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 9

Eigenvalues and Eigenvectors

在這一章中, 我們介紹有關於 $n\times n$ matrices 的重要概念 eigenvalue 和 eigenvector. 我們首先介紹如何找到 eigenvalue 和 eigenvector 接著才介紹它們的性質. 最後我們探討 $n\times n$ matrix 對角化的問題以及其相關應用.

9.1. Characteristic Polynomial

給定 $n\times n$ matrix A, 以及 $\mathbf{v}\in\mathbb{R}^n$, 在數學上常會探討 $A^k\mathbf{v}$ 其中 k 為任意正整數的問題. 例如 Fibonacci sequence $F_1,F_2,\ldots$. 是一組滿足 $F_{k+1}=F_k+F_{k-1}$ 的遞迴數列. 若我們令 $A=\begin{bmatrix}1&1\\1&0\end{bmatrix}$ 且對任意 $k\geq 2$ 令 $\mathbf{v}_k=\begin{bmatrix}F_k\\F_{k-1}\end{bmatrix}$, 則

$$A\mathbf{v}_k=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}F_k\\F_{k-1}\end{bmatrix}=\begin{bmatrix}F_k+F_{k-1}\\F_k\end{bmatrix}=\begin{bmatrix}F_{k+1}\\F_k\end{bmatrix}=\mathbf{v}_{k+1}.$$

因此我們有 $\mathbf{v}_3=A\mathbf{v}_2$, $\mathbf{v}_4=A\mathbf{v}_3=A(A\mathbf{v}_2)=A^2\mathbf{v}_2,\ldots$, 這樣一直下去可得 $\mathbf{v}_{k+1}=A^{k-1}\mathbf{v}_2$. 也就是說對於任意 $k\geq 2$, 我們只要能算出 $A^{k-1}\mathbf{v}_2$, 就能求出 F_{k+1} 為何.

一搬來說當 k 越大時, 計算 $A^k\mathbf{v}$ 就越困難. 不過在一種特殊其況, 即當存在 $\lambda\in\mathbb{R}$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$ 時, 我們有 $A^2\mathbf{v}=A(A\mathbf{v})=A(\lambda\mathbf{v})=\lambda(A\mathbf{v})=\lambda^2\mathbf{v}$. 同理我們會有 $A^3\mathbf{v}=\lambda^3\mathbf{v},\ldots$, $A^k\mathbf{v}=\lambda^k\mathbf{v}$. 也就是說在這種情況之下, 就很容易計算出 $A^k\mathbf{v}$. 因此我們對於怎樣的 $\mathbf{v}$ 會存在 $\lambda\in\mathbb{R}$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$ 特別有興趣. 所以有以下的定義.

Definition 9.1.1. 假設 A 為 $n\times n$ matrix. 若對於非零向量 $\mathbf{v}\in\mathbb{R}^n$, 存在 $\lambda\in\mathbb{R}$ 使得 $A\mathbf{v}=\lambda\mathbf{v}$, 則稱 $\mathbf{v}$ 為 A 的一個 *eigenvector*, 而此 λ 稱為 A 的一個 *eigenvalue*.

注意, 依定義 A 的 eigenvector 一定是非零向量. 又若 $\mathbf{v}$ 是 A 的一個 eigenvector, 且 $\lambda,\lambda'\in\mathbb{R}$ 滿足 $A\mathbf{v}=\lambda\mathbf{v}=\lambda'\mathbf{v}$, 則由 $(\lambda-\lambda')\mathbf{v}=\mathbf{O}$ 以及 $\mathbf{v}\neq\mathbf{O}$, 可得 $\lambda=\lambda'$. 因此對於 A 的一個 eigenvector $\mathbf{v}$ 一定有也僅有一個實數 λ 會滿足 $A\mathbf{v}=\lambda\mathbf{v}$. 此時我們稱 eigenvector $\mathbf{v}$ 所對應的 eigenvalue 為 λ.

Question 9.1. 假設 A 為 $n \times n$ *matrix,* $\mathbf{v} \in \mathbb{R}^n$ 為非零向量滿足 $A\mathbf{v} = \mathbf{O}$. 是否 $\mathbf{v}$ 為 *eigenvector?* 其所對應的 *eigenvalue* 為何?

要如何找到一個 $n \times n$ matrix 的 eigenvector 及其對應的 eigenvalue 呢? 其實一般的找法是先找到 eigenvalue, 然後再找出與其對應的 eigenvector. 首先觀察若 $\lambda \in \mathbb{R}$ 是 A 的 eigenvalue, 表示存在一個非零向量 $\mathbf{v} \in \mathbb{R}^n$ 使得 $A\mathbf{v} = \lambda\mathbf{v}$. 由於 $I_n\mathbf{v} = \mathbf{v}$, 所以看成矩陣的運算 $\lambda\mathbf{v} = (\lambda I_n)\mathbf{v}$. 因此 $A\mathbf{v} = \lambda\mathbf{v}$ 就等同於 $(A - \lambda I_n)\mathbf{v} = \mathbf{O}$. 換言之, λ 是 A 的 eigenvalue 等同於由 $n \times n$ matrix $A - \lambda I_n$ 所對應的 linear system $(A - \lambda I_n)\mathbf{x} = \mathbf{O}$ 有 nontrivial solution $\mathbf{x} = \mathbf{v}$. 由 Theorem 3.5.9, 這又等同於 $A - \lambda I_n$ 不是 invertible, 再由 Theorem 8.2.6(1) 知這也等同於 $\det(A - \lambda I_n) = 0$. 總言之, 要找到 A 的 eigenvalue λ 就是要找到 λ 滿足 $\det(A - \lambda I_n) = 0$.

要怎樣找到 λ 滿足 $\det(A - \lambda I_n) = 0$ 呢? 假設 $A = [a_{ij}]$, 若我們將 t 視為變數, 考慮 $\det(A - tI_n)$. 由於

$$A - tI_n = \begin{bmatrix} a_{11} - t & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - t & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - t \end{bmatrix}$$

利用數學歸納法, 我們可以證明 $\det(A - tI_n)$ 會是一個以 t 為變數的 n 次實係數多項式. 而若 $t = \lambda$ 為此多項式的一實數根, 則 λ 就會滿足 $\det(A - \lambda I_n) = 0$, 也就是說 λ 就會是 A 的一個 eigenvalue. 反之, 若 λ 就會是 A 的一個 eigenvalue, 就表示 $t = \lambda$ 會是多項式 $\det(A - tI_n)$ 的一個根. 由此可知多項式 $\det(A - tI_n)$ 可以讓我們完全掌握 A 的 eigenvalue, 我們因而給它一個特別的定義.

Definition 9.1.2. 假設 A 為 $n \times n$ matrix, 考慮以 t 為變數的多項式 $p_A(t) = \det(A - tI_n)$. 我們稱 $p_A(t)$ 為 A 的 *characteristic polynomial* (特徵多項式)..

從上面的討論我們知道 $t = \lambda$ 為 characteristic polynomial $p_A(t)$ 的一個實根若且唯若 λ 為 A 的 eigenvalue. 要注意因為這裡我們談的是實矩陣 A 在 $\mathbb{R}^n$ 的 eigenvectors, 因此若 $t = \tau$ 是 $p_A(t)$ 的一個虛根 (即非實數的根), 此時假設存在 $\mathbf{v} \in \mathbb{R}^n$ 使得 $A\mathbf{v} = \tau\mathbf{v}$. 由於 $A\mathbf{v} \in \mathbb{R}^n$, 但 $\tau\mathbf{v} \notin \mathbb{R}^n$, 所以 $A\mathbf{v} = \tau\mathbf{v}$ 不可能成立. 因此在這個課程裡, 我們考慮 eigenvalue 僅考慮 characteristic polynomial 的實根.

Example 9.1.3. 考慮 $A = \begin{bmatrix} 2 & 0 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix}$, 此時 A 的 characteristic polynomial 為 $p_A(t) = \det(A - tI_3) = \det\begin{bmatrix} 2-t & 0 & 1 \\ 1 & 3-t & 1 \\ 1 & 1 & 2-t \end{bmatrix}$. 對第一個 row 降階求行列式得

$$p_A(t) = (2-t)\det\begin{bmatrix} 3-t & 1 \\ 1 & 2-t \end{bmatrix} + \det\begin{bmatrix} 1 & 3-t \\ 1 & 1 \end{bmatrix} = (2-t)(t^2 - 5t + 6 - 1) + (1 - (3-t)).$$

化簡可得 $p_A(t) = (2-t)(t^2 - 5t + 4) = (2-t)(t-1)(t-4)$. 由此知 $t = 1, 2, 4$ 為 A 的 characteristic polynomial 的三實根, 也因此得 A 有三個 eigenvalues $1, 2, 4$.

接下來我們說明當 A 為 $n\times n$ matrix 時, 其 characteristic polynomial $\det(A-tI_n)$ 確實是 t 的多項式. 首先觀察當我們在利用降階求 determinant 時, 其實是一些乘積之和. 利用數學歸納法可得這些乘積是由每一個 column 中的某個元素相乘而得而且它們都不會在同一個 row. 例如當我們計算 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ 的 characteristic polynomial $\det\begin{bmatrix} a-t & b \\ c & d-t \end{bmatrix}$ 時不難發現會貢獻 t 的最高次項乘積的是 $(a-t)(d-t)$ 而另一個乘積 bc 就僅影響到常數項, 因此其最高次項 t^2 與次高次項 t 的係數就完全由 $(a-t)(d-t)$ 的 t^2 與 t 的係數即 $at^2-(a+d)t$ 所決定. 現考慮 3×3 matrix $A=\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ 的 characteristic polynomial. 利用對第一個 row 降階的方式我們有

$$\det\begin{bmatrix} a-t & b & c \\ d & e-t & f \\ g & h & i-t \end{bmatrix}=(a-t)\det\begin{bmatrix} e-t & f \\ h & i-t \end{bmatrix}-b\det\begin{bmatrix} d & f \\ g & i-t \end{bmatrix}+c\det\begin{bmatrix} d & e-t \\ g & h \end{bmatrix}.$$

從前面 2×2 的情形我們看出 $\det\begin{bmatrix} e-t & f \\ h & i-t \end{bmatrix}$ 的 t^2 與次高次項 t 的係數就完全由 $(e-t)(i-t)$ 的 t^2 與 t 的係數所決定, 因此 $(a-t)(e-t)(i-t)$ 貢獻出 t^3 和 t^2 的係數. 而 $\det\begin{bmatrix} d & f \\ g & i-t \end{bmatrix}$ 和 $\det\begin{bmatrix} d & e-t \\ g & h \end{bmatrix}$ 最多僅有 t 的一次出現, 因此得 $\det(A-tI_3)$ 的 t^3 和 t^2 的係數完全由 $(a-t)(e-t)(i-t)$ 所決定. 也就是說 A 的 chacteristic polynomial $p_A(t)$ 為 3 次多項式且其最高次的兩項為 $(-1)^3t^3+(-1)^2(a+e+i)$. 這裡 a,e,i 為 A 的 diagonal entries, 它們之和 $a+e+i$ 我們稱為 A 的 trace, 用 $\mathrm{tr}(A)$ 來表示. 利用數學歸納法, 我們可得當 $A=[a_{ij}]$ 為 $n\times n$ matrix 時, A 的 characteristic polynomial $p_A(t)=\det(A-tI_n)$ 為 t 的 n 次實係數多項式, 且其最高次的兩項是由 $(a_{11}-t)(a_{22}-t)\cdots(a_{nn}-t)$ 所貢獻因此為 $(-1)^nt^n+(-1)^{n-1}(a_{11}+\cdots+a_{nn})t^{n-1}$. 由於 A 的 diagonal entries 之和 $a_{11}+\cdots+a_{nn}$ 我們定為 $\mathrm{tr}(A)$, 因此有以下之結論.

Proposition 9.1.4. 假設 A 為 $n\times n$ *matrix.* 則 A 的 *characteristic polynomial* 為 t 的 n 次實係數多項式. 其 t^n 項係數為 $(-1)^n$, t^{n-1} 項係數為 $(-1)^{n-1}\mathrm{tr}(A)$ 而常數項係數為 $\det(A)$.

Proof. 令 $p_A(t)=\det(A-tI_n)$, 由前面的討論我們僅剩討論 $p_A(t)$ 的常數項. 由於 $p_A(t)$ 是多項式所以它的常數項是 $p_A(0)=\det(A-0I_n)=\det(A)$. □

Question 9.2. 假設 A 為 $n\times n$ *matrix.* 試問 A 最多會有幾個相異的 *eigenvalues?*

接下來我們介紹一個和 eigenvalue 有關的定義. 若 $\lambda\in\mathbb{R}$ 是 A 的 eigenvalue. 由於 $t=\lambda$ 會是 A 的 characteristic polynomial $p_A(t)=\det(A-tI_n)$ 的一個根. 由因式定理知 $t-\lambda$ 會整除 $p_A(t)$. 若 $(t-\lambda)^m$ 可整除 $p_A(t)$, 但 $(t-\lambda)^{m+1}$ 不能整除 $p_A(t)$, 則我們稱 eigenvalue λ 的 *algebraic multiplicity* (代數重根數) 為 m. 當然了當 $t=\lambda$ 是 $p_A(t)$ 的一個單根, 我們就說 λ 的 algebraic multiplicity 為 1.

Question 9.3. *Identity matrix* I_n 的 *eigenvalue* 有哪些? 其 *algebraic multiplicity* 為何?

最後我們介紹一些和 characteristic polynomial 有關的性質. 一般來說兩個 $n\times n$ matrices 的 characteristic polynomial 可能不相同. 不過在一種特殊情況之下, 它們的

characteristic polynomial 會一樣. 當 A, B 為 $n \times n$ matrices, 若存在 $n \times n$ 的 invertible matrix U, 使得 $B = U^{-1}AU$, 則我們稱 A, B 為 *similar* (關於這個定義的原因我們以後會再詳述). 此時我們可得 A 和 B 的 characteristic polynomial 是相同的.

