0}$, 故由 Lemma 6.1.2 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 $T(\mathbf{0})=\mathbf{0}$, 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 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformation 都是這樣的形式.

**Lemma 6.1.5.** *假設 $\mathbb{F}$ 為 field 且 $A \in M_{m\times n}(\mathbb{F})$. 考慮 $T : \mathbb{F}^n \to \mathbb{F}^m$ 定義為: $T(\mathbf{v}) = A\mathbf{v}$, $\forall \mathbf{v} \in V$. 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 $\mathbb{F}^n$ 映到 $\mathbb{F}^m$ 的函數. 任取 $\mathbf{v} \in \mathbb{F}^n$, 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 $\mathbf{v} \in \mathbb{F}^n$ 為 $n\times 1$ matrix), 故 $A\mathbf{v} \in \mathbb{F}^m$. 得 $T$ 確實是一個從 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 function.

現要證明 $T$ 為 linear, 亦即對任意 $\mathbf{u},\mathbf{v} \in \mathbb{F}^n$ 以及 $r \in \mathbb{F}$, 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 2.1.9 和 Proposition 2.1.10) 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 6.1.5 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 $V$ 到 $W$ 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 $V$ 到 $W$ 的函數. 對於任意 $\mathbf{v} \in V$, 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 $V$ 的向量映射到 $W$ 中. 這是因為依假設, 對任意 $\mathbf{v} \in V$, 我們有 $T_1(\mathbf{v}) \in W$ 以及 $T_2(\mathbf{v}) \in W$, 所以自然有 $(T_1+T_2)(\mathbf{v}) = T_1(\mathbf{v})+T_2(\mathbf{v}) \in W$. 所以 $T_1+T_2 : V \to W$ 確實是 well-defined function. 接下來我們要說明若 $T_1 : V \to W$, $T_2 : V \to W$ 皆為 linear transformation, 則 $T_1+T_2 : V \to W$ 亦為 linear transformation. 也就是說對於任意 $\mathbf{u},\mathbf{v} \in V$ 以及 $r \in \mathbb{F}$, 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 $V$ 到 $W$ 的 linear transformation.

**Question 6.1.** *若 $A_1,A_2$ 皆為 $m\times n$ matrix. 考慮 $T_1 : \mathbb{F}^n \to \mathbb{F}^m$, $T_2 : \mathbb{F}^n \to \mathbb{F}^m$, 分別定義為 $T_1(\mathbf{v}) = A_1\mathbf{v}$, $T_2(\mathbf{v}) = A_2\mathbf{v}$, $\forall \mathbf{v} \in \mathbb{F}^n$. 則 $T_1+T_2$ 是怎樣的函數?*

給定一個 linear transformation $T:V\to W$, 以及 $c\in\mathbb{F}$, 我們也可定義函數 $cT:V\to W$, 其定義為 $(cT)(\mathbf{v})=c(T(\mathbf{v}))$, $\forall\mathbf{v}\in V$ (也就是說它把每一個 $V$ 的向量 $\mathbf{v}$ 對應到 $c$ 倍的 $T(\mathbf{v})$). 很容易看出 $rT:V\to W$ 確實是一個 function. 事實上, 它也是 linear transformation. 這是因為對於任意 $\mathbf{u},\mathbf{v}\in V$ 以及 $r\in\mathbb{F}$, 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 $cT:V\to W$ 為 linear transformation.

**Question 6.2.** 設 $A$ 為 $m\times n$ *matrix.* 考慮 $T:\mathbb{F}^n\to\mathbb{F}^m$ 定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$. 則對於 $c\in\mathbb{F}$, $cT$ 是怎樣的函數?

設 $T_1,\cdots,T_k$ 為 $V$ 到 $W$ 的 linear transformations. $c_1,\dots,c_k\in\mathbb{F}$, 則由前知 $c_1T_1,\dots,c_kT_k$ 皆為 $V$ 到 $W$ 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 6.1.6.** 設 $T_1,\cdots,T_k$ 為 $V$ 到 $W$ 的 *linear transformations,* $c_1,\dots,c_k\in\mathbb{F}$ 則 $c_1T_1+\cdots+c_kT_k$ 為 $V$ 到 $W$ 的 *linear transformation.*

另一個產生 linear transformation 的方法就是利用“合成函數”. 若 $U,V,W$ 為 vector spaces 且 $T:U\to V$ 和 $T':V\to W$ 為函數, 由於對任意 $\mathbf{u}\in U$, 依定義 $T(\mathbf{u})\in V$, 也就是說 $T(\mathbf{u})$ 會落在 $T'$ 的定義域中. 所以我們可以將 $T(\mathbf{u})$ 代入 $T'$ 中, 亦即得 $T'(T(\mathbf{u}))\in W$. 這樣的方法幫我們定義出一個從 $U$ 到 $W$ 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 $T'\circ T:U\to W$ 的定義為 $T'\circ T(\mathbf{u})=T'(T(\mathbf{u}))$, $\forall\mathbf{u}\in U$. 我們有下面之結果.

**Proposition 6.1.7.** 假設 $T:U\to V$ 和 $T':V\to W$ 為 *linear transformation,* 則 $T'\circ T:U\to W$ 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 $\mathbf{u},\mathbf{v}\in U$ 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 6.3.** 設 $A$ 為 $m\times n$ *matrix,* $B$ 為 $k\times m$ *matrix.* 考慮 $T:\mathbb{F}^n\to\mathbb{F}^m$, $T':\mathbb{F}^m\to\mathbb{F}^k$, 分別定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$ 且 $T'(\mathbf{w})=B\mathbf{w}$, $\forall\mathbf{w}\in\mathbb{F}^m$. 則 $T'\circ T$ 是怎樣的函數?

當 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是 $V$ 的一組 basis 時, 對任意 $\mathbf{v}\in V$, 皆存在唯一的一組 $c_1,\dots,c_n\in\mathbb{F}$, 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 現若 $T:V\to W$ 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 $W$ 中的哪些向量, 則對於任意 $\mathbf{v}\in V$, 我們都可以知道 $T(\mathbf{v})$ 為何. 因此我們有以下的定理.

**Theorem 6.1.8.** 假設 $V,W$ 為 *vector spaces over* $\mathbb{F}$ 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$, 為 $V$ 的一組 *basis.* 令 $\mathbf{w}_1,\ldots,\mathbf{w}_n\in W$, 則存在唯一的 *linear transformation* $T:V\to W$, 滿足 $T(\mathbf{v}_1)=\mathbf{w}_1,\ldots,T(\mathbf{v}_i)=\mathbf{w}_i,\ldots,T(\mathbf{v}_n)=\mathbf{w}_n$.

**Proof.** 首先證明存在性. 定義 $T:V\to W$ 為 $T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n$, $\forall c_1,\ldots,c_n\in F$. 我們需說明這是 well-defined function. 也就是說對任意 $\mathbf{v}\in V$, $T(\mathbf{v})$ 皆有定義且 $T(\mathbf{v})\in W$. 然而因 $\mathbf{v}_1,\ldots,\mathbf{v}_n$, 為 $V$ 的一組 basis, 故存在唯一的一組 $c_1,\ldots,c_n\in\mathbb{F}$, 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 故此時得 $T(\mathbf{v})=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n\in W$. 接著我們要說明 $T$ 為 linear transformation, 也就是說對任意 $\mathbf{u},\mathbf{v}\in V$ 以及 $r\in\mathbb{F}$, 我們要證明 $T(\mathbf{u}+r\mathbf{v})=T(\mathbf{u})+rT(\mathbf{v})$. 由於 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 $V$ 的一組 basis, 存在 $c_1,\ldots,c_n$ 以及 $d_1,\ldots,d_n\in\mathbb{F}$ 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

$$\begin{aligned}T(\mathbf{u})+rT(\mathbf{v})&=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)+rT(d_1\mathbf{v}_1+\cdots+d_n\mathbf{v}_n)=\\&(c_1\mathbf{w}_1+\cdots+c_n\mathbf{w}_n)+r(d_1\mathbf{w}_1+\cdots+d_n\mathbf{w}_n)=(c_1+rd_1)\mathbf{w}_1+\cdots+(c_n+rd_n)\mathbf{w}_n.\end{aligned}$$

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 $T':V\to W$ 是另一個 linear transformation 滿足 $T'(\mathbf{v}_1)=\mathbf{w}_1,\ldots,T'(\mathbf{v}_n)=\mathbf{w}_n$, 且 $T'\neq T$, 則會造成矛盾. 依定義, $T'\neq T$ 表示存在 $\mathbf{v}\in V$ 使得 $T'(\mathbf{v})\neq T(\mathbf{v})$. 此時因存在 $c_1,\ldots,c_n$, 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 6.1.8 中的 $\mathbf{w}_1,\ldots,\mathbf{w}_n\in W$ 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 $V$ 的一組 basis 後, 我們可以將這組 basis 裡的向量對應到 $W$ 中任意的向量, 就會得到一個 $V$ 到 $W$ 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 6.1.8 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.

## 6.2. Range and Null Space

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, “null space” 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 $T: V \to W$, 任取定義域 $V$ 中的子集合 $S$, 我們很自然的會考慮 $T(S) = \{T(\mathbf{v}) \in W \mid \mathbf{v} \in S\}$ 這一個集合, 它就是將所有 $S$ 中的元素利用 $T$ 映射到 $W$ 後的元素收集起來所得的集合. 同樣的, 對於對應域 $W$ 中的子集合 $S'$, 我們也會考慮 $T^{-1}(S') = \{\mathbf{v} \in V \mid T(\mathbf{v}) \in S'\}$ 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 $W$ 中的子集合, 而 $T^{-1}(S')$ 會是定義域 $V$ 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 $T(S) = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}), \text{ for some } \mathbf{v} \in S\}$, 有時為了強調 $T(S)$ 為 $W$ 的子集合, 我們也會用這種表示法.

由於 linear transformation 的特色, 當 $T: V \to W$ 為 linear transformation 時, 我們專注於考慮 $V'$ 為 $V$ 的 subspace 時的情況. 也就是說我們要了解

$$T(V') = \{T(\mathbf{v}) \in W \mid \mathbf{v} \in V'\} = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}), \text{ for some } \mathbf{v} \in V'\}$$

的特性. 同樣的若 $W'$ 為 $W$ 的 subspace, 我們也要了解

$$T^{-1}(W') = \{\mathbf{v} \in V \mid T(\mathbf{v}) \in W'\}$$

的特性. 事實上, 我們有以下之結果.

**Proposition 6.2.1.** 假設 $V, W$ 皆為 *vector space over* $\mathbb{F}$ 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.*

**Proof.** 依定義我們知 $T(V')$ 會是 $W$ 的子集合, 而 $T^{-1}(W')$ 會是 $V$ 的子集合. 故現僅需證明它們為 subspaces, 即利用 Corollary 3.3.3 我們要證明, 若 $\mathbf{w}, \mathbf{w}' \in T(V')$ 且 $r \in \mathbb{F}$, 則 $\mathbf{w} + r\mathbf{w}' \in T(V')$ 以及若 $\mathbf{v}, \mathbf{v}' \in T^{-1}(W')$ 且 $r \in \mathbb{F}$, 則 $\mathbf{v} + r\mathbf{v}' \in T^{-1}(W')$.

首先再強調一次, 當我們說 $W$ 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 $r \in \mathbb{F}$, 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 $V$ 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 $r \in \mathbb{F}$, 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 $W$ 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 $V' = V$ 和 $W' = \{\mathbf{0}\}$ 這兩個特殊情況時, 即

$$T(V) = \{\mathbf{w} \in W \mid \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\} \quad \text{and} \quad T^{-1}(\{\mathbf{0}\}) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}$$

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 6.2.2.** 假設 $V, W$ 皆為 vector space over $\mathbb{F}$ 且 $T: V \to W$ 為 linear transformation.

(1) 我們稱 $W$ 的 subspace $T(V)$ 為 $T$ 的 *range* (也稱為 *image*). 通常我們用 $\mathsf{R}(T)$ (或 $\mathrm{im}(T)$) 來表示 $T$ 的 range.

(2) 我們稱 $V$ 的 subspace $T^{-1}(\{\mathbf{0}\})$ 為 $T$ 的 *null space* (也稱為 *kernel*), 通常我們用 $\mathsf{N}(T)$ (或 $\ker(T)$) 來表示 $T$ 的 null space.

首先我們來看 linear transformation 的 range. 假設 $T : V \to W$ 為 linear transformation. 由 Proposition 6.2.1 我們知 $T$ 的 range $\mathsf{R}(T)$ 是 $W$ 的 subspace, 故知 $\dim(\mathsf{R}(T)) \leq \dim(W)$. 若 $\dim(\mathsf{R}(T)) = \dim(W)$, 則依 Proposition 3.6.10 知此時 $\mathsf{R}(T) = W$. 也就是說對於任意的 $\mathbf{w} \in W$, 由於 $\mathbf{w} \in \mathsf{R}(T)$ 依定義存在 $\mathbf{v} \in V$ 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 $\mathbf{w} \in W$, 存在 $\mathbf{v} \in V$ 使得 $T(\mathbf{v}) = \mathbf{w}$ 故得 $\mathbf{w} \in \mathsf{R}(T)$, 得證 $W \subseteq \mathsf{R}(T)$. 再利用已知 $\mathsf{R}(T) \subseteq W$, 得證 $\mathsf{R}(T) = W$. 我們有以下的性質.

**Proposition 6.2.3.** *假設 $T : V \to W$ 為 linear transformation. 則 $T$ 為 onto 若且唯若 $\dim(\mathsf{R}(T)) = \dim(W)$.*

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 6.2.3 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 6.2.4.** *假設 $V, W$ 皆為 vector space over $\mathbb{F}$ 且 $T : V \to W$ 為 linear transformation 且 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 為 $V$ 的一組 spanning vectors. 則*

$$\mathsf{R}(T) = \mathrm{Span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)).$$

**Proof.** 設 $\mathbf{w} \in \mathsf{R}(T)$, 表示存在 $\mathbf{v} \in V$ 使得 $\mathbf{w} = T(\mathbf{v})$. 又因 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 $V$ 的一組 spanning vectors, 故存在 $c_1, \cdots, c_n \in \mathbb{F}$, 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 $\mathsf{R}(T) \subseteq \mathrm{Span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n))$.

另一方面, 設 $\mathbf{w} \in \mathrm{Span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n))$, 表示存在 $c_1, \cdots, c_n \in \mathbb{F}$, 使得 $\mathbf{w} = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)$. 因此利用 $T$ 為 linear 得

$$\mathbf{w} = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n) = T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \in \mathsf{R}(T),$$

得證 $\mathrm{Span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)) \subseteq \mathsf{R}(T)$. 因此證明了 $\mathsf{R}(T) = \mathrm{Span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n))$. $\square$

**Example 6.2.5.** (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 6.1.4 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 6.2.4 我們有 $\mathsf{R}(T) = \mathrm{Span}(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\-1\end{bmatrix}) = \mathbb{R}^2$. 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 6.2.4 我們有 $\mathsf{R}(T) = \mathrm{Span}(\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\-1\end{bmatrix})$. 由於 $\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}0\\1\\-1\end{bmatrix}$ 為 linearly independent, 故得 $\dim(\mathsf{R}(T)) = 2$. 由 Proposition 6.2.3 知 $T$ 不是 onto.

**Question 6.4.** 假設 $T : V \to W$ 為 *linear transformation.* 若 $\dim(W) > \dim(V)$, 則 $T$ 有可能是 *onto* 嗎?

**Question 6.5.** 假設 $\mathbb{F}$ 是 *field* 且 $A \in M_{m\times n}(\mathbb{F})$, 考慮 *linear transformation* $T_A : \mathbb{F}^n \to \mathbb{F}^m$ 定義為 $T_A(\mathbf{v}) = A\mathbf{v}$, $\forall \mathbf{v} \in \mathbb{F}^n$. 考慮 $\mathbb{F}^n$ 的 *standard basis* $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. 試說明 $\mathsf{R}(T_A)$ *(*即 $T_A$ 的 *rang)* 和 $\mathrm{Col}(A)$ *(*即 $A$ 的 *column space)* 之間的關係.

要注意是有可能一個 linear transformation $T$ 的 range 為 $\{\mathbf{0}\}$. 此表示 $T$ 將所有定義域的向量皆映射到 $\mathbf{0}$, 亦即 $T(\mathbf{v}) = \mathbf{0}$, $\forall \mathbf{v} \in V$. 這樣的 linear transformation, 我們依慣例, 仍稱之為 *zero mapping.*

**Question 6.6.** 假設 $T : V \to W$ 為 *linear transformation* 且 $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ 為 $V$ 的一組 *basis.* 試證明 $T$ 為 *zero mapping* 若且唯若 $T(\mathbf{v}_1) = \cdots = T(\mathbf{v}_n) = \mathbf{0}$.

接下來我們看 null space 與 linear transformation 的關係. 假設 $T : V \to W$ 為 linear transformation. 若 $T$ 為 one-to-one, 由於已知 $T(\mathbf{0}) = \mathbf{0}$, 故知不可能有非零的向量 $\mathbf{v}$ 使得 $T(\mathbf{v}) = \mathbf{0}$. 因此可得 $\mathsf{N}(T) = \{\mathbf{0}\}$. 其實反過來 $\mathsf{N}(T) = \{\mathbf{0}\}$, 也會使得 $T$ 為 one-to-one, 我們有以下的結果.

**Proposition 6.2.6.** 假設 $T : V \to W$ 為 *linear transformation.* 則 $T$ 為 *one-to-one* 若且唯若 $\dim(\mathsf{N}(T)) = 0$, 亦即 $\mathsf{N}(T) = \{\mathbf{0}\}$.

**Proof.** 我們已知當 $T$ 為 one-to-one 時, 不會有非零向量映射到 $\mathbf{0}$, 故知 $\mathsf{N}(T) = \{\mathbf{0}\}$, 得 $\dim(\mathsf{N}(T)) = 0$. 反之, 當 $\dim(\mathsf{N}(T)) = 0$, 即 $\mathsf{N}(T) = \{\mathbf{0}\}$, 此時若 $T$ 不是 one-to-one, 表示存在 $\mathbf{v}, \mathbf{v}' \in V$ 滿足 $\mathbf{v} \neq \mathbf{v}'$ 但 $T(\mathbf{v}) = T(\mathbf{v}')$. 由於 $T$ 為 linear, 得 $T(\mathbf{v} - \mathbf{v}') = T(\mathbf{v}) - T(\mathbf{v}') = \mathbf{0}$, 亦即 $\mathbf{v} - \mathbf{v}' \in T^{-1}(\{\mathbf{0}\}) = \mathsf{N}(T) = \{\mathbf{0}\}$. 得到 $\mathbf{v} = \mathbf{v}'$ 之矛盾, 故證得 $T$ 為 one-to-one. □

**Example 6.2.7.** 我們探討 Example 6.2.5 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 $\mathbf{v}=\begin{bmatrix} a\\ b\\ c\end{bmatrix}\in \mathsf{N}(T)$, 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 $\begin{bmatrix} 1\\ -1\\ 1\end{bmatrix}\in\mathsf{N}(T)$, 知 $\mathsf{N}(T)\neq\{\mathbf{0}\}$ (事實上 $\mathsf{N}(T)=\mathrm{Span}(\begin{bmatrix} 1\\ -1\\ 1\end{bmatrix})$). 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 $\mathbf{v}=\begin{bmatrix} a\\ b\end{bmatrix}\in\mathsf{N}(T)$, 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 $\mathsf{N}(T)=\{\mathbf{0}\}$, 所以由 Proposition 6.2.6 知 $T$ 是 one-to-one.

要注意 Proposition 6.2.6 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 6.7.** *Proposition 6.2.6* 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 $\mathsf{N}(T)=\{\mathbf{0}\}$ 還是由 $\mathsf{N}(T)=\{\mathbf{0}\}$ 推得 $T$ 為 *one-to-one*?

**Question 6.8.** 假設 $T:V\to W$ 為 *linear transformation*. 在 *Question 6.6* 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 *null space* 的等價關係嗎?

**Question 6.9.** 假設 $\mathbb{F}$ 是 *field* 且 $A\in M_{m\times n}(\mathbb{F})$, 考慮 *linear transformation* $T_A:\mathbb{F}^n\to\mathbb{F}^m$ 定義為 $T_A(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$. 試說明 $\mathsf{N}(T_A)$ (即 $T_A$ 的 *null space*) 和 $N(A)$ (即 $A$ 的 *null space*) 之間的關係.

對於一般的函數, 要判斷其是否為 onto 或是 one-to-one 並不容易, 但當 $T$ 為 linear transformation 時, Proposition 6.2.3 和 Proposition 6.2.6 提供我們一個簡便的方法判斷 $T$ 是否為 onto 或是 one-to-one. 也就是僅要確認 $\mathsf{R}(T)$ 和 $\mathsf{N}(T)$ 的維度即可. 既然 $\mathsf{R}(T)$ 和 $\mathsf{N}(T)$ 的維度這麼重要, 我們有以下的定義.

**Definition 6.2.8.** 假設 $V,W$ 皆為 vector space over $\mathbb{F}$ 且 $T:V\to W$ 為 linear transformation.

(1) 我們稱 $T$ 的 range $\mathsf{R}(T)$ 的維度為 $T$ 的 *rank*, 記做 $\mathrm{rank}(T)$.

(2) 我們稱 $T$ 的 null space $\mathsf{N}(T)$ 的維度為 $T$ 的 *nullity*, 記做 $\mathrm{nullity}(T)$.

**Question 6.10.** 假設 $\mathbb{F}$ 是 *field* 且 $A\in M_{m\times n}(\mathbb{F})$, 考慮 *linear transformation* $T_A:\mathbb{F}^n\to\mathbb{F}^m$ 定義為 $T_A(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$. 試說明 $\mathrm{rank}(T_A)$ 和 $\mathrm{rank}(A)$ 以及 $\mathrm{nullity}(T_A)$ 和 $\mathrm{nullity}(A)$ 之間的關係.

在 Question 6.10 中我們看到 matrix 的 rank 和 nullity 和 linear transformation 的 rank 和 nullity 關係密切. 對於 matrix 的 rank 和 nullity 我們有所謂的 Dimension Theorem (參見 Theorem 3.7.14), 對於 linear transformation 我們也有以下的定理.

**Theorem 6.2.9** (Dimension Theorem)**.** 假設 $V,W$ 皆為 *vector space over* $\mathbb{F}$*,* 其中 $V$ 為 *finite dimensional.* 若 $T:V\to W$ 為 *linear transformation* 則

$$\mathrm{rank}(T)+\mathrm{nullity}(T)=\dim(V).$$

**Proof.** 假設 $\mathbf{u}_1,\ldots,\mathbf{u}_n\in\mathsf{N}(T)$, 為 $T$ 的 null space 的一組 basis. 因為 $\mathbf{u}_1,\ldots,\mathbf{u}_n$ 在 $V$ 中且為 linearly independent, 故由 Proposition 3.6.5 知存在 $\mathbf{v}_1,\ldots,\mathbf{v}_m\in V$ 使得 $\{\mathbf{u}_1,\ldots,\mathbf{u}_n,\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ 為 $V$ 的一組 basis. 注意此時 $\mathrm{nullity}(T)=\dim(\mathsf{N}(T))=n$ 且 $\dim(V)=m+n$. 我們要證明 $\{T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)\}$ 會是 $\mathsf{R}(T)$ 的一組 basis.

首先證明 $\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m))=\mathsf{R}(T)$. 由 Proposition 6.2.4, 我們知

$$\mathsf{R}(T)=\mathrm{Span}(T(\mathbf{u}_1),\ldots,T(\mathbf{u}_n),T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)).$$

然而 $\mathbf{u}_1,\ldots,\mathbf{u}_n\in\mathsf{N}(T)$, 亦即 $T(\mathbf{u}_1),\ldots,T(\mathbf{u}_n)$ 皆為 $W$ 中的零向量, 故得

$$\mathsf{R}(T)=\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)).$$

接下來我們證明 $\{T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)\}$ 為 linearly independent. 假設 $\{T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)\}$ 不是 linearly independent, 亦即存在 $c_1,\ldots,c_m\in\mathbb{F}$ 不全為 0 使得 $c_1T(\mathbf{v}_1)+\cdots+c_mT(\mathbf{v}_m)=\mathbf{0}$. 此時由 $T$ 為 linear transformation 知 $T(c_1\mathbf{v}_1+\cdot+c_m\mathbf{v}_m)=\mathbf{0}$, 因此得 $c_1\mathbf{v}_1+\cdot+c_m\mathbf{v}_m\in\mathsf{N}(T)$. 然而 $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ 為 $\mathsf{N}(T)$ 的 basis, 故存在 $d_1,\ldots,d_n\in\mathbb{F}$ 使得 $c_1\mathbf{v}_1+\cdot+c_m\mathbf{v}_m=d_1\mathbf{u}_1+\cdot+d_n\mathbf{u}_n$, 故得

$$d_1\mathbf{u}_1+\cdots+d_n\mathbf{u}_n-c_1\mathbf{v}_1-\cdots-c_m\mathbf{v}_m=\mathbf{0}.$$

然而 $\{\mathbf{u}_1,\ldots,\mathbf{u}_n,\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ 為 linearly independent, 故得 $d_1=\cdots=d_m=c_1=\cdots=c_m=0$. 此與 $c_1,\ldots,c_m$ 不全為 0 相矛盾, 故得證 $\{T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)\}$ 為 linearly independent.

既然 $\{T(\mathbf{v}_1),\ldots,T(\mathbf{v}_m)\}$ 是 $\mathsf{R}(T)$ 的一組 basis, 我們有 $\mathrm{rank}(T)=\dim(\mathsf{R}(T))=m$, 故得證 $\mathrm{rank}(T)+\mathrm{nullity}(T)=\dim(V)$. □

前面提過對於一般的函數, 要探討是否為 onto 或是 one-to-one 並不容易. 而對於 linear transformation, 我們可以藉由求其 range 及 null space 這兩個 subspaces 來了解這些問題. 而 Dimension Theorem 告訴我們, 只要了解 range 及 null space 這兩個 subspaces 中其中一個, 就可以了解另一個了.

**Question 6.11.** 假設 $V,W$ 皆為 *vector space over* $\mathbb{F}$ 且 $\dim(V)=\dim(W)$*.* 若 $T:V\to W$ 為 *linear transformation,* 證明 $T$ 是 *one-to-one* 若且唯若 $T$ 為 *onto.*

## 6.3. Matrix Representation

給定一個 matrix $A\in M_{m\times n}(\mathbb{F})$, 前面我們已知可以定義出一個 linear transformation $T:\mathbb{F}^n\to\mathbb{F}^m$, 其定義為 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$. 在這一節中, 我們要說明所有的 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformations 都可以寫成這樣的形式, 並將此概念推廣到一般 linear spaces 之間的 linear transformation. 也就是說, 我們將 linear transformation 和 matrix 相連結並推得一些重要的性質.

**6.3.1. $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformations.** 前面 Theorem 6.1.8 告訴我們給定一個 linear transformation, 只要知道此 linear transformation 將一組 $\mathbb{F}^n$ 的 basis 對應到哪些向量, 就可以唯一確定這一個 linear transformation. 在 $\mathbb{F}^n$ 中, 我們有一個最簡單的 basis, 即 standard basis $\{\mathbf{e}_1,\dots,\mathbf{e}_n\}$. 若 $T:\mathbb{F}^n\to\mathbb{F}^m$ 是一個 linear transformation, 由前面所述, 我們僅要知道 $T(\mathbf{e}_1),\dots,T(\mathbf{e}_n)$ 是哪些 $\mathbb{F}^m$ 的 vectors, 就可以知道任意 $\mathbf{v}\in\mathbb{F}^n$, $T(\mathbf{v})$ 為何了.

事實上對於每一個 $\mathbf{v}\in\mathbb{F}^n$, 都可以找到 $c_1,\dots,c_n\in\mathbb{F}$ 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 $\mathbf{v}\in\mathbb{F}^n$, 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 6.3.1.** *給定一個 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 function $T$. 則 $T$ 為 linear transformation 若且唯若存在一個 $m\times n$ matrix $A$ 使得 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$. 此 $m\times n$ matrix $A$ 是唯一的, 事實上對任意 $i=1,\dots,n$, $A$ 的 $i$-th column 為 $T(\mathbf{e}_i)$, 其中 $\{\mathbf{e}_1,\dots,\mathbf{e}_n\}$ 為 $\mathbb{F}^n$ 的 standard basis.*

**Proof.** 由 Lemma 6.1.5 我們知道, 若 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$, 則 $T$ 為 linear transformation. 反之, 若 $T:\mathbb{F}^n\to\mathbb{F}^m$ 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 $T(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$.

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 $i=1,\dots,n$, $B\mathbf{e}_i$ 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 6.3.1 告訴我們從 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 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 6.3.2.** 假設 $T:\mathbb{F}^n\to\mathbb{F}^m$ 為 linear transformation 且 $\{\mathbf{e}_1,\dots,\mathbf{e}_n\}$ 為 $\mathbb{F}^n$ 的 standard basis. 則對於 $i=1,\dots,n$, 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 $\mathbf{v}\in\mathbb{F}^n$, 我們有 $T(\mathbf{v})=[T]\mathbf{v}$.

**Example 6.3.3.** 我們探討 Example 6.2.5 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它的的 range 和 null space.

**Proposition 6.3.4.** 假設 $T:\mathbb{F}^n\to\mathbb{F}^m$ 為 *linear transformation* 且令 $[T]\in M_{m\times n}(\mathbb{F})$ 為其 *standard matrix representation.* 則 $T$ 的 *range* 等於 $[T]$ 的 *column space,* 而 $T$ 的 *null space* 等於 $[T]$ 的 *null space.*

**Proof.** 由於 $\mathbf{e}_1,\dots,\mathbf{e}_n$ 為 $\mathbb{F}^n$ 的一組 basis, 由 Proposition 6.2.4 我們知 $T$ 的 range, 即 $\mathsf{R}(T)=\mathrm{Span}(T(\mathbf{e}_1),\dots,T(\mathbf{e}_n))$. 然而 $T(\mathbf{e}_1),\dots,T(\mathbf{e}_n)$ 剛好就是 $[T]$ 的 $n$ 個 column, 故由定義 $\mathrm{Span}(T(\mathbf{e}_1),\dots,T(\mathbf{e}_n))$ 就是 $[T]$ 的 column space. 得證 $T$ 的 range 就是 $[T]$ 的 column space.

另一方面, 若 $\mathbf{v} \in \mathsf{N}(T)$, 表示 $T(\mathbf{v}) = \mathbf{0}$. 然而依 standard matrix representation 之定義 $T(\mathbf{v}) = [T]\mathbf{v}$, 故得 $[T]\mathbf{v} = \mathbf{0}$, 亦即 $\mathbf{v}$ 屬於 $[T]$ 的 null space. 得證 $\mathsf{N}(T)$ 包含於 $[T]$ 的 null space. 反之, 若 $\mathbf{v}$ 屬於 $[T]$ 的 null space, 表示 $[T]\mathbf{v} = \mathbf{0}$, 亦即 $T(\mathbf{v}) = \mathbf{0}$, 得證 $\mathbf{v} \in \mathsf{N}(T)$. 證明了 $[T]$ 的 null space 包含於 $\mathsf{N}(T)$, 因此 $T$ 的 null space 等於 $[T]$ 的 null space. □

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 null space 的維度稱為 $A$ 的 nullity, 用 $\mathrm{nullity}(A)$ 來表示 (參見 Definition 3.7.13). 由 Proposition 6.3.4, 我們知道 $T$ 的 range 的維度等於 $\mathrm{rank}([T])$, 而 $T$ 的 null space 的維度等於 $\mathrm{nullity}([T])$, 也就是說我們有以下的結果.

**Corollary 6.3.5.** 假設 $T : \mathbb{F}^n \to \mathbb{F}^m$ 為 *linear transformation* 且令 $[T] \in M_{m\times n}(\mathbb{F})$ 為其 *standard matrix representation.* 則

$$\mathrm{rank}(T) = \dim(\mathsf{R}(T)) = \mathrm{rank}([T]) \quad \text{and} \quad \mathrm{nullity}(T) = \dim(\mathsf{N}(T)) = \mathrm{nullity}([T]).$$

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 null space 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. 例如我們也很容易利用 linear transformation 的 Dimension Theorem (Theorem 6.2.9) 推得矩陣的 Dimension Theorem (Theorem 3.7.14).

**Example 6.3.6.** 我們利用 Example 6.3.3 中的 linear transformation 及其 standard matrix representation 探討其 range 和 null space.

(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 6.3.4 我們有 $\mathsf{R}(T) = \mathbb{R}^2$ (此與 Example 6.2.5(1) 一致). 另一方面, $[T]$ 的 null space 為聯立方程組 $[T]\mathbf{x} = \mathbf{0}$, 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 6.3.4 我們有 $\mathsf{N}(T) = \mathrm{Span}(\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix})$ (此與 Example 6.2.7(1) 一致).

(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 6.3.4 我們有 $\mathsf{R}(T) = \mathrm{Span}(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix})$ (此與 Example 6.2.5(2) 一致). 另一方面, $[T]$ 的 null

space 為聯立方程組 $[T]\mathbf{x}=\mathbf{0}$, 即

$$\left\{\begin{array}{lllll} x_1 & & = & 0 \\ x_1 & +x_2 & = & 0 \\ x_1 & -x_2 & = & 0 \end{array}\right. \quad \sim \quad \left\{\begin{array}{llll} x_1 & & = & 0 \\ & x_2 & = & 0 \end{array}\right.$$

的解集合. 因此由 Proposition 6.3.4 我們有 $\mathsf{N}(T)=\{\begin{bmatrix}0\\0\end{bmatrix}\}=\{\mathbf{0}\}$ (此與 Example 6.2.7(2) 一致).

當 $T_1,T_2$ 皆為 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformation 時, 對任意 $c_1,c_2\in\mathbb{F}$, 我們可以利用它們得到一個新的 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformation $c_1T_1+c_2T_2$ (參見 Proposition 6.1.6). 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 $\mathbb{F}^m$ 到 $\mathbb{F}^k$ 的 linear transformation, 我們可得合成函數 $T\circ T_1$ 為 $\mathbb{F}^n$ 到 $\mathbb{F}^k$ 的 linear transformation (參見 Proposition 6.1.7). 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 6.3.7.** 設 $T_1,T_2$ 為 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 *linear transformations,* 而 $T$ 為 $\mathbb{F}^m$ 到 $\mathbb{F}^k$ 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 $c_1,c_2\in\mathbb{F}$, 皆有 $c_1T_1+c_2T_2:\mathbb{F}^n\to\mathbb{F}^m$ 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) $T\circ T_1:\mathbb{F}^n\to\mathbb{F}^k$ 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 $\mathbf{v}\in\mathbb{F}^n$, 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 $c_1T_1+c_2T_2:\mathbb{F}^n\to\mathbb{F}^m$ 的 standard matrix representation 之要求, 故由 standard matrix representation 的唯一性 (Theorem 6.3.1) 知 $[c_1T_1+c_2T_2]=c_1[T_1]+c_2[T_2]$.

(2) 依定義對任意 $\mathbf{v}\in\mathbb{F}^n$, 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 $\mathbf{w}\in\mathbb{F}^m$ 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 $T\circ T_1:\mathbb{F}^n\to\mathbb{F}^k$ 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 6.3.1) 知 $[T\circ T_1]=[T][T_1]$. □

**Example 6.3.8.** 我們利用 Example 6.3.3 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 $T:\mathbb{F}^n\to\mathbb{F}^m$, $T':\mathbb{F}^m\to\mathbb{F}^k$ 為 linear transformations. 由於 $T$ 的 range 是 $\mathbb{F}^m$ 的 subspace, 即 $\mathsf{R}(T)=T(\mathbb{R}^n)\subseteq\mathbb{R}^m$, 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, 即 $\mathsf{R}(T'\circ T)\subseteq\mathsf{R}(T')$. 利用 subspace 之間 dimension 的關係 (Proposition 3.6.10(4)), 我們得 $\operatorname{rank}(T'\circ T)\le\operatorname{rank}(T')$. 因此當 $A\in M_{m\times n}(\mathbb{F})$, $B\in M_{k\times m}(\mathbb{F})$, 我們可以推得 $\operatorname{Col}(BA)\subseteq\operatorname{Col}(B)$ 且 $\operatorname{rank}(BA)\le\operatorname{rank}(B)$ (Proposition 3.7.16(1)).

**Question 6.12.** 假設 $T:\mathbb{F}^n\to\mathbb{F}^m$, $T':\mathbb{F}^m\to\mathbb{F}^k$ 為 *linear transformations.* 證明 $\mathsf{N}(T)\subseteq\mathsf{N}(T'\circ T)$, 並依此證明若 $A\in M_{m\times n}(\mathbb{F})$, $B\in M_{k\times m}(\mathbb{F})$, 則 $\mathsf{N}(A)\subseteq\mathsf{N}(BA)$ 且 $\operatorname{rank}(BA)\le\operatorname{rank}(A)$.

**6.3.2. Coordinatization.** 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 $\mathbb{F}^n$ 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* (有序基底). 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 $(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 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 $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 為 $V$ 的一組 ordered basis, 我們就會 $\beta$ 來表示這一組

ordered basis $(\mathbf{v}_1,\ldots,\mathbf{v}_n)$. 對於 $\mathbb{F}^n$ 的 standard basis, 我們通常用 $\varepsilon$ 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 $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 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 $\beta$ 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 $\beta$ 將 $\mathbf{v}$ 坐標化後所得的坐標表示法. 為了方便, 我們就用 $[\mathbf{v}]_\beta$ 來表示利用 $\beta$ 將 $\mathbf{v}$ 坐標化後所得的坐標. 坐標化的好處是, 我們可以將 $[\mathbf{v}]_\beta$ 看成是 $\mathbb{F}^n$ 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 $\mathbb{F}^n$ 中的向量來處理.

**Example 6.3.9.** 我們看看前面提過的幾個 vector space 坐標化的情形.

(A) 考慮 $M_{2\times 2}(\mathbb{F})$ 及其 ordered basis

$$\varepsilon=\left(\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix},\begin{bmatrix}0&0\\1&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix}\right)$$

(通常我們稱這一組 basis 為 $M_{2\times 2}(\mathbb{F})$ 的 standard basis). 對於任意 $M_{2\times 2}(\mathbb{F})$ 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 $\varepsilon$ 所得的坐標表示法為

$$\left[\begin{bmatrix}a&b\\c&d\end{bmatrix}\right]_\varepsilon=\begin{bmatrix}a\\b\\c\\d\end{bmatrix}.$$

例如在 $M_{2\times 2}(\mathbb{R})$, 我們有

$$\left[\begin{bmatrix}1&-2\\-3&4\end{bmatrix}\right]_\varepsilon=\begin{bmatrix}1\\-2\\-3\\4\end{bmatrix}.$$

(B) 在 $P_2(\mathbb{F})$ 中通常我們會稱 $1,x,x^2$ 這組 basis 為 standard basis. 考慮 $\varepsilon=(1,x,x^2)$ 這組 ordered basis. 很容易看出在 $P_2(\mathbb{R})$ 中, $2x^2-3x+4$ 用 $\varepsilon$ 所得的坐標為 $\begin{bmatrix}4\\-3\\2\end{bmatrix}$, 所以我們有

$$[2x^2-3x+4]_\varepsilon=\begin{bmatrix}4\\-3\\2\end{bmatrix}.$$

我們也可考慮 ordered basis $\beta=(p_1(x),p_2(x),p_3(x))$ 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 3.6.11). 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

$$[2x^2-3x+4]_\beta=\begin{bmatrix}4\\3\\9\end{bmatrix}.$$

(C) 我們也可將 $\mathbb{F}^n$ 中的向量用不同的 ordered basis 坐標化. 例如在 $\mathbb{R}^3$ 中考慮 ordered basis $\beta=(\begin{bmatrix}1\\1\\1\end{bmatrix},\begin{bmatrix}0\\1\\1\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix})$. 若要求向量 $\begin{bmatrix}1\\2\\3\end{bmatrix}$ 以 $\beta$ 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

$$\begin{bmatrix}1\\2\\3\end{bmatrix}_\beta=\begin{bmatrix}0\\2\\1\end{bmatrix}.$$

要注意這裡我們有 $\begin{bmatrix}1\\2\\3\end{bmatrix}_\varepsilon=\begin{bmatrix}1\\2\\3\end{bmatrix}$, 其中 $\varepsilon$ 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 $\varepsilon$ 來將所有 $\mathbb{R}^3$ 的向量的坐標化.

給定 $V$ 的一組 ordered basis $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_n)$, 將 $V$ 中的元素利用 $\beta$ 坐標化, 其實就定出了一個從 $V$ 到 $\mathbb{F}^n$ 的函數 $T_\beta:V\to\mathbb{F}^n$, 其中對任意 $\mathbf{v}\in V$, 我們有 $T_\beta(\mathbf{v})=[\mathbf{v}]_\beta$. 為什麼這是一個函數呢? 因為 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 $V$ 的一組 basis, 所以由 $\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$ 是 $V$ 的 spanning set, 可得任意的 $\mathbf{v}\in V$ 確實都存在 $c_1,\dots,c_n\in\mathbb{F}$ 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 所以 $T_\beta$ 確實可以將每個定義域中的元素 $\mathbf{v}$ 對應到對應域 $\mathbb{F}^n$ 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 $\mathbb{F}^n$ 中兩個不同向量的情況. 這是因為 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 為 linearly independent, 所以每個 $\mathbf{v}\in V$, 僅有一組 $c_1,\dots,c_n\in\mathbb{F}$ 會使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$.

既然 $T_\beta:V\to\mathbb{F}^n$ 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 $c_1,\dots,c_n$ 與 $d_1,\dots,d_n$ 皆屬於 $\mathbb{F}$. 依定義我們有

$$T_\beta(\mathbf{v})=[\mathbf{v}]_\beta=\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\quad\text{and}\quad T_\beta(\mathbf{w})=[\mathbf{w}]_\beta=\begin{bmatrix}d_1\\\vdots\\d_n\end{bmatrix}.$$

對於任意 $r\in\mathbb{F}$, 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

$$T_\beta(\mathbf{v}+r\mathbf{w})=[\mathbf{v}+r\mathbf{w}]_\beta=\begin{bmatrix}c_1+rd_1\\ \vdots\\ c_n+rd_n\end{bmatrix}=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}+r\begin{bmatrix}d_1\\ \vdots\\ d_n\end{bmatrix}=T_\beta(\mathbf{v})+rT_\beta(\mathbf{w}).$$

得證 $T_\beta:V\to\mathbb{F}^n$ 為 linear transformation.

$T_\beta:V\to\mathbb{R}^n$ 不只是 linear transformation, 事實上 $T_\beta:V\to\mathbb{R}^n$ 是 one-to-one 且 onto. 要檢查 $T_\beta$ 為 one-to-one, 我們僅要檢查 $\mathsf{N}(T_\beta)=\{\mathbf{0}\}$ 即可 (參見 Proposition 6.2.6). 然而若 $\mathbf{v}\in\mathsf{N}(T_\beta)$, 表示 $\mathbf{v}$ 用 $\beta$ 的坐標表示法是 $\mathbb{F}^n$ 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 $\mathbf{v}=\mathbf{0}$, 故知 $\mathsf{N}(T_\beta)=\{\mathbf{0}\}$. 要檢查 $T_\beta$ 為 onto, 我們可以利用 $T_\beta(\mathbf{v}_i)=\mathbf{e}_i$, $\forall i=1,\ldots,n$, 故得證 $\mathsf{R}(T_\beta)=\mathrm{Span}(T_\beta(\mathbf{v}_1),\ldots,T_\beta(\mathbf{v}_n))=\mathrm{Span}(\mathbf{e}_1,\ldots,\mathbf{e}_n)=\mathbb{F}^n$ (參見 Proposition 6.2.4), 即 $T_\beta$ 為 onto. 我們證得了以下的定理.

**Theorem 6.3.10.** 假設 $V$ 為 *vector space over* $\mathbb{F}$, $\dim(V)=n$ 且 $\beta$ 為 $V$ 的一組 *ordered basis.* 考慮 $T_\beta:V\to\mathbb{F}^n$ 定義為 $T_\beta(\mathbf{v})=[\mathbf{v}]_\beta$, $\forall\mathbf{v}\in V$. 則 $T_\beta$ 為 *linear transformation* 且是 *one-to-one* 以及 *onto.*

一般來說當一個 linear transformation $T:V\to W$ 是 one-to-one 且 onto 時, 為了方便以及強調其特殊性, 我們會稱 $T$ 為一個 *isomorphism*. 知道 $T_\beta:V\to\mathbb{F}^n$ 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 $T_\beta$, 將問題轉換成大家熟悉的 $\mathbb{F}^n$ 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 $\beta$, 然後考慮 $[\mathbf{w}_1]_\beta,\ldots,[\mathbf{w}_k]_\beta$, 這一組 $\mathbb{F}^n$ 中的向量. 利用我們熟悉的判斷 $\mathbb{F}^n$ 中向量是否為 linearly independent 的方法判斷 $[\mathbf{w}_1]_\beta,\ldots,[\mathbf{w}_k]_\beta$ 是否為 linearly independent. 由於對於 $i=1,\ldots,k$, $[\mathbf{w}_i]_\beta=T_\beta(\mathbf{w}_i)$, 因此由 $T_\beta$ 為 isomorphism 以後我們會知道 $\mathbf{w}_1,\ldots,\mathbf{w}_k$ 為 linearly independent 若且唯若 $[\mathbf{w}_1]_\beta,\ldots,[\mathbf{w}_k]_\beta$ 為 linearly independent (參見 Proposition 6.4.4). 因此我們可以由 $[\mathbf{w}_1]_\beta,\ldots,[\mathbf{w}_k]_\beta$ 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 6.3.11.** 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 $f_2(x),f_1(x),f_0(x)$, 其次數分別為為 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 $\varepsilon=(1,x,x^2)$, 可得

$$[f_2(x)]_\varepsilon=\begin{bmatrix}c\\ b\\ a\end{bmatrix},\quad [f_1(x)]_\varepsilon=\begin{bmatrix}e\\ d\\ 0\end{bmatrix}\quad\text{and}\quad [f_0(x)]_\varepsilon=\begin{bmatrix}r\\ 0\\ 0\end{bmatrix}.$$

由於 $a,d,r$ 皆不等於 0, 很容易看出矩陣

$$\begin{bmatrix}c&e&r\\ b&d&0\\ a&0&0\end{bmatrix}$$

的 rank 為 3, 亦即 $[f_2(x)]_\varepsilon, [f_1(x)]_\varepsilon, [f_0(x)]_\varepsilon$ 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 $\varepsilon = (1, x, \ldots, x^n)$ 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 over $\mathbb{R}$, 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 $\mathbf{w}_1, \ldots, \mathbf{w}_5 \in U$, 我們知 $W$ 為 $U$ 的 subspace. 我們的想法是利用 $\beta = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$ 這組 $U$ 的 ordered basis 將 $U$ 的元素坐標化. 由於 $\mathbf{w}_1, \ldots, \mathbf{w}_5 \in U$, 我們可以將 $\mathbf{w}_1, \ldots, \mathbf{w}_5$ 坐標化, 得 $[\mathbf{w}_1]_\beta, \ldots, [\mathbf{w}_5]_\beta$ 這 5 個 $\mathbb{R}^4$ 中的向量. 利用過去我們知道求 $\mathbb{R}^4$ 中 $\mathrm{Span}([\mathbf{w}_1]_\beta, \ldots, [\mathbf{w}_5]_\beta)$ 的 basis 的方法求出一組 basis. 再將它們還原成 $U$ 中的元素, 就得到 $W$ 的一組 basis.

現由於

$$[\mathbf{w}_1]_\beta = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 1 \end{bmatrix}, [\mathbf{w}_2]_\beta = \begin{bmatrix} 2 \\ -2 \\ 4 \\ 2 \end{bmatrix}, [\mathbf{w}_3]_\beta = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, [\mathbf{w}_4]_\beta = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, [\mathbf{w}_5]_\beta = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

我們考慮以它們為 column 的 $4 \times 5$ matrix 並利用 elementary row operations 將之化為 echelon form 得

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

由於 echelon form 的 1-st, 3-rd, 4-th column 為 pivot 所在位置, 故知 $[\mathbf{w}_1]_\beta, [\mathbf{w}_3]_\beta, [\mathbf{w}_4]_\beta$ 為 $\mathrm{Span}([\mathbf{w}_1]_\beta, \ldots, [\mathbf{w}_5]_\beta)$ 的一組 basis (參見 Proposition 3.7.8). 由於 $T_\beta$ 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 6.13.** 考慮 *Example 6.3.11 (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.*

**6.3.3. Matrix Representation of general linear transformation.** 當 $V, W$ 分別為 dimension 為 $n, m$ 的 vector space over $\mathbb{F}$. 我們可以透過 $V, W$ 的 ordered basis, 將 $V, W$ 的元素轉換成 $\mathbb{F}^n$ 和 $\mathbb{F}^m$ 的 vector. 因此我們可以將 $V$ 到 $W$ 的 linear transformation $T$ 視為 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformation, 而談論 $T$ 的 matrix representation.

分別給定 $V,W$ 的一組 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 以及 $\gamma=(\mathbf{w}_1,\ldots,\mathbf{w}_m)$. 令 $T_\beta:V\to\mathbb{F}^n$ 與 $T_\gamma:W\to\mathbb{F}^m$ 分別為利用 $\beta$ 以及 $\gamma$ 將 $V$, $W$ 的元素坐標化的 linear transformation. 回顧一下, 這裡 $T_\beta,T_\gamma$ 皆為 isomorphism. 由於 isomorphism 是 one-to-one 且 onto, 故其反函數是存在的, 且將來我們會證明此反函數仍為 linear transformation (參見 Theorem 6.4.3). 現對於任意 $V$ 到 $W$ 的 linear transformation $T$, 我們考慮合成函數 $T_\gamma\circ T\circ T_\beta^{-1}:\mathbb{F}^n\to\mathbb{F}^m$. 由於 $T_\gamma$, $T$ 以及 $T_\beta^{-1}$ 皆為 linear transformation, 所以 $T_\gamma\circ T\circ T_\beta^{-1}$ 是一個 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformation (Proposition 6.1.7). 因此 $T_\gamma\circ T\circ T_\beta^{-1}$ 有一個 standard matrix representation, 我們定義這個 standard matrix representation 為 $T$ 相對於 $\beta$, $\gamma$ 這兩組 ordered basis 所得的 *matrix representation*, 並用 $[T]_\beta^\gamma$ 來表示. 到底 $[T]_\beta^\gamma$ 是怎樣的矩陣呢? 依定義它是一個 $m\times n$ matrix, 且對於 $i=1,\ldots,n$, $[T]_\beta^\gamma$ 的 $i$-th column 應為 $T_\gamma\circ T\circ T_\beta^{-1}(\mathbf{e}_i)$. 由於 $T_\beta(\mathbf{v}_i)=\mathbf{e}_i$, 所以 $T_\beta^{-1}(\mathbf{e}_i)=\mathbf{v}_i$. 因此得

$$T_\gamma\circ T\circ T_\beta^{-1}(\mathbf{e}_i)=T_\gamma(T(T_\beta^{-1}(\mathbf{e}_i)))=T_\gamma(T(\mathbf{v}_i))=[T(\mathbf{v}_i)]_\gamma.$$

也就是說 $[T]_\beta^\gamma$ 的 $i$-th column 就是將 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 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 $\gamma$ 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

$$[T]_\beta^\gamma=\begin{bmatrix} | & | & & | \\ [T(\mathbf{v}_1)]_\gamma & [T(\mathbf{v}_2)]_\gamma & \cdots & [T(\mathbf{v}_n)]_\gamma \\ | & | & & | \end{bmatrix}.$$

**Example 6.3.12.** 考慮假設 $V$ 為 finite dimensional vector space over $\mathbb{F}$. 考慮 $V$ 上的 identity map $\mathrm{id}:V\to V$, 亦即對任意 $\mathbf{v}\in V$, 我們定義 $\mathrm{id}(\mathbf{v})=\mathbf{v}$. 很容易看出 id 是一個 linear transformation. 現對任意 $V$ 的 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 我們想知道 $\mathrm{id}:V\to V$ 對定義域和對應域都用 $\beta$ 這個 ordered basis 所得的 matrix representation $[\mathrm{id}]_\beta^\beta$ 為何? 依照前面的討論, 我們知道 $[\mathrm{id}]_\beta^\beta$ 的 1-st column 就是 $\mathrm{id}(\mathbf{v}_1)$ 利用 $\beta$ 坐標化所得的 $\mathbb{F}^n$ 中的向量. 由於 $\mathrm{id}(\mathbf{v}_1)=\mathbf{v}_1$, 而 $\mathbf{v}_1$ 又是 $\beta$ 中第一個向量, 故知 $[\mathrm{id}(\mathbf{v}_1)]_\beta=[\mathbf{v}_1]_\beta=\mathbf{e}_1$, 也就是說 $[\mathrm{id}]_\beta^\beta$ 的 1-st column 就是 $\mathbf{e}_1$. 同理 $[\mathrm{id}]_\beta^\beta$ 的 $i$-th column 就是 $\mathbf{e}_i$. 也就是說 $[\mathrm{id}]_\beta^\beta$ 就是 $n\times n$ 的 identity matrix $I_n$.