Proposition 9.1.5. 假設 A, B 為 $n \times n$ *matrices* 且存在 $n \times n$ 的 *invertible matrix* U 滿足 $B = U^{-1}AU$. 則 A 和 B 有相同的 *characteristic polynomial.*

Proof. 依假設 B 的 characteristic polynomial 為 $\det(B - tI_n) = \det(U^{-1}AU - tI_n)$. 然而依矩陣乘法性質

$$U^{-1}(A - tI_n)U = U^{-1}AU - U^{-1}(tI_n)U = U^{-1}AU - tU^{-1}I_nU = U^{-1}AU - tI_n.$$

因此再由 determinant 的性質 (Theorem 8.2.6) 得

$$\det(B - tI_n) = \det(U^{-1}(A - tI_n)U) = \det(U^{-1})\det(A - tI_n)\det(U) = \det(A - tI_n).$$

得證 A 和 B 有相同的 characteristic polynomial. □

另一個會有相同的 characteristic polynomial 的情況就是 A 和 A^{T} 有相同的 characteristic polynomial.

Proposition 9.1.6. 假設 A 為 $n \times n$ *matrix,* 則 A 和 A^{T} 有相同的 *characteristic polynomial*

Proof. 利用 trnaspose 的性質 $(A - tI_n)^{\mathrm{T}} = A^{\mathrm{T}} - tI_n^{\mathrm{T}} = A^{\mathrm{T}} - tI_n$ (Proposition 3.2.4), 故利用 Theorem 8.2.6 (3), 我們有

$$P_{A^{\mathrm{T}}}(t) = \det(A^{\mathrm{T}} - tI_n) = \det((A - tI_n)^{\mathrm{T}}) = \det(A - tI_n) = P_A(t).$$

□

Question 9.4. 試說明 A 和 A^{T} 有相同的 *eigenvalues* 且對每個 *eigenvalue* 其在 A 和 A^{T} 的 *algebraic multiplicity* 也相同.

9.2. Eigenspace

我們了解了如何找到一個 $n \times n$ matrix 的 eigenvalue 之後, 接下來便是要找出這些 eigenvalue 所對應的 eigenvectors.

假設 A 為 $n \times n$ matrix 且 $\lambda \in \mathbb{R}$ 為 A 的一個 eigenvalue. 由於 $\det(A - \lambda I_n) = 0$, 我們知聯立方程組 $(A - \lambda I_n)\mathbf{x} = \mathbf{O}$ 存在非零的 nontrivial solution. 現假設 $\mathbf{v} \in \mathbb{R}^n$ 為非零向量且 $\mathbf{x} = \mathbf{v}$ 為 $(A - \lambda I_n)\mathbf{x} = \mathbf{O}$ 的一組解. 此即表示 $\mathbf{v}$ 滿足 $(A - \lambda I_n)\mathbf{v} = \mathbf{O}$, 亦即 $A\mathbf{v} = \lambda\mathbf{v}$. 故此時 $\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector. 反之, 若 $\mathbf{v}$ 為 A 的一個以 λ 為 eigenvalue 的 eigenvector, 則 $\mathbf{x} = \mathbf{v}$ 必為 $(A - \lambda I_n)\mathbf{x} = \mathbf{O}$ 的一組 nontrivial solution. 因此我們只要掌握 $n \times n$ matrix $A - \lambda I_n$ 的 nullspace (即 $\{\mathbf{v} \in \mathbb{R}^n \mid (A - \lambda I_n)\mathbf{v} = \mathbf{O}\}$) 中的非零向量就會是 A 相對於 λ 的 eigenvector. 由於 nullspace 是 vector space, 因此我們有以下的定義.

Definition 9.2.1. 假設 A 為 $n \times n$ matrix 且 $\lambda \in \mathbb{R}$ 為 A 的一個 eigenvalue. 則 $A - \lambda I_n$ 的 nullspace 稱為 A 對於 eigenvalue λ 的 *eigenspace*. 我們用 $E_A(\lambda)$ 來表示.

要注意對於 λ 的 eigenspace 並不是由以 λ 為 eigenvalue 的 eigenvectors 所組成. 這是因為零向量 **O** 不是 eigenvector, 但 vector space 必須包含 **O**. 所以對於 λ 的 eigenspace 應該是由所有以 λ 為 eigenvalue 的 eigenvectors 和 **O** 所組成. 那為什麼要讓它形成 vector space 呢? 因為 vector space 有其方便性, 例如有了 vector space 我們就可以利用 dimension 來知道它的大小. 因此我們定義 $E_A(\lambda)$ 的 dimension 為 eigenvalue λ 的 *geometric multiplicity* (幾何重根數). 要注意 eigenvalue λ 的 algebraic multiplicity 無法讓我們知道 λ 所對應的 eigenvectors 的多寡, 而是 λ 的 geometric multiplicity 可以提供這一個訊息.

Example 9.2.2. 考慮 $A=\begin{bmatrix}-1&4&2\\-1&3&1\\-1&2&2\end{bmatrix}$, $B=\begin{bmatrix}0&3&1\\-1&3&1\\0&1&1\end{bmatrix}$. 可計算出 A 和 B 有相同的 characteristic polynomial $-(t-1)^2(t-2)$. 也因此 1 和 2 皆為 A 和 B 的 eigenvalues. 而且對於矩陣 A 和 B, eigenvalue 1 的 algebraic multiplicity 皆為 2, 而 eigenvalue 2 的 algebraic multiplicity 皆為 1. 接下來我們分節計算 A 和 B 的 eigenspace.

首先考慮 A 對於 eigenvalue 1 的 eigenspace, 亦即找出 $A-I_3=\begin{bmatrix}-2&4&2\\-1&2&1\\-1&2&1\end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix}1&-2&-1\\0&0&0\\0&0&0\end{bmatrix}$. 可得 $E_A(1)=\mathrm{Span}(\begin{bmatrix}2\\1\\0\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 1 的 eigenvector 就是那些由 $\begin{bmatrix}2\\1\\0\end{bmatrix}$ 和 $\begin{bmatrix}1\\0\\1\end{bmatrix}$ 的 linear combination 所得的 nonzero vector. 例如 $\mathbf{v}=\begin{bmatrix}4\\1\\2\end{bmatrix}=\begin{bmatrix}2\\1\\0\end{bmatrix}+2\begin{bmatrix}1\\0\\1\end{bmatrix}$ 就滿足

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

由於 $\dim(E_A(1))=2$, 我們也得到 A 對於 eigenvalue 1 的 geometric multiplicity 為 2. 至於 A 對於 eigenvalue 2 的 eigenspace, 亦即找出 $A-2I_3=\begin{bmatrix}-3&4&2\\-1&1&1\\-1&2&0\end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix}1&-2&0\\0&1&-1\\0&0&0\end{bmatrix}$. 因此得 $E_A(2)=\mathrm{Span}(\begin{bmatrix}2\\1\\1\end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 2 的 eigenvector 就是那些和 $\begin{bmatrix}2\\1\\1\end{bmatrix}$ 平行的 nonzero vector, 我們也得到 A 對於 eigenvalue 2 的 geometric multiplicity 為 1.

至於 B 對於 eigenvalue 1 的 eigenspace, 即 $B-I_3=\begin{bmatrix}-1&3&1\\-1&2&1\\0&1&0\end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix}1&0&-1\\0&1&0\\0&0&0\end{bmatrix}$. 可得 $E_B(1)=\mathrm{Span}(\begin{bmatrix}1\\0\\1\end{bmatrix})$. 也就是說 B 對於 eigenvalue 為 1 的 eigenvector 就是那些和 $\begin{bmatrix}1\\0\\1\end{bmatrix}$ 平行的 nonzero vector, 我們也得到 B 對於 eigenvalue 1 的 geometric multiplicity 為 1. 至於 B 對於 eigenvalue 2 的 eigenspace, 亦即找出 $B-2I_3=\begin{bmatrix}-2&3&1\\-1&1&1\\0&1&-1\end{bmatrix}$ 的 null space. 經由 elementary row operations, 可化為 echelon form $\begin{bmatrix}1&0&-2\\0&1&-1\\0&0&0\end{bmatrix}$. 因此得 $E_A(2)=\mathrm{Span}(\begin{bmatrix}2\\1\\1\end{bmatrix})$. 也就是說 A 對於 eigenvalue 為 2 的 eigenvector 就是那些和 $\begin{bmatrix}2\\1\\1\end{bmatrix}$ 平行的 nonzero vector, 我們也得到 A 對於 eigenvalue 2 的 geometric multiplicity 為 1.

從 Example 9.2.2 我們知道即使兩個矩陣有相同的 characteristic polynomial, 他們的 eigenvectors 有可能有很大的差異, 甚至連 eigenspace 的 dimension 也可能不同. 因此有可能兩個矩陣有相同的 eigenvalues 且相對應的 algebraic multiplicity 也相同, 但其 geometric multiplicity 卻相異.

Question 9.5. *Identity matrix I_n 的 eigenvalue 的 geometric multiplicity 為何?*

由 Proposition 9.1.6 我們知道 A 和 A^{T} 有相同的 characteristic polynomial 所以他們有相同的 eigenvalue 而且這些 eigenvalue 在 A 和 A^{T} 的 algebraic multiplicity 會相同. 這對 geometric multiplicity 也成立, 我們有以下的結果.

Proposition 9.2.3. *假設 A 為 $n\times n$ matrix 且 $\lambda\in\mathbb{R}$ 為 A 的一個 eigenvalue. 則 λ 對於 A 的 geometric multiplicity 與 λ 對於 A^{T} 的 geometric multiplicity 相等.*

Proof. 我們要說明 $\dim(E_A(\lambda))=\dim(E_{A^{\mathrm{T}}}(\lambda))$, 亦即 $\dim(N(A-\lambda I_n))=\dim(N(A^{\mathrm{T}}-\lambda I_n))$. 由 Theorem 4.4.13 我們知 $\dim(N(A-\lambda I_n))=\mathrm{null}(A-\lambda I_n)=n-\mathrm{rank}(A-\lambda I_n)$, 同理由 $A\in\mathscr{M}_{n\times n}$ 得 $\dim(N(A^{\mathrm{T}}-\lambda I_n))=n-\mathrm{rank}(A^{\mathrm{T}}-\lambda I_n)$. 因為 $A^{\mathrm{T}}-\lambda I_n=(A-\lambda I_n)^{\mathrm{T}}$ 以及 $\mathrm{rank}((A-\lambda I_n)^{\mathrm{T}})=\mathrm{rank}(A-\lambda I_n)$ (Proposition 4.4.14), 得證 $\dim(N(A-\lambda I_n))=\dim(N(A^{\mathrm{T}}-\lambda I_n))$. □

當 $\mathbf{v}$ 為 A 的 eigenvector, 若其 eigenvalue 為 λ, 則和 $\mathbf{v}$ 平行的 nonzero vector 皆為 eigenvalue 為 λ 的 eigenvector. 也因此若 $\mathbf{v},\mathbf{w}$ 為 A 的 eigenvectors 而他們所對應的 eigenvalue 是相異時, $\mathbf{v},\mathbf{w}$ 不可能平行. 也就是說 $\mathbf{v},\mathbf{w}$ 為 linearly independent. 這個結果可推廣到更一般的狀況.

Proposition 9.2.4. 假設 A 為 $n\times n$ *matrix* 且 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 A 的 *eigenvectors.* 若 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 所對應的 *eigenvalues* 皆相異, 則 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 *linearly independent.*

Proof. 我們利用數學歸納法證明. 前面已知 $k=2$ 的情形成立, 接著我們假設有 $k-1$ 個 eigenvectors 的情形也成立. 現考慮 k 個 eigenvectors 的情形. 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_k$ 為 A 的 eigenvectors 且其對應的 eigenvalue 分別為 $\lambda_1,\ldots,\lambda_k$ (亦即 $A\mathbf{v}_i=\lambda_i\mathbf{v}_i$, for $i=1,\ldots,n$). 依歸納法之假設 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$ 為 linearly independent. 現用反證法, 假設 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1},\mathbf{v}_k$ 為 linearly dependent. 依 Lemma 4.2.4, 這表示 $\mathbf{v}_k\in\mathrm{Span}(\mathbf{v}_1,\ldots,\mathbf{v}_{k-1})$. 也就是說存在 $c_1,\ldots,c_{k-1}\in\mathbb{R}$ 使得

$$\mathbf{v}_k=c_1\mathbf{v}_1+\cdots+c_{k-1}\mathbf{v}_{k-1} \tag{9.1}$$

利用 eigenvector 的定義我們得

$$\lambda_k\mathbf{v}_k=A\mathbf{v}_k=A(c_1\mathbf{v}_1+\cdots+c_{k-1}\mathbf{v}_{k-1})=c_1A\mathbf{v}_1+\cdots+c_{k-1}A\mathbf{v}_{k-1}=c_1\lambda_1\mathbf{v}_1+\cdots c_{k-1}\lambda_{k-1}\mathbf{v}_{k-1}. \tag{9.2}$$

將式子 (9.1) 乘上 λ_k 與式子 (9.2) 相減得

$$c_1(\lambda_k-\lambda_1)\mathbf{v}_1+\cdots+c_{k-1}(\lambda_k-\lambda_{k-1})\mathbf{v}_{k-1}=\mathbf{O}. \tag{9.3}$$

由於 $\mathbf{v}_k\neq\mathbf{O}$, 我們知 $c_1,\ldots,c_{k-1}$ 不全為 0. 而由 eigenvalue 皆相異, 我們知對任意 $i=1,\ldots,k-1$, 皆有 $\lambda_k-\lambda_i\neq 0$. 因此 $c_1(\lambda_k-\lambda_1),\ldots,c_{k-1}(\lambda_k-\lambda_{k-1})$ 為不全為 0 的實數. 換句話說, 式子 (9.3) 告訴我們 $\mathbf{v}_1,\ldots,\mathbf{v}_{k-1}$ 為 linearly dependent, 此與歸納法之假設相矛盾, 得證本定理. □

9.3. Diagonalization

假設 A 為 $n\times n$ matrix, 我們已經知道當 $\mathbf{v}$ 為 A 的 eigenvector 且其對應的 eigenvalue 為 λ, 則我們可以很容易計算 $A^k\mathbf{v}=\lambda^k\mathbf{v}$. 不過若 $\mathbf{v}$ 不是 A 的 eigenvector 怎麼辦? 在一種特殊的情況之下, 我們仍然可以很輕鬆的計算 $A^k\mathbf{v}$ 為何. 事實上, 當 $\mathbf{v}_1,\ldots,\mathbf{v}\in\mathbb{R}^n$ 是 $\mathbb{R}^n$ 的一組 basis 且皆為 A 的 eigenvectors 時, 我們就可以很容易的計算 $A^k\mathbf{v}$. 這是因為此時若 $\mathbf{v}\in\mathbb{R}^n$, 則由 basis 的定義知存在 $c_1,\ldots,c_n\in\mathbb{R}$ 使得 $\mathbf{v}=c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n$. 現若 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 所對應的 eigenvalues 分別為 $\lambda_1,\ldots,\lambda_n$, 此時

$$A\mathbf{v}=A(c_1\mathbf{v}_1+\cdots+c_n\mathbf{v}_n)=c_1A\mathbf{v}_1+\cdots+c_nA\mathbf{v}_n=c_1\lambda_1\mathbf{v}_1+\cdots+c_n\lambda_n\mathbf{v}_n.$$

若左邊再乘上 A 得

$$A^2\mathbf{v}=A(A\mathbf{v})=A(c_1\lambda_1\mathbf{v}_1+\cdots+c_n\lambda_n\mathbf{v}_n)=c_1\lambda_1A\mathbf{v}_1+\cdots+c_n\lambda_nA\mathbf{v}_n=c_1\lambda_1^2\mathbf{v}_1+\cdots+c_n\lambda_n^2\mathbf{v}_n.$$

因此利用數學歸納法可得

$$A^k\mathbf{v}=c_1\lambda_1^k\mathbf{v}_1+\cdots+c_n\lambda_n^k\mathbf{v}_n.$$

我們對有這樣特點的 matrix 給了以下的定義.