**Example 6.3.13.** 考慮 $P_2(\mathbb{R})$ 上的 standard ordered basis $\varepsilon_2=(1,x,x^2)$ 以及 $P_3(\mathbb{R})$ 上的 standard ordered basis $\varepsilon_3=(1,x,x^2,x^3)$. 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 $\varepsilon_2,\varepsilon_3$ 的 matrix representation $[T]_{\varepsilon_2}^{\varepsilon_3}$. 依照前面的探討, 我們知 $[T]_{\varepsilon_2}^{\varepsilon_3}$ 的 1-st column 應該就是將 1 代入 $T$, 得 $T(1)=(x+1)\cdot 1$ 再利用 $\varepsilon_3$ 將 $x+1$ 坐標化寫成 $\mathbb{R}^4$ 的向量. 由於 $x+1=1+x+0x^2+0x^3$, 所以得 $[T]_{\varepsilon_2}^{\varepsilon_3}$ 的

1-st column 為

$$[T(1)]_{\varepsilon_3} = [x+1]_{\varepsilon_3} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

同理我們有 $[T]_{\varepsilon_2}^{\varepsilon_3}$ 的 2-nd, 3-rd column 分別為

$$[T(x)]_{\varepsilon_3} = [(x+1)(x-1)]_{\varepsilon_3} = [x^2-1] = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix},$$

$$[T(x^2)]_{\varepsilon_3} = [(x+1)(x-1)^2]_{\varepsilon_3} = [x^3-x^2-x+1]_{\varepsilon_3} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}.$$

因此得

$$[T]_{\varepsilon_2}^{\varepsilon_3} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

我們更換 $P_2(\mathbb{R})$ 以及 $P_3(\mathbb{R})$ 的 ordered basis. 回顧在 Example 3.6.11 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 $\beta = (p_1(x), p_2(x), p_3(x))$ 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 $\gamma = (q_1(x), q_2(x), q_3(x), q_4(x))$ 為 $P_3(\mathbb{R})$ 的 ordered basis. 我們要得到 $T$ 對於 $\beta, \gamma$ 的 matrix representation $[T]_\beta^\gamma$. 首先 $[T]_\beta^\gamma$ 的 1-st column 為 $T(p_1(x)) = (x+1)p_1(x-1)$ 利用 $\gamma$ 坐標化所得 $\mathbb{R}^4$ 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 $[T]_\beta^\gamma$ 的 1-st column 為

$$[T(p_1(x))]_\gamma = [(x+1)p_1(x-1)]_\gamma = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

同樣的方法我們可以得到 $[T]_\beta^\gamma$ 的 2-nd, 3-rd column 分別為

$$[T(p_2(x))]_\gamma = [(x+1)p_2(x-1)]\gamma = \begin{bmatrix} 0 \\ 0 \\ 2 \\ 0 \end{bmatrix}, [T(p_3(x))]_\gamma = [(x+1)p_3(x-1)]_\gamma = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 3 \end{bmatrix}.$$

因此得

$$[T]_\beta^\gamma = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

由 Example 6.3.13 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 $\mathbb{F}^n$ 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 $\beta,\gamma$, 以及一個 $V$ 到 $W$ 的 linear transformation $T$, 我們可以圖示如下:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_\beta"'] & W \arrow[d, shift left, "T_\gamma"] \\ \mathbb{F}^n \arrow[r] \arrow[u, shift right, "T_\beta^{-1}"'] & \mathbb{F}^m \arrow[u, shift left, "T_\gamma^{-1}"] \end{tikzcd}$$

這樣的圖示一般稱為 *commutative diagram*. 它表示底下 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的函數為 $T_\gamma \circ T \circ T_\beta^{-1}$, 亦即先利用 $T_\beta^{-1}$ 將 $\mathbb{F}^n$ 映射到 $V$, 再利用 $T$ 將 $V$ 映射到 $W$, 最後利用 $T_\gamma$ 將 $W$ 映射到 $\mathbb{F}^m$. 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 $T_\beta$ 到 $\mathbb{F}^n$, 接著藉由 $T_\gamma \circ T \circ T_\beta^{-1}$ 從 $\mathbb{F}^n$ 到 $\mathbb{F}^m$, 最後從 $\mathbb{F}^m$ 藉由 $T_\beta^{-1}$ 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 $T_\beta$ 得到 $T_\beta(\mathbf{v})$, 接著利用 $T_\gamma \circ T \circ T_\beta^{-1}$ 將 $T_\beta(\mathbf{v})$ 送至 $T_\gamma \circ T \circ T_\beta^{-1}(T_\beta(\mathbf{v})) = T_\gamma(T(\mathbf{v}))$, 最後再利用 $T_\gamma^{-1}$ 將 $T_\gamma(T(\mathbf{v}))$ 送至 $T_\gamma^{-1}(T_\gamma(T(\mathbf{v}))) = T(\mathbf{v})$.

利用 $V \to \mathbb{F}^n$ 接 $\mathbb{F}^n \to \mathbb{F}^m$ 再接 $\mathbb{F}^m \to W$ 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 $\mathbb{F}^n \to \mathbb{F}^m$ 這一段的路徑, 有 standard matrix representation, 即 $[T]_\beta^\gamma$. 也就是說對於任意 $\mathbb{F}^n$ 的向量, 我們只要左邊乘上 $[T]_\beta^\gamma$ 就可以知道會被映射到哪一個 $\mathbb{F}^m$ 的向量. 所以對任意 $\mathbf{v} \in V$, 我們可以先利用 $\beta$ 將 $\mathbf{v}$ 坐標化得 $T_\beta(\mathbf{v}) = [\mathbf{v}]_\beta$. 接著由於 $[\mathbf{v}]_\beta \in \mathbb{F}^n$, 故將之代入 $T_\gamma \circ T \circ T_\beta^{-1}$ 就是將 $[\mathbf{v}]_\beta$ 左邊乘上 $[T]_\beta^\gamma$ 這一個 matrix. 也就是說 $T_\gamma \circ T \circ T_\beta^{-1}([\mathbf{v}]_\beta)$ 就是 $[T]_\beta^\gamma[\mathbf{v}]_\beta$. 最後再利用 $W$ 的 ordered basis $\gamma$ 將 $[T]_\beta^\gamma[\mathbf{v}]_\beta$ 這個 $\mathbb{F}^m$ 中的向量還原回 $W$ 的元素 $T_\gamma^{-1}([T]_\beta^\gamma[\mathbf{v}]_\beta)$, 就是 $T(\mathbf{v})$ 之值. 因此我們有以下之結果.

**Proposition 6.3.14.** 假設 $V,W$ 為 *vector space over* $\mathbb{F}$ 且 $\beta = (\mathbf{v}_1,\ldots,\mathbf{v}_n)$, $\gamma = (\mathbf{w}_1,\ldots,\mathbf{w}_m)$ 分別為 $V,W$ 的 *ordered basis.* 設 $T : V \to W$ 為 *linear transformation* 且 $[T]_\beta^\gamma \in M_{m\times n}(\mathbb{F})$ 為

$T$ 相對於 $\beta,\gamma$ 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

$$[T]_\beta^\gamma\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}=\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix},$$

則 $T(\mathbf{v})=d_1\mathbf{w}_1+\cdots+d_m\mathbf{w}_m$. 亦即

$$[T]_\beta^\gamma[\mathbf{v}]_\beta=[T(\mathbf{v})]_\gamma.$$

**Proof.** 因 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 依定義 $\mathbf{v}$ 利用 $\beta$ 坐標化所得 $\mathbb{F}^n$ 的向量 $T_\beta(\mathbf{v})$ 為 $\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$. 故由前面所述 $T_\gamma\circ T\circ T_\beta^{-1}(T_\beta(\mathbf{v}))=T_\gamma(T(\mathbf{v}))$ 就是將 $\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$ 左邊乘上 $[T]_\beta^\gamma$ 所得的 $\mathbb{F}^m$ 中向量 $\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}$. 故由 $T_\gamma(T(\mathbf{v}))=\begin{bmatrix}d_1\\ \vdots\\ d_m\end{bmatrix}$ 可得 $T(\mathbf{v})=d_1\mathbf{w}_1+\cdots+d_m\mathbf{w}_m$. □

**Example 6.3.15.** 我們考慮 Example 6.3.13 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 $\varepsilon_2=(1,x,x^2)$ 以及 $P_3(\mathbb{R})$ 的 ordered basis $\varepsilon_3=(1,x,x^2,x^3)$, 我們知道 $T$ 對於 $\varepsilon_2,\varepsilon_3$ 的 matrix representation 為 $[T]_{\varepsilon_2}^{\varepsilon_3}=\begin{bmatrix}1&-1&1\\1&0&-1\\0&1&-1\\0&0&1\end{bmatrix}$ 故由 $[p(x)]_{\varepsilon_2}=[x^2-1]_{\varepsilon_2}=\begin{bmatrix}-1\\0\\1\end{bmatrix}$ 得

$$[T(p(x))]_{\varepsilon_3}=\begin{bmatrix}1&-1&1\\1&0&-1\\0&1&-1\\0&0&1\end{bmatrix}\begin{bmatrix}-1\\0\\1\end{bmatrix}=\begin{bmatrix}0\\-2\\-1\\1\end{bmatrix},$$

即 $T(p(x))=x^3-x^2-2x$.

當然我們也可以利用 Example 6.3.13 中 $V$ 的 ordered basis $\beta=(p_1(x),p_2(x),p_3(x))$ 以及 $W$ 的 ordered basis $\gamma=(q_1(x),q_2(x),q_3(x),q_4(x))$ 求出 $T(x^2-1)$. 此時若 $p(x)=x^2-1=c_1p_1(x)+c_2p_2(x)+c_3p_3(x)$, 則代 $x=-1,0,1$ 得 $c_1=0$, $c_2=-1$, $c_3=0$, 亦即 $[p(x)]_\beta=\begin{bmatrix}0\\-1\\0\end{bmatrix}$.

故將 $[p(x)]_\beta$ 左邊乘上 $T$ 對於 $\beta,\gamma$ 的 representation matrix $[T]_\beta^\gamma$ 得

$$[T(p(x))]_\gamma=\begin{bmatrix}0&0&0\\1&0&0\\0&2&0\\0&0&3\end{bmatrix}\begin{bmatrix}0\\-1\\0\end{bmatrix}=\begin{bmatrix}0\\0\\-2\\0\end{bmatrix}.$$

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 $T:\mathbb{F}^n\to\mathbb{F}^m$ 是 linear transformation 時, 我們可以利用 $T$ 的 standard matrix representation $[T]$ 的 null space 來決定 $T$ 的 null space, 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 null space 和 range. 分別選定 $V$ 和 $W$ 的 ordered basis $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_n)$ 和 $\gamma=(\mathbf{w}_1,\dots,\mathbf{w}_m)$. 由前面的 commutative diagram, 利用 $V\to W$ 接著 $W\to\mathbb{F}^m$ 的路徑以及 $V\to\mathbb{F}^n$ 接著 $\mathbb{F}^n\to\mathbb{F}^m$ 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

$$[T(\mathbf{v})]_\gamma=[T]_\beta^\gamma\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}. \tag{6.1}$$

現若 $\mathbf{v}\in\mathsf{N}(T)$, 表示 $T(\mathbf{v})=\mathbf{0}$, 因此由 $[T(\mathbf{v})]_\gamma=[\mathbf{0}]_\gamma=\begin{bmatrix}0\\\vdots\\0\end{bmatrix}$, 得 $[T]_\beta^\gamma\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=\begin{bmatrix}0\\\vdots\\0\end{bmatrix}$, 亦即 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=[\mathbf{v}]_\beta$ 為 $[T]_\beta^\gamma$ 的 null space 的向量. 反之, 若 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\in\mathbb{F}^n$ 為 $[T]_\beta^\gamma$ 的 null space 的向量, 表示 $[T]_\beta^\gamma\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=\begin{bmatrix}0\\\vdots\\0\end{bmatrix}$. 故由式子 (6.1) 知, 當 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$ 時, 我們有 $[T(\mathbf{v})]_\gamma=\begin{bmatrix}0\\\vdots\\0\end{bmatrix}$. 此即表示 $T(\mathbf{v})=\mathbf{0}$, 得證 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in\mathsf{N}(T)$.

另一方面, 若 $\mathbf{w}=d_1\mathbf{w}_1+\cdots+d_m\mathbf{w}_m\in\mathsf{R}(T)$, 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 (6.1) 知, $\begin{bmatrix}d_1\\\vdots\\d_m\end{bmatrix}=[T(\mathbf{v})]_\gamma=[T]_\beta^\gamma\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}$, 亦即 $\begin{bmatrix}d_1\\\vdots\\d_m\end{bmatrix}$ 是 $[T]_\beta^\gamma$ 的 column space 的向量. 反之, 若 $\begin{bmatrix}d_1\\\vdots\\d_m\end{bmatrix}\in\mathbb{F}^m$ 為 $[T]_\beta^\gamma$ 的 column space 的向量, 表示存在 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\in\mathbb{F}^n$ 使得 $\begin{bmatrix}d_1\\\vdots\\d_m\end{bmatrix}=[T]_\beta^\gamma\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}$. 故由式子 (6.1) 知, 若令 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 我們有 $[T(\mathbf{v})]_\gamma=\begin{bmatrix}d_1\\\vdots\\d_m\end{bmatrix}$, 得證 $d_1\mathbf{w}_1+\cdots+d_m\mathbf{v}_m=T(\mathbf{v})\in\mathsf{R}(T)$. 我們證得了以下的結果.

**Proposition 6.3.16.** 假設 $V,W$ 為 *vector space* 且 $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_n)$, $\gamma=(\mathbf{w}_1,\dots,\mathbf{w}_m)$ 分別為 $V,W$ 的 *ordered basis*. 設 $T:V\to W$ 為 *linear transformation* 且 $[T]_\beta^\gamma\in M_{m\times n}(\mathbb{F})$ 為 $T$ 相

對於 $\beta,\gamma$ 的 *matrix representation.* 則 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in\mathsf{N}(T)$ 若且唯若 $\begin{bmatrix} c_1\\ \vdots\\ c_n\end{bmatrix}$ 屬於 $[T]_\beta^\gamma$ 的 *null space.* 而 $d_1\mathbf{w}_1+\cdots+d_m\mathbf{w}_m\in\mathsf{R}(T)$ 若且唯若 $\begin{bmatrix} d_1\\ \vdots\\ d_m\end{bmatrix}$ 屬於 $[T]_\beta^\gamma$ 的 *column space.*

**Example 6.3.17.** 我們考慮 Example 6.3.13 的例子, 當考慮 $P_2(\mathbb{R})$ 的 ordered basis $\varepsilon_2=(1,x,x^2)$ 以及 $P_3(\mathbb{R})$ 的 ordered basis $\varepsilon_3=(1,x,x^2,x^3)$, 我們知道 $T$ 對於 $\varepsilon_2,\varepsilon_3$ 的 matrix representation $T_{\varepsilon_2}^{\varepsilon_3}$ 利用 elementary row operations 化為 echelon form 可得

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

由於 pivot 的個數等於 column 的個數, 我們知 $T_{\varepsilon_2}^{\varepsilon_3}$ 的 null space 為 $\{\begin{bmatrix}0\\0\\0\end{bmatrix}\}$, 故知 $\mathsf{N}(T)=\{\mathbf{0}\}$, 亦即 $T$ 為 one-to-one. 另一方面 $[T]_{\varepsilon_2}^{\varepsilon_3}$ 的 rank 為 3, 故 $\{\begin{bmatrix}1\\1\\0\\0\end{bmatrix},\begin{bmatrix}-1\\0\\1\\0\end{bmatrix},\begin{bmatrix}1\\-1\\-1\\1\end{bmatrix}\}$ 為 column space 的一組 basis. 因此得 $\{x+1,x^2-1,x^3-x^2-x+1\}$ 為 $\mathsf{R}(T)$ 的一組 basis. 由於 $\dim(P_3(\mathbb{R}))=4\neq\dim(\mathsf{R}(T))=3$, 我們知 $\mathsf{R}(T)\neq P_3(\mathbb{R})$, 故 $T$ 不是 onto.

**Example 6.3.18.** 考慮 $M_{2\times2}(\mathbb{R})$ 所形成的 vector space (參見 Example 3.2.2 (A)). 考慮函數 $T:M_{2\times2}(\mathbb{R})\to M_{2\times2}(\mathbb{R})$ 定義為 $T(A)=A-A^{\mathrm{t}}$, $\forall A\in M_{2\times2}(\mathbb{R})$. 我們可得 $T$ 為 linear transformation. 這是因為對任意 $A,B\in M_{2\times2}(\mathbb{R})$ 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 $M_{2\times2}(\mathbb{R})$ 的 ordered basis $\varepsilon=\left(\begin{bmatrix}1&0\\0&0\end{bmatrix},\begin{bmatrix}0&1\\0&0\end{bmatrix},\begin{bmatrix}0&0\\1&0\end{bmatrix},\begin{bmatrix}0&0\\0&1\end{bmatrix}\right)$. 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 $\varepsilon,\varepsilon$ 的 matrix representation 為 $[T]_\varepsilon^\varepsilon=\begin{bmatrix}0&0&0&0\\0&1&-1&0\\0&-1&1&0\\0&0&0&0\end{bmatrix}$. 利用 elementary row operation 將 $[T]_\varepsilon^\varepsilon$ 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 $[T]_\varepsilon^\varepsilon$ 的 null space 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 $\mathsf{N}(T)$ 的一組 basis. 又 $[T]_\varepsilon^\varepsilon$ 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 $\mathsf{R}(T)$ 的一組 basis. 注意我們有 $\dim(\mathsf{R}(T))+\dim(\mathsf{N}(T))=1+3=4=\dim(M_{2\times 2}(\mathbb{R}))$. 另外若 $A\in\mathsf{N}(T)$ 表示 $T(A)=A-A^{\mathrm{t}}=\mathbf{0}$, 亦即 $A=A^{\mathrm{t}}$. 反之亦然, 也就是說 $A\in\mathsf{N}(T)$ 若且唯若 $A$ 為 symmetric matrix. 因此由 $\dim(\mathsf{N}(T))=3$, 我們知所有 $2\times 2$ 的 symmetric matrices 所成的 subspace 的維度為 3.

**Question 6.14.** 試求所有 $3\times 3$ 的 *symmetric matrices* 所成的 *subspace* 的維度為何?

當 $T_1,T_2$ 皆為 $\mathbb{F}^n$ 到 $\mathbb{F}^m$ 的 linear transformation 時, 對任意 $c_1,c_2\in\mathbb{F}$, 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 6.3.7). 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, $\beta,\gamma$, 我們會有

$$[c_1T_1+c_2T_2]_\beta^\gamma=c_1[T_1]_\beta^\gamma+c_2[T_2]_\beta^\gamma.$$

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 $\alpha,\beta,\gamma$, 我們有以下的圖示

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_\alpha"'] & W \arrow[d, shift left, "T_\beta"] \\ \mathbb{F}^n \arrow[r] \arrow[u, shift right, "T_\alpha^{-1}"'] & \mathbb{F}^m \arrow[u, shift left, "T_\beta^{-1}"] \end{tikzcd} \qquad \begin{tikzcd} W \arrow[r, "T'"] \arrow[d, shift right, "T_\beta"'] & U \arrow[d, shift left, "T_\gamma"] \\ \mathbb{F}^m \arrow[r] \arrow[u, shift right, "T_\beta^{-1}"'] & \mathbb{F}^k \arrow[u, shift left, "T_\gamma^{-1}"] \end{tikzcd}$$

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 $\beta$, 所以兩邊 $W$ 到 $\mathbb{F}^m$ 的之間的函數相同 (皆為 $T_\beta$). 因此我們可以將上面兩個 commutative diagrams 合併成一個 commutative diagram 如下:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift right, "T_\alpha"'] & W \arrow[r, "T'"] \arrow[d, shift left, "T_\beta"] & U \arrow[d, shift left, "T_\gamma"] \\ \mathbb{F}^n \arrow[r] \arrow[u, shift right, "T_\alpha^{-1}"'] & \mathbb{F}^m \arrow[r] \arrow[u, shift left, "T_\beta^{-1}"] & \mathbb{F}^k \arrow[u, shift left, "T_\gamma^{-1}"] \end{tikzcd}$$

由於底部 $\mathbb{F}^n\to\mathbb{F}^m$ 的 matrix representation 為 $[T]_\alpha^\beta$, 而 $\mathbb{F}^m\to\mathbb{F}^k$ 的 matrix representation 為 $[T']_\beta^\gamma$, 因此由 Lemma 6.3.7 知, 它們的合成所對應的 matrix representation 為 $[T']_\beta^\gamma[T]_\alpha^\beta$. 因此我們有

$$[T'\circ T]_\alpha^\gamma=[T']_\beta^\gamma[T]_\alpha^\beta.$$

綜合以上的討論, 我們有以下有關 Lemma 6.3.7 的推廣.

**Theorem 6.3.19.** 假設 $V,W,U$ 為 *finite dimensional vector space over* $\mathbb{F}$ 且令 $\alpha,\beta,\gamma$ 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 $c_1,c_2\in\mathbb{F}$, 我們有

$$[c_1T_1+c_2T_2]_\alpha^\beta = c_1[T_1]_\alpha^\beta + c_2[T_2]_\alpha^\beta.$$

(2) 設 $T:V\to W$ 及 $T':W\to U$ 為 *linear transformation.* 則

$$[T'\circ T]_\alpha^\gamma = [T']_\beta^\gamma[T]_\alpha^\beta.$$

假設 $V,W$ 為 finite dimensional vector space over $\mathbb{F}$, 其中 $\dim(V)=n,\dim(W)=m$. 令 $\mathscr{L}(V,W)$ 為所有 $V$ 到 $W$ 的 linear transformations 所成的集合. Proposition 6.1.6 告訴我們 $\mathscr{L}(V,W)$ 有加法和係數積的封閉性. 很容易證明 $\mathscr{L}(V,W)$ 是一個 over $\mathbb{F}$ 的 vector space. 令 $\beta$, $\gamma$ 分別為 $V,W$ 上的 ordered basis, 我們可以訂一出一個由 $\mathscr{L}(V,W)$ 到 $M_{m\times n}(\mathbb{F})$ 的函數 $\mathscr{M}:\mathscr{L}(V,W)\to M_{m\times n}(\mathbb{F})$, 其定義為對任意 $T\in\mathscr{L}(V,W)$, $\mathscr{M}(T)=[T]_\beta^\gamma$. Theorem 6.3.19 告訴我們 $\mathscr{M}:\mathscr{L}(V,W)\to M_{m\times n}(\mathbb{F})$ 是一個 linear transformation. 我們有以下的結果.

**Theorem 6.3.20.** 假設 $V,W$ 為 *finite dimensional vector space over* $\mathbb{F}$, 其中 $\dim(V)=n,\dim(W)=m$. 給定 $\beta$, $\gamma$ 分別為 $V,W$ 上的 *ordered basis,* 定義函數 $\mathscr{M}:\mathscr{L}(V,W)\to M_{m\times n}(\mathbb{F})$, 其中 $\forall T\in\mathscr{L}(V,W)$, $\mathscr{M}(T)=[T]_\beta^\gamma$. 則 $\mathscr{M}:\mathscr{L}(V,W)\to M_{m\times n}(\mathbb{F})$ 是一個 *one-to-one* 且 *onto* 的 *linear transformation.* 並可得 $\dim_{\mathbb{F}}(\mathscr{L}(V,W))=mn$.

**Proof.** 對任意 $T_1,T_2\in\mathscr{L}(V,W)$ 以及 $c\in\mathbb{F}$, 我們有 $\mathscr{M}(T_1+cT_2)=[T_1+cT_2]_\beta^\gamma$, 而 $\mathscr{M}(T_1)+c\mathscr{M}[T_2]=[T_1]_\beta^\gamma+c[T_2]_\beta^\gamma$, 故由 Theorem 6.3.19(1) 知 $\mathscr{M}(T_1+cT_2)=\mathscr{M}(T_1)+c\mathscr{M}(T_2)$.

假設 $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 以及 $\gamma=(\mathbf{w}_1,\ldots,\mathbf{w}_m)$, 對任意 $A=[a_{ij}]\in M_{m\times n}(\mathbb{F})$, 由於 $A$ 的 $i$-th column 為 $\begin{bmatrix}a_{1i}\\ \vdots\\ a_{mi}\end{bmatrix}$, 對任意 $i=1,\ldots,n$ 我們考慮 $T\in\mathscr{L}(V,W)$ 為唯一的 linear transformation 滿足 $T(\mathbf{v}_i)=a_{1i}\mathbf{w}_1+\cdots+a_{mi}\mathbf{w}_m$ (參見 Theorem 6.1.8). 依定義 $[T]_\beta^\gamma$ 的 $i$-th column 為 $[T(\mathbf{v}_i)]_\gamma$ 與 $A$ 的 $i$-th column 相同, 故證得 $\mathscr{M}(T)=[T]_\beta^\gamma=A$. 這證得了 $\mathscr{M}:\mathscr{L}(V,W)\to M_{m\times n}(\mathbb{F})$ 是 onto, 也證得它是 one-to-one, 因為這樣的 $T\in\mathscr{L}(V,W)$ 是唯一的.

最後利用 Dimension Theorem (Theorem 6.2.9), 我們知道 $\operatorname{rank}(\mathscr{M})+\operatorname{nullity}(\mathscr{M})=\dim_{\mathbb{F}}(\mathscr{L}(V,W))$. 由於 $\mathscr{M}$ 是 one-to-one, $\operatorname{nullity}(\mathscr{M})=0$. 又由於 $\mathscr{M}$ 是 onto, 我們知 $\operatorname{rank}(\mathscr{M})=\dim_{\mathbb{F}}(M_{m\times n}(\mathbb{F}))=mn$. 故得證 $\dim_{\mathbb{F}}(\mathscr{L}(V,W))=mn$. □

Theorem 6.3.20, 告訴我們線性映射和矩陣間的對應關係. 也就是說我們可以利用表現矩陣來了解線性映射, 也可以用線性映射來了解矩陣. 兩者之間互相的關係大家應充分體會.

## 6.4. Invertible Linear Transformation

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. 這一節中我們將學得, 對於 invertible linear transformation, 利用 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 $T : V \to W$ 為 invertible linear transformation. 由於 $T$ 為 onto, 我們需要 $\mathrm{rank}(T) = \dim(W)$ (Proposition 6.2.3). 然而 $T$ 為 one-to-one, 故知 $\dim(\mathsf{N}(T)) = \mathrm{nullity}(T) = 0$ (Proposition 6.2.6). 利用 Dimension Theorem for linear transformation (Theorem 6.2.9) 我們得

$$\dim(V) = \mathrm{rank}(T) + \mathrm{nullity}(T) = \dim(W) + 0 = \dim(W).$$

這告訴我們只有在 $\dim(V) = \dim(W)$ 時, $T$ 才有可能為 invertible. 現假設 $\dim(V) = \dim(W)$ 且 $T : V \to W$ 為 one-to-one, 由 $\mathrm{nullity}(T) = 0$, 我們得 $\mathrm{rank}(T) = \dim(V) = \dim(W)$, 亦即 $T$ 為 onto. 同樣的, 若 $T$ 為 onto, 則由 $\mathrm{rank}(T) = \dim(W) = \dim(V)$ 得 $\mathrm{nullity}(T) = 0$, 亦即 $T$ 為 one-to-one. 這告訴我們當 $\dim(V) = \dim(W)$ 時, 對於 linear transformation $T : V \to W$, $T$ 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 6.4.1.** 假設 $V, W$ 為 *finite dimensional vector spaces over* $\mathbb{F}$ 且 $T : V \to W$ 為 *linear transformation.* 則僅有當 $\dim(V) = \dim(W)$ 時, $T$ 才有可能是 *invertible.* 又當 $\dim(V) = \dim(W)$ 時, $T$ 為 *invertible* 和 $T$ 為 *onto* 是等價的也和 $T$ 為 *one-to-one* 等價.

由 Lemma 6.4.1 我們知道若 $T : V \to W$ 為 invertible, 則 $\dim(V) = \dim(W)$. 換句話說若 $\dim(V) \neq \dim(W)$, 則不可能存在一個 $V$ 到 $W$ 的 invertible linear transformation. 這個性質的反向也是對的, 我們有以下的定理.

**Proposition 6.4.2.** 設 $V, W$ 為 *finite dimensional vector spaces over* $\mathbb{F}$. 則存在 $T : V \to W$ 為 *invertible linear transformation* 若且唯若 $\dim(V) = \dim(W)$.

**Proof.** 由 Lemma 6.4.1, 我們僅要證明若 $\dim(V) = \dim(W)$ 則存在 linear transformation $T : V \to W$ 為 invertible. 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 6.1.8). 我們要說明 $T$ 為 invertible. 然而由 Proposition 6.2.4 知 $\mathsf{N}(T) = \mathrm{Span}(T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)) = \mathrm{Span}(\mathbf{w}_1, \ldots, \mathbf{w}_n)$. 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 $\mathsf{N}(T) = W$, 即 $T$ 為 onto. 再由 Lemma 6.4.1 得證 $T$ 為 invertible. $\square$

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* (反函數). 我們都知道 $T^{-1} : W \to V$ 仍為 one-to-one 且 onto, 不過令人好奇的是它是否仍為 linear transformation? 我們有以下之結果.

**Theorem 6.4.3.** 假設 $V,W$ 為 *vector spaces over* $\mathbb{F}$ 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 $\mathbf{w}_1,\mathbf{w}_2\in W$, $r\in\mathbb{F}$, 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 $T^{-1}(\mathbf{w}_1)+rT^{-1}(\mathbf{w}_2)$ 就等同於檢查是否 $T^{-1}(\mathbf{w}_1+r\mathbf{w}_2)=\mathbf{v}_1+r\mathbf{v}_2$, 也就說是否 $T(\mathbf{v}_1+r\mathbf{v}_2)=\mathbf{w}_1+r\mathbf{w}_2$. 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. $\square$

在 Proposition 6.2.4 我們知道當 $T:V\to W$ 是 onto 時會保持 spanning set. 也就是說若 $\mathbf{v}_1,\dots,\mathbf{v}_n$ 是 $V$ 的一組 spanning set, 則 $T(\mathbf{v}_1),\dots,T(\mathbf{v}_n)$ 會是 $W$ 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 6.4.4.** 假設 $V,W$ 為 *vector spaces over* $\mathbb{F}$ 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 $c_1,\dots,c_n\in\mathbb{F}$ 不全為 0 使得 $c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=\mathbf{0}$. 現因 $T$ 為 linear, 我們有 $\mathbf{0}=c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)$. 因此利用 $T$ 為 one-to-one, 知僅有在 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$ 時才有可能使得 $T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=\mathbf{0}$. 故由 $\mathbf{v}_1,\dots,\mathbf{v}_n\in V$ 為 linearly independent 得知 $c_1=\cdots=c_n=0$. 此和 $c_1,\dots,c_n\in\mathbb{F}$ 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 $c_1,\dots,c_n\in\mathbb{F}$ 不全為 0 使得 $c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n=\mathbf{0}$. 現因 $T$ 為 linear, 我們有 $c_1T(\mathbf{v}_1)+\cdots+c_nT(\mathbf{v}_n)=T(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=T(\mathbf{0})=\mathbf{0}$. 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 invertible 的 linear transformation 時, 我們稱 $T$ 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 6.4.5.** 設 $V,W$ 為 *vector spaces over* $\mathbb{F}$ 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 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 $V$ 的一組 basis. 因為 $T$ 為 onto, 我們有 $\mathsf{R}(T)=W$. 又因 $\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ 為 $V$ 的 spanning set, 利用 Proposition 6.2.4 知 $\mathrm{Span}(T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n))=\mathsf{R}(T)=W$, 亦即 $\{T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)\}$ 為 $W$ 的 spanning set. 又因為 $T$ 為 one-to-one 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n\in V$ 為 linearly independent, 利用 Proposition 6.4.4 知 $T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)\in W$ 為 linearly independent. 故得 $T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)$ 為 $W$ 的一組 basis.

($\Leftarrow$) 假設 $T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)$ 為 $W$ 的一組 basis. 由於 $T^{-1}:W\to V$ 為 linear transformation (Theorem 6.4.3), 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 $T(\mathbf{v}_1),\ldots,T(\mathbf{v}_n)$ 為 $W$ 的一組 basis, 套用前面所證可得 $T^{-1}(T(\mathbf{v}_1)),\ldots,T^{-1}(T(\mathbf{v}_n))$ 為 $V$ 的一組 basis. 由於對於 $i=1,\ldots,n$ 皆有 $T^{-1}(T(\mathbf{v}_i))=\mathbf{v}_i$, 故知 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 $V$ 的一組 basis. □

假設 $V$ 為 vector space over $\mathbb{F}$ 且 $\beta$ 為 $V$ 的一組 ordered basis. 令 $T_\beta:V\to\mathbb{R}^n$ 定義為 $T_\beta(\mathbf{v})=[\mathbf{v}]_\beta$, $\forall\mathbf{v}\in V$, 在 Theorem 6.3.10 我們知道 $T_\beta$ 是 isomorphism. 而在 Proposition 6.3.16 中我們知道若 $V,W$ 為 vector space over $\mathbb{F}$ 且 $\beta,\gamma$ 分別為 $V,W$ 的一組 ordered basis, 則對於 linear transformation $T:V\to W$, $\mathbf{v}\in\mathsf{N}(T)$ 若且唯若 $T_\beta(\mathbf{v})\in\mathsf{N}([T]_\beta^\gamma)$, 而 $\mathbf{w}\in\mathsf{R}(T)$ 若且唯若 $T_\gamma(\mathbf{w})\in\mathrm{Col}([T]_\beta^\gamma)$. 因此 $T_\beta$ 定義出了一個從 $\mathsf{N}(T)$ 到 $\mathsf{N}([T]_\beta^\gamma)$ 的 isomorphism 且 $T_\gamma$ 定義出了一個從 $\mathsf{R}(T)$ 到 $\mathrm{Col}([T]_\beta^\gamma)$ 的 isomorphism. 利用 Theorem 6.4.5, 我們得到以下的結果.

**Corollary 6.4.6.** *$V,W$ 為 vector space over $\mathbb{F}$ 且 $\beta,\gamma$ 分別為 $V,W$ 的一組 ordered basis. 則 $\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}$ 為 $\mathsf{N}(T)$ 的一組 basis 若且唯若 $\{[\mathbf{v}_1]_\beta,\ldots,[\mathbf{v}_r]_\beta\}$ 為矩陣 $[T]_\beta^\gamma$ 的 null space $\mathsf{N}([T]_\beta^\gamma)$ 的一組 basis. 而 $\{\mathbf{w}_1,\ldots,\mathbf{w}_s\}$ 為 $\mathsf{R}(T)$ 的一組 basis 若且唯若 $\{[\mathbf{w}_1]_\gamma,\ldots,[\mathbf{w}_s]_\gamma\}$ 為矩陣 $[T]_\beta^\gamma$ 的 column space $\mathrm{Col}([T]_\beta^\gamma)$ 的一組 basis. 也因此得到*

$$\mathrm{nullity}(T)=\mathrm{nullity}([T]_\beta^\gamma),\quad \mathrm{rank}(T)=\mathrm{rank}([T]_\beta^\gamma).$$

當 $T:V\to W$ 為 isomorphism 時, 我們已知 $T$ 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 $\beta,\gamma$ 後, 我們有以下的 commutative diagram.

$$\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, shift right, "T_\beta"'] & W \arrow[r, "T^{-1}"] \arrow[d, shift left, "T_\gamma"] & V \arrow[d, shift left, "T_\beta"] \\
\mathbb{F}^n \arrow[r] \arrow[u, shift right, "T_\beta^{-1}"'] & \mathbb{F}^n \arrow[r] \arrow[u, shift left, "T_\gamma^{-1}"] & \mathbb{F}^n \arrow[u, shift left, "T_\beta^{-1}"]
\end{tikzcd}$$

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 $\mathrm{id}:V\to V$. 在 Example 6.3.12, 我們知道 $[\mathrm{id}]_\beta^\beta=I_n$, 故利用 Theorem 6.3.19 推得

$$[T^{-1}]_\gamma^\beta[T]_\beta^\gamma=[T^{-1}\circ T]_\beta^\beta=[\mathrm{id}]_\beta^\beta=I_n.$$

同理, 由於 $T\circ T^{-1}$ 為 $W$ 到 $W$ 的 identity map, 我們得 $[T]_\beta^\gamma[T^{-1}]_\gamma^\beta=I_n$. 得證 $[T^{-1}]_\gamma^\beta$ 為 $[T]_\beta^\gamma$ 的反矩陣.

綜合以上的討論, 我們有以下的結論.

**Theorem 6.4.7.** 假設 $V,W$ 為 *finite dimensional vector spaced over* $\mathbb{F}$ 且令 $\beta,\gamma$ 分別為 $V,W$ 的 *ordered basis.* 設 $T:V\to W$ 為 *linear transformation.* 則 $T$ 為 *isomorphism* 若且唯若 $[T]_\beta^\gamma$ 為 *invertible matrix.* 又此時 $T^{-1}:W\to V$ 對應於 $\beta,\gamma$ 的 *matrix representation* 為

$$[T^{-1}]_\gamma^\beta=([T]_\beta^\gamma)^{-1}.$$

**Proof.** 由前面的討論已知當 $T$ 為 isomorphism 時其對應於 ordered basis $\beta,\gamma$ 的 matrix representation $[T]_\beta^\gamma$ 為 invertible matrix. 我們僅須證明當 $[T]_\beta^\gamma$ 為 invertible matrix 時, $T:V\to W$ 為 isomorphism, 亦即證明 $T$ 為 one-to-one 且 onto. 然而由 $[T]_\beta^\gamma$ 為 invertible matrix, 我們知 $[T]_\beta^\gamma$ 為方陣, 亦即 $\dim(V)=\dim(W)$. 因為 invertible matrix 的 null space 是 $\{\mathbf{0}\}$, 故由 Proposition 6.3.16 知 $\mathsf{N}(T)=\{\mathbf{0}\}$, 亦即 $T$ 為 one-to-one. 最後由 $\dim(V)=\dim(W)$ 以及 Lemma 6.4.1 我們得證 $T$ 亦為 onto. $\square$

**Example 6.4.8.** 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 2.5.8 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.

## 6.5. 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.

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

# Linear Operators

在這一章中, 我們探討特別的一種但很常用的 linear transformation, 稱為 linear operator. 它是定義域與對應域相同的 linear transformation. 這一章我們介紹其基本性質，下一章再更進一步探討對角化問題。

## 7.1. Change of Basis

在這一節中, 我們介紹 change of basis 的概念, 了解到一個 linear operator 換了 ordered basis 後其表現矩陣的關係. 這個概念能幫助我們以後處理矩陣對角化的問題.

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 $\beta,\beta'$ 為 $V$ 的兩組 ordered bases, 而 $\gamma,\gamma'$ 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 $[T]_\beta^\gamma$ 和 $[T]_{\beta'}^{\gamma'}$ 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 $\beta=\{\mathbf{v}_1,\dots,\mathbf{v}_n\}$, 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 $\beta$ 這一組 ordered basis, 而對應域選的是 $\beta'=\{\mathbf{v}'_1,\dots,\mathbf{v}'_n\}$ 這一組 ordered basis, identity map 對應於 $\beta,\beta'$ 的 matrix representation $[\mathrm{id}]_\beta^{\beta'}$ 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 $[\mathbf{v}_i]_{\beta'}$. 所以當 $\beta$ 和 $\beta'$ 相異時, $[\mathrm{id}]_\beta^{\beta'}$ 不是 identity matrix. 現對任意 $\mathbf{v}\in V$, 因 $\mathbf{v}$ 對於 $\beta$ 的坐標表示法為 $[\mathbf{v}]_\beta$, 依 matrix representation 的性質 (Proposition 6.3.14) 可得

$$[\mathrm{id}]_\beta^{\beta'}[\mathbf{v}]_\beta=[\mathrm{id}(\mathbf{v})]_{\beta'}=[\mathbf{v}]_{\beta'}.$$

也就是說, 矩陣 $[\mathrm{id}]_\beta^{\beta'}$ 可以將 $V$ 中元素對於 $\beta$ 的坐標表示轉換成對於 $\beta'$ 的坐標表示, 也因此我們稱 $[\mathrm{id}]_\beta^{\beta'}$ 為 *change-of-basis matrix.*

要注意 $\mathrm{id}: V \to V$ 是 isomorphism, 所以由 Theorem 6.3.19 (3), 我們得 $[\mathrm{id}]_{\beta}^{\beta'}$ 為 invertible 且

$$([\mathrm{id}]_{\beta}^{\beta'})^{-1} = [\mathrm{id}^{-1}]_{\beta'}^{\beta} = [\mathrm{id}]_{\beta'}^{\beta} \tag{7.1}$$

也就是說將 $\beta$ 的坐標表示轉換成對於 $\beta'$ 的坐標表示的 change-of-basis matrix 的 inverse 就是 $\beta'$ 的坐標表示轉換成對於 $\beta$ 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 $\beta, \beta'$ 為 $V$ 的兩組 ordered bases, 而 $\gamma, \gamma'$ 為 $W$ 的兩組 ordered basis. 我們要探討 $[T]_{\beta}^{\gamma}$ 和 $[T]_{\beta'}^{\gamma'}$ 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 6.3.19 (2) 得

$$[\mathrm{id}_W]_{\gamma}^{\gamma'}[T]_{\beta}^{\gamma}[\mathrm{id}_V]_{\beta'}^{\beta} = [T]_{\beta'}^{\gamma'}.$$

這就是所謂的 change-of basis formula, 我們將之完整敘述如下.

**Theorem 7.1.1** (Change-of-basis Formula)**.** 假設 $T: V \to W$ 為 *linear transformation* 且 $\beta, \beta'$ 為 $V$ 的兩組 *ordered bases,* 而 $\gamma, \gamma'$ 為 $W$ 的兩組 *ordered basis,* 則存在 *invertible matrix* $P, Q$ 使得 $[T]_{\beta'}^{\gamma'} = Q([T]_{\beta}^{\gamma})P$, 其中 $P$ 為將 $\beta'$ 的坐標表示轉換成 $\beta$ 的坐標表示的 *change-of-basis matrix* $[\mathrm{id}_V]_{\beta'}^{\beta}$, 而 $Q$ 為將 $\gamma$ 的坐標表示轉換成 $\gamma'$ 的坐標表示的 *change-of-basis matrix* $[\mathrm{id}_W]_{\gamma}^{\gamma'}$.

**Example 7.1.2.** 在 Example 6.3.13 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 $\varepsilon = (x^2, x, 1)$, $\beta = (p_1(x), p_2(x), p_3(x))$ 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 6.3.13 中我們得到

$$[T]_{\varepsilon}^{\varepsilon'} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \quad [T]_{\beta}^{\beta'} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

因 $[p_1(x)]_{\varepsilon} = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \end{bmatrix}$, $[p_2(x)]_{\varepsilon} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, $[p_3(x)]_{\varepsilon} = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$ 依定義 $\beta$ 到 $\varepsilon$ 的 change-of-basis matrix 為 $[\mathrm{id}_{P_2(\mathbb{R})}]_{\beta}^{\varepsilon} = \begin{bmatrix} 1/2 & -1 & 1/2 \\ -1/2 & 0 & 1/2 \\ 0 & 1 & 0 \end{bmatrix}$. 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 $[x^3]_{\beta'} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 8 \end{bmatrix}$. 用同樣方法求 $x^2, x, 1$ 對於 $\beta'$ 的坐標表示法,

我們得 $\varepsilon'$ 到 $\beta'$ 的 change-of-basis matrix 為 $[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon'}^{\beta'} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{bmatrix}$. 我們也可以先寫下 $\beta'$ 到 $\varepsilon'$ 的 change-of-basis matrix $[\mathrm{id}_{P_3(\mathbb{R})}]_{\beta'}^{\varepsilon'} = \begin{bmatrix} -1/6 & 1/2 & -1/2 & 1/6 \\ 1/2 & -1 & 1/2 & 0 \\ -1/3 & -1/2 & 1 & -1/6 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ 再取 inverse 得 $[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon'}^{\beta'}$. 最後我們驗算

$$[\mathrm{id}_{P_3(\mathbb{R})}]_{\varepsilon'}^{\beta'} [T]_{\varepsilon}^{\varepsilon'} [\mathrm{id}_{P_2(\mathbb{R})}]_{\beta}^{\varepsilon} = \begin{bmatrix} -1 & 1 & -1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 8 & 4 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -1 & \frac{1}{2} \\ \frac{-1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} = [T]_{\beta}^{\beta'}.$$

前面提過, 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 7.1.1, 我們得以下之結果.

**Corollary 7.1.3.** 假設 $T : V \to V$ 為 *linear transformation* 且 $\beta, \beta'$ 為 $V$ 的兩組 *ordered bases.* 則存在 *invertible matrix* $P$ 使得 $[T]_{\beta'}^{\beta'} = P^{-1}([T]_{\beta}^{\beta})P$, 其中 $P$ 為將 $\beta'$ 的坐標表示轉換成 $\beta$ 的坐標表示的 *change-of-basis matrix* $[\mathrm{id}_V]_{\beta'}^{\beta}$.

**Proof.** 考慮 Theorem 7.1.1 其中 $W = V$, $\gamma = \beta$ 以及 $\gamma' = \beta'$ 的情形. 此時 $Q = [\mathrm{id}_V]_{\beta}^{\beta'}$ 由式子 (7.1), 知 $Q = ([\mathrm{id}_V]_{\beta'}^{\beta})^{-1} = P^{-1}$, 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. 例如 Orthonormal basis 也可幫助我們處理 linear operator 的問題. 考慮 $T : V \to V$ 為 linear operator. 當我們給定 $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 為 $V$ 的 ordered basis, 我們便可得到 $T$ 對 $\beta$ 的表現矩陣 $A = [T]_{\beta}$. 其中 $A$ 的 $j$-th column, 就是 $T(\mathbf{v}_j)$ 用 $\beta$ 寫下的坐標. 也就是說若 $T(\mathbf{v}_j) = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 則 $A$ 的 $j$-th column 就是 $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$. 特別的, 當 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 是 $V$ 的一組 orthonormal basis, 我們很容易將 $T(\mathbf{v}_j)$ 寫成 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 的線性組合, 事實上 Proposition 4.3.6 告訴我們 $T(\mathbf{v}_j) = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$, 其中 $c_i = \langle T(\mathbf{v}_j), \mathbf{v}_i \rangle = \langle \mathbf{v}_i, T(\mathbf{v}_j) \rangle$. 也就是說 $A$ 的 $j$-th column 其 $i$-th entry 為 $c_i = \langle \mathbf{v}_i, T(\mathbf{v}_j) \rangle$, 因此 $[T]_{\beta}$ 的 $(i, j)$-th entry 就是 $\langle \mathbf{v}_i, T(\mathbf{v}_j) \rangle$.

**Proposition 7.1.4.** *假設 $V$ 為 inner product space 且 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 $V$ 的一組 orthonormal basis. 若 $T:V\to V$ 為 linear operator 且考慮 $V$ 的 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 則 $T$ 用 $\beta$ 所得的 matrix representation $[T]_\beta$ 其 $(i,j)$-th entry 為 $\langle \mathbf{v}_i, T(\mathbf{v}_j)\rangle$.*

**Question 7.1.** *在 Proposition 7.1.4 中若 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$ 是由 orthogonal basis 所形成, 則 $[T]_\beta$ 的 $(i,j)$-th entry 應為何?*

另外有的 linear operator 可以找到好的基底使其 matrix representation 為對角矩陣. 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 7.1.5.** 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 $[T]_\varepsilon^\varepsilon=\frac{1}{25}\begin{bmatrix} 9 & 12\\ 12 & 16\end{bmatrix}$. 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 $[T]_\beta^\beta=\begin{bmatrix}1&0\\0&0\end{bmatrix}$. 我們很容易由

$$[T\circ T]_\beta^\beta=([T]_\beta^\beta)([T]_\beta^\beta)=\begin{bmatrix}1&0\\0&0\end{bmatrix}^2=\begin{bmatrix}1&0\\0&0\end{bmatrix}=[T]_\beta^\beta$$

推得 $T\circ T=T$. 事實上從 $\beta$ 這組 ordered basis 我們很容易看出 $T$ 就是將 $\mathbb{R}^2$ 上的向量對 $\begin{bmatrix}3\\4\end{bmatrix}$ 的投影. 另外令 $P=[\mathrm{id}]_\beta^\varepsilon=\begin{bmatrix}3&-4\\4&3\end{bmatrix}$, 我們得

$$\begin{bmatrix}1&0\\0&0\end{bmatrix}=[T]_\beta^\beta=([\mathrm{id}]_\varepsilon^\beta)([T]_\varepsilon^\varepsilon)([\mathrm{id}]_\beta^\varepsilon)=P^{-1}(\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix})P,$$

所以 $\begin{bmatrix}1&0\\0&0\end{bmatrix}$ 和 $\frac{1}{25}\begin{bmatrix}9&12\\12&16\end{bmatrix}$ 為 similar.