Definition 9.3.1. 假設 A 為 $n\times n$ matrix. 若存在 $\mathbb{R}^n$ 的一組 basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 A 的 eigenvectors, 則稱 A 為 *diagonalizable* (可對角化).

為甚麼稱為 diagonalizable 呢? 這是因為若 $\mathbf{v}_1,\ldots,\mathbf{v}\in\mathbb{R}^n$ 是 $\mathbb{R}^n$ 的一組 basis 且皆為 A 的 eigenvectors, 又假設它們所對應的 eigenvalues 分別為 $\lambda_1,\ldots,\lambda_n$. 亦即 $A\mathbf{v}_1=\lambda_1\mathbf{v}_1,\ldots,A\mathbf{v}_n=\lambda_n\mathbf{v}_n$. 此時由矩陣乘法的定義我們有

$$A\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}=\begin{bmatrix}| & | & & |\\ A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n\\ | & | & & |\end{bmatrix}=\begin{bmatrix}| & | & & |\\ \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n\\ | & | & & |\end{bmatrix}.$$

另一方面若考慮 (i,i)-th entry 為 λ_i 的 $n\times n$ diagonal matrix (即對角線第 i 個位置為 λ_i 而對角線外其餘位置皆為 0), 則我們有

$$\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}D=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \lambda_n\end{bmatrix}=\begin{bmatrix}| & | & & |\\ \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_3\mathbf{v}_n\\ | & | & & |\end{bmatrix}.$$

因此若令 $C=\begin{bmatrix}| & | & & |\\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\\ | & | & & |\end{bmatrix}$, 則我們有 $AC=CD$. 現又因 C 的 column 之間為 linearly independent 且有 n 個 column, 我們得 C 的 rank 為 n, 因此由 C 為 $n\times n$ matrix 得知 C 為 invertible (參見 Theorem 3.5.2). 因此我們可將 $AC=CD$ 改寫成 $D=C^{-1}AC$. 反之, 若存在一個 $n\times n$ invertible matrix 使得 $C^{-1}AC$ 為 diagonal matrix D, 則因 C 為 $n\times n$ invertible matrix, 所以 C 的 n 個 column vectors 形成 $\mathbb{R}^n$ 的一組 basis. 又因為 $AC=CD$, 由上面矩陣乘法的性質知 C 的 i-th column 就會是 A 以 D 的 (i,i)-th entry 為 eigenvalue 的 eigenvector. 所以 C 的 column vectors 就是 $\mathbb{R}^n$ 的一組 basis 且為 A 的 eigenvectors, 也就是說 A 為 diagonalizable. 前面曾經提過, 形如 $U^{-1}AU$ (其中 U 為 $n\times n$ invertible matrix) 這樣的 matrix 就稱為和 A 為 similar 的 matrix. 因此由這裡的討論, 我們知道 A 為 diagonalizable 就等同於 A 和一個 diagonal matrix 是 similar. 這也就是 diagonalizable 這個名稱的原因.

要如何知道一個 $n\times n$ matrix A 是否為 diagonalizable 呢? 從其定義, 我們知道它必須要有夠多的 eigenvectors. 以下我們要看一種特殊的情況可以確保 A 有夠多的 eigenvectors 從而得到 A 為 diagonalizable. 首先要有夠多的 eigenvectors 就表示要有夠多的 eigenvalues, 所以我們假設 A 的 characteristic polynomial 可以在 $\mathbb{R}$ 中完全分解. 也就是存在 $\lambda_1,\ldots,\lambda_k\in\mathbb{R}$ 皆相異且滿足 $p_A(t)=(-1)^n(t-\lambda_1)^{m_1}\cdots(t-\lambda_k)^{m_k}$. 依定義對於 $i=1,\ldots,k$, m_i 就是 λ_i 的 algebraic multiplicity 而且因 $p_A(t)$ 的次數為 n, 我們有 $m_1+\cdots+m_k=n$. 等一下我們會證明對於每個 eigenvalue, 其 geometric multiplicity 會小於等於其 algebraic multiplicity. 所以這裡 A 的 eigenvectors 要夠多, 最好的狀況就是每一個 eigenvalue 其 geometric multiplicity 等於其 algebraic multiplicity. 所以這裡我們假設對於 $i=1,\ldots,k$, λ_i 的 geometric multiplicity 等於其 algebraic multiplicity, 亦即 $\dim(E_A(\lambda_i))=m_i$. 此時我們令 $\mathbf{v}_{i,1},\ldots,\mathbf{v}_{i,m_i}$ 為 $E_A(\lambda_i)$ 的一組 basis. 將這 k 組 vectors 收集在一起後, 我們要說明它們 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly independent. 此時由於它們在 $\mathbb{R}^n$ 中且共有 $m_1+\cdots+m_k=n$ 個向量, 因此由 Corollary 4.3.5, 知它們是 $\mathbb{R}^n$ 中的一組 basis. 再加上它們皆為 A 的 eigenvectors, 所以可知此時 A 為 diagonalizable.

現假設 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly dependent. 亦即存在不全為 0 的實數 $c_{1,1},\ldots,c_{1,m_1},\ldots,c_{k,1},\ldots,c_{k,m_k}$ 使得

$$c_{1,1}\mathbf{v}_{1,1}+\cdots+c_{1,m_1}\mathbf{v}_{1,m_1}+\cdots+c_{k,1}\mathbf{v}_{k,1}+\cdots+c_{k,m_k}\mathbf{v}_{k,m_k}=\mathbf{O}.$$

此時對任意 $i\in\{1,\ldots,k\}$, 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 $\mathbf{w}_i\neq\mathbf{O}$. 但由於 $w_i\in E_A(\lambda_i)$, 故此時 $\mathbf{w}_i$ 為 eigenvalue 為 λ_i 的 eigenvector. 也就是說, 若存在某些 $c_{i,j}\neq 0$, 則對於那些 i, $\mathbf{w}_i$ 會是 eigenvalue 為 λ_i 的 eigenvectors 滿足 $\mathbf{w}_1+\cdots+\mathbf{w}_k=\mathbf{O}$. 此與 Proposition 9.2.4 所述, 不同 eigenvalue 的 eigenvectors 之間是 linearly independent 的結果相矛盾, 故得證 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,m_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,m_k}$ 是 linearly independent. 我們因此證得了當 A 的 characteristic polynomial 可以在 $\mathbb{R}$ 中完全分解且 A 的每一個 eigenvalue 的 geometric multiplicity 等於其 algebraic multiplicity, 則 A 為 diagonalizable.

其實反過來也是對的, 也就是說若 A 為 diagonalizable, 則 A 的 characteristic polynomial 可以在 $\mathbb{R}$ 中完全分解而且 A 的每一個 eigenvalue 的 geometric multiplicity 等於其 algebraic multiplicity. 不過在證明之前我們先證明前面提過的一般來說一個 eigenvalue 的 geometric multiplicity 會小於等於其 algebraic multiplicity.

Proposition 9.3.2. *假設 A 為 $n\times n$ matrix. 若 λ 為 A 的一個 eigenvalue 且其 geometric multiplicity 為 d 以及 algebraic multiplicity 為 m, 則 $d\leq m$.*

Proof. 依假設 $\dim(E_A(\lambda))=d$, 故令 $\mathbf{v}_1,\ldots,\mathbf{v}_d$ 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 $\mathbb{R}^n$ 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 $E=\left[\begin{array}{c|c}\lambda I_d & M_1\\ \hline \mathbf{O} & M_2\end{array}\right]$. 對於 1-st column 展開, 利用數學歸納法我們可以證明 $\det(E-tI_n)=(\lambda-t)^d\det(M_2-tI_{n-d})$. 換言之, E 的 characteristic polynomial 可以被 $(t-\lambda)^d$ 所整除. 然而 A 和 E 為 similar (因為 $E=C^{-1}AC$), 所以它們有相同的 characteristic polynomial (參見 Proposition 9.1.5), 因此得 $(t-\lambda)^d$ 可整除 $p_A(t)$. 然而 λ 的 algebraic multiplicity 為 m, 表示 m 為 $t-\lambda$ 可以整除 $p_A(t)$ 的最高次數, 因此得證 $d\leq m$. □

現假設 $n\times n$ matrix A 是 diagonalizable. 依定義令 $\mathbf{v}_{1,1},\ldots,\mathbf{v}_{1,d_1},\ldots,\mathbf{v}_{k,1},\ldots,\mathbf{v}_{k,d_k}$ 是 $\mathbb{R}^n$ 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 λ_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 λ_i 的 geometric multiplicity $\dim(E_A(\lambda_i))\geq d_i$. 現又假設每個 λ_i 的 algebraic multiplicity 為 m_i, 由 Proposition 9.3.2 我們有

$$m_i\geq\dim(E_A(\lambda_i))\geq d_i,\forall i=1,\ldots,k. \tag{9.4}$$

由於 $m_1+\cdots+m_k$ 表示 A 的 characteristic polynomial $p_A(t)$ 的實根個數 (含重根), 其值會小於等於 $p_A(t)$ 的次數 n. 而 $m_1+\cdots+m_k$ 表示 $\mathbb{R}^n$ 的 dimension, 即 n. 因此將式子 (9.4) 中 $i=1,\ldots,k$ 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 9.3.3. 假設 A 為 $n\times n$ *matrix.* 以下敘述是等價的.

(1) $\mathbb{R}^n$ 中存在一組 *basis* 是由 A 的 *eigenvectors* 所組成.

(2) 存在一個 $n\times n$ *invertible matrix* C 使得 $C^{-1}AC$ 為 *diagonal matrix.*

(3) A 的 *characteristic polynomial* 可在實數中完全分解且 A 的每個 *eigenvalue* 的 *geometric multiplicity* 等於其 *algebraic multiplicity.*

依照 diagonalizable matrix 的定義, 我們可以將 Theorem 9.3.3 中任一項當成檢驗矩陣是否為 diagonalizable 的方法, 其中以 (3) 是最常見的.

Question 9.6. 假設 A 為 $n\times n$ *matrix.* 試說明 A 為 *diagonalizable* 若且唯若 A^{T} 為 *diagonalizable.*

Example 9.3.4. 我們考慮 Example 9.2.2 中的矩陣 $A=\begin{bmatrix}-1&4&2\\-1&3&1\\-1&2&2\end{bmatrix}, B=\begin{bmatrix}0&3&1\\-1&3&1\\0&1&1\end{bmatrix}$. 前面已知它們有相同的 characteristic polynomial $-(t-1)^2(t-2)$. 也因此 A,B 的 eigenvalue 1 其 algebraic multiplicity 皆為 2, 而 eigenvalue 2 的 algebraic multiplicity 皆為 1. 不過在 Example 9.2.2 中我們知道 B 在 eigenvalue 1 的 geometric multiplicity 為 1, 所以 B 不是 diagonalizable matrix. 反之, A 在 eigenvalue 1 和 eigenvalue 2 的 geometric multiplicity 皆等於其 algebraic multiplicity, 所以 A 為 diagonalizable matrix. 我們看如何將 A 對角化.

由於 A 對於 eigenvalue 為 1 和 2 的 eigenspace 分別為 $E_A(1)=\mathrm{Span}(\begin{bmatrix}2\\1\\0\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix})$ 和 $E_A(2)=\mathrm{Span}(\begin{bmatrix}2\\1\\1\end{bmatrix})$, 可得 $\begin{bmatrix}2\\1\\0\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix},\begin{bmatrix}2\\1\\1\end{bmatrix}$ 就是一組由 A 的 eigenvectors 所形成的 $\mathbb{R}^3$ 的 basis. 因此若令 $C=\begin{bmatrix}2&1&2\\1&0&1\\0&1&1\end{bmatrix}$ 以及 $D=\begin{bmatrix}1&0&0\\0&1&0\\0&0&2\end{bmatrix}$, 則

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

再由 C 為 invertible, 得 $C^{-1}AC=D$.