Corollary 7.1.3 告訴我們，一個 linear operator 選取不同的 ordered basis, 其表現矩陣會有 similar 的關係. 反過來, 當給定 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$, 我們可以考慮 $L_A:\mathbb{F}^n\to\mathbb{F}^n$, 其定義為 $L_A(\mathbf{v})=A\mathbf{v}$, $\forall\mathbf{v}\in\mathbb{F}^n$ 這一個 linear operator. 此時 $L_A$ 對 $\mathbb{F}^n$ 的 standard ordered basis $\varepsilon=(\mathbf{e}_1,\ldots,\mathbf{e}_n)$ 的表現矩陣 $[L_A]_\varepsilon^\varepsilon$ 就是 $A$. 現若 $B\in\mathsf{M}_{n\times n}(\mathbb{F})$ 且 $B$ 和 $A$ similar, 亦即存在 invertible matrix $U$ 滿足 $B=U^{-1}AU$. 現考慮 ordered basis $\beta=(\mathbf{v}_1,\ldots,\mathbf{v}_n)$, 其中 $\mathbf{v}_i$ 是 $U$ 的 $i$-th column, 則依定義 $U=[\mathrm{id}_{\mathbb{F}^n}]_\beta^\varepsilon$, 也因此由 Corollary 7.1.3 知 $B$ 是 $L_A$ 用 $\beta$ 所得的表現矩陣, 即 $B=[L_A]_\beta^\beta$. 從這裡我們知道, 以後要探討兩個相似矩陣的問題, 我們都可以將之視為是同一個 linear operator 利用不同的 ordered basis 所得的表現矩陣.

## 7.2. Characteristic Polynomial

給定 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$, 以及 $\mathbf{v}\in\mathbb{F}^n$, 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 7.2.1.** 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. 若對於非零向量 $\mathbf{v} \in \mathbb{F}^n$, 存在 $\lambda \in \mathbb{F}$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$, 則稱 $\mathbf{v}$ 為 $A$ 的一個 *eigenvector*, 而此 $\lambda$ 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 $(\lambda - \lambda')\mathbf{v} = \mathbf{0}$ 以及 $\mathbf{v} \neq \mathbf{0}$, 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 $\lambda$ 會滿足 $A\mathbf{v} = \lambda\mathbf{v}$. 此時我們稱 eigenvector $\mathbf{v}$ 所對應的 eigenvalue 為 $\lambda$.

**Question 7.2.** 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$, $\mathbf{v} \in \mathbb{F}^n$ 為非零向量滿足 $A\mathbf{v} = \mathbf{0}$. 是否 $\mathbf{v}$ 為 *eigenvector?* 其所對應的 *eigenvalue* 為何?

**Example 7.2.2.** 考慮

$$A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}.$$

我們有

$$A\mathbf{v}_1 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2\mathbf{v}_1.$$

所以 $\mathbf{v}_1$ 是 $A$ 的一個 eigenvector, 而 $-2$ 是其對應的 eigenvalue. 同樣的

$$A\mathbf{v}_2 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 20 \end{bmatrix} = 5 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 5\mathbf{v}_2.$$

所以 $\mathbf{v}_2$ 是 $A$ 的一個 eigenvector, 而 5 是其對應的 eigenvalue. 然而

$$A\mathbf{v}_3 = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 13 \\ 22 \end{bmatrix} \notin \mathrm{Span}(\begin{bmatrix} 4 \\ 3 \end{bmatrix}).$$

所以 $\mathbf{v}_3$ 不是 $A$ 的一個 eigenvector.

首先我們看一些 eigenvector 和 eigenvalue 的性質.

**Proposition 7.2.3.** 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$, $\mathbf{v} \in \mathbb{F}^n$ 是 $A$ 的 *eigenvector* 且其 *eigenvalue* 為 $\lambda$.

(1) 若 $c \in \mathbb{F}$ 且 $c \neq 0$, 則 $c\mathbf{v}$ 亦為 $A$ 的一個 *eigenvalue* 為 $\lambda$ 的 *eigenvector.*

(2) 若 $\mathbf{v}' \in \mathbb{F}^n$ 亦為 $A$ 的一個 *eigenvalue* 為 $\lambda$ 的 *eigenvector* 且 $\mathbf{v} + \mathbf{v}' \neq 0$, 則 $\mathbf{v} + \mathbf{v}'$ 亦為 $A$ 的一個 *eigenvalue* 為 $\lambda$ 的 *eigenvector.*

**Proof.** 依假設我們知道 $\mathbf{v} \neq \mathbf{0}$ 且 $A\mathbf{v} = \lambda\mathbf{v}$.

(1) 令 $\mathbf{w}=c\mathbf{v}$, 由於 $c\neq 0$ 且 $\mathbf{v}\neq\mathbf{0}$, 我們知 $\mathbf{w}\neq\mathbf{0}$. 現考慮

$$A\mathbf{w}=A(c\mathbf{v})=c(A\mathbf{v})=c(\lambda\mathbf{v})=\lambda(c\mathbf{v})=\lambda\mathbf{w}.$$

得證 $\mathbf{w}=c\mathbf{v}$ 為 $A$ 的一個以 $\lambda$ 為 eigenvalue 的 eigenvector.

(2) 令 $\mathbf{u}=\mathbf{v}+\mathbf{v}'$. 依假設 $A\mathbf{v}'=\lambda\mathbf{v}'$ 且 $\mathbf{u}\neq\mathbf{0}$. 現考慮

$$A\mathbf{u}=A(\mathbf{v}+\mathbf{v}')=A\mathbf{v}+A\mathbf{v}'=\lambda\mathbf{v}+\lambda\mathbf{v}'=\lambda(\mathbf{v}+\mathbf{v}')=\lambda\mathbf{u}.$$

得證 $\mathbf{u}=\mathbf{v}+\mathbf{v}'$ 為 $A$ 的一個以 $\lambda$ 為 eigenvalue 的 eigenvector.

□

**Question 7.3.** 假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$ 且 $\mathbf{v}_1,\mathbf{v}_2\in\mathbb{F}^n$ 皆為 $A$ 的一個以 *eigenvalue* 為 $\lambda$ 的 *eigenvector.* 證明若 $\mathbf{w}\in\mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$ 且 $\mathbf{w}\neq\mathbf{0}$, 則 $\mathbf{w}$ 也是 $A$ 的一個以 *eigenvalue* 為 $\lambda$ 的 *eigenvector.*

要注意, Question 7.3 並不是說任意兩個 eigenvector 的線性組合仍為 eigenvector. 必須是它們所對應的 eigenvalue 是一樣的才會對. 例如在 Example 7.2.2 中雖然 $\mathbf{v}_1,\mathbf{v}_2$ 都是 $A$ 的 eigenvector, 但 $\mathbf{v}_3=\mathbf{v}_1+\mathbf{v}_2$ 就不是 $A$ 的 eigenvector.

在 Example 7.2.2 中 $A\in\mathsf{M}_{2\times 2}(\mathbb{R})$, 而 $\mathbf{v}_1,\mathbf{v}_2$ 不平行, 所以 $\mathbf{v}_1,\mathbf{v}_2$ 形成 $\mathbb{R}^2$ 的一組 basis. 另一方面 $\mathbf{v}_1,\mathbf{v}_2$ 都是 $A$ 的 eigenvector. 這樣的矩陣 $A$ 是很特別的, 我們對有這樣特點的 matrix 給了以下的定義.

**Definition 7.2.4.** 假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$. 若存在 $\mathbb{F}^n$ 的一組 basis $\mathbf{v}_1,\dots,\mathbf{v}_n$ 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 $\mathbf{v}_1,\dots,\mathbf{v}\in\mathbb{F}^n$ 是 $\mathbb{F}^n$ 的一組 basis 且皆為 $A$ 的 eigenvectors, 又假設它們所對應的 eigenvalues 分別為 $\lambda_1,\dots,\lambda_n$. 亦即 $A\mathbf{v}_1=\lambda_1\mathbf{v}_1,\dots,A\mathbf{v}_n=\lambda_n\mathbf{v}_n$. 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 $\lambda_i$ 的 $n\times n$ diagonal matrix $D$(即對角線第 $i$ 個位置為 $\lambda_i$ 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 2.5.2). 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 $\mathbb{F}^n$ 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 $\mathbb{F}^n$ 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.

**Example 7.2.5.** 考慮 Example 7.2.2 中
$$A=\begin{bmatrix}1&3\\4&2\end{bmatrix},\mathbf{v}_1=\begin{bmatrix}1\\-1\end{bmatrix},\mathbf{v}_2=\begin{bmatrix}3\\4\end{bmatrix}.$$
由於 $\mathbf{v}_1,\mathbf{v}_2$ 為 $A$ 的 eigenvectors 且可形成 $\mathbb{R}^2$ 的一組 basis, 我們知 $A$ 為 diagonalizable. 事實上若令 $C=\begin{bmatrix}1&3\\-1&4\end{bmatrix}$, 則由
$$AC=\begin{bmatrix}1&3\\4&2\end{bmatrix}\begin{bmatrix}1&3\\-1&4\end{bmatrix}=\begin{bmatrix}-2&15\\2&20\end{bmatrix}=\begin{bmatrix}1&3\\-1&4\end{bmatrix}\begin{bmatrix}-2&0\\0&5\end{bmatrix}=C\begin{bmatrix}-2&0\\0&5\end{bmatrix}.$$
故得 $C^{-1}AC=\begin{bmatrix}-2&0\\0&5\end{bmatrix}$ 為 diagonal matrix.

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 $\lambda\in\mathbb{F}$ 是 $A$ 的 eigenvalue, 表示存在一個非零向量 $\mathbf{v}\in\mathbb{F}^n$ 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 $(A-\lambda I_n)\mathbf{v}=\mathbf{0}$. 換言之, $\lambda$ 是 $A$ 的 eigenvalue 等同於由 $n\times n$ matrix $A-\lambda I_n$ 所對應的 linear system $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 有 nontrivial solution $\mathbf{x}=\mathbf{v}$. 由 Theorem 2.5.9, 這又等同於 $A-\lambda I_n$ 不是 invertible, 再由 Theorem 5.2.6(1) 知這也等同於 $\det(A-\lambda I_n)=0$. 總言之, 要找到 $A$ 的 eigenvalue $\lambda$ 就是要找到 $\lambda$ 滿足 $\det(A-\lambda I_n)=0$.

要怎樣找到 $\lambda$ 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 $\lambda$ 就會滿足 $\det(A-\lambda I_n)=0$, 也就是說 $\lambda$ 就會是 $A$ 的一個 eigenvalue. 反之, 若 $\lambda$ 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 7.2.6.** 假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$, 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 $\lambda \in \mathbb{F}$ 為 characteristic polynomial $p_A(t)$ 的一個根若且唯若 $\lambda$ 為 $A$ 的 eigenvalue. 這裡要注意要談論 eigenvalue 是必須強調在哪一個 field 的 eigenvalue. 例如當 $A \in \mathsf{M}_{n\times n}(\mathbb{R})$, 其 characteristic polynomial $p_A(t)$ 是一個實係數多項式, 不過 $p_A(t)$ 有可能有非實數的虛根. 此時這個虛根不會是 $A$ 在 $\mathbb{R}^n$ 中的 eigenvector 所對應的 eigenvalue. 事實上如果 $\lambda \in \mathbb{C}\setminus\mathbb{R}$ 是 $p_A(t)$ 的一個虛根, 此時假設存在 $\mathbf{v} \in \mathbb{R}^n$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$. 由於 $A\mathbf{v} \in \mathbb{R}^n$, 但 $\lambda\mathbf{v} \notin \mathbb{R}^n$, 所以 $A\mathbf{v} = \lambda\mathbf{v}$ 不可能成立. 不過依前面的探討我們知道一定會有 $\mathbf{w} \in \mathbb{C}^n$ 滿足 $A\mathbf{w} = \lambda\mathbf{w}$. 在這個課程裡, 當我們探討矩陣 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 的 eigenvalue 時, 若沒有特別說明, 都僅討論在 $\mathbb{F}$ 的 eigenvalue. 例如當我們討論實矩陣時, 我們考慮 eigenvalue 僅考慮 characteristic polynomial 的實根.

**Example 7.2.7.** 考慮 $B = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$, 此時 $A$ 的 characteristic polynomial 為

$$p_B(t) = \det(B - tI_3) = \det \begin{bmatrix} -1-t & 4 & 2 \\ 1 & 3-t & 1 \\ -1 & 2 & 2-t \end{bmatrix}.$$

對第一個 row 降階求行列式得

$$p_A(t) = (-1-t)\det\begin{bmatrix} 3-t & 1 \\ 2 & 2-t \end{bmatrix} - 4\det\begin{bmatrix} -1 & 1 \\ -1 & 2-t \end{bmatrix} + 2\det\begin{bmatrix} -1 & 3-t \\ -1 & 2 \end{bmatrix}.$$

化簡可得 $p_A(t) = -(t-1)^2(t-2)$. 也因此 $t=1$ 和 $t=2$ 為 $A$ 的 characteristic polynomial 的二實根, 也因此得 $A$ 有兩個 eigenvalues $1,2$.

接下來我們說明當 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 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 $2\times 2$ 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 $3\times 3$ 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 $2\times 2$ 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 $(-1)^3t^3 + (-1)^2(a+e+i)t^2$. 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 $(-1)^n t^n+(-1)^{n-1}(a_{11}+\cdots+a_{nn})t^{n-1}$. 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 7.2.8.** 假設 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$. 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 7.4.** 假設 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$. 試問 $A$ 最多會有幾個相異的 *eigenvalues?*

**Example 7.2.9.** 考慮 Example 7.2.2 中 $A=\begin{bmatrix}1&3\\4&2\end{bmatrix}$ 的 characteristic polynomial $p_A(t)$. 由於 $\mathrm{tr}(A)=1+2=3$ 以及 $\det(A)=2-12=-10$, 利用 Proposition 7.2.8 可得

$$p_A(t)=(-1)^2t^2+(-1)3t+(-10)=t^2-3t-10.$$

事實上利用 characteristic polynomial 的定義直接計算可得

$$p_A(t)=\det\begin{bmatrix}1-t&3\\4&2-t\end{bmatrix}=(1-t)(2-t)-12=t^2-3t-10.$$

分解後可得 $-2,5$ 為 $A$ 的 eigenvalues.

接下來我們介紹一個和 eigenvalue 有關的定義. 若 $\lambda\in\mathbb{F}$ 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 $\lambda$ 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 $\lambda$ 的 algebraic multiplicity 為 1. 例如 Example 7.2.7 中 $B$ 有兩個 eigenvalue 1 和 2, 其中 eigenvalue 1 的 algebraic multiplicity 為 2, 而 eigenvalue 2 的 algebraic multiplicity 為 1. 而 Example 7.2.9 中 $A$ 的兩個 eigenvalue $-2,5$ 其 algebraic multiplicity 皆為 1. 有關 algebraic multiplicity 的性質, 以後我們還會進一步討論.

**Question 7.5.** *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. 前面提過當 $A,B$ 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 7.2.10.** 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 5.2.6) 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 7.2.11.** 假設 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$, 則 $A$ 和 $A^t$ 有相同的 *characteristic polynomial*

**Proof.** 利用 trnaspose 的性質 $(A-tI_n)^t=A^t-tI_n^t=A^t-tI_n$ (Proposition 2.2.4), 故利用 Theorem 5.2.6 (3), 我們有

$$P_{A^t}(t)=\det(A^t-tI_n)=\det((A-tI_n)^t)=\det(A-tI_n)=P_A(t).$$

□

**Question 7.6.** 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.

## 7.3. Eigenspace 和 Eigenvector

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 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$ 且 $\lambda\in\mathbb{F}$ 為 $A$ 的一個 eigenvalue. 由於 $\det(A-\lambda I_n)=0$, 我們知聯立方程組 $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 存在非零的 nontrivial solution. 現假設 $\mathbf{v}\in\mathbb{F}^n$ 為非零向量且 $\mathbf{x}=\mathbf{v}$ 為 $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 的一組解. 此即表示 $\mathbf{v}$ 滿足 $(A-\lambda I_n)\mathbf{v}=\mathbf{0}$, 亦即 $A\mathbf{v}=\lambda\mathbf{v}$. 故此時 $\mathbf{v}$ 為 $A$ 的一個以 $\lambda$ 為 eigenvalue 的 eigenvector. 反之, 若 $\mathbf{v}$ 為 $A$ 的一個以 $\lambda$ 為 eigenvalue 的 eigenvector, 則 $\mathbf{x}=\mathbf{v}$ 必為 $(A-\lambda I_n)\mathbf{x}=\mathbf{0}$ 的一組 nontrivial solution. 因此我們只要掌握 $n\times n$ matrix $A-\lambda I_n$ 的 nullspace (即 $\{\mathbf{v}\in\mathbb{F}^n\mid (A-\lambda I_n)\mathbf{v}=\mathbf{0}\}$) 中的非零向量就會是 $A$ 相對於 $\lambda$ 的 eigenvector. 由於 nullspace 是 vector space, 因此我們有以下的定義.

**Definition 7.3.1.** 假設 $A\in \mathsf{M}_{n\times n}(\mathbb{F})$ 且 $\lambda\in\mathbb{F}$ 為 $A$ 的一個 eigenvalue. 則 $A-\lambda I_n$ 的 nullspace 稱為 $A$ 對於 eigenvalue $\lambda$ 的 *eigenspace*. 我們用 $E_A(\lambda)$ 來表示.

要注意對於 $\lambda$ 的 eigenspace 並不是由以 $\lambda$ 為 eigenvalue 的 eigenvectors 所組成. 這是因為零向量 $\mathbf{0}$ 不是 eigenvector, 但 vector space 必須包含 $\mathbf{0}$. 所以對於 $\lambda$ 的 eigenspace 應該是由所有以 $\lambda$ 為 eigenvalue 的 eigenvectors 和 $\mathbf{0}$ 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 $\lambda$ 的 *geometric multiplicity*

(幾何重根數). 要注意 eigenvalue $\lambda$ 的 algebraic multiplicity 無法讓我們知道 $\lambda$ 所對應的 eigenvectors 的多寡, 而是 $\lambda$ 的 geometric multiplicity 可以提供這一個訊息.

**Example 7.3.2.** 考慮 $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$. 由前面 Example 7.2.7, Example 7.2.9 我們已計算出 $A$ 和 $B$ 的 characteristic polynomial 分別為 $p_A(t) = (x+2)(x-5)$, $p_B(t) = -(t-1)^2(t-2)$. 接下來我們分別計算 $A$ 和 $B$ 的 eigenspace.

首先考慮 $A$ 對於 eigenvalue $-2$ 的 eigenspace, 亦即找出 $A-(-2I_2) = \begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$. 可得 $E_A(1) = \mathrm{Span}(\begin{bmatrix} 1 \\ -1 \end{bmatrix})$. 也就是說 $A$ 對於 eigenvalue 為 $-2$ 的 eigenvector 就是那些和 $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ 平行的 nonzero vector. 由於 $\dim(E_A(-2)) = 1$, 我們也得到 $A$ 對於 eigenvalue $-2$ 的 geometric multiplicity 為 1. 至於 $A$ 對於 eigenvalue 5 的 eigenspace, 亦即找出 $A-5I_2 = \begin{bmatrix} -4 & 3 \\ 4 & -3 \end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix} -4 & 3 \\ 0 & 0 \end{bmatrix}$. 因此得 $E_A(5) = \mathrm{Span}(\begin{bmatrix} 3 \\ 4 \end{bmatrix})$. 也就是說 $A$ 對於 eigenvalue 為 5 的 eigenvector 就是那些和 $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ 平行的 nonzero vector, 我們也得到 $A$ 對於 eigenvalue 5 的 geometric multiplicity 為 1. 在 Example 7.2.2 中我們舉出 $A$ 的 eigenvector 的例子其實是這樣得到的.

接著考慮 $B$ 對於 eigenvalue 1 的 eigenspace, 亦即找出 $B - I_3 = \begin{bmatrix} -2 & 4 & 2 \\ -1 & 2 & 1 \\ -1 & 2 & 1 \end{bmatrix}$ 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 $E_B(1) = \mathrm{Span}(\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix})$. 也就是說 $B$ 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

$$B\mathbf{v} = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \mathbf{v}.$$

由於 $\dim(E_B(1)) = 2$, 我們也得到 $B$ 對於 eigenvalue 1 的 geometric multiplicity 為 2. 至於 $B$ 對於 eigenvalue 2 的 eigenspace, 亦即找出 $B-2I_3 = \begin{bmatrix} -3 & 4 & 2 \\ -1 & 1 & 1 \\ -1 & 2 & 0 \end{bmatrix}$ 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 $E_B(2) =$

Span($\begin{bmatrix}2\\1\\1\end{bmatrix}$). 也就是說 $B$ 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 $B$ 對於 eigenvalue 2 的 geometric multiplicity 為 1.

在 Proposition 7.2.3 中我們知道兩個有相同 eigenvalue 的 eigenvectors 其線性組合只要不是 $\mathbf{0}$, 就會是有同樣 eigenvalue 的 eigenvector. 所以一般在探討一個 $n\times n$ 矩陣的 eigenvector 時, 我們只要寫下其 eigenspace 的一組基底即可.

以前我們提過有關 matrix 的問題都可以轉換成 linear transformation 的問題, 反之亦然. 回顧一下當 $V$ 是 over $\mathbb{F}$ 的 vector space, 則一個 linear transformation $T:V\to V$, 稱為一個 linear operator. 特別地當 $\dim_{\mathbb{F}}(V)=n$, 且 $\beta$ 是 $V$ 的一組 ordered basis, 則 $T$ 利用這組 ordered basis 所得的 matrix representation $[T]_\beta$ 會是一個 $n\times n$ matrix. 注意這裡因為定義域和對應域都是 $V$ 所以兩邊是選同樣的 ordered basis, 因此我們將原本 matrix representation 的表示法 $[T]_\beta^\beta$ 省略寫成 $[T]_\beta$. 方陣 $[T]_\beta$ 的 eigenvalue 和 eigenvector 會和 $T$ 有甚麼關係呢? 假設 $\beta=(\mathbf{v}_1,\dots,\mathbf{v}_n)$, 而 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\in\mathbb{F}^n$ 是 $[T]_\beta$ 的一個 eigenvector 且其 eigenvalue 為 $\lambda$, 此時我們有

$$[T]_\beta\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=\lambda\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}=\begin{bmatrix}\lambda c_1\\\vdots\\\lambda c_n\end{bmatrix}.$$

若令 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$, 回顧一下在 Proposition 6.3.14 中告訴我們這表示 $T(\mathbf{v})$ 用 $\beta$ 這組 ordered basis 的坐標表示應該是 $\begin{bmatrix}\lambda c_1\\\vdots\\\lambda c_n\end{bmatrix}$. 也就是說

$$T(\mathbf{v})=\lambda c_1\mathbf{v}_1+\cdots+\lambda c_n\mathbf{v}_n=\lambda(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=\lambda\mathbf{v}.$$

反之, 若 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n\in V$ 滿足 $T(\mathbf{v})=\lambda\mathbf{v}$, 則由 Proposition 6.3.14 知 $\begin{bmatrix}c_1\\\vdots\\c_n\end{bmatrix}\in\mathbb{F}^n$ 是 $[T]_\beta$ 的一個 eigenvector 且其 eigenvalue 為 $\lambda$. 也因此我們有以下的定義.

**Definition 7.3.3.** 假設 $V$ 是一個 vector space over $\mathbb{F}$ 且 $T:V\to V$ 是一個 linear operator. 若對 $\mathbf{v}\in V$, 存在 $\lambda\in\mathbb{F}$ 滿足 $T(\mathbf{v})=\lambda\mathbf{v}$, 則稱 $\mathbf{v}$ 為 $T$ 的一個 *eigenvector*, 且 $\lambda$ 為其 *eigenvalue*.

**Example 7.3.4.** 考慮 Linear operator $T:P_2(\mathbb{R})\to P_2(\mathbb{R})$ 定義為

$$T(f(x))=f(x)+(x+1)f'(x),\quad \forall f(x)\in P_2(\mathbb{R}).$$

此時令 $g(x)=x^2+2x+1$, 則

$$T(g(x))=(x^2+2x+1)+(x+1)(2x+2)=3(x^2+2x+1)=3g(x).$$

故 $x^2+2x+1$ 是 $T$ 的一個 eigenvector 且其 eigenvalue 為 3.

在 Proposition 7.2.3, 我們提到關於方陣的 eigenvalue 和 eigenvector 的性質. 事實上這性質對 linear operator 也是對的, 我們有以下的性質. 由於證明方法和矩陣的情形一致, 我們就不再證明了.

**Proposition 7.3.5.** 假設 $V$ 是一個 *vector space over* $\mathbb{F}$ 且 $T: V \to V$ 是一個 *linear operator.* 又假設 $\mathbf{v}_1, \mathbf{v}_2$ 為 $T$ 的 *eigenvectors* 且其 *eigenvalue* 皆為 $\lambda \in \mathbb{F}$. 若 $c_1, c_2 \in \mathbb{F}$ 且 $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \neq \mathbf{0}$, 則 $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ 也會是 $T$ 的一個以 $\lambda$ 為 *eigenvalue* 的 *eigenvector.*

回顧一下在 $\mathsf{M}_{n\times n}(\mathbb{F})$ 的情形, 我們有所謂 diagonalizable matrix, 也就是說這樣的矩陣可以在 $\mathbb{F}^n$ 找到一組由 eigenvectors 所組成的 basis. 同樣的對於 linear operator, 我們也有以下的定義.

**Definition 7.3.6.** 假設 $V$ 是一個 vector space over $\mathbb{F}$ 且 $T: V \to V$ 是一個 linear operator. 若 $V$ 中存在一組 basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 $T$ 的 eigenvectors, 則稱 $T$ 為 *diagonalizable* (可對角化).

為何這樣子的 linear operator 會稱為 diagonalizable 呢? 其原因比矩陣的情況更容易讓人理解. 事實上如果 $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ 是 $V$ 的一組 ordered basis 且 $\mathbf{v}_i$ 皆為 $T$ 的 eigenvector. 假設 $\lambda_i$ 就是 $\mathbf{v}_i$ 所對應的 eigenvalue, 亦即 $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$, $\forall i \in \{1, \ldots, n\}$. 此時考慮 $T$ 利用 $\beta$ 所得的 matrix representation $[T]_\beta$. 回顧一下 $[T]_\beta$ 的 1-st column 是 $T(\mathbf{v}_1) = \lambda_1$ 用 $\beta$ 寫下的坐標表示, 即 $\begin{bmatrix} \lambda_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \lambda_1 \mathbf{e}_1$. 而對一般的 $i \in \{1, \ldots, n\}$, $[T]_\beta$ 的 $i$-th column 就是 $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_1$ 用 $\beta$ 寫下的坐標表示, 即 $\lambda_i \mathbf{e}_i$. 也因此 $[T]_\beta$ 就是 $\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}$ 這樣的 diagonal matrix.

要怎樣找一個 linear operator $T: V \to V$ 的 eigenvalue 和 eigenvector 呢? 從前面一開始的說明可以知道, 任取 $V$ 的一組 ordered basis $\beta$, 只要考慮其 matrix representation $[T]_\beta$ 的 eigenvalue 和 eigenvector 就可以還原成 $T$ 的 eigenvalue 和 eigenvector 了, 我們看以下的例子.

**Example 7.3.7.** 考慮 Example 7.3.4 中的 linear operator $T: P_2(\mathbb{R}) \to P_2(\mathbb{R})$, 以及 $P_2(\mathbb{R})$ 的 standard basis $\varepsilon = (1, x, x^2)$. 由於依定義 $T(1) = 1$, $T(x) = 2x + 1$, $T(x^2) = 3x^2 + 2x$, 我們得 $[T]_\varepsilon = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$. 因為 $[T]_\varepsilon$ 是上三角矩陣, 很容易求得其 characteristic polynomial 為 $(1-t)(2-t)(3-t)$. 得知 $[T]_\varepsilon$ 的 eigenvalue 為 $1, 2, 3$ (事實上這也是 $T$ 的 eigenvalue). 接下來我們利用解 $[T]_\varepsilon$ 的 eigenspace 得 $[T]_\varepsilon$ 的 eigenvectors. 對於 eigenvalue 1 所得的 eigenspace 就是 $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix}$ 的 null space, 即 $\mathrm{Span}(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix})$. 然而 $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ 是 1 在 $P_2(\mathbb{R})$ 利用 $\varepsilon$ 所

得的坐標表示. 故知 Span(1) 中的非 0 元素是 $T$ 的 eigenvector 且其 eigenvalue 為 1. 事實上我們有 $T(1)=1$, 對於 $[T]_{\varepsilon}$ 的 eigenvalue 2 所得的 eigenspace 就是 $\begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix}$ 的 null space, 即 Span($\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$). 然而 $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ 是 $x+1$ 在 $P_2(\mathbb{R})$ 利用 $\varepsilon$ 所得的坐標表示. 故知 Span$(x+1)$ 中的非 0 元素是 $T$ 的 eigenvector 且其 eigenvalue 為 2. 事實上我們有 $T(x+1)=2(x+1)$, 對於 $[T]_{\varepsilon}$ 的 eigenvalue 3 所得的 eigenspace 就是 $\begin{bmatrix} -2 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ 的 null space, 即 Span($\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$). 然而 $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ 是 $x^2+2x+1$ 在 $P_2(\mathbb{R})$ 利用 $\varepsilon$ 所得的坐標表示. 故知 Span$(x^2+2x+1)$ 中的非 0 元素是 $T$ 的 eigenvector 且其 eigenvalue 為 3. 事實上在 Example 7.3.4 中我們算過 $T(x^2+2x+1)=3(x^2+2x+1)$,

因為 $\{1,x+1,x^2+2x+1\}$ 是 $T$ 的 eigenvectors 且是 $P_2(\mathbb{R})$ 的一組 basis, 所以我們知 $T$ 是 diagonalizable. 事實上若考慮 ordered basis $\beta=(1,x+1,x^2+2x+1)$, 則 $[T]_{\beta}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

最後我們要強調, 在求 linear operator $T:V\to V$ 的 eigenvalue 和 eigenvector 時, 不必擔心選取 $V$ 的 ordered basis 為何. 這是因為 eigenvalue 和 eigenvector 的定義和 $T$ 有關, 而和 $V$ 的 ordered basis 無關. 所以即使選取 $V$ 的 ordered basis 不同會造成不同的矩陣表示, 所得的 eigenvalue 和 eigenvector 都可得到同樣 $T$ 的 eigenvalue 和 eigenvector. 事實上我們知道, 當 $V$ 選取不同的 ordered basis $\beta,\beta'$, 雖然 $[T]_{\beta}$ 和 $[T]_{\beta'}$ 會不同, 但它們會是 similar, 所以它們會有同樣的 characteristic polynomial (Proposition 7.2.10), 因此有同樣的 eigenvalues. 最後要提醒的是, 在選取 $V$ 的 ordered basis 使用表現矩陣來求 eigenvalue 和 eigenvector 時, 定義域和對應域都要使用同樣的 ordered basis, 否則這樣的表現矩陣所求得的 eigenvalue 和 eigenvector 和 $T$ 的 eigenvalue 和 eigenvector 的定義是不吻合的.

基於上面的探討, 我們可以定義一個 linear operator $T:V\to V$ 的 characteristic polynomial $P_T(t)$. 其定義的方法就是任取一個 $V$ 的 ordered basis $\beta$, 若 $A=[T]_{\beta}^{\beta}$, 則定義 $P_T(t)=P_A(t)$. 注意這樣定義出來的 characteristic polynomial 和 $\beta$ 的選取無關. 主要的原因是若取 $V$ 的另一組 ordered basis, 其表現矩陣會和 $A$ 是 similar. 所以利用 Proposition 7.2.10 知, similar matrix 的 characteristic polynomial 是一樣的, 所以 $P_T(t)$ 不會因選取的 ordered basis 不同而有所不同. 注意, 前面我們提過, $T$ 的 eigenvalue 就是其表現矩陣 $A$ 的 characteristic polynomial 的根, 所以依此定義我們也可以說 $T$ 的 eigenvalue 就是 $T$ 的 characteristic polynomial 的根. 這裡唯一要注意的是 $T$ 的 eigenvector 並不是 $A$ 的 eigenvector. 事實上 $T$ 的 eigenvector 並需用 ordered basis $\beta$ 寫成 $\mathbb{F}^n$ 上的坐標表示法後, 才會是 $A$ 的 eigenvector.

## 7.4. Cayley-Hamilton Theorm

在這節中我們將介紹 Cayley-Hamilton Theorem. 首先我們先介紹 linear operator 的 invariant subspace, 再利用 invariant subspace 的概念證明 linear operator 的 Cayley-Hamilton Theorem, 再因此推得矩陣的 Cayley-Hamilton Theorem.

在探討函數的理論時, 通常當定義域很大時, 我們可以透過所謂的 restriction 將函數限制在較小的範圍來了解該函數. 給定一個函數 $f: X \to Y$, 以及 $X$ 中的子集合 $S$, 所謂 $f$ 的 restriction on $S$, 用 $f|_S$ 表示, 就是將 $f$ 的定義域縮小到 $S$, 其他對於 $f$ 的映射方式都沒有改變. 也就是說 $f|_S$ 是一個定義域為 $S$ 的函數 $f|_S: S \to Y$, 且對於任意 $s \in S$, $f|_S(s) = f(s)$, 不過若 $x \in X$ 但 $x \notin S$, 則 $f|_S(x)$ 是無定義的. 現若 $T: V \to V$ 是 linear operator, $W$ 為 $V$ 的 subspace, 則 $T|_W$ 依然會是一個 linear transformation (只是定義域在 $W$ 上). 不過 $T|_W$ 未必會是一個 linear operator, 因為 $T$ 未必會將 $W$ 中的元素映射到 $W$. 如此一來, 我們就不能將過去探討 linear operator 的理論運用在 $T|_W$ 上了. 為了達到 $T|_W$ 仍為 linear operator 的目的, 我們必須選有特殊性質的 $W$ (即 $T$ 會將 $W$ 的元素映射到 $W$), 這樣就能套用 linear operator 的理論了. 因此我們有以下的定義.

**Definition 7.4.1.** 假設 $V$ 是一個 vector space over $\mathbb{F}$ 且 $T: V \to V$ 是 linear operator. 若 $W$ 是 $V$ 的 subspace 且滿足 $T(W) \subseteq W$ (即 $T(\mathbf{w}) \in W, \forall \mathbf{w} \in W$), 則稱 $W$ 為一個 $T$-invariant subspace.

要注意當 $T: V \to V$ 是 linear operator„ Definition 7.4.1, 告訴我們 $W$ 是 $T$-invariant, 表示 $T(W) \subseteq W$, 並不是說 $T(W) = W$, 也不是說 $T(\mathbf{w}) = \mathbf{w}$, $\forall \mathbf{w} \in W$. 請大家不要誤解. 也就是說要檢查 $V$ 的 subspace $W$ 是否為 $T$-invariant subspace, 我們僅要檢查是否所有 $W$ 的元素 $\mathbf{w}$ 經由 $T$ 的映射 (即 $T(\mathbf{w})$) 依然在 $W$ 中. 當然了, 因 $T$ 為 linear operator, 對任意 $\mathbf{v} \in V$, 皆有 $T(\mathbf{v}) \in V$, 故 $V$ 本身是 $T$-invariant. 還有因為 $T$ 是 linear tansformation, 我們知道 $T(\mathbf{0}) = \mathbf{0}$, 所以 zero space $\{\mathbf{0}\}$ 也是 $T$-invariant. 另外若 $\lambda \in \mathbb{F}$ 是 $T$ 的 eigenvector, 則 $\lambda$ 所對應的 eigenspace $E_T(\lambda) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda\mathbf{v}\}$ 也會是 $T$-invariant subspace. 這是因為若 $\mathbf{v} \in E_T(\lambda)$, 則 $T(\mathbf{v}) = \lambda\mathbf{v}$. 由於 $E_T(\lambda)$ 是 $V$ 的 subspace 且 $\mathbf{v} \in E_T(\lambda)$, 自然有 $\lambda\mathbf{v} \in E_T(\lambda)$, 亦即 $T(\mathbf{v}) \in E_T(\lambda)$. 故 $E_T(\lambda)$ 亦為 $T$-invariant. 另外我們過去熟悉的 $T$ 的 range $\mathsf{R}(T)$, 也是 $T$-invariant, 這是因為對任意 $\mathbf{v} \in V$, 自然有 $T(\mathbf{v}) \in T(V) = \mathsf{R}(T)$. 當然當 $\mathbf{v} \in \mathsf{R}(T)$, 由於 $T: V \to V$ 是 linear operator, 故 $\mathsf{R}(T) \subseteq V$, 因此我們依然有 $T(\mathbf{v}) \in \mathsf{R}(T)$. $T$ 的 null space $\mathsf{N}(T)$ 也是 $T$-invariant. 這是因為對任意 $\mathbf{v} \in \mathsf{N}(T)$, 由於 $T(\mathbf{v}) = \mathbf{0}$ 且 $\mathbf{0} \in \mathsf{N}(T)$ (別忘了 $T(\mathbf{0}) = \mathbf{0}$), 故 $T(\mathbf{v}) \in \mathsf{N}(T)$.

除了前面舉的幾個例子外, 還有哪些 $T$-invariant subspace 呢? 前面提過考慮 $T$-invariant subspace 就是想將 $T$ 的定義域縮小. 所以給定 $\mathbf{v} \in V$, 我們很想知道甚麼是包含 $\mathbf{v}$ 最小的 $T$-invariant subspace. 假設 $W$ 是包含 $\mathbf{v}$ 的 $T$-invariant subspace. 當然了, 我們有 $\mathbf{v} \in W$. 不過由 $W$ 是 $T$-invariant, 我們自然要有 $T(\mathbf{v}) \in W$ (因 $\mathbf{v} \in W$). 再由 $T(\mathbf{v}) \in W$ 以及 $W$ 是 $T$-invariant, 我們有 $T(T(\mathbf{v})) = T^2(\mathbf{v}) \in W$. 如此一直下去我們知 $T^m(\mathbf{v}) \in W, \forall m \in \mathbb{N}$. 由此我們知, 若 $W$ 是包含 $\mathbf{v}$ 的 $T$-invariant subspace, 則 $W$ 必須包含 $\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots, T^m(\mathbf{v}), \ldots\}$

這個集合 (即 $\{T^i(\mathbf{v}) \mid i \in \mathbb{N}\}$) 中所有的元素. 不過 $W$ 是 subspace, 所以也必須包含所有這些元素所 span 的 subspace, 所以我們有以下的定義.

**Definition 7.4.2.** 假設 $T: V \to V$ 是 linear operator. 對任意 $\mathbf{v} \in V$, 令

$$C(T,\mathbf{v}) = \mathrm{Span}\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots, T^m(\mathbf{v}), \ldots\}.$$

我們稱 $C(T,\mathbf{v})$ 為 the $T$-cyclic space generated by $\mathbf{v}$.

要注意若令 $S = \{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \ldots, T^m(\mathbf{v}), \ldots\}$, 雖然 $S$ 中可能有無窮多個元素, 不過依 span 的定義, $C(T,\mathbf{v}) = \mathrm{Span}(S)$ 中的元素是 $S$ 中有限多個元素的線性組合. 因此若 $\mathbf{w} \in C(T,\mathbf{v}) = \mathrm{Span}(S)$, 表示存在 $c_0, c_1, \ldots, c_m \in \mathbb{F}$ 使得 $\mathbf{w} = c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_mT^m(\mathbf{v})$ (其中可能有些 $c_i = 0$). 故得 $T(\mathbf{w}) = c_0T(\mathbf{v}) + c_1T^2(\mathbf{v}) + \cdots + c_mT^{m+1}(\mathbf{v}) \in \mathrm{Span}(S) = C(T,\mathbf{v})$. 也因此得證 $c(T,\mathbf{v})$ 是 $T$-invariant subspace. 前面提過包含 $\mathbf{v}$ 的 $T$-invariant subspace 必包含 $S$, 故我們得到下面的定理.

**Proposition 7.4.3.** 假設 $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 則 $C(T,\mathbf{v})$ 是包含 $\mathbf{v}$ 最小的 $T$*-invariant subspace.*

**Question 7.7.** 假設 $V$ 是 *vector space over* $\mathbb{F}$, $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 證明對任意 $\mathbf{w} \in C(T,\mathbf{v})$, 皆存在係數在 $\mathbb{F}$ 的多項式 $f(x)$ 使得 $\mathbf{w} = f(T)(\mathbf{v})$. 依此得 $C(T,\mathbf{v}) = \{f(T)(\mathbf{v}) \mid f(x) \in \mathbb{F}[x]\}$.

當 $V$ 是 finite dimensional vector space over $\mathbb{F}$, $C(T,\mathbf{v})$ 為其 subspace, 故 $C(T,\mathbf{v})$ 也是 finite dimensional. 如何知道 $C(T,\mathbf{v})$ 的維度呢? 當然了, 若 $\mathbf{v} = \mathbf{0}$, 則 $T^i(\mathbf{v}) = T^i(\mathbf{0}) = \mathbf{0}$, $\forall i \in \mathbb{N}$, 故此時 $C(T,\mathbf{v}) = \{\mathbf{0}\}$, 即 $\dim(C(T,\mathbf{v})) = 0$. 因此我們僅考慮 $\mathbf{v} \neq \mathbf{0}$ 的情況. 首先我們考慮 $\mathbf{v}, T(\mathbf{v})$ 是否為 linearly independent. 若 $\mathbf{v}, T(\mathbf{v})$ 不是 independent, 由於 $\mathbf{v} \neq \mathbf{0}$, 故知存在 $c \in \mathbb{F}$ 使得 $T(\mathbf{v}) = c\mathbf{v}$. 由此知 $T^i(\mathbf{v}) = c^i\mathbf{v} \in \mathrm{Span}(\mathbf{v})$, $\forall i \in \mathbb{N}$. 因此得 $C(T,\mathbf{v}) = \mathrm{Span}(\mathbf{v})$, 即 $\dim(C(T,\mathbf{v})) = 1$. 而若 $\mathbf{v}, T(\mathbf{v})$ 為 independent, 則我們考慮 $\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v})$ 是否為 independent. 若它們不是 independent, 則由 $\mathbf{v}, T(\mathbf{v})$ 為 independent 知 $T^2(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$ (Lemma 3.5.4). 因此存在 $c, d \in \mathbb{F}$ 使得 $T^2(\mathbf{v}) = c\mathbf{v} + dT(\mathbf{v})$. 此時

$$T^3(\mathbf{v}) = T(T^2(\mathbf{v})) = cT(\mathbf{v}) + dT^2(\mathbf{v}) = cT(\mathbf{v}) + d(c\mathbf{v} + dT(\mathbf{v})) = dc\mathbf{v} + (c + d^2)T(\mathbf{v}).$$

因此得 $T^3(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$. 再利用數學歸納法, 我們可以證明 $T^i(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$, $\forall i \in \mathbb{N}$, 因此知此時 $C(T,\mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}))$, 即 $\dim(C(T,\mathbf{v})) = 2$. 我們可以一直這樣探討下去得到以下的定理.

**Proposition 7.4.4.** 假設 $V$ 是 *vector space over* $\mathbb{F}$, $T: V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 則 $\dim(C(T,\mathbf{v})) = m$ 若且唯若 $m$ 是最大的 $i$ 使得 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{i-1}(\mathbf{v})$ 為 *linear independent:* 也就是說 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 *linear independent* 但 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}), T^m(\mathbf{v})$ 不是 *linearly independent.*

事實上若 $\dim(C(T,\mathbf{v})) = m$, 則 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 $C(T,\mathbf{v})$ 的一組 *basis.*

**Proof.** 首先我們用數學歸納法證明若 $T^m(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 則

$$T^i(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})), \forall i \in \mathbb{N}.$$

由於已知 $i \le m$ 成立, 我們直接歸納假設 $T^k(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$ 成立. 現考慮 $T^{k+1}(\mathbf{v})$. 由於 $T^k(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 存在 $c_0, c_1, \ldots, c_{m-1} \in \mathbb{F}$ 使得 $T^k(\mathbf{v}) = c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_{m-1}T^{m-1}(\mathbf{v})$. 故

$$T^{k+1}(\mathbf{v}) = T(T^k(\mathbf{v})) = c_0T(\mathbf{v}) + c_1T^2(\mathbf{v}) + \cdots + c_{m-2}T^{m-1}(\mathbf{v}) + c_{m-1}T^m(\mathbf{v}).$$

由於 $T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 和 $T^m(\mathbf{v})$ 皆屬於 $\mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 得證

$$T^{k+1}(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})).$$

也因此證明了 $T^i(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})), \forall i \in \mathbb{N}$. 因而得知

$$C(T, \mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})).$$

現假設 $\dim(C(T, \mathbf{v})) = m$, 我們先說明 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 linearly independent. 若它們不是 independent, 表示存在 $k \le m-1$ 使得 $T^k(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{k-1}(\mathbf{v}))$, 依前面討論, 此表示 $C(T, \mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{k-1}(\mathbf{v}))$, 亦即 $C(T, \mathbf{v})$ 是由 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{k-1}(\mathbf{v})$ 這 $k$ 個向量所展成. 然而 $k \le m-1$, 此與 $\dim(C(T, \mathbf{v})) = m$ 的假設相矛盾, 故知 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 linearly independent. 既然 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 independent 又可展成 $C(T, \mathbf{v})$, 故知它們形成 $C(T, \mathbf{v})$ 的一組 basis. 然而 $T^m(\mathbf{v}) \in C(T, \mathbf{v})$, 故知 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}), T^m(\mathbf{v})$ 不是 linearly independent.

反之, 假設 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 是 linearly independent, 但 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}), T^m(\mathbf{v})$ 不是 linearly independent. 由 Lemma 3.5.4, 我們知 $T^m(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$. 因此再由前面討論得 $C(T, \mathbf{v}) = \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$. 亦即 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 不只是 independent 且可展成 $C(T, \mathbf{v})$. 故 $\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v})$ 形成 $C(T, \mathbf{v})$ 的一組 basis, 得證 $\dim(C(T, \mathbf{v})) = m$. □

**Question 7.8.** 證明 $\dim(C(T, \mathbf{v})) = 1$ 若且唯若 $\mathbf{v}$ 是 $T$ 的 *eigenvector.*

既然 $C(T, \mathbf{v})$ 是 $T$-invariant, 我們知 $T|_{C(T,\mathbf{v})} : C(T, \mathbf{v}) \to C(T, \mathbf{v})$ 是 linear operator. 那麼 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 會是甚麼呢? 要求 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial, 我們要先找到 $T|_{C(T,\mathbf{v})}$ 的定義域 $C(T, \mathbf{v})$ 的一組 ordered basis, 在利用這組 ordered basis 得到 $T|_{C(T,\mathbf{v})}$ 的表現矩陣, 再求該矩陣的 characteristic polynomial. 根據 Proposition 7.4.4, 若 $\dim(C(T, \mathbf{v})) = m$, 我們很自然的選 $(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$ 這一組 $C(T, \mathbf{v})$ 的 ordered basis.

接著我們來看 $T|_{C(T,\mathbf{v})}$ 用 $\beta = (\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$ 這一組 ordered basis 其表現矩陣 $[T|_{C(T,\mathbf{v})}]_\beta$ 為何? 首先 $[T|_{C(T,\mathbf{v})}]_\beta$ 的 1-st column 是 $\beta$ 的第一個向量 (即 $\mathbf{v}$) 經由 $T$ 映射後所得的向量 (即 $T(\mathbf{v})$) 用 ordered basis $\beta$ 所得的坐標表示. 由於 $T(\mathbf{v})$ 恰好是 $\beta$ 的第二個向量, 故其坐標表示為 $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}$. 同理得 $[T|_{C(T,\mathbf{v})}]_\beta$ 的 $i$-th column 為 $\mathbf{e}_{i+1}$, 其中

$1 \leq i \leq m-1$. 至於 $[T|_{C(T,\mathbf{v})}]_\beta$ 的最後一個 column, 應該是 $\beta$ 的最後一個向量 (即 $T^{m-1}(\mathbf{v})$) 經由 $T$ 映射後所得的向量 (即 $T(T^{m-1}(\mathbf{v})) = T^m(\mathbf{v})$) 用 ordered basis $\beta$ 所得的坐標表示. 然而 $T^m(\mathbf{v}) \in \mathrm{Span}(\mathbf{v}, T(\mathbf{v}), \ldots, T^{m-1}(\mathbf{v}))$, 若假設 $T^m(\mathbf{v}) = c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_{m-1}T^{m-1}(\mathbf{v})$, 其中 $c_0, c_1, \ldots, c_{m-1} \in \mathbb{F}$, 則 $[T|_{C(T,\mathbf{v})}]_\beta$ 的最後一個 column 就是 $\begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{m-1} \end{bmatrix}$. 因此得

$$[T|_{C(T,\mathbf{v})}]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & \cdots & 0 & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1} \end{bmatrix}.$$

如何求這樣的矩陣的 characteristic polynomial 呢? 我們首先考慮 $2 \times 2$ 矩陣的情形. 若 $A_2 = \begin{bmatrix} 0 & c_0 \\ 1 & c_1 \end{bmatrix}$, 直接計算可得 $A_2$ 的 characteristic polynomial 為

$$\det(A_2 - tI_2) = \det \begin{bmatrix} -t & c_0 \\ 1 & c_1 - t \end{bmatrix} = t^2 - c_1t - c_0.$$

而若 $A_3 = \begin{bmatrix} 0 & 0 & c_0 \\ 1 & 0 & c_1 \\ 0 & 1 & c_2 \end{bmatrix}$, 考慮矩陣 $A_3 - tI_3 = \begin{bmatrix} -t & 0 & c_0 \\ 1 & -t & c_1 \\ 0 & 1 & c_2 - t \end{bmatrix}$ 的 determinant, 由於我們可以用數學歸納法處理, 所以不直接計算而是採用降階的方式處理. 對 $A_3 - tI_3$ 的 1-st row 降階求 determinant 得

$$\det(A_3 - tI_3) = (-t)\det \begin{bmatrix} -t & c_1 \\ 1 & c_2 - t \end{bmatrix} + c_0 \det \begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix}.$$

其中第一個矩陣是前面 $2 \times 2$ 的情況可得其 determinant 為 $t^2 - c_2t - c_1$, 而第二個矩陣式上三角矩陣故其 determinant 為 1. 因此可得

$$\det(A_3 - tI_3) = (-t)(t^2 - c_2t - c_1) + c_1 = -(t^3 - c_2t^2 - c_1t - c_0).$$

利用數學歸納法我們可以得到以下之結果.

**Proposition 7.4.5.** 假設 $V$ 是 *vector space over* $\mathbb{F}$, $T : V \to V$ 是 *linear operator* 且 $\mathbf{v} \in V$. 若 $\dim(C(T,\mathbf{v})) = m$ 且 $T^m(\mathbf{v}) = c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_{m-1}T^{m-1}(\mathbf{v})$, 則 $T|_{C(T,\mathbf{v})}$ 的 *characteristic polynomial* 為

$$(-1)^m(t^m - c_{m-1}t^{m-1} - \cdots - c_1t - c_0).$$

**Proof.** 我們繼續剛才的討論, 利用數學歸納法求 $m \times m$ 矩陣

$$[T|_{C(T,\mathbf{v})}]_\beta = \begin{bmatrix} 0 & 0 & \cdots & 0 & c_0 \\ 1 & 0 & \cdots & 0 & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1} \end{bmatrix}.$$

的 characteristic polynomial. 對 1-st row 展開得

$$\det\begin{bmatrix} -t & 0 & \cdots & 0 & c_0 \\ 1 & -t & \cdots & 0 & c_1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -t & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1}-t \end{bmatrix} =$$

$$(-t)\det\begin{bmatrix} -t & 0 & \cdots & 0 & c_1 \\ 1 & -t & \cdots & 0 & c_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -t & c_{m-2} \\ 0 & 0 & \cdots & 1 & c_{m-1}-t \end{bmatrix} + (-1)^{m+1}c_0\det\begin{bmatrix} 1 & -t & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -t \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

注意等式右邊的矩陣都是降階後的 $(m-1)\times(m-1)$ matrix. 其中第一個是歸納假設成立的 $(m-1)\times(m-1)$ matrix, 所以其 determinant 為 $(-1)^{m-1}(t^{m-1}-c_{m-1}t^{m-2}-\cdots-c_2t-c_1)$. 而第二個矩陣式對角線為 1 的 upper triangular matrix 故其 determinant 為 1. 因此得證其 characteristic polynomial 為 $(-1)^m(t^m-c_{m-1}t^{m-1}-\cdots-c_2t^2-c_1t-c_0)$. □

利用 Proposition 7.4.5, 我們馬上可得以下結論.