最後我們要強調的是, 由於 A 的 eigenvalue λ 的 geometric multiplicity 必大於 0 (因對應 λ 的 eigenvector 必存在) 且其值必小於等於其 algebraic multiplicity (Proposition 9.3.2). 因此若 λ 是 A 的 characteristic polynomial 的單根 (即 λ 的 algebraic multiplicity 為 1), 其 geometric multiplicity 一定等於其 algebraic multiplicity (皆為 1). 所以在檢查一個矩陣

是否為 diagonalizable 時, 對於 algebraic multiplicity 為 1 的 eigenvalue 我們就不必檢查其 geometric multiplicity 了. 舉例來說, 若 A 的 characteristic polynomial 可在 $\mathbb{R}$ 中完全分解且其根皆為單根 (無重根), 則 A 一定為 diagonalizable. 另外還有一種矩陣不必檢查就知道一定是 diagonalizable, 就是 symmetric matrix. 以後我們將會證明所有的 symmetric matrix 皆為 diagonalizable.

9.4. The Spectral Theorem

在這一節中我們要探討 symmetric matrix. 我們將證明 symmetric matrix 皆為 diagonalizable, 更重要的是它們都是所謂的 *orthogonal diagonalizable*. 這個結果在數學和物理方面都有很重要的應用, 不過我們不會深入探討它的應用, 而著重於說明如何將 symmetric matrix 對角化.

首先我們來看 2×2 symmetric matrix 的情形. 假設 $A=\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, 其中 $b\neq 0$ (因為若 $b=0$, 此時 A 已為 diagonal matrix 不必對角化). 此時 A 的 characteristic polynomial 為 $P_A(t)=t^2-(a+c)t+(ac-b^2)$. 由於 $p_A(t)$ 的判別式 $(a+c)^2-4(ac-b^2)=(a-c)^2+4b^2>0$, 我們得 $P_A(t)=0$ 有兩相異實根 λ_1,λ_2. 也就是說 λ_1,λ_2 為 A 的兩相異 eigenvealue, 故知 A 為 diagonalizable. 事實上若令 $\mathbf{v}_1=\begin{bmatrix} b \\ \lambda_1-a \end{bmatrix}$, 我們有

$$A\mathbf{v}_1=\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} b \\ \lambda_1-a \end{bmatrix}=\begin{bmatrix} \lambda_1 b \\ b^2+\lambda_1 c-ac \end{bmatrix}=\lambda_1\begin{bmatrix} b \\ \lambda_1-a \end{bmatrix}=\lambda_1\mathbf{v}_1.$$

注意這裡我們用到了 $\lambda_1^2-(a+c)\lambda_1+(ac-b^2)=0$. 由於 $b\neq 0$, 我們知 $\mathbf{v}_1\neq\mathbf{O}$, 故 $\mathbf{v}_1$ 是 A 的 eigenvector 其 eigenvalue 為 λ_1. 同理令 $\mathbf{v}_2=\begin{bmatrix} b \\ \lambda_2-a \end{bmatrix}$, 我們可得 $\mathbf{v}_2$ 為 A 的 eigenvector 其 eigenvalue 為 λ_2. 重要的是, 我們有 $\langle\mathbf{v}_1,\mathbf{v}_2\rangle=b^2+\lambda_1\lambda_2-a(\lambda_1+\lambda_2)+a^2$. 利用根與係數關係, 即 $\lambda_1\lambda_2=ac-b^2$ 以及 $\lambda_1+\lambda_2=a+c$, 我們得 $\langle\mathbf{v}_1,\mathbf{v}_2\rangle=0$. 也就是說 $\mathbf{v}_1,\mathbf{v}_2$ 這組 $\mathbb{R}^2$ 的 basis 不只是由 A 的 eigenvectors 所組成, 而且它們倆倆互相垂直. 這種比一般 diagonalizable 更強的條件我們便稱之為 orthogonal diagonalizable. 其正式的定義如下.

Definition 9.4.1. 假設 $A\in\mathscr{M}_{n\times n}$, 若存在一組 $\mathbb{R}^n$ 的 orthogonal basis $\mathbf{v}_1,\dots,\mathbf{v}_n$ 其中每個 $\mathbf{v}_i$ 皆為 A 的 eigenvectors, 則稱 A 為 *orthogonal diagonalizable.*

當然了, 在 Definition 9.4.1 中若令 $\mathbf{u}_i=\frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$ 則 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 $\mathbb{R}^n$ 的一組 orthonormal basis 且皆為 A 的 eigenvectors. 所以 A 為 orthogonal diagonalizable 也等同於 $\mathbb{R}^n$ 中有一組 orthonormal basis 是由 A 的 eigenvector 所組成. 此時若 $\mathbf{u}_i$ 所對應的 eigenvalue 為 λ_i 且令 $Q=\begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix}$ 則可得 $AQ=QD$ 其中 D 為 (i,i)-th entry 為 λ_i 的 diagonal matrix, 也就是說我們可以將 A 對角化成 $Q^{-1}AQ=D$. 一般由 eigenvectors 所形成的 basis 都可以達到這個對角化的目的, 為何特別考慮 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 orthonormal basis 的情形呢? 這是因為當 $\mathbf{u}_1,\dots,\mathbf{u}_n$ 為 $\mathbb{R}^n$ 的 orthonormal basis 時, 我們會有 $Q^{\mathrm{T}}Q=I_n$, 也因此由 inverse matrix 的唯一性, 我們知 $Q^{\mathrm{T}}=Q^{-1}$. 也就是說當 Q 的 column vectors 為 $\mathbb{R}^n$ 的 orthonormal basis

時, 我們可以馬上得知 $Q^{-1}=Q^{\mathrm{T}}$. 就因為這個特性, 當一個 $n\times n$ matrix 其 column vectors 是由 $\mathbb{R}^n$ 的 orthonormal basis 所組成時, 我們特別稱之為 *orthogonal matrix* (注意不是稱為 orthonormal matrix). 也因此我們可以將 A 對角化成 $Q^{\mathrm{T}}AQ=D$, 故稱 A 為 orthogonal diagonalizable.

Question 9.7. 假設 $Q\in\mathscr{M}_{n\times n}$, 是否 $Q^{-1}=Q^{\mathrm{T}}$ 即表示 Q 為 *orthogonal matrix?*

反之, 若存在 Q 為 $n\times n$ orthogonal matrix 以及 $D=\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \end{bmatrix}$ 為 $n\times n$ diagonal matrix 使得 $Q^{\mathrm{T}}AQ=D$. 此時由 $AQ=QD$, 知 Q 的 i-th column 為 A 的 eigenvalue 為 λ_i 的 eigenvector, 也因此由 Q 的 column vectors 形成 $\mathbb{R}^n$ 的 orthonormal basis, 我們有以下之結果.

Proposition 9.4.2. 假設 $A\in\mathscr{M}_{n\times n}$. 則 A 為 *orthogonal diagonalizable* 若且唯若存在 $n\times n$ 的 *orthogonal matrix* Q 使得 $Q^{\mathrm{T}}AQ$ 為 *diagonal matrix.*

利用 Proposition 9.4.2, 我們知當 A 為 orthogonal diagonalizable 時存在 $Q,D\in\mathscr{M}_{n\times n}$ 其中 Q 為 orthogonal matrix, D 為 diagonal matrix 使得 $A=QDQ^{\mathrm{T}}$. 此時 $A^{\mathrm{T}}=(QDQ^{\mathrm{T}})^{\mathrm{T}}=(Q^{\mathrm{T}})^{\mathrm{T}}D^{\mathrm{T}}Q^{\mathrm{T}}$. 由於 $(Q^{\mathrm{T}})^{\mathrm{T}}=Q$ 且 $D^{\mathrm{T}}=D$ (因為 D 為 diagonal matrix), 我們得 $A^{\mathrm{T}}=QDQ^{\mathrm{T}}=A$, 亦即 A 為 symmetric. 得證了以下結果.

Corollary 9.4.3. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *orthogonal diagonalizable,* 則 A 為 *symmetric matrix.*

所謂 Spectral Theorem 指的就是 Corollary 9.4.3 的反向也是對的. 也就是說我們要證明當 A 為 symmetric 時, A 必為 orthogonal diagonalizable. 首先我們需要知道 symmetric matrix 和內積之間的關係.

Lemma 9.4.4. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *symmetric,* 則對於任意 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ 皆有 $\langle A\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{v},A\mathbf{w}\rangle$.

Proof. 回顧一下, 若將內積寫成矩陣乘法的形式, 對於任意 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ 我們有 $\langle\mathbf{v},\mathbf{w}\rangle=\mathbf{v}^{\mathrm{T}}\mathbf{w}$ (注意此處 $\mathbf{v},\mathbf{w}$ 皆視為 $n\times 1$ matrix). 因此得

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

最後由 $A^{\mathrm{T}}=A$ 之假設得證 $\langle A\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{v},A\mathbf{w}\rangle$. □

一個 $n\times n$ matrix 是否為 diagonalizable 第一個要檢查的條件就是其 characteristic polynomial 須在實數中完全分解. 接下來我們便是要說明一個 symmetric matrix 其 characteristic polynomial 確實可以在實數中完全分解.

Lemma 9.4.5. 假設 $A\in\mathscr{M}_{n\times n}$ 為 *symmetric,* 則 A 的 *characteristic polynomial* $p_A(t)$ 的根皆為實根.

Proof. 假設 $\lambda = a + b\imath$ (此處 $\imath$ 為虛數滿足 $\imath^2 = -1$) 為 $p_A(t)$ 的一個虛根, 即 $a, b \in \mathbb{R}$ 且 $b \neq 0$. 接下來我們要考慮複數矩陣, 極其 entry 為複數的矩陣. 要注意複述矩陣的運算以及行列式和實數矩陣有相同的規則. 所以依 $a + b\imath$ 為 $p_A(t)$ 的一根, 矩陣 $A - (a + b\imath)I_n$ 的行列式值為 0. 現將矩陣 $A - (a + b\imath)I_n$ 和矩陣 $A - (a - b\imath)I_n$ 相乘得

$$(A - (a + b\imath)I_n)(A - (a - b\imath)I_n) = A^2 - 2aA + (a^2 + b^2)I_n.$$

注意由於 $a, b \in \mathbb{R}$ 以及 A 為實數矩陣, 所以 $A^2 - 2aA + (a^2 + b^2)I_n$ 亦為實數矩陣. 另外由於 $\det(A - (a + b\imath)I_n) = 0$, 故有

$$\det(A^2 - 2aA + (a^2 + b^2)I_n) = \det(A - (a + b\imath)I_n)\det(A - (a - b\imath)I_n) = 0.$$

也就是說 $A^2 - 2aA + (a^2 + b^2)I_n$ 為 singular, 亦即存在 $\mathbf{v} \in \mathbb{R}^n$ 且 $\mathbf{v} \neq 0$ 使得

$$(A^2 - 2aA + (a^2 + b^2)I_n)\mathbf{v} = A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v} = \mathbf{O}.$$

然而

$$\langle A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v}, \mathbf{v}\rangle = \langle A^2\mathbf{v}, \mathbf{v}\rangle - 2a\langle A\mathbf{v}, \mathbf{v}\rangle + a^2\langle \mathbf{v}, \mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle.$$

又利用 A 為 symmetric, Lemma 9.4.4 告訴我們 $\langle A^2\mathbf{v}, \mathbf{v}\rangle = \langle A(A\mathbf{v}), \mathbf{v}\rangle = \langle A\mathbf{v}, A\mathbf{v}\rangle$, 故得

$$\langle A\mathbf{v} - a\mathbf{v}, A\mathbf{v} - a\mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle = \langle A^2\mathbf{v}, \mathbf{v}\rangle - 2a\langle A\mathbf{v}, \mathbf{v}\rangle + a^2\langle \mathbf{v}, \mathbf{v}\rangle + b^2\langle \mathbf{v}, \mathbf{v}\rangle,$$

亦即

$$\|A\mathbf{v} - a\mathbf{v}\|^2 + b^2\|\mathbf{v}\|^2 = \langle A^2\mathbf{v} - 2aA\mathbf{v} + (a^2 + b^2)\mathbf{v}, \mathbf{v}\rangle = \langle \mathbf{O}, \mathbf{v}\rangle = 0.$$

因為 $\|A\mathbf{v} - a\mathbf{v}\| \geq 0$, $\|\mathbf{v}\| > 0$, 我們得 $b = 0$. 此與當初假設 $b \neq 0$ 相矛盾, 故知 $p_A(t) = 0$ 沒有虛根, 即所有的根都是實根. $\square$

知道一個 symmetric matrix 的 characteristic polynomial 的根皆為實根, 我們便可以證明 symmetric matrix 皆為 orthogonal diagonalizable. 這裡我們要用數學歸納法, 也就是因為已證得 2×2 symmetric matrix 皆為 orthogonal diagonalizable. 現假設 $(n-1) \times (n-1)$ symmetric matrix 皆為 orthogonal diagonalizable. 我們要利用此證明當 A 為 $n \times n$ symmetric matrix 時亦為 orthogonal diagonalizable. 首先由 Lemma 9.4.5 知存在實數 λ 為 A 的一個 eigenvalue. 令 $\mathbf{u}_1$ 為 A 對於 λ 的 eigenvector 且 $\|\mathbf{u}_1\| = 1$. 利用 Gram-Schmidt process, 我們可以將 $\mathbf{u}_1$ 拓展成 $\mathbb{R}^n$ 的一組 orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. 現考慮 orthogonal matrix $Q = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix}$, 對於 $j = 1, \ldots, n$ 若 $A\mathbf{u}_j = c_{1\,j}\mathbf{u}_1 + \cdots + c_{n\,j}\mathbf{u}_n$, 則依舉陣乘法定應我們有 $AQ = QC$, 其中 $C = [c_{i\,j}]$. 因 Q 為 orthogonal matrix, 我們得 $C = Q^{-1}AQ = Q^{\mathrm{T}}AQ$. 因此再由 A 為 symmetric 得 $C^{\mathrm{T}} = Q^{\mathrm{T}}AQ = C$, 亦即 C 亦為 symmetric. 另一方面依假設 $A\mathbf{u}_1 = \lambda\mathbf{u}_1$, 我們知 C 的 1-st column 為 $\begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, 故由 C 為 symmetric 知 C 的 1-st row 為