**Corollary 7.4.6.** 假設 $T:V\to V$ 為 *linear operator* 且 $\mathbf{v}\in V$ 為非零向量. 令 $g(x)$ 為 $T|_{C(T,\mathbf{v})}:C(T,\mathbf{v})\to C(T,\mathbf{v})$ 的 *characteristic polynomial.* 則 $g(T)(\mathbf{v})=\mathbf{0}$.

**Proof.** 假設 $\dim(C(T,\mathbf{v}))=m$ 且 $T^m(\mathbf{v})=c_0\mathbf{v}+c_1T(\mathbf{v})+\cdots+c_{m-1}T^{m-1}(\mathbf{v})$. Proposition 7.4.5 告訴我們 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 為 $g(x)=(-1)^m(x^m-c_{m-1}x^{m-1}-\cdots-c_1x-c_0)$. 因此

$$\begin{aligned} g(T)(\mathbf{v}) &= (-1)^m(T^m-c_{m-1}T^{m-1}-\cdots-c_1T-c_0\mathrm{id}_{C(T,\mathbf{v})})(\mathbf{v}) \\ &= (-1)^m(T^m(\mathbf{v})-c_{m-1}T^{m-1}(\mathbf{v})-\cdots-c_1T(\mathbf{v})-c_0\mathbf{v})=\mathbf{0}. \end{aligned}$$

□

**Question 7.9.** 假設 $T:V\to V$ 為 *linear operator* 且 $\mathbf{v}\in V$ 為非零向量. 令 $g(x)$ 為 $T|_{C(T,\mathbf{v})}$ 的 *characteristic polynomial.* 證明 $g(T)|_{C(T,\mathbf{v})}=\mathbf{0}$.

Linear operator 限制在較小的 $T$-invariant subspace 基本上和原來的 operator 是相同的, 因此它們的 characteristic polynomial 之間應該有關係. 接下來, 我們便是要探討它們之間的關係. 假設 $T:V\to V$ 是 linear operator, 且 $W$ 為 $T$-invariant subspace. 要討論 $T$ 和 $T|_W$ 的 characteristic polynomial, 我們需找 $W$ 和 $V$ 的 ordered basis, 然後得到相對應的表現矩陣, 再得到它們的 characteristic polynomial. 因為 $W$ 是 $V$ 的 subspace, 我們又期待它們之間的 characteristic polynomial 相關. 自然的, 我們可以先找 $W$ 的一組 ordered basis $\beta=(\mathbf{w}_1,\ldots,\mathbf{w}_k)$ 再將 $\beta$ 擴大成 $V$ 的一組 ordered basis $\gamma=(\mathbf{w}_1,\ldots,\mathbf{w}_k,\mathbf{w}_{k+1},\ldots,\mathbf{w}_n)$. 現假設 $A=[T|_W]_\beta$, 注意 $A$ 的 $i$-th column 就是 $T(\mathbf{w}_i)$ 用 $\beta$ 所得的坐標表示 (即 $\mathbb{F}^k$ 中的向量). 現考慮 $T$ 用 ordered basis $\gamma$ 所得的矩陣表示 $[T]_\gamma$. 要注意當 $1\le i\le k$ 時, $[T]_\gamma$ 的 $i$-th column 和 $A$ 的 $i$-th column 一樣是 $T(\mathbf{w}_i)$, 不同的是它應該是 $T(\mathbf{v}_i)$ 用 $\gamma$ 的坐標表示 (即 $\mathbb{F}^n$

中的向量). 由於 $T(\mathbf{w}_i)\in W$, 其用 $\gamma$ 寫下的線性組合, 僅需要用到前面 $k$ 個向量, 即 $\beta$ 的線性組合, 所以 $T(\mathbf{w}_i)$ 用 $\gamma$ 的坐標表示基本上和用 $\beta$ 坐標表示相同, 只是後面 $k+1,\dots,m$ 這 $m-k$ 個 entry 須補上 0. 至於 $[T]_\gamma$ 在 $k_1$-th column 之後的 columns 我們就不知道會是怎樣, 不過這並不會影響我們要探討的問題. 總而言之, 若 $[T|_W]\beta=A$, 我們可以將 $[T]_\gamma$ 寫成 $\left[\begin{array}{c|c} A & M_1 \\ \hline \mathbf{0} & M_2\end{array}\right]$ 這樣的形式. 注意這裡 $A$ 是 $k\times k$ matrix, $\mathbf{0}$ 是 $(n-k)\times k$ matrix, 而 $M_1,M_2$ 分別為 $k\times(n-k)$ 和 $(n-k)\times(n-k)$ matrix. 因此 $[T]_\gamma$ 的 characteristic polynomial 應為

$$\det\left[\begin{array}{c|c} A-tI_k & M_1 \\ \hline \mathbf{0} & M_2-tI_{n-k}\end{array}\right].$$

回顧前面提過在計算形如 $\left[\begin{array}{c|c} A & B \\ \hline \mathbf{0} & C\end{array}\right]$ 這樣的矩陣的 determinant 時, 當 $A$, $C$ 分別為 $k\times k$ 和 $(n-k)\times(n-k)$ matrix 時我們可以先將前 $k$ 個 row 用 elementary row operations 將 $A$ 化成 echelon form, 再將後面 $n-k$ 個 row 用 elementary row operations 將 $C$ 化為 echelon form. 因為這樣最後是一個 upper triangular matrix, 其 determinant 就是對角線相乘, 因此知 $\det\left[\begin{array}{c|c} A & B \\ \hline \mathbf{0} & C\end{array}\right]=(\det A)(\det C)$. 因此前面算 $T$ 的 characteristic polynomial 便會是 $\det(A-tI_k)\det(M_2-tI_{n-k})$. 因為 $\det(A-tI_k)$ 就是 $T|_W$ 的 characteristic polynomial, 因此我們有以下的結果.

**Proposition 7.4.7.** 假設 $V$ 為 *vector space over* $\mathbb{F}$, $T:V\to V$ 為 *linear operator* 且 $W$ 為 $T$*-invariant subspace,* 則 $T|_W$ 的 *characteristic polynomial* 會是 $T$ 的 *characteristic polynomial* 的因式. 也就是說, 若 $f(x)$ 是 $T$ 的 *characteristic polynomial* 且 $g(t)$ 是 $T|_W$ 的 *characteristic polynomial,* 則存在係數在 $\mathbb{F}$ 的多項式 $h(x)$ 滿足 $f(x)=h(x)g(x)$.

特別地, 當 $\mathbf{v}\in V$ 我們可以考慮 Proposition 7.4.7 中 $W=C(T,\mathbf{v})$ 的情形, 也就是說若 $g(x)$ 是 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 而 $p_T(x)$ 是 $T$ 的 characteristic polynomial, 則存在多項式 $h(x)$ 滿足 $p_T(x)=h(x)g(x)$. 利用這個多項式的等式, 我們會想要將 $T$ 代入兩邊的多項式, 不過我們必須先釐清 $h(T)g(T)$ 是甚麼意思. 當然了, 依定義 $h(T)$, $g(T)$ 都是定義域為 $V$ 的 linear operator. 我們要知道的是, 如果兩個多項式 $h(x)$, $g(x)$ 相乘會是 $f(x)$, 那麼 $f(T),h(T),g(T)$ 這三個 linear operator 會有甚麼關係. 首先在定義將 linear operator $T$ 代入一個多項式時, 我們用到了將 $T$ 代入 $x^i$ 會得到 $T^i$. 這裡 $x^i$ 是多項式的乘法, 而 $T^i$ 是 linear operator 的合成. 簡單來說就是多項式相乘代入 linear operator 後會得到的是 linear operator 的合成. 我們剛剛是觀察單項式, 不過多項式只是單項式相加, 而多項式的加法和乘法有所謂的分配律 (即 $f_1(x)(f_2(x)+f_3(x))=f_1(x)f_2(x)+f_1(x)f_3(x)$). linear operators 的合成也有所謂的分配律 (即若 $T_1,T_2,T_3$ 皆為定義域為 $V$ 的 linear operators. 則 $T_1\circ(T_2+T_3)=T_1\circ T_2+T_1\circ T_3$.) 也就是說多項式的乘法和 linear operator 的合成是符合同樣的運算規則的. 例如 $h(x)=x^2-2x$, $g(x)=x+1$ 且 $f(x)=h(x)g(x)=x^3-x^2-2x$. 我們檢查 $h(T)$ 和 $g(T)$ 的合成. 依定義

$$h(T)\circ g(T)=(T^2-2T)\circ(T+\mathrm{id}_V)=T^2\circ(T+\mathrm{id}_V)-2T\circ(T+\mathrm{id}_V).$$

上式整理後可得 $T^3+T^2-2T^2-2T=T^3-T^2-2T=f(T)$.

由前面的討論我們知若 $g(x)$ 是 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial 而 $p_T(x)$ 是 $T$ 的 characteristic polynomial, 則存在多項式 $h(x)$ 滿足 $p_T(x) = h(x)g(x)$. 將 $T$ 代入 $p_T(x) = h(x)g(x)$ 這個等式可得 $p_T(T) = h(T) \circ g(T)$. 有了這個 linear operator 的關係式, 我們就可以證明 Cayley-Hamilton Theorem.

**Theorem 7.4.8** (Cayley-Hamilton Theorem)**.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $T : V \to V$ 為 *linear operator.* 考慮 $T$ 的 *characteristic polynomial* $p_T(x)$, 我們有 $p_T(T) : V \to V$ 為 *zero operator,* 亦即對任意 $\mathbf{v} \in V$, $p_T(T)(\mathbf{v}) = \mathbf{0}$.

**Proof.** 由於當 $\mathbf{v} = \mathbf{0}$, 時自然有 $p_T(T)(\mathbf{v}) = \mathbf{0}$ (因 $p_T(T)$ 是 linear transformation), 故僅考慮 $\mathbf{v} \in V$ 且 $\mathbf{v} \neq O$. 此時考慮 $\mathbf{v}$ 所產生的 $T$-cyclic space $C(T,\mathbf{v})$, 並令 $g(x)$ 為 $T|_{C(T,\mathbf{v})}$ 的 characteristic polynomial. 由 Proposition 7.4.7 知存在 $h(x)$ 係數在 $\mathbb{F}$ 的 polynomial 使得 $p_T(x) = h(x)g(x)$. 因此知 $p_T(T) = h(T) \circ g(T)$. 此時 $p_T(T)(\mathbf{v}) = (h(T) \circ g(T))(\mathbf{v}) = h(T)(g(T)(\mathbf{v}))$. 然而由 Corollary 7.4.6 知 $g(T)(\mathbf{v}) = \mathbf{0}$, 故得證 $p_T(T)(\mathbf{v}) = h(T)(g(T)(\mathbf{v})) = h(T)(\mathbf{0}) = \mathbf{0}$. □

假設 $V$ 是 over $\mathbb{F}$ 的 vector space 且 $\dim(V) = n$, 我們知道定義在 $V$ 上的 linear operators 的性質和 $M_{n\times n}(\mathbb{F})$ 上的矩陣的性質有相對照的關係. 所以 Cayley-Hamilton Theorem 對矩陣應該也是對的. 亦即若 $p_A(x)$ 是 $A$ 的 characteristic polynomial, 則 $p_A(A)$ 會是零矩陣 $\mathbf{0}$. 或許大家會覺得這很好證明, 因為依定義 $p_A(x) = \det(A - xI_n)$, 所以 $p_A(A) = \det(A - AI_n) = \det(\mathbf{0}) = 0$. 這樣的證明是不對的. 依定義 $p_A(A)$ 是將 $A$ 代入 $p_A(x)$ 這個多項式所得的 $n \times n$ 矩陣, 因此定理是說 $p_A(A)$ 是 $n \times n$ 的零矩陣. 而這個證明裡所得的 $\det(A - A) = 0$ 是 $\mathbb{F}$ 的零元素. 兩個談的是不同的事 (僅在 $n = 1$ 時相同). 這個證明的問題是當我們考慮 $\det(A - xI_n)$ 時, 是將 $x$ 看待成未知的 $\mathbb{F}$ 中的元素, 因此 $\det(A - xI_n)$ 可寫成 $\det \begin{bmatrix} a_{11} - x & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} - x \end{bmatrix}$, 再將之展開得 $x$ 的多項式. 因此在這一步驟當 $x$ 代任何 $\mathbb{F}$ 中的元素進入上述的矩陣是行得通的, 但是 $x$ 不能帶入矩陣 $A$. 因為 $\det \begin{bmatrix} a_{11} - A & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} - A \end{bmatrix}$ 是沒意義的. 也就是 $A - xI_n$ 我們視為是 $A$ 這個矩陣減去對角線皆為 $x$ 的對角矩陣 $xI_n$ 所得的矩陣. 巧的是當 $x$ 代 $A$ 時 $xI_n$ 是 $AI_n = A$ 有定義, 所以會讓人誤以為這樣可以證明 Cayley-Hamilton Theorem.

要證明矩陣形式的 Cayley-Hamilton Theorem, 可以利用 adjoint matrix 和 determinant 的關係處理, 不過由於我們已證明 linear operator 的情形, 所以這裡我們利用 linear operator 和 matrix 的關係處理, 順便讓大家熟悉 linear operator 和 matrix 之間的轉換關係. 首先給定 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$, 考慮和 $T$ 有關的 linear transformation $T : \mathbb{F}^n \to \mathbb{F}^n$, 其定義為 $T(\mathbf{v}) = A\mathbf{v}$, $\forall \mathbf{v} \in \mathbb{F}^n$. 在此定義之下, 使用 $\mathbb{F}^n$ 的 standard ordered basis $\varepsilon$ 我們會有 $T$ 的表現矩陣 $[T]_\varepsilon = A$. 因此依定義 $T$ 的 characteristic polynomial $p_T(x)$ 就是 $A$ 的 characteristic polynomial $p_A(x)$. 利用 linear operator 的 Cayley-Hamilton Theorem (Theorem 7.4.8), 我們知 $p_T(T) = p_A(T)$ 是 zero operator, 所以 $p_A(T)$ 用 $\varepsilon$ 所得的表現矩陣 $[p_A(T)]_\varepsilon$ 就是零矩陣. 然而 $[p_A(T)]_\varepsilon$ 和 $p_A(A)$ 有什麼關係呢? 回顧一下, 當 $T : V \to W$, $T' : W \to U$ 為 linear transformation 時考慮 $\alpha,\beta,\gamma$ 分別為 $V,W,U$ 的 ordered basis, 則我們有 $[T'\circ T]_\alpha^\gamma=[T']_\beta^\gamma[T]_\alpha^\beta$. 現在 $T:\mathbb{F}^n\to\mathbb{F}^n$ 為 linear operator, 故考慮 $\alpha=\beta=\gamma=\varepsilon$ 以及 $T'=T$, 的情形我們會有

$$[T^2]_\varepsilon=[T\circ T]_\varepsilon=[T]_\varepsilon[T]_\varepsilon=A^2.$$

同理我們會有 $[T^k]_\varepsilon=A^k$. 另外我們也知兩個 linear transformations 的線性組合, 用相同的 ordered basis 其表現矩陣, 就是它們個別的表現矩陣做相同的線性組合. 因此若 $p_A(x)=(-1)^n(x^n+\cdots+c_1x+c_0)$, 則 $p_A(T)=(-1)^n(T^n+\cdots+c_1T+c_0\mathrm{id}_{\mathbb{F}^n})$. 因此 $p_A(T)$ 利用 standard ordered basis $\varepsilon$ 所得的表現矩陣 $[p_A(T)]_\varepsilon$ 就會是

$$(-1)^n([T]_\varepsilon^n+\cdots+c_1[T]_\varepsilon+c_0I_n)=(-1)^n(A^n+\cdots+c_1A+c_0I_n)=p_A(A).$$

得證 $p_A(A)=\mathbf{0}$, 因此有以下的定理.

**Theorem 7.4.9** (Cayley-Hamilton Theorem)**.** 假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$ 且 $p_A(x)$ 為 $A$ 的 *characteristic polynomial,* 則 $p_A(A)$ 為零矩陣.

當 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$ 為 diagonalizable 時, 我們可以將 $A$ 對角化求得 $A$ 的高次方 $A^k$, 而不必真的將多個 $A$ 相乘 (參見 Example 8.1.7). 當 $A$ 不是 diagonalizable, 要計算 $A^k$ 我們可以利用 Theorem 7.4.9 以及長除法將其次數降下來, 這樣也可很快算出 $A^k$, 我們看下面的 Example.

**Example 7.4.10.** 考慮 $A=\begin{bmatrix}1&2\\-2&1\end{bmatrix}$. 由於 $A$ 的 characteristic polynomial $p_A(x)=x^2-2x+5$ 在 $\mathbb{R}$ 中無法分解, 故知 $A$ 在 $\mathbb{R}$ 不是 diagonalizable. 不過由 Cayley-Hamilton Theorem, 我們知 $A^2-2A+5I_2=\mathbf{0}$. 我們可以利用此式計算 $A$ 的高次方. 例如計算 $A^5$. 首先利用長除法, 將 $x^5$ 除以 $p_A(x)=x^2-2x+5$, 可得 $x^5=(x^3+2x^2-x-12)(x^2-2x+5)-19x+60$. 兩邊代入 $A$ 得 $A^5=(A^3+2A^2-A-12I_2)(A^2-2A+5I_2)-19A+60I_2$. 由於 $A^2-2A+5I_2=\mathbf{0}$, 故得 $A^5=-19A+60I_2=\begin{bmatrix}41&-38\\38&41\end{bmatrix}$.

Chapter 8

# Diagonalizable Matrices and Their Applications

在這一章中, 我們探討 $n \times n$ matrix 對角化的問題以及其相關應用.

## 8.1. Diagonalizability

我們曾經提過, 當 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 若存在 $\mathbb{F}^n$ 的一組 basis $\mathbf{v}_1,\dots,\mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 $A$ 的 eigenvector, 則稱 $A$ 為 *diagonalizable*. 另一方面當 $V$ 是一個 vector space over $\mathbb{F}$ 且 $T: V \to V$ 是一個 linear operator. 若 $V$ 中存在一組 basis $\mathbf{v}_1,\dots,\mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 $T$ 的 eigenvectors, 則稱 $T$ 為 *diagonalizable*. 在這一節中我們將探討如何判斷一個方陣或一個 linear operator 是否是 diagonalizable.

要如何知道 $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ 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 $\mathbb{F}$ 中完全分解. 也就是存在 $\lambda_1,\dots,\lambda_k \in \mathbb{F}$ 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 $i = 1,\dots,k$, $m_i$ 就是 $\lambda_i$ 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 $i = 1,\dots,k$, $\lambda_i$ 的 geometric multiplicity 等於其 algebraic multiplicity, 亦即 $\dim(E_A(\lambda_i)) = m_i$. 此時我們令 $\mathbf{v}_{i,1},\dots,\mathbf{v}_{i,m_i}$ 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 $\mathbf{v}_{1,1},\dots,\mathbf{v}_{1,m_1},\dots,\mathbf{v}_{k,1},\dots,\mathbf{v}_{k,m_k}$ 是 linearly independent. 因為當它們是 linearly independent 時再加上它們是在 $\mathbb{F}^n$ 中且共有 $m_1+\cdots+m_k = n$ 個向量, 所以由 Corollary 3.6.10, 知它們是 $\mathbb{F}^n$ 中的一組 basis. 又因為它們皆為 $A$ 的 eigenvectors, 所以可知此時 $A$ 為 diagonalizable.

要說明 eigenvector 之間的線性關係, 我們先探討兩個 eigenvectors 的情況. 當 $\mathbf{v}$ 為 $A$ 的 eigenvector, 若其 eigenvalue 為 $\lambda$, 則和 $\mathbf{v}$ 平行的 nonzero vector 皆為 eigenvalue 為 $\lambda$ 的 eigenvector (參見 Proposition 7.2.3 (1)). 也因此若 $\mathbf{v},\mathbf{w}$ 為 $A$ 的 eigenvectors 而他們所對應的 eigenvalue 是相異時, 則 $\mathbf{v},\mathbf{w}$ 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 8.1.1.** 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 3.5.4, 這表示 $\mathbf{v}_k\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{k-1})$. 也就是說存在 $c_1,\ldots,c_{k-1}\in\mathbb{F}$ 使得

$$\mathbf{v}_k=c_1\mathbf{v}_1+\cdots+c_{k-1}\mathbf{v}_{k-1} \tag{8.1}$$

利用 eigenvector 的定義我們得

$$\lambda_k\mathbf{v}_k=A\mathbf{v}_k=A(c_1\mathbf{v}_1+\cdots+c_{k-1}\mathbf{v}_{k-1})=c_1A\mathbf{v}_1+\cdots+c_{k-1}A\mathbf{v}_{k-1}=c_1\lambda_1\mathbf{v}_1+\cdots c_{k-1}\lambda_{k-1}\mathbf{v}_{k-1}. \tag{8.2}$$

將式子 (8.1) 乘上 $\lambda_k$ 與式子 (8.2) 相減得

$$c_1(\lambda_k-\lambda_1)\mathbf{v}_1+\cdots+c_{k-1}(\lambda_k-\lambda_{k-1})\mathbf{v}_{k-1}=\mathbf{0}. \tag{8.3}$$

由於 $\mathbf{v}_k\neq\mathbf{0}$, 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 $\lambda_k-\lambda_i\neq0$. 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 (8.3) 告訴我們 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$ 為 linearly dependent, 此與歸納之假設相矛盾, 得證本定理. □

如何說明 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly independent 呢? 照慣例, 我們先假設 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly dependent. 亦即存在不全為 0 的 $c_{1,1},\ldots,c_{1,m_1},\ldots,c_{k,1},\ldots,c_{k,m_k}\in\mathbb{F}$ 使得

$$c_{1,1}\mathbf{v}_{1,1}+\cdots+c_{1,m_1}\mathbf{v}_{1,m_1}+\cdots+c_{k,1}\mathbf{v}_{k,1}+\cdots+c_{k,m_k}\mathbf{v}_{k,m_k}=\mathbf{0}.$$

此時對任意 $i\in\{1,\ldots,k\}$, 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 $\mathbf{v}_{i,1},\ldots,\mathbf{v}_{i,m_i}$ 為 linearly independent, 如果 $c_{i,1},\ldots,c_{i,m_i}$ 不全為 0, 可得 $\mathbf{w}_i\neq\mathbf{0}$. 但由於 $\mathbf{w}_i\in E_A(\lambda_i)$, 故此時 $\mathbf{w}_i$ 為 eigenvalue 為 $\lambda_i$ 的 eigenvector. 也就是說, 若存在某些 $c_{i,j}\neq0$, 則對於那些 $i$, $\mathbf{w}_i$ 會是 eigenvalue 為 $\lambda_i$ 的 eigenvectors 滿足 $\mathbf{w}_1+\cdots+\mathbf{w}_k=\mathbf{0}$. 此與 Proposition 8.1.1 所述, 不同 eigenvalue 的 eigenvectors 之間是 linearly independent 的結果相矛盾, 故得證 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly independent. 我們因此證得了當 $A$ 的 characteristic polynomial 可以在 $\mathbb{F}$ 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 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$ 為 diagonalizable, 則 $A$ 的 characteristic polynomial 可以在 $\mathbb{F}$ 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 8.1.2.** 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. 若 $\lambda \in \mathbb{F}$ 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 $\mathbf{v}_1,\ldots,\mathbf{v}_d$ 為 $E_A(\lambda)$ 的一組 basis. 由於 $\mathbf{v}_1,\ldots,\mathbf{v}_d$ 為 linearly independent, 我們可以將之拓展成 $\mathbb{F}^n$ 中的一組 basis $\mathbf{v}_1,\ldots,\mathbf{v}_d,\mathbf{v}_{d+1},\ldots,\mathbf{v}_n$. 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 $E = \left[\begin{array}{c|c} \lambda I_d & M_1 \\ \hline \mathbf{0} & M_2 \end{array}\right]$. 由於 $E - tI_n = \left[\begin{array}{c|c} (\lambda-t)I_d & M_1 \\ \hline \mathbf{0} & M_2 - tI_{n-d} \end{array}\right]$, 我們可得 $\det(E-tI_n) = (\lambda-t)^d \det(M_2 - tI_{n-d})$. 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 7.2.10), 因此得 $(t-\lambda)^d$ 可整除 $p_A(t)$. 然而 $\lambda$ 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 $d \leq m$. □

利用 Proposition 8.1.2 可以得到一個有趣的結果. 由於 $A$ 的 eigenvalue $\lambda$ 的 geometric multiplicity 必大於 0 (因對應 $\lambda$ 的 eigenvector 必存在) 且其值必小於等於其 algebraic multiplicity (Proposition 8.1.2). 因若 $\lambda$ 是 $A$ 的 characteristic polynomial 的單根 (即 $\lambda$ 的 algebraic multiplicity 為 1), 其 geometric multiplicity 一定等於其 algebraic multiplicity (皆為 1).