$[\lambda \quad 0 \quad \cdots \quad 0]$. 也就是說 C 可以寫成以下的形式

$$C=\left[\begin{array}{c|ccc}\lambda & 0 & \cdots & 0\\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right].$$

由於 C 為 symmetric, 這裡 B 是 $(n-1)\times(n-1)$ symmetric matrix. 依歸納假設, 我們知 B 為 orthogonal diagonalizable, 亦即存在 $\mathbf{w}_1,\ldots,\mathbf{w}_{n-1}$ 為 $\mathbb{R}^{n-1}$ 的一組 orthonomal basis 且為 B 的 eigenvectors. 此時令 $R=\begin{bmatrix}| & | & & |\\ \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_{n-1}\\ | & | & & |\end{bmatrix}$, 我們得 R 為 $(n-1)\times(n-1)$ orthogonal matrix 且存在 $(n-1)\times(n-1)$ digonal matrix D 滿足 $R^{\mathrm{T}}BR=D$. 現在令 $P=\left[\begin{array}{c|ccc}1 & 0 & \cdots & 0\\ \hline 0 & & & \\ \vdots & & R & \\ 0 & & & \end{array}\right]$. 依矩陣乘法, 我們有

$$P^{\mathrm{T}}CP=\left[\begin{array}{c|ccc}\lambda & 0 & \cdots & 0\\ \hline 0 & & & \\ \vdots & & R^{\mathrm{T}}BR & \\ 0 & & & \end{array}\right]=\left[\begin{array}{c|ccc}\lambda & 0 & \cdots & 0\\ \hline 0 & & & \\ \vdots & & D & \\ 0 & & & \end{array}\right].$$

也就是說 $P^{\mathrm{T}}CP$ 為 diagonal matrix, 也因此得 $(QP)^{\mathrm{T}}A(QP)=P^{\mathrm{T}}(Q^{\mathrm{T}}AQ)P=P^{\mathrm{T}}CP$ 為 diagonal matrix. 注意由於 Q,P 皆為 orthogonal matrix, $(QP)^{\mathrm{T}}(QP)=P^{\mathrm{T}}(Q^{\mathrm{T}}Q)P=P^{\mathrm{T}}P=I_n$, 也就是說 QP 亦為 orthogonal matrix. 因此由 Proposition 9.4.2, 得 A 為 orthogonal diagonalizable, 也因此證明了 Spectral Theorem.

Theorem 9.4.6 (Spectral Theorem)**.** 假設 A 為 $n\times n$ *symmetric matrix,* 則 A 為 *orthogonal diagonalizable.*

接下來我們來探討, 給定一個 $n\times n$ symmetric matrix A, 如何找到 orthogonal matrix Q 使得 $Q^{\mathrm{T}}AQ$ 為 diagonal matrix. 當然了, 我們可以如 Theorem 9.4.6 的證明, 利用數學歸納法一步一步地將 Q 找到. 不過這要重複做好幾次的 Gram-Schmidt process, 頗為複雜. 利用以下的 Proposition, 我們可以將步驟簡化許多.

Proposition 9.4.7. 假設 A 為 $n\times n$ *symmetric matrix.* 若 $\mathbf{v},\mathbf{w}\in\mathbb{R}^n$ 為 A 的 *eigenvectors* 且其對應的 *eigenvalue* 為相異實數, 則 $\langle\mathbf{v},\mathbf{w}\rangle=0$.

Proof. 假設 $\mathbf{v},\mathbf{w}$ 所對應的 eigenvalue 分別為 λ,λ'. 也就是說 $A\mathbf{v}=\lambda\mathbf{v}, A\mathbf{w}=\lambda'\mathbf{w}$. 考慮 $\langle A\mathbf{v},\mathbf{w}\rangle=\langle\lambda\mathbf{v},\mathbf{w}\rangle=\lambda\langle\mathbf{v},\mathbf{w}\rangle$. 同理我們有 $\langle\mathbf{v},A\mathbf{w}\rangle=\lambda'\langle\mathbf{v},\mathbf{w}\rangle$. 然而 Lemma 9.4.4 告訴我們 $\langle A\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{v},A\mathbf{w}\rangle$, 故得 $(\lambda-\lambda')\langle\mathbf{v},\mathbf{w}\rangle=0$. 因此由題設 $\lambda\neq\lambda'$ 推得 $\langle\mathbf{v},\mathbf{w}\rangle=0$. □

當 A 為 $n\times n$ symmetric matrix, 我們簡單說明一下如何找到一組 A 的 eigenvectors 形成 $\mathbb{R}^n$ 的 orthonormal basis. 首先我們列出 A 的所有相異的 eigenvalues $\lambda_1,\ldots,\lambda_k$, 然後求出它們所對應的 eigenspace $E_A(\lambda_1),\ldots,E_A(\lambda_k)$. 若我們將每個 $E_A(\lambda_i)$ 的 basis 放在一起, 由於

A 為 diagonalizable 它們會是 $\mathbb{R}^n$ 的一組 basis. 雖然 Proposition 9.4.7 告訴我們, 當 $\lambda_i \neq \lambda_j$ 時, $E_A(\lambda_i)$ 和 $E_A(\lambda_j)$ 之間的向量是相互垂直的. 不過 $E_A(\lambda_i)$ 本身的那一組 basis 之間的向量未必倆倆相互垂直. 所以我們必須利用 Gram-Schmidt process 分別找到 A 每個 eigenspace $E_A(\lambda_i)$ 的一組 orthonormal basis. 再將這些 eigenspace 的 basis 放在一起它們自然兩兩互相垂直, 也因此它們就是由 A 的 eigenvectors 所組成的 $\mathbb{R}^n$ 的一組 orthonormal basis. 我們看以下的例子.

Example 9.4.8. (1) 考慮 symmetric matrix $A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. 我們有 A 的 characteristic polynomial 為 $p_A(t) = -(t+1)(t-1)(t-2)$. 所以 A 有三個相異的 eigenvalues, $-1, 1, 2$. 知道 A 必能對角化, 而且由 Proposition 9.4.7 知它們所對應的 eigenvector 會倆倆互相垂直. 事實上我們可求出對應到 $-1, 1, 2$ 的 eigenvector 分別為

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

很容易檢查它們確實倆倆互相垂直. 此時對於 $i = 1, 2, 3$ 令 $\mathbf{u}_i = \frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i$, 我們得

$$\mathbf{u}_1 = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

為 $\mathbb{R}^3$ 的一組 orthonromal basis. 故可將 A 對角化成

$$\begin{bmatrix} -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(2) 考慮 symmetric matrix $B = \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix}$. 我們有 B 的 characteristic polynomial 為 $p_B(t) = -t(t-9)^2$. 所以 B eigenvalues 為 $0, 9$. 知道 B 必能對角化, 我們知 $\dim(E_B(0)) = 1$, $\dim(E_B(9)) = 9$. 事實上 $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}$ 為 $E_B(0) = N(B)$ 的 basis, 而 $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$ 為 $E_B(9)$ 的 basis. 而且由 Proposition 9.4.7 知 $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle = 0$, 事實上很容易檢查它們確實成立. 不過 $\langle \mathbf{v}_2, \mathbf{v}_3 \rangle = 1 \neq 0$, 我們必須利用 Gram-Schmidt process 將 $\mathbf{v}_2, \mathbf{v}_3$ 換成 $E_B(9)$ 的一組 orthogonal basis. 令 $\mathbf{w}_2 = \mathbf{v}_2$ 且

$$\mathbf{w}_3 = \mathbf{v}_3 - \mathrm{Proj}_{\mathbf{w}_2}\mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}.$$

此時令 $\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1$, $\mathbf{u}_2 = \frac{1}{\|\mathbf{w}_2\|}\mathbf{w}_2$, $\mathbf{u}_3 = \frac{1}{\|\mathbf{w}_3\|}\mathbf{w}_3$ 我們得

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}2\\2\\1\end{bmatrix}, \quad \mathbf{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\end{bmatrix}, \quad \mathbf{u}_3 = \frac{1}{3\sqrt{2}}\begin{bmatrix}-1\\-1\\4\end{bmatrix}$$

為 $\mathbb{R}^3$ 的一組 orthonromal basis. 故可將 B 對角化成

$$\begin{bmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{3\sqrt{2}} & -\frac{1}{3\sqrt{2}} & \frac{4}{3\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \\ \frac{1}{3} & 0 & \frac{4}{3\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

9.5. Application: Conics and Quadric Surfaces

我們將利用 symmetric matrix 是 orthogonal diagonalizable 的特性將坐標平面上的二次曲線以及坐標空間上的二次曲面的方程式化成標準式, 以方便我們判別它們是哪一類的圖形.

一般來說我們是利用平移和旋轉的方法將二次曲線和二次曲面的方程式化成標準式. 其中旋轉的部分牽涉到對角化, 我們首先利用 quadratic form 來談對角化的問題. 所謂 n 個變數的 quadratic form 指的就是形如

$$\sum_{i,j=1}^{n} a_{ij}x_ix_j$$

這樣的二次式. 例如 $x^2+3xy-y^2$, $3x^2+y^2-z^2+5xy+xz+3yz$ 就是分別是兩個變數和三個變數的 quadratic form. 令 $\mathbf{x} = \begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}$, 所有 n 個變數的 quadratic form 都可以用矩陣表示成 $\mathbf{x}^{\mathrm{T}}A\mathbf{x}$ 的形式, 其中 A 為 $n\times n$ symmetric matrix. 例如兩個變數的 quadratic form $ax_1^2+bx_1x_2+cx_2^2$ 就可以寫成

$$ax_1^2+bx_1x_2+cx_2^2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}.$$

而三個變數的 quadratic form $ax_1^2+bx_2^2+cx_3^2+rx_1x_2+sx_1x_3+tx_2x_3$ 就可以寫成

$$ax_1^2+bx_2^2+cx_3^2+rx_1x_2+sx_1x_3+tx_2x_3 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} a & r/2 & s/2 \\ r/2 & b & t/2 \\ s/2 & t/2 & c \end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}.$$

將 quadratic form 寫成這樣的矩陣表示的好處是因為 A 是 symmetric, 故存在 orthogonal matrix Q 使得 $Q^{\mathrm{T}}AQ$ 為 diagonal matrix $\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$. 因此如果我們將變數 $\mathbf{x} = \begin{bmatrix}x_1\\ \vdots \\ x_n\end{bmatrix}$ 變換成 $\mathbf{t} = \begin{bmatrix}t_1\\ \vdots \\ t_n\end{bmatrix}$ 其中 $\mathbf{t} = Q^{\mathrm{T}}\mathbf{x}$ (注意因 $Q^{\mathrm{T}} = Q^{-1}$, 這等同於令 $\mathbf{x} = Q\mathbf{t}$), 則

$$\mathbf{x}^{\mathrm{T}}A\mathbf{x} = (Q\mathbf{t})^{\mathrm{T}}A(Q\mathbf{t}) = \mathbf{t}^{\mathrm{T}}(Q^{\mathrm{T}}AQ)\mathbf{t} = \begin{bmatrix} t_1 & \cdots & t_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\begin{bmatrix}t_1\\ \vdots \\ t_n\end{bmatrix} = \lambda_1t_1^2+\cdots+\lambda_nt_n^2.$$

也就是說, 我們可以藉由變換變數將一個 quadratic form 變成只有平方項. 我們看以下的例子.

Example 9.5.1. 考慮 quadratic form $x_1^2+4x_1x_2-2x_2^2$. 我們先寫下其矩陣形式
$$x_1^2+4x_1x_2-2x_2^2=\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$
由於 $\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$ 為 symmetric matrix, 故為 orthogonal diagonalizable, 事實上我們有
$$\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} 2/\sqrt{5} & -1/\sqrt{5} \\ 1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}=\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}.$$
因此若令 $\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}=\begin{bmatrix} 2/\sqrt{5} & 1/\sqrt{5} \\ -1/\sqrt{5} & 2/\sqrt{5} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ 則
$$\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}=\begin{bmatrix} t_1 & t_2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}=2t_1^2-3t_2^2.$$

對於 quadratic form $x_2^2+x_3^2+2x_1x_2+2x_1x_3$, 其矩陣形式為
$$x_2^2+x_3^2+2x_1x_2+2x_1x_3=\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$
我們曾在 Example 9.4.8 計算過 $Q^{\mathrm{T}}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}Q=\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ 其中 Q 為 orthogonal matrix $\begin{bmatrix} -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix}$. 因此若令 $\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}=\begin{bmatrix} -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ 則
$$\begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}\begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}=\begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}=-t_1^2+t_2^2+2t_3^2.$$

現在我們回到二次曲線的情況. 對於坐標平面上的二次曲線其一般的通式為 $ax^2+bxy+cy^2+dx+ey+f=0$. 我們可以將此式表為矩陣形式, 即
$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} d & e \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+f=0. \tag{9.5}$$
假設 symmetric matrix $A=\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$ 可對角化成 $Q^{\mathrm{T}}AQ=\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$. 此時考慮變換變數 $\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}=Q^{\mathrm{T}}\begin{bmatrix} x \\ y \end{bmatrix}$ (也就是 $\begin{bmatrix} x \\ y \end{bmatrix}=Q\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}$), 則式子 (9.5) 可寫成
$$\begin{bmatrix} \overline{x} & \overline{y} \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}+\begin{bmatrix} d & e \end{bmatrix}Q\begin{bmatrix} \overline{x} \\ \overline{y} \end{bmatrix}+f=0.$$
寫回方程式的樣子就是
$$\lambda_1\overline{x}^2+\lambda_2\overline{y}^2+d'\overline{x}+e'\overline{y}+f=0, \tag{9.6}$$
其中 $\begin{bmatrix} d' & e' \end{bmatrix}=\begin{bmatrix} d & e \end{bmatrix}Q$.

首先我們考慮 λ_1,λ_2 皆不為 0 的情形, 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+\lambda_2(\bar{y}-k)^2=f'.$$

我們分成下面幾種情形討論.

(A) λ_1,λ_2 同號:

(1) f' 與 λ_1,λ_2 同號: 此時圖形為 *ellipse* (橢圓). 注意當 $\lambda_1=\lambda_2$ 時會是圓, 不過這裡我們將之視為橢圓的一種.

(2) $f'=0$: 此時很容易看出圖形為 $(\bar{x},\bar{y})=(h,k)$ 這一點.

(3) f' 與 λ_1,λ_2 異號: 此時很容易看出圖形為空集合.

(B) λ_1,λ_2 異號:

(1) $f'\neq 0$: 此時圖形為 *hyperbola* (雙曲線).

(2) $f'=0$: 此時圖形為兩相交直線.

Example 9.5.2. 考慮二次曲線方程式 $2xy+\sqrt{2}x+\sqrt{2}y=1$. 此方程式可用矩陣表示成

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=1.$$

由於

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}=\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

考慮變數變換 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$, 我們得

$$\begin{bmatrix} \bar{x} & \bar{y} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}+\begin{bmatrix} \sqrt{2} & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}=1.$$

因此此曲線用新的變數其方程式為 $\bar{x}^2-\bar{y}^2+2\bar{x}=1$. 利用配方法得 $(\bar{x}+1)^2-\bar{y}^2=2$, 故其圖形為雙曲線.

同理若原方程式為 $2xy+\sqrt{2}x+\sqrt{2}y=-1$, 變換變數後的方程式為 $(\bar{x}+1)^2-\bar{y}^2=0$ 其圖形便會是兩相交直線 $\bar{x}+\bar{y}+1=0$ 和 $\bar{x}-\bar{y}+1=0$.

另一種情況是 λ_1,λ_2 其中有一個為 0. 注意 λ_1,λ_2 不可能同時為 0, 否則會是一次方程式. 不失一般性, 我們假設 $\lambda_1\neq 0,\lambda_2=0$ 的情形. 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+e'\bar{y}=f'.$$

我們分成下面幾種情形討論.

(C) λ_1,λ_2 其中有一個為 0 (不失一般性假設 $\lambda_1\neq 0,\lambda_2=0$):

(1) $e'\neq 0$: 此時圖形為 *parabola* (拋物線).

(2) $e'=0$ 且 λ_1,f' 同號: 此時圖形為兩平行直線 (與直線 $\bar{x}=0$ 平行).

(3) $e'=0$ 且 $f'=0$: 此時圖形為一直線 $\bar{x}=h$.

(4) $e'=0$ 且 λ_1,f' 異號: 此時圖形為空集合.

Example 9.5.3. 考慮二次曲線方程式 $x^2-2xy+y^2+4\sqrt{2}x=4$. 此方程式可用矩陣表示成

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} 4\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=4.$$

由於

$$\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}=\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

考慮變數變換 $\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$, 我們得

$$\begin{bmatrix} \bar{x} & \bar{y} \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}+\begin{bmatrix} 4\sqrt{2} & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}=4.$$

因此此曲線用新的變數其方程式為 $2\bar{x}^2+4\bar{x}+4\bar{y}=4$. 利用配方法得 $2(\bar{x}+1)^2+4\bar{y}=6$, 故其圖形為拋物線.

總而言之, 我們可以從二次曲線的 quadratic form 部分得到其 eigenvalue λ_1,λ_2, 然後由 λ_1,λ_2 的正負號判斷此二次曲線應歸類於哪一種曲線. 若 λ_1,λ_2 同號, 則為橢圓類; 而 λ_1,λ_2 異號, 則為雙曲線類; 而若 λ_1,λ_2 有一個為 0, 則為拋物線類. 不過最後我們還是得經由配方法求得其一次項與常數項, 這樣才能確認此曲線是否為 degenerated (退化) 的情形 (即直線, 點或空集合).

Question 9.8. *假設二次曲線方程式的 quadratic form 的部分可表成* $\begin{bmatrix} x & y \end{bmatrix}A\begin{bmatrix} x \\ y \end{bmatrix}$, *中 A 為 2×2 symmetric matrix. 試問是否可由 $\det(A)$ 來判斷此曲線是橢圓類, 雙曲線類還是拋物線類 (不考慮退化情形)?*

對於坐標空間的二次曲面我們也是用同樣方法處理. 首先寫成矩陣的形式

$$\begin{bmatrix} x & y & z \end{bmatrix}A\begin{bmatrix} x \\ y \\ z \end{bmatrix}+\begin{bmatrix} c & d & e \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}+f=0,$$

其中 A 為 3×3 symmetric matrix. 再將 A 對角化然後變換變數成

$$\lambda_1\bar{x}^2+\lambda_2\bar{y}^2+\lambda_3\bar{z}^2+c'\bar{x}+d'\bar{y}+d'\bar{z}+f=0. \tag{9.7}$$

二次曲面的分類頗為複雜, 大家不必記下這些分類. 不過為了完整性, 我們還是列出這些分類供同學參考. 由於此處無法利用圖形來解釋, 建議有興趣的同學參考課本上的圖形.

首先我們考慮 $\lambda_1,\lambda_2,\lambda_3$ 皆不為 0 的情形, 此時可以利用配方法將式子 (9.7) 改寫成

$$\lambda_1(\bar{x}-h)^2+\lambda_2(\bar{y}-k)^2+\lambda_3(\bar{z}-l)^2=f'.$$

我們分成下面幾種情形討論.

(A) $\lambda_1,\lambda_2,\lambda_3$ 同號:

(1) f' 與 $\lambda_1,\lambda_2,\lambda_3$ 同號: 此時曲面為有界的, 且與 $\bar{x}=h$, $\bar{y}=k$ 和 $\bar{z}=l$ 三個平面所交的圖形為橢圓. 曲面有點像橄欖球表面一樣, 我們稱之為 *ellipsoid*. 注意當 $\lambda_1=\lambda_2=\lambda_3$ 時會是球面, 不過這裡我們將之視為 ellipsoid 的一種.

(2) $f'=0$: 此時很容易看出圖形為 $(\bar{x},\bar{y},\bar{z})=(h,k,l)$ 這一點.

(3) f' 與 $\lambda_1,\lambda_2,\lambda_3$ 異號: 此時很容易看出圖形為空集合.

(B) $\lambda_1,\lambda_2,\lambda_3$ 異號 (不失一般性假設 λ_1,λ_2 同號):

(1) f' 與 λ_1,λ_2 同號: 此時曲面與 $\bar{z}=l$ 所交的圖形為橢圓, 而分別和 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為雙曲線. 因為曲面整體上只有一片, 我們稱之為 *hyperboloid* of one sheet.

(2) f' 與 λ_1,λ_2 異號: 此時曲面與平面 $\bar{z}=l$ 不相交, 不過若將平面往上或往下移動夠多的話會交出橢圓. 此曲面分別和 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為雙曲線. 因為曲面整體上有兩片, 我們稱之為 *hyperboloid* of two sheets.

(3) $f'=0$: 此時曲面與平面 $\bar{z}=l$ 交於一點, 不過若將平面往上或往下移的話會交出橢圓. 此區面分別和 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為兩相交直線. 圖形有點像甜筒, 我們稱之為 *elliptic cone*.

另一種情況是 $\lambda_1,\lambda_2,\lambda_3$ 其中有一個為 0. 注意 $\lambda_1,\lambda_2,\lambda_3$ 不可能皆為 0, 否則會是一次方程式. 不失一般性, 我們假設 $\lambda_1\neq 0$. 我們又可分成下面幾種情形討論.

(C) λ_2,λ_3 僅有一個為 0 (不失一般性假設 $\lambda_2\neq 0$): 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+\lambda_2(\bar{y}-k)^2+e'\bar{z}=f'.$$

(1) $e'\neq 0$ 且 λ_1,λ_2 同號: 此曲面會完全在平面 $e'\bar{z}=f'$ 之上方或下方, 不過若將平面往上或往下移動會交出橢圓. 而此曲面分別與 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為凹向一致的拋物線. 我們稱之為 *elliptic paraboloid*.

(2) $e'\neq 0$ 且 λ_1,λ_2 異號: 此曲面與平面 $e'\bar{z}=f'$ 交於兩相交直線, 不過若將平面往上或往下移動會交出雙曲線. 此曲面分別與 $\bar{x}=h$, $\bar{y}=k$ 所交的圖形為凹向相反的拋物線. 我們稱之為 *elliptic paraboloid*. 此曲面上的一點 $(\bar{x},\bar{y},\bar{z})=(h,k,f'/e')$ 就是所謂的 saddle point (鞍點).

(3) $e'=0$ 且 λ_1,λ_2,f' 同號: 此時曲面與任何的水平平面 $\bar{z}=s$ 所交的圖形為橢圓. 圖形像橢圓柱面, 稱為 *elliptic cylinder*.

(4) $e'=0$ 且 λ_1,λ_2 異號又 $f'\neq 0$: 此時曲面與任何的水平平面 $\bar{z}=s$ 所交的圖形為雙曲線. 圖形像雙曲柱面, 稱為 *hyperbolic cylinder*.

(5) $e'=0$ 且 λ_1,λ_2 同號但與 f' 異號: 此時是空集合.

(6) $e'=0$ 且 λ_1,λ_2 同號又 $f'=0$: 此時曲面與任何的水平平面 $\bar{z}=s$ 僅交於一點. 圖形為一鉛直線.

(7) $e'=0$ 且 λ_1,λ_2 異號又 $f'=0$: 此時曲面與任何的水平平面 $\bar{z}=s$ 交於兩相交直線. 圖形為兩相交平面.

(D) λ_2,λ_3 皆為 0: 此時可以利用配方法將式子 (9.6) 改寫成

$$\lambda_1(\bar{x}-h)^2+d'\bar{y}+e'\bar{z}=f'.$$

(1) d', e' 不全為 0: 此時令 $r=\sqrt{(d')^2+(e')^2}$ 利用變換變數

$$\begin{bmatrix}\bar{x}\\ \bar{y}\\ \bar{z}\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & d'/r & -e'/r\\ 0 & e'/r & d'/r\end{bmatrix}\begin{bmatrix}t_1\\ t_2\\ t_3\end{bmatrix}$$

我們又可將上式改寫成

$$\lambda_1(t_1-h)^2+rt_2=f'.$$

此曲面與任何的水平平面 $t_3=s$ 所交的圖形為拋物線. 圖形像拋物柱面, 稱為 *parabolic cylinder*.

(2) $d'=e'=0$ 且 f' 與 λ_1 同號: 此時圖形為兩平行平面 (與 $\bar{x}=0$ 平行).

(3) $d'=e'=0$ 且 $f'=0$: 此時圖形為平面 $\bar{x}=h$.

(4) $d'=e'=0$ 且 f' 與 λ_1 異號: 此時為空集合.

Example 9.5.4. 考慮坐標空間中曲面 $5x^2+5x^2+8z^2-8xy-4xz-4yz+2x+2y+z=9$. 寫成矩陣形式為

$$\begin{bmatrix}x & y & z\end{bmatrix}\begin{bmatrix}5 & -4 & -2\\ -4 & 5 & -2\\ -2 & -2 & 8\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}+\begin{bmatrix}2 & 2 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix}=9.$$

由於

$$\begin{bmatrix}\frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\\ \frac{-1}{3\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{4}{3\sqrt{2}}\\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3}\end{bmatrix}\begin{bmatrix}5 & -4 & -2\\ -4 & 5 & -2\\ -2 & -2 & 8\end{bmatrix}\begin{bmatrix}\frac{-1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3}\\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3}\\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3}\end{bmatrix}=\begin{bmatrix}9 & 0 & 0\\ 0 & 9 & 0\\ 0 & 0 & 0\end{bmatrix}$$

(請參考 Example 9.4.8 (2)). 考慮變數變換 $\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}\frac{-1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3}\\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3}\\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3}\end{bmatrix}\begin{bmatrix}\bar{x}\\ \bar{y}\\ \bar{z}\end{bmatrix}$, 我們得

$$\begin{bmatrix}\bar{x} & \bar{y} & \bar{z}\end{bmatrix}\begin{bmatrix}9 & 0 & 0\\ 0 & 9 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\bar{x}\\ \bar{y}\\ \bar{z}\end{bmatrix}+\begin{bmatrix}2 & 2 & 1\end{bmatrix}\begin{bmatrix}\frac{-1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3}\\ \frac{1}{\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3}\\ 0 & \frac{4}{3\sqrt{2}} & \frac{1}{3}\end{bmatrix}\begin{bmatrix}\bar{x}\\ \bar{y}\\ \bar{z}\end{bmatrix}=9.$$

因此此曲面用新的變數其方程式為 $9\bar{x}^2+9\bar{y}^2+3\bar{z}=9$, 為前面列出的 (C)(1) 這個情形, 故知其為 elliptic paraboloid.

Question 9.9. 空間中曲面 $5x^2+5x^2+8z^2-8xy-4xz-4yz+2x+2y+z=0$ 會是怎樣的圖形?

9.6. Application: Markov Processes

當 $A\in\mathscr{M}_{n\times n}$ 為 diagonalizable 時, 對於 $k\in\mathbb{N}$ 我們可以利用對角舉陣很容易求出 A^k. 進而對於任意 $\mathbf{v}\in\mathbb{R}^n$, 推算出 $A^k\mathbf{v}$. 其實還有一種情況 (即使不是 diagonalizable), 當 k 很大時我們也能 "估計" $A^k\mathbf{v}$ 大約為何. 這就是本節要探討的課題.

首先我們看 $A \in \mathscr{M}_{n\times n}$ 為 diagonalizable 的情形. 此時由於存在 diagonal matrix $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ 以及 invertible matrix $P \in \mathscr{M}_{n\times n}$ 使得 $D = P^{-1}AP$, 因此 $A = PDP^{-1}$. 依此我們可以推得

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1} = P\begin{bmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{bmatrix}P^{-1},$$

然後用數學歸納法推得

$$A^k = P\begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix}P^{-1}.$$

Example 9.6.1. 我們利用 Fibonacci sequence $0,1,1,2,3,5,8,13,\ldots,$ 來說明如何利用對角化. Fibonacci sequence 是一組滿足 $F_{k+1} = F_k + F_{k-1}$ 的遞迴數列, 其中 $F_0 = 0, F_1 = 1$. 我們令 $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ 且對任意 $k \geq 1$ 令 $\mathbf{v}_k = \begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix}$, 則

$$A\mathbf{v}_k = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} F_k \\ F_{k-1} \end{bmatrix} = \begin{bmatrix} F_k + F_{k-1} \\ F_k \end{bmatrix} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = \mathbf{v}_{k+1}.$$

因此我們有 $\mathbf{v}_{k+1} = A^k\mathbf{v}_1$. 也就是說對於任意 $k \geq 1$, 我們只要能算出 $A^k\mathbf{v}_1$, 就能求出 F_{k+1} 為何. 然而 A 的 characteristic polynomial 為 $P_A(t) = t^2 - t - 1$, 故得 A 的 eigenvalues 為 $\lambda_1 = (1-\sqrt{5})/2$, $\lambda_2 = (1+\sqrt{5})/2$. 因 A 是 2×2 matrix, 所以由 A 兩個相異的 eigenvalues 得 A 為 diagonalizable. 事實上 A 對於 λ_1, λ_2 的 eigenvector 分別為 $\mathbf{v}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$, $\lambda_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$. 因此若令 $P = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}$, 我們有

$$P = \begin{bmatrix} \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \\ 1 & 1 \end{bmatrix}, \quad P^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} -1 & \frac{1+\sqrt{5}}{2} \\ 1 & \frac{-1+\sqrt{5}}{2} \end{bmatrix}.$$

因此將 A 對角化得 $A = P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}P^{-1}$, 也因此求出對任意 $k \in \mathbb{N}$,

$$A^k = P\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}P^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \lambda_1^k & 0 \\ 0 & \lambda_2^k \end{bmatrix}\begin{bmatrix} -1 & \lambda_2 \\ 1 & -\lambda_1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} \lambda_2^{k+1} - \lambda_1^{k+1} & \lambda_2^k - \lambda_1^k \\ \lambda_2^k - \lambda_1^k & \lambda_2^{k-1} - \lambda_1^{k-1} \end{bmatrix}.$$

所以由 $\mathbf{v}_1 = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, 我們得

$$\mathbf{v}_{k+1} = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = A^k\mathbf{v}_1 = A^k\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} \lambda_2^{k+1} - \lambda_1^{k+1} \\ \lambda_2^k - \lambda_1^k \end{bmatrix},$$

故得

$$F_{k+1} = \frac{1}{5}(\lambda_2^{k+1} - \lambda_1^{k+1}) = \frac{1}{5}\left((\frac{1+\sqrt{5}}{2})^{k+1} - (\frac{1-\sqrt{5}}{2})^{k+1} \right).$$

接下來我們要探討的是, 有時即使 A 不是 diagonalizable, 但我們仍能估計 $A^k\mathbf{v}$. 這裡要探討的情況是所謂 *Markov Peocesses*, 是機率統計上的課題. 由於我們僅專注於線性代數的部分, 在這裡就不多談它的由來和例子, 直接切入主題.

Definition 9.6.2. 對於一 $\mathbb{R}^n$ 上的 vector $\mathbf{v} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, 若 $c_1 + \cdots + c_n = 1$ 且對於所有 $i = 1, \ldots, n$, 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 9.6.3. $A = \begin{bmatrix} 1/2 & 1 \\ 1/2 & 0 \end{bmatrix}$ 和 $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ 皆為 stochastic matrix. 而且 A 為 regular, 因為 $A^2 = \begin{bmatrix} 3/4 & 1/2 \\ 1/4 & 1/2 \end{bmatrix}$, 每個 entry 皆為正. 然而 I_2 不是 regular, 因為對於任意 $r \in \mathbb{N}$ 皆有 $I_2^n = I_2$ (除了對角線, 其他位置的 entry 皆為 0).

接下來我們看幾個有關 stochastic matrix 的性質.

Lemma 9.6.4. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix* 且 $\mathbf{v} \in \mathbb{R}^n$ 為 *probability vector*. 則 $A\mathbf{v}$ 亦為 *probability vector*. 另外若 A 的每一個 *entry* 皆為正實數, 則 $A\mathbf{v}$ 的每個 *entry* 亦皆為正實數.

Proof. 令 $A = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, 則 $A\mathbf{v} = c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$. 因此 $A\mathbf{v}$ 所有 entries 之和就是 $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$ 所有 entries 之和. 這等同於個別算出每個 $c_i\mathbf{a}_i$ 的所有 entries 之和再全部加起來. 然而因 $\mathbf{a}_i$ 為 probability vector, $c_i\mathbf{a}_i$ 的所有 entries 之和為 c_i, 所以 $c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$ 所有 entries 之和為 $c_1 + \cdots + c_n = 1$. 又因為 $c_1, \ldots, c_n$ 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. $\square$

現若 $A = \begin{bmatrix} | & & | \\ \mathbf{a}_1 & \cdots & \mathbf{a}_n \\ | & & | \end{bmatrix}$ 為 stochastic matrix, 則依矩陣乘法定義 A^2 的 i-th column 為 $A\mathbf{a}_i$, 故由 Lemma 9.6.4 知, A^2 的每個 column 皆為 probability vector, 亦即 A^2 亦為 stochastic matrix. 同理對任意 $k \geq 2$, A^k 的 i-th column 為 $A^{k-1}\mathbf{a}_i$, 因此利用數學歸納法以及 Lemma 9.6.4, 我們得證 A^k 亦為 stochastic matrix. 同樣的利用數學歸納法以及 Lemma 9.6.4, 我們可以證明若 A^r 的每一個 entry 皆為正實數, 則對於所有 $k \in \mathbb{N}$, $A^{r+k} = A^{r+k-1}A$ 的每個 entry 亦皆為正實數. 因此有以下的定理 (證明從略).

Proposition 9.6.5. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix*, 則對所有 $k \in \mathbb{N}$, A^k 亦為 *stochastic matrix*. 又若 A 為 *regular* 且 A^r 的每個 *entry* 皆為正實數, 則對所有 $k \in \mathbb{N}$, A^{r+k} 的每個 *entry* 亦皆為正實數.

接下來我們要談論 stochastic matrix 的 eigenvalues 以及 eigenvectors. 不像前面的情況, 由於我門探討的是一般的 stochastic matrix 而不是具體的矩陣, 所以我們無法從它的 characteristic polynomial 來處理. 這裡我們需要特定的技巧, 首先我們從轉置矩陣出發.

Lemma 9.6.6. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix*. 則 1 為 A^{T} 的一個 *eigenvalue* 且 $\mathbf{v} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ 為其 *eigenvector*. 另外若 A 的每個 *entry* 皆為正實數, 則對於 A^{T}, 其 *eigenvalue* 1 的 *geometric multiplicity* 為 1.

Proof. 由於 A 為 stochastic matrix, A 每一個 column vector $\mathbf{a}_i$ 皆為 probability vector, 亦即 $\langle \mathbf{a}_i, \mathbf{v} \rangle = 1$. 因此我們有 $A^{\mathrm{T}}\mathbf{v} = \begin{bmatrix} \langle \mathbf{a}_1, \mathbf{v} \rangle \\ \vdots \\ \langle \mathbf{a}_n, \mathbf{v} \rangle \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{v}$. 得證 $\mathbf{v}$ 為 A^{T} 的 eigenvector 且其 eigenvalue 為 1.

現假設 A 的每個 entry 皆為正實數且 $\mathbf{w} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 為 eigenvalue 為 1 的 eigenvector. 注意 $\mathbf{w} \neq \mathbf{O}$, 因此不失一般性, 我們可假設 $c_1, \ldots, c_n$ 的最大值不為 0 (因為若最大值為 0, 表示每個 $c_i \leq 0$, 故此時考慮 $-\mathbf{w}$, 其仍為 A^{T} 的一個 eigenvalue 為 1 的 eigenvector 且此時 $-\mathbf{w}$ 每個 entry 的最大值為正實數). 假設 c_j 為 $c_1, \ldots, c_n$ 的最大值. 考慮 $A\mathbf{w}$ 的 j-th entry, 依定義其值為 $\langle \mathbf{a}_j, \mathbf{w} \rangle = a_{j1}c_1 + \cdots + a_{ji}c_i + \cdots + a_{jn}c_n$. 因為 $a_{j1}, \ldots, a_{jn}$ 皆為正實數且 $c_j > 0$ 為 $c_1, \ldots, c_n$ 的最大值, 我們有

$$a_{j1}c_1 + \cdots + a_{ji}c_i + \cdots + a_{jn}c_n \leq a_{j1}c_j + \cdots + a_{ji}c_j + \cdots + a_{jn}c_j = (a_{j1} + \cdots + a_{jn})c_j = c_j. \quad (9.8)$$

由於依假設 $A^{\mathrm{T}}\mathbf{w} = \mathbf{w}$, 所以 A^{T} 的 j-th entry 應為 c_j, 也就是說式子 (9.8) 中的小於等於的符號應為等號, 也因此證得了 $c_1 = \cdots = c_j = \cdots = c_n = r$. 這說明了 $\mathbf{w} = r\mathbf{v}$, 亦即所有 A^{T} 的 eigenvalue 為 1 的 eigenvector 皆在 $\mathrm{Span}(\mathbf{v})$ 中. 因此得證 A^{T} 其 eigenvalue 1 的 geometric multiplicity 為 1 □

回顧 Proposition 9.1.6 和 Proposition 9.2.3 告訴我們 A 和 A^{T} 有相同的 eigenvalues 而且每個 eigenvalue 對於 A 和 A^{T} 的 geometric multiplicity 相同. 因此我們有以下的結果.

Proposition 9.6.7. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *stochastic matrix*. 則 1 為 A 的一個 *eigenvalue*. 另外若 A 為 *regular*, 則對於 A, 其 *eigenvalue* 1 的 *geometric multiplicity* 為 1.

Proof. 因 A 為 stochastic matrix, 由 Lemma 9.6.6 知 1 為 A^{T} 的一個 eigenvalue. 故由 Proposition 9.1.6 知 1 亦為 A 的一個 eigenvalue. 另外, 若 A 為 regular 且假設 $r \in \mathbb{N}$ 使得 A^r 的每個 entry 皆為正實數, 則由 Lemma 9.6.6 知 $(A^r)^{\mathrm{T}}$ 的 eigenvalue 1 其 geometric multiplicity 為 1. 也因此由 Proposition 9.2.3 知 A^r 的 eigenvalue 1, 其 geometric multiplicity 亦為 1, 也就是說 $\dim(E_{A^r}(1)) = 1$. 現對於任意 $\mathbf{v} \in E_A(1)$, 由於 $A\mathbf{v} = \mathbf{v}$, 我們得 $A^r\mathbf{v} = \mathbf{v}$, 亦即 $\mathbf{v} \in E_{A^r}(1)$. 因此得 $E_A(1) \subseteq E_{A^r}(1)$, 所以 $\dim(E_A(1)) \leq \dim(E_{A^r}(1)) = 1$. 然而前面已知 1 為 A 的一個 eigenvalue, 因此 $\dim(E_A(1)) > 0$. 得證 $\dim(E_A(1)) = 1$, 亦即 1 對於 A 的 geometric multiplicity 為 1. $\square$

現在我們來探討一個 stochastic matrix $A \in \mathscr{M}_{n\times n}$ 其 eigenvector 有何特性. 對任意 $\mathbf{x} \in \mathbb{R}^n$ 假設 $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ 且 $A\mathbf{x} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, 對於 $i = 1, \ldots, n$ 我們有 $y_i = \sum_{j=1}^n a_{ij}x_j$ 且因為對於任意 $i, j \in \{1, \ldots, n\}$ 皆有 $a_{ij} \geq 0$, 我們得

$$\begin{array}{rcl} |y_1| = |a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n| & \leq & a_{11}|x_1| + a_{12}|x_2| + \cdots + a_{1n}|x_n| \\ & \vdots & \\ |y_i| \ = \ |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n| & \leq & a_{i1}|x_1| + a_{i2}|x_2| + \cdots + a_{in}|x_n| \\ & \vdots & \\ |y_n| = |a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n| & \leq & a_{n1}|x_1| + a_{n2}|x_2| + \cdots + a_{nn}|x_n| \end{array} \tag{9.9}$$

將式子 (9.9) 由上往下加起來且將右式同樣的 $\left|x_j\right|$ 加在一起, 由於 A 為 stochastic matrix, 對於 $j = 1, \ldots, n$, 我們有 $a_{1j} + \cdots + a_{ij} + \cdots + a_{nj} = 1$, 故得

$$|y_1| + |y_2| + \cdots + |y_n| \leq |x_1| + |x_2| + \cdots + |x_n|. \tag{9.10}$$

現若 $\mathbf{x}$ 是 A 的一個 eigenvalue 為 λ 的 eigenvector, 我們有 $A\mathbf{x} = \lambda\mathbf{x}$, 也就是 $y_i = \lambda x_i$, $\forall i = 1, \ldots, n$. 代入式子 (9.10) 得 $|\lambda|(|x_1| + |x_2| + \cdots + |x_n|) \leq |x_1| + |x_2| + \cdots + |x_n|$. 由於 $x_1, \ldots, x_n$ 不全為 0 得證 $|\lambda| \leq 1$. 也就是說當 A 為 stochastic matrix, 它的 eigenvalue λ 必須滿足 $|\lambda| \leq 1$.

我們先考慮 $\lambda \neq 1$ 的情形. 此時對於 $i = 1, \ldots, n$ 我們有 $\sum_{j=1}^n a_{ij}x_j = y_i = \lambda x_i$. 將此 n 個等式全部加起來且等式左邊同樣的 x_i 項合併, 由於 A 為 stochastic matrix 我們得 $x_1 + \cdots + x_n = \lambda(x_1 + \cdots + x_n)$. 又由於 $\lambda \neq 1$ 得證 $x_1 + \cdots + x_n = 0$.

現在我們回到探討 A 為 regular 的情形. 首先考慮 eigenvalue 為 -1 的情形. 假設 $\mathbf{w} \in \mathbb{R}^n$ 為 A 的 eigenvalue 為 -1 的 eigenvector. 此時由於 $A\mathbf{w} = -\mathbf{w}$, 我們得 $A^k\mathbf{w} = (-1)^k\mathbf{w}$. 也就是說當 $k > 0$ 為偶數時 $\mathbf{w}$ 就會是 A^k 的 eigenvalue 為 1 的 eigenvector. 又 A 必有 eigenvalue 為 1 的 eigenvector $\mathbf{v}$, 且當 A 為 regular 時我們知必存在 $k > 0$ 為偶數使得 A^k 的每個 entry 皆為正實數 (Proposition 9.6.5), 故由 Proposition 9.6.7 的證明我們知此時 $E_A(1) = E_{A^k}(1) = \mathrm{Span}(\mathbf{v})$. 然而 $\mathbf{w} \in E_{A^k}(1) = \mathrm{Span}(\mathbf{v})$, 此與 $\mathbf{v}, \mathbf{w}$ 為 linearly independent (因 $\mathbf{v}, \mathbf{w}$ 分別為 A 對應到 $1, -1$ 的 eigenvector, 由 Proposition 9.2.4 知它們為 linearly independent) 相矛盾. 故知當 A 為 regular stochastic matrix 時 -1 不可能會是 A 的 eigenvalue.

最後我們考慮 A 為 regular 且 eigenvalue 為 1 的情形. 要注意由 Proposition 9.6.7 的證明我們知道, 若 A^r 的每個 entry 皆為正實數, 則 eigenspace $E_A(1) = E_{A^r}(1)$. 所以我們只要考慮 A 為 stochastic matrix 且 $A = (a_{ij})$ 的每個 entry a_{ij} 皆為正實數的情形即可. 假設 $\mathbf{x}$ 為 A 的 eigenvalue 為 1 的 eigenvector. 此時由於 $A\mathbf{x} = \mathbf{x}$, 我們有 $y_i = x_i$, $\forall i = 1, \ldots, n$, 因此得 $|y_1| + |y_2| + \cdots + |y_n| = |x_1| + |x_2| + \cdots + |x_n|$. 也就是說式子 (9.10) 的等式必須成立. 然而式子 (9.10) 的等式成立若且唯若式子 (9.9) 的每一項等式皆成立. 而又對於任意 $i, j \in \{1, \ldots, n\}$ 皆有 $a_{ij} > 0$ 故式子 (9.9) 的每一項等式皆成立若且唯若 $x_1, \ldots, x_n$ 皆同時大於等於 0 或同時小於等於 0. 由於 $x_1, \ldots, x_n$ 不全為 0, 我們有 $x_1 + \cdots + x_n \neq 0$, 故令 $\mathbf{v} = \frac{1}{x_1 + \cdots + x_n}\mathbf{x}$, 我們得 $\mathbf{v}$ 為 probability vector 且為 A 的 eigenvalue 為 1 的 eigenvector. 因為 $\dim(E_A(1)) = 1$, 我們知道 $\mathbf{v}$ 會是 $\mathbb{R}^n$ 中唯一同時符合這兩個要求的 vector. 綜合以上的討論, 我們有以下的結論.

Proposition 9.6.8. *假設 $A \in \mathscr{M}_{n\times n}$ 為 stochastic matrix, 則 A 的任一 eigenvalue λ 皆需滿足 $|\lambda| \leq 1$. 若 $\mathbf{w} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^n$ 為 A 的 eigenvector 且其 eigenvalue 不等於 1, 則 $c_1 + \cdots + c_n = 0$.*

另外若 A 為 regular, 則 -1 不可能會是 A 的 eigenvalue 且在 $\mathbb{R}^n$ 中存在唯一的 probability vector 會是 A 的 eigenvalue 為 1 的 eigenvector.

Question 9.10. *假設 A 為 regular stochastic matrix 且 $\mathbf{v}$ 為 A 的 eigenvalue 為 1 的 eigenvector. 試證明 $\mathbf{v}$ 每一個 entry 皆同時為正或同時為負 (即沒有一個 entry 會是 0).*

現在我們假設 $A \in \mathscr{M}_{n\times n}$ 為 regular stochastic matrix 且為 diagonalizable. 令 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 為 $\mathbb{R}^n$ 的一組 basis 且為 A 的 eigenvectors. 假設 $\mathbf{v}_i$ 所對應的 eigenvalue 為 λ_i, 其中 $\lambda_1 = 1$. 由 Proposition 9.6.8 我們可設 $\mathbf{v}_1$ 為 probability vector, 又對於 $i = 2, \ldots, n$, $|\lambda_i| < 1$ 故若 $\mathbf{v}_i = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ 我們有 $c_1 + \cdots + c_n = 0$. 現令 $P = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix}$, 我們得 $A = P\begin{bmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}P^{-1}$.

由於矩陣的乘法僅牽涉到每個 entry 之間的加法與乘法, 所以當取極限時可以分別取極限再乘在一起. 因此我們有

$$\lim_{k\to\infty} A^k = \lim_{k\to\infty} P\begin{bmatrix} 1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}^k P^{-1} = \lim_{k\to\infty} P\begin{bmatrix} 1 & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{bmatrix}P^{-1} = P\begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}P^{-1}.$$

注意由於 $\mathbf{v}_1$ 的每個 entry 之和為 1, 而對於 $i = 2, \ldots, n$, $\mathbf{v}_i$ 每一個 entry 之和為 0, 利用 $P^{-1}P = I_n$, 我們很容易看出 P^{-1} 的 1-st row 的每一個 entry 皆為 1. 故得

$$P\begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix}P^{-1} = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} 1 & \cdots & 1 \\ — & \mathbf{O} & — \\ & \vdots & \\ — & \mathbf{O} & — \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_1 & \cdots & \mathbf{v}_1 \\ | & | & & | \end{bmatrix}.$$

也就是說當 A 為 regular stochastic matrix 且為 diagonalizable, 則當 k 越來越大時 A^k 會趨近一個每個 column 皆為 $\mathbf{v}_1$ 的矩陣, 其中 $\mathbf{v}_1$ 為 $\mathbb{R}^n$ 中唯一的 probability vector 滿足 $A\mathbf{v}_1 = \mathbf{v}_1$. 這個結果其實對於一般的 regular stochastic matrix (不需 diagonalizable 之假設) 皆成立. 它的證明並沒有用到太多線性代數的技巧, 而著重於數列極限的估計. 不過為了完整性, 我們仍簡略的證明一下. 同學們可以忽略它的證明, 不過我們希望大家仍能充分了解此定理的敘述以及其衍伸的結果 (Corollary 9.6.10).

Theorem 9.6.9. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *regular stochastic matrix* 且 $\mathbf{v} \in \mathbb{R}^n$ 為唯一的 *probability vector* 滿足 $A\mathbf{v} = \mathbf{v}$. 則

$$\lim_{k\to\infty} A^k = \begin{bmatrix} | & | & & | \\ \mathbf{v} & \mathbf{v} & \cdots & \mathbf{v} \\ | & | & & | \end{bmatrix}.$$

Proof. 當 $n = 1$ 時 $A = [1]$ 是唯一的 stochastic matrix, 此時本定理自然成立, 所以我們考慮 $n \geq 2$ 的情形. 此時我們要證明 $\lim_{k\to\infty} A^k$ 的每一個 column 是相同的 vector. 也就是說我們要先證明存在 $\mathbf{w} \in \mathbb{R}^n$ 使得 $\lim_{k\to\infty} A^k = \begin{bmatrix} | & | & & | \\ \mathbf{w} & \mathbf{w} & \cdots & \mathbf{w} \\ | & | & & | \end{bmatrix}$. 若證明了這部分, 則因 $A\mathbf{v} = \mathbf{v}$, 我們有 $A^k\mathbf{v} = \mathbf{v}$, $\forall k \in \mathbb{N}$, 故得 $\lim_{k\to\infty}(A^k\mathbf{v}) = \mathbf{v}$. 然而若 $\mathbf{v} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$, 因 $\mathbf{v}$ 為 probability vector, 我們有 $c_1 + \cdots + c_n = 1$. 因此利用 $\lim_{k\to\infty} A^k = \begin{bmatrix} | & | & & | \\ \mathbf{w} & \mathbf{w} & \cdots & \mathbf{w} \\ | & | & & | \end{bmatrix}$. 得

$$\lim_{k\to\infty}(A^k\mathbf{v}) = (\lim_{k\to\infty} A^k)\mathbf{v} = \begin{bmatrix} | & | & & | \\ \mathbf{w} & \mathbf{w} & \cdots & \mathbf{w} \\ | & | & & | \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = c_1\mathbf{w} + c_2\mathbf{w} + \cdots + c_n\mathbf{w} = \mathbf{w}.$$

得證 $\mathbf{w} = \mathbf{v}$. 要如何證明 $\lim_{k\to\infty} A^k$ 的每一個 column 是相同的 vector 呢? 給定 $i \in \{1, \ldots, n\}$, 我們考慮所有 A^k 的 i-th row. 若令 M_k, m_k 分別為 A^k 的 i-th row 的 entry 中的最大值與最小值, 我們要證明 $\lim_{k\to\infty}(M_k - m_k) = 0$, 因此得證 $\lim_{k\to\infty} A^k$ 的 i-th row 的每個 entry 皆為同一個數. 因為這是對所有 $i = 1, \ldots, n$ 皆成立, 也因此得證 $\lim_{k\to\infty} A^k$ 的每一個 column 皆為相同的 vector.

我們令 a_{ij} 為 A 的 (i,j)-th entry 且當 $k > 1$ 時令 $a_{ij}^{(k)}$ 為 A^k 的 (i,j)-th entry. 依矩陣乘法的定義以及 $A^{k+1} = A^kA$, 我們有 $a_{ij}^{(k+1)} = \sum_{q=1}^n a_{iq}^{(k)} a_{qj}$. 假設 A^k 的 i-th row 的最大值和最小值分別發生在 A^k 的 (i,s)-th entry 以及 (i,t)-th entry (即 $M_k = a_{is}^{(k)}$, $m_k = a_{it}^{(k)}$), 又假設 A^k 的 i-th row 的最大值和最小值分別發生在 A^{k+1} 的 (i,j)-th entry 以及 (i,l)-th entry (即 $M_{k+1} = a_{ij}^{(k+1)}$, $m_{k+1} = a_{il}^{(k+1)}$). 此時我們有

$$M_{k+1} = a_{ij}^{(k+1)} = a_{it}^{(k)} a_{tj} + \sum_{q\neq t} a_{iq}^{(k)} a_{qj} \leq m_k a_{tj} + M_k \sum_{q\neq t} a_{qj} = m_k a_{tj} + M_k(1 - a_{tj}). \tag{9.11}$$

$$m_{k+1} = a_{il}^{(k+1)} = a_{is}^{(k)} a_{sl} + \sum_{q\neq s} a_{iq}^{(k)} a_{ql} \geq M_k a_{sl} + m_k \sum_{q\neq s} a_{ql} = M_k a_{sl} + m_k(1 - a_{sl}). \tag{9.12}$$

注意式子 (9.11), (9.12) 用到了 A 為 stochastic matrix 之假設, 即 $\sum_{q=1}^{n} a_{qj} = \sum_{q=1}^{n} a_{ql} = 1$. 現在將式子 (9.11) 減去式子 (9.12) 得

$$M_{k+1} - m_{k+1} \leq M_k(1 - a_{tj} - a_{sl}) + m_k(a_{tj} - (1 - a_{sl})) = (M_k - m_k)(1 - a_{tj} - a_{sl}). \tag{9.13}$$

由於 $a_{tj} \geq 0, a_{sl} \geq 0$, 我們有 $1 - a_{tj} - a_{sl} \leq 1$, 故由式子 (9.13) 得 $M_{k+1} - m_{k+1} \leq M_k - m_k$. 也就是說數列 $(M_k - m_k)_{k=1}^{\infty}$ 是一個遞減的數列.

如何證明 $(M_k - m_k)_{k=1}^{\infty}$ 這個遞減的數列其極限為 0 呢? 我們只要找到它的一個 subsequence 趨近於 0 即可. 這裡就需要用到 A 為 regular 的假設. 也就是說存在 $r \in \mathbb{N}$ 使得 A^r 的每個 entry 皆為正實數. 為了方便起見, 我們令 $B = A^r$ 且令 b_{ij} 為 B 的 (i,j)-th entry 且當 $k > 1$ 時令 $b_{ij}^{(k)}$ 為 $B^k = A^{rk}$ 的 (i,j)-th entry. 注意依假設我們有 $b_{ij} > 0$ 且 $b_{ij}^{(k)} > 0$. 又給定 $i \in \{1, \ldots, n\}$, 我們有 M_{rk}, m_{rk} 分別為 $B^k = A^{rk}$ 的 i-th row 的 entry 中的最大值與最小值. 因此用相同的方法, 套用式子 (9.11), (9.12) 在 B^k 和 B^{k+1} 的情形, 可以推得式子 (9.13) 在 B^k 與 B^{k+1} 相對應的結果

$$M_{r(k+1)} - m_{r(k+1)} \leq M_{rk}(1 - b_{tj} - b_{sl}) + m_{rk}(b_{tj} - (1 - b_{sl})) = (M_{rk} - m_{rk})(1 - b_{tj} - b_{sl}). \tag{9.14}$$

要注意式子 (9.14) 與式子 (9.13) 最大的不同處在於 $b_{tj} > 0$ 且 $b_{sl} > 0$. 現令 β 為 B 的所有 entry 中的最小值. 依 β 的定義, 我們有 $b_{tj} \geq \beta, b_{sl} \geq \beta$, 故得 $1 - b_{tj} - b_{sl} \leq 1 - 2\beta$. 因此由式子 (9.13) 得

$$M_{r(k+1)} - m_{r(k+1)} \leq (M_{rk} - m_{rk})(1 - 2\beta). \tag{9.15}$$

因 β 所在的那個 column 所有的 entry