現假設 $n\times n$ matrix $A$ 是 diagonalizable. 依定義令 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,d_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,d_k}$ 是 $\mathbb{F}^n$ 的一組 basis, 且對任意 $i \in \{1,\ldots,k\}$, $\mathbf{v}_{i,1},\ldots,\mathbf{v}_{i,d_i}$ 為 $A$ 以 $\lambda_i$ 為 eigenvalue 的 eigenvector, 其中 $\lambda_1,\ldots,\lambda_k$ 皆相異. 由於 $\mathbf{v}_{i,1},\ldots,\mathbf{v}_{i,d_i} \in E_A(\lambda_i)$ 且為 linearly independent, 我們知 $\lambda_i$ 的 geometric multiplicity $\dim(E_A(\lambda_i)) \geq d_i$. 現又假設每個 $\lambda_i$ 的 algebraic multiplicity 為 $m_i$, 由 Proposition 8.1.2 我們有

$$m_i \geq \dim(E_A(\lambda_i)) \geq d_i, \forall i = 1,\ldots,k. \tag{8.4}$$

由於 $m_1+\cdots+m_k$ 表示 $A$ 的 characteristic polynomial $p_A(t)$ 根的個數 (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 $\mathbb{F}^n$ 的 dimension, 即 $n$. 因此將式子 (8.4) 中 $i = 1,\ldots,k$ 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 8.1.3.** 假設 $A \in \mathrm{M}_{n\times n}(\mathbb{F})$. 以下敘述是等價的.

(1) $\mathbb{F}^n$ 中存在一組 *basis* 是由 $A$ 的 *eigenvectors* 所組成.

(2) 存在一個 *invertible matrix* $C \in \mathsf{M}_{n\times n}(\mathbb{F})$ 使得 $C^{-1}AC$ 為 *diagonal matrix.*

(3) $A$ 的 *characteristic polynomial* 可在 $\mathbb{F}$ 中完全分解且 $A$ 的每個 *eigenvalue* 的 *geometric multiplicity* 等於其 *algebraic multiplicity.*

**Example 8.1.4.** 我們考慮矩陣 $A = \begin{bmatrix} 0 & 3 & 1 \\ -1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}, B = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$. 經計算可得它們有相同的 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. 由於 eigenvalue 2 的 algebraic multiplicity 為 1, 我們知其 geometric multiplicity 亦為 1, 所以我們僅要檢查 eigenvalue 1 的 geometric multiplicity 即可.

矩陣 $A$ 對於 eigenvalue 1 的 eigenspace, 即 $A - I_3 = \begin{bmatrix} -1 & 3 & 1 \\ -1 & 2 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ 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 $E_A(1) = \mathrm{Span}(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix})$. 也就是說 $A$ 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 $A$ 對於 eigenvalue 1 的 geometric multiplicity 為 1. 因其 geometric multiplicity 不等於 algebraic multiplicity, 可得 $A$ 不是 diagonalizable matrix. 回顧在 Example 7.3.2 中我們計算過 $B$ 在 eigenvalue 1 和 eigenvalue 2 的 geometric multiplicity 皆等於其 algebraic multiplicity, 所以 $B$ 為 diagonalizable matrix. 我們看如何將 $B$ 對角化.

由於 $B$ 對於 eigenvalue 為 1 和 2 的 eigenspace 分別為 $E_B(1) = \mathrm{Span}(\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix})$ 和 $E_B(2) = \mathrm{Span}(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix})$, 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 $B$ 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

$$BC = \begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 4 \\ 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = CD.$$

再由 $C$ 為 invertible, 得 $C^{-1}BC = D$.

依照 diagonalizable matrix 的定義, 我們可以將 Theorem 8.1.3 中任一項當成檢驗矩陣是否為 diagonalizable 的方法.

**Question 8.1.** 假設 $A$ 為 $n \times n$ *matrix.* 試利用 *Theorem 8.1.3* (2) 說明 $A$ 為 *diagonalizable* 若且唯若 $A^{\mathrm{t}}$ 為 *diagonalizable.*

由 Proposition 7.2.11 我們知道 $A$ 和 $A^{\mathrm{t}}$ 有相同的 characteristic polynomial 所以他們有相同的 eigenvalue 而且這些 eigenvalue 在 $A$ 和 $A^{\mathrm{t}}$ 的 algebraic multiplicity 會相同. 而 Question 8.1 似乎暗示這對 geometric multiplicity 也成立, 事實上我們有以下的結果.

**Proposition 8.1.5.** 假設 $A$ 為 $n\times n$ *matrix* 且 $\lambda\in\mathbb{R}$ 為 $A$ 的一個 *eigenvalue.* 則 $\lambda$ 對於 $A$ 的 *geometric multiplicity* 與 $\lambda$ 對於 $A^{\mathrm{t}}$ 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 3.7.14 我們知 $\dim(N(A-\lambda I_n))=\mathrm{nullity}(A-\lambda I_n)=n-\mathrm{rank}(A-\lambda I_n)$, 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 3.7.15), 得證 $\dim(N(A-\lambda I_n))=\dim(N(A^{\mathrm{t}}-\lambda I_n))$. □

**Question 8.2.** 假設 $A$ 為 $n\times n$ *matrix.* 試利用 *Theorem 8.1.3* (3) 說明 $A$ 為 *diagonalizable* 若且唯若 $A^{\mathrm{t}}$ 為 *diagonalizable.*

前一節我們提過, 對於 linear operator 的 eigenvalue, eigenvector 和其表現矩陣的 eigenvalue, eigenvector 之間的關係, 換言之一個 linear operator 是否為 diagonalizable 取決於其表現矩陣是否為 diagonalizable. 所以 Theorem 8.1.3 對於 linear operator 也是對的, 因此我們有以下結果.

**Theorem 8.1.6.** 假設 $V$ 為 *vector space over* $\mathbb{F}$ 且 $T:V\to V$ 為 *linear operator.* 以下敘述是等價的.

(1) $V$ 中存在一組 *basis* 是由 $T$ 的 *eigenvectors* 所組成.

(2) 存在一個 $V$ 的 *ordered basis* $\beta$ 使得 $[T]_\beta^\beta$ 為 *diagonal matrix.*

(3) $T$ 的 *characteristic polynomial* 可在 $\mathbb{F}$ 中完全分解且 $T$ 的每個 *eigenvalue* 的 *geometric multiplicity* 等於其 *algebraic multiplicity.*

最後我們再次強調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 $\mathbb{F}$ 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. 下一節我們將會證明所有的 symmetric matrix 皆為 diagonalizable.

假設 $A\in\mathsf{M}_{n\times n}(\mathbb{F})$ 為 diagonalizable, 我們知存在 invertible matrix $Q$ 使得 $Q^{-1}AQ$ 為 diagonal matrix $D$. 換言之, 我們可以將 $A$ 寫成 $A=QDQ^{-1}$. 也因此我們可得

$$A^2=(QDQ^{-1})(QDQ^{-1})=QD^2Q^{-1}.$$

同理對任意 $m\in\mathbb{N}$, 我們有 $A^m=QD^mQ^{-1}$. 寫成這樣有什麼好處呢? 因為 $D$ 為對角矩陣 $\begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix}$, 我們很可以容易算出 $D^m$, 即 $\begin{bmatrix}\lambda_1^m & & 0\\ & \ddots & \\ 0 & & \lambda_n^m\end{bmatrix}$. 因此只要知道 $Q$ 和 $Q^-1$, 我們就可以很輕易算出 $A^m$ (即 $QD^mQ^{-1}$), 而不必真正將 $A$ 乘到 $m$ 次方了.

**Example 8.1.7.** 考慮實矩陣 $B=\begin{bmatrix} -1 & 4 & 2 \\ -1 & 3 & 1 \\ -1 & 2 & 2 \end{bmatrix}$. 在 Example 8.1.4 我們算出 $Q^{-1}BQ=D$, 其中 $Q=\begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ 以及 $D=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$. 由於 $Q^{-1}=\begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & 0 \\ -1 & 2 & 1 \end{bmatrix}$, 我們得

$$B^5=QD^5Q^{-1}=\begin{bmatrix} 2 & 1 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 32 \end{bmatrix}\begin{bmatrix} 1 & -1 & -1 \\ 1 & -2 & 0 \\ -1 & 2 & 1 \end{bmatrix}=\begin{bmatrix} -61 & 124 & 62 \\ -31 & 63 & 31 \\ -31 & 62 & 32 \end{bmatrix}.$$

## 8.2. 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 $2\times 2$ 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 $\lambda_1,\lambda_2$. 也就是說 $\lambda_1,\lambda_2$ 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 $\mathbf{v}_1\neq\mathbf{0}$, 故 $\mathbf{v}_1$ 是 $A$ 的 eigenvector 其 eigenvalue 為 $\lambda_1$. 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 $\lambda_2$. 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 8.2.1.** 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 8.2.1 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 $\lambda_i$ 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 $\lambda_i$ 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 8.3.** 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 $\lambda_i$ 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 8.2.2.** 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 8.2.2, 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 8.2.3.** 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 8.2.3 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 8.2.4.** 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 8.2.5.** 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, 即其 entry 為複數的矩陣. 要注意複數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 $\mathbf{v} \neq \mathbf{0}$ 使得

$$(A^2 - 2aA + (a^2 + b^2)I_n)\mathbf{v} = A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v} = \mathbf{0}.$$

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 8.2.4 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

$$\|A\mathbf{v} - a\mathbf{v}\|^2 + b^2\|\mathbf{v}\|^2 = \langle A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{0}, \mathbf{v}\rangle = 0.$$

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 $2 \times 2$ 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 8.2.5 知存在實數 $\lambda$ 為 $A$ 的一個 eigenvalue. 令 $\mathbf{u}_1$ 為 $A$ 對於 $\lambda$ 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 $A\mathbf{u}_j = c_{1j}\mathbf{u}_1 + \cdots + c_{nj}\mathbf{u}_n$, 則依舉陣乘法定應我們有 $AQ = QC$, 其中 $C = [c_{ij}]$. 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 8.2.2, 得 $A$ 為 orthogonal diagonalizable, 也因此證明了 Spectral Theorem.

**Theorem 8.2.6** (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 8.2.6 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 8.2.7.** 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 $\lambda,\lambda'$. 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 8.2.4 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 8.2.7 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 8.2.8.** (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 8.2.7 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 $\dim(E_B(0)) = 1$, $\dim(E_B(9)) = 2$. 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 8.2.7 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

$$\begin{bmatrix}\frac{2}{3}&\frac{2}{3}&\frac{1}{3}\\-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}&0\\-\frac{1}{3\sqrt{2}}&-\frac{1}{3\sqrt{2}}&\frac{4}{3\sqrt{2}}\end{bmatrix}\begin{bmatrix}5&-4&-2\\-4&5&-2\\-2&-2&8\end{bmatrix}\begin{bmatrix}\frac{2}{3}&-\frac{1}{\sqrt{2}}&-\frac{1}{3\sqrt{2}}\\\frac{2}{3}&\frac{1}{\sqrt{2}}&-\frac{1}{3\sqrt{2}}\\\frac{1}{3}&0&\frac{4}{3\sqrt{2}}\end{bmatrix}=\begin{bmatrix}0&0&0\\0&9&0\\0&0&9\end{bmatrix}.$$

## 8.3. 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 8.3.1.** 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 8.2.8 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

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} d & e \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+f=0. \tag{8.5}$$

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 (8.5) 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

$$\lambda_1\overline{x}^2+\lambda_2\overline{y}^2+d'\overline{x}+e'\overline{y}+f=0, \tag{8.6}$$

其中 $\begin{bmatrix} d' & e' \end{bmatrix}=\begin{bmatrix} d & e \end{bmatrix}Q$.

首先我們考慮 $\lambda_1,\lambda_2$ 皆不為 0 的情形, 此時可以利用配方法將式子 (8.6) 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) $\lambda_1,\lambda_2$ 同號:

(1) $f'$ 與 $\lambda_1,\lambda_2$ 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 $\lambda_1,\lambda_2$ 異號: 此時很容易看出圖形為空集合.

(B) $\lambda_1,\lambda_2$ 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 8.3.2.** 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 $\lambda_1,\lambda_2$ 其中有一個為 0. 注意 $\lambda_1,\lambda_2$ 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 (8.6) 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) $\lambda_1,\lambda_2$ 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 $\lambda_1,f'$ 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 $\lambda_1,f'$ 異號: 此時圖形為空集合.

**Example 8.3.3.** 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 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$, 我們得

$$\begin{bmatrix} \overline{x} & \overline{y} \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}+\begin{bmatrix} 4\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}=4.$$

因此此曲線用新的變數其方程式為 $2\overline{x}^2+4\overline{x}+4\overline{y}=4$. 利用配方法得 $(\overline{x}+1)^2=-4\cdot\frac{1}{2}(\overline{y}-\frac{3}{2})$, 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 $\lambda_1,\lambda_2$, 然後由 $\lambda_1,\lambda_2$ 的正負號判斷此二次曲線應歸類於哪一種曲線. 若 $\lambda_1,\lambda_2$ 同號, 則為橢圓類; 而 $\lambda_1,\lambda_2$ 異號, 則為雙曲線類; 而若 $\lambda_1,\lambda_2$ 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 8.4.** *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* $2\times 2$ *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 $3\times 3$ symmetric matrix. 再將 $A$ 對角化然後變換變數成

$$\lambda_1\overline{x}^2+\lambda_2\overline{y}^2+\lambda_3\overline{z}^2+c'\overline{x}+d'\overline{y}+d'\overline{z}+f=0. \tag{8.7}$$

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 (8.7) 改寫成

$$\lambda_1(\overline{x}-h)^2+\lambda_2(\overline{y}-k)^2+\lambda_3(\overline{z}-l)^2=f'.$$

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 $\overline{x}=h$, $\overline{y}=k$ 和 $\overline{z}=l$ 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 $(\overline{x},\overline{y},\overline{z})=(h,k,l)$ 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 $\lambda_1,\lambda_2$ 同號):

(1) $f'$ 與 $\lambda_1,\lambda_2$ 同號: 此時曲面與 $\overline{z}=l$ 所交的圖形為橢圓, 而分別和 $\overline{x}=h$, $\overline{y}=k$ 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 $\lambda_1,\lambda_2$ 異號: 此時曲面與平面 $\overline{z}=l$ 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 $\overline{x}=h$, $\overline{y}=k$ 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 $\overline{z}=l$ 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 $\overline{x}=h$, $\overline{y}=k$ 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) $\lambda_2,\lambda_3$ 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 (8.6) 改寫成

$$\lambda_1(\overline{x}-h)^2+\lambda_2(\overline{y}-k)^2+e'\overline{z}=f'.$$

(1) $e'\neq 0$ 且 $\lambda_1,\lambda_2$ 同號: 此曲面會完全在平面 $e'\overline{z}=f'$ 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 $\overline{x}=h$, $\overline{y}=k$ 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 $\lambda_1,\lambda_2$ 異號: 此曲面與平面 $e'\overline{z}=f'$ 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 $\overline{x}=h$, $\overline{y}=k$ 所交的圖形為凹向相反的拋物線. 我們稱之為 *hyperbolic paraboloid*. 此曲面上的一點 $(\overline{x},\overline{y},\overline{z})=(h,k,f'/e')$ 就是所謂的 saddle point (鞍點).

(3) $e'=0$ 且 $\lambda_1,\lambda_2,f'$ 同號: 此時曲面與任何的水平平面 $\overline{z}=s$ 所交的圖形為橢圓. 圖形像橢圓柱面, 稱為 *elliptic cylinder*.

(4) $e'=0$ 且 $\lambda_1,\lambda_2$ 異號又 $f'\neq 0$: 此時曲面與任何的水平平面 $\overline{z}=s$ 所交的圖形為雙曲線. 圖形像雙曲柱面, 稱為 *hyperbolic cylinder*.

(5) $e'=0$ 且 $\lambda_1,\lambda_2$ 同號但與 $f'$ 異號: 此時是空集合.

(6) $e'=0$ 且 $\lambda_1,\lambda_2$ 同號又 $f'=0$: 此時曲面與任何的水平平面 $\overline{z}=s$ 僅交於一點. 圖形為一鉛直線.

(7) $e'=0$ 且 $\lambda_1,\lambda_2$ 異號又 $f'=0$: 此時曲面與任何的水平平面 $\overline{z}=s$ 交於兩相交直線. 圖形為兩相交平面.

(D) $\lambda_2,\lambda_3$ 皆為 0: 此時可以利用配方法將式子 (8.6) 改寫成

$$\lambda_1(\overline{x}-h)^2+d'\overline{y}+e'\overline{z}=f'.$$

(1) $d', e'$ 不全為 0: 此時令 $r=\sqrt{(d')^2+(e')^2}$ 利用變換變數

$$\begin{bmatrix}\overline{x}\\ \overline{y}\\ \overline{z}\end{bmatrix}=\begin{bmatrix}1&0&0\\ 0&d'/r&-e'/r\\ 0&e'/r&d'/r\end{bmatrix}\begin{bmatrix}t_1\\ t_2\\ t_3\end{bmatrix}$$

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 $\lambda_1$ 同號: 此時圖形為兩平行平面 (與 $\overline{x}=0$ 平行).

(3) $d'=e'=0$ 且 $f'=0$: 此時圖形為平面 $\overline{x}=h$.

(4) $d'=e'=0$ 且 $f'$ 與 $\lambda_1$ 異號: 此時為空集合.

**Example 8.3.4.** 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 8.2.8 (2)). 考慮變數變換 $\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}\frac{-1}{\sqrt{2}}&\frac{-1}{3\sqrt{2}}&\frac{2}{3}\\ \frac{1}{\sqrt{2}}&\frac{-1}{3\sqrt{2}}&\frac{2}{3}\\ 0&\frac{4}{3\sqrt{2}}&\frac{1}{3}\end{bmatrix}\begin{bmatrix}\overline{x}\\ \overline{y}\\ \overline{z}\end{bmatrix}$, 我們得

$$\begin{bmatrix}\overline{x}&\overline{y}&\overline{z}\end{bmatrix}\begin{bmatrix}9&0&0\\ 0&9&0\\ 0&0&0\end{bmatrix}\begin{bmatrix}\overline{x}\\ \overline{y}\\ \overline{z}\end{bmatrix}+\begin{bmatrix}2&2&1\end{bmatrix}\begin{bmatrix}\frac{-1}{\sqrt{2}}&\frac{-1}{3\sqrt{2}}&\frac{2}{3}\\ \frac{1}{\sqrt{2}}&\frac{-1}{3\sqrt{2}}&\frac{2}{3}\\ 0&\frac{4}{3\sqrt{2}}&\frac{1}{3}\end{bmatrix}\begin{bmatrix}\overline{x}\\ \overline{y}\\ \overline{z}\end{bmatrix}=9.$$

因此此曲面用新的變數其方程式為 $9\overline{x}^2+9\overline{y}^2+3\overline{z}=9$, 為前面列出的 (C)(1) 這個情形, 故知其為 elliptic paraboloid.

**Question 8.5.** 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?

## 8.4. 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 8.4.1.** 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 $2\times 2$ matrix, 所以由 $A$ 兩個相異的 eigenvalues 得 $A$ 為 diagonalizable. 事實上 $A$ 對於 $\lambda_1, \lambda_2$ 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 $\mathbf{v}_1 = \begin{bmatrix} F_1 \\ F_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 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 Processes*, 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 8.4.2.** 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 $i=1,\ldots,n$, 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 8.4.3.** $A=\begin{bmatrix} 1/2 & 1 \\ 1/2 & 0 \end{bmatrix}$ 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 8.4.4.** *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 $c_1,\ldots,c_n$ 以及 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 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 $\mathbf{a}_1,\ldots,\mathbf{a}_n$ 的每一個 entry 皆為正實數, 此時由於 $c_1,\ldots,c_n$ 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 $c_1,\ldots,c_n$ 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. □

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 8.4.4 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 8.4.4, 我們得證 $A^k$ 亦為 stochastic matrix. 同樣的利用數學歸納法以及 Lemma 8.4.4, 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 8.4.5.** 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 8.4.6.** 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix.* 則 1 為 $A^{\mathrm{t}}$ 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 $A^{\mathrm{t}}$, 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 $A^{\mathrm{t}}$ 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 $\mathbf{w} \neq \mathbf{0}$, 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 $A^{\mathrm{t}}$ 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

$$a_{j1}c_1 + \cdots + a_{ji}c_i + \cdots + a_{jn}c_n \leq a_{j1}c_j + \cdots + a_{ji}c_j + \cdots + a_{jn}c_j = (a_{j1} + \cdots + a_{jn})c_j = c_j. \quad (8.8)$$

由於依假設 $A^{\mathrm{t}}\mathbf{w} = \mathbf{w}$, 所以 $A^{\mathrm{t}}$ 的 $j$-th entry 應為 $c_j$, 也就是說式子 (8.8) 中的小於等於的符號應為等號, 也就是說對每一個 $i = 1, \ldots, n$ 皆有 $a_{ji}c_i = a_{ji}c_j$。故由 $a_{ji} \neq 0$ 的假設證得了 $c_1 = \cdots = c_j = \cdots = c_n = r$. 這說明了 $\mathbf{w} = r\mathbf{v}$, 亦即所有 $A^{\mathrm{t}}$ 的 eigenvalue 為 1 的 eigenvector 皆在 $\mathrm{Span}(\mathbf{v})$ 中. 因此得證 $A^{\mathrm{t}}$ 其 eigenvalue 1 的 geometric multiplicity 為 1 □

回顧 Proposition 7.2.11 和 Proposition 8.1.5 告訴我們 $A$ 和 $A^{\mathrm{t}}$ 有相同的 eigenvalues 而且每個 eigenvalue 對於 $A$ 和 $A^{\mathrm{t}}$ 的 geometric multiplicity 相同. 因此我們有以下的結果.

**Proposition 8.4.7.** 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 8.4.6 知 1 為 $A^{\mathrm{t}}$ 的一個 eigenvalue. 故由 Proposition 7.2.11 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 8.4.6 知 $(A^r)^{\mathrm{t}}$ 的 eigenvalue 1 其 geometric multiplicity 為 1. 也因此由 Proposition 8.1.5 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

$$\begin{array}{rcl}
|y_1| = |a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n| & \leq & a_{11}|x_1| + a_{12}|x_2| + \cdots + a_{1n}|x_n| \\
& \vdots & \\
|y_i| = |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n| & \leq & a_{i1}|x_1| + a_{i2}|x_2| + \cdots + a_{in}|x_n| \\
& \vdots & \\
|y_n| = |a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n| & \leq & a_{n1}|x_1| + a_{n2}|x_2| + \cdots + a_{nn}|x_n|
\end{array} \tag{8.9}$$

將式子 (8.9) 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

$$|y_1| + |y_2| + \cdots + |y_n| \leq |x_1| + |x_2| + \cdots + |x_n|. \tag{8.10}$$

現若 $\mathbf{x}$ 是 $A$ 的一個 eigenvalue 為 $\lambda$ 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 (8.10) 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 $\lambda$ 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 8.4.5), 故由 Proposition 8.4.7 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 8.1.1 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 8.4.7 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 (8.10) 的等式必須成立. 然而式子 (8.10) 的等式成立若且唯若式子 (8.9) 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 (8.9) 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 8.4.8.** *假設 $A\in\mathscr{M}_{n\times n}$ 為 stochastic matrix, 則 $A$ 的任一 eigenvalue $\lambda$ 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 8.6.** *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 $\lambda_i$, 其中 $\lambda_1=1$. 由 Proposition 8.4.8 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

$$P\begin{bmatrix}1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{bmatrix}P^{-1}=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}\begin{bmatrix}1 & \cdots & 1\\ - & \mathbf{0} & -\\ & \vdots & \\ - & \mathbf{0} & -\end{bmatrix}=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_1 & \cdots & \mathbf{v}_1\\ | & | & & |\end{bmatrix}.$$

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 8.4.10).

**Theorem 8.4.9.** 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

$$M_{k+1} = a_{ij}^{(k+1)} = a_{it}^{(k)} a_{tj} + \sum_{q\neq t} a_{iq}^{(k)} a_{qj} \leq m_k a_{tj} + M_k \sum_{q\neq t} a_{qj} = m_k a_{tj} + M_k(1 - a_{tj}). \tag{8.11}$$

$$m_{k+1}=a_{il}^{(k+1)}=a_{is}^{(k)}a_{sl}+\sum_{q\neq s}a_{iq}^{(k)}a_{ql}\geq M_k a_{sl}+m_k\sum_{q\neq s}a_{ql}=M_k a_{sl}+m_k(1-a_{sl}). \tag{8.12}$$

注意式子 (8.11), (8.12) 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 (8.11) 減去式子 (8.12) 得

$$M_{k+1}-m_{k+1}\leq M_k(1-a_{tj}-a_{sl})+m_k(a_{tj}-(1-a_{sl}))=(M_k-m_k)(1-a_{tj}-a_{sl}). \tag{8.13}$$

由於 $a_{tj}\geq 0, a_{sl}\geq 0$, 我們有 $1-a_{tj}-a_{sl}\leq 1$, 故由式子 (8.13) 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 (8.11), (8.12) 在 $B^k$ 和 $B^{k+1}$ 的情形, 可以推得式子 (8.13) 在 $B^k$ 與 $B^{k+1}$ 相對應的結果

$$M_{r(k+1)}-m_{r(k+1)}\leq M_{rk}(1-b_{tj}-b_{sl})+m_{rk}(b_{tj}-(1-b_{sl}))=(M_{rk}-m_{rk})(1-b_{tj}-b_{sl}). \tag{8.14}$$

要注意式子 (8.14) 與式子 (8.13) 最大的不同處在於 $b_{tj}>0$ 且 $b_{sl}>0$. 現令 $\beta$ 為 $B$ 的所有 entry 中的最小值. 依 $\beta$ 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 (8.13) 得

$$M_{r(k+1)}-m_{r(k+1)}\leq(M_{rk}-m_{rk})(1-2\beta). \tag{8.15}$$

因 $\beta$ 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 (8.15) 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 8.4.9 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 8.4.10.** 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 8.4.9 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 8.7.** *Corollary 8.4.10* 中 $\mathbf{w}$ 需要假設是 *probability vector* 才會成立嗎?

## 8.5. 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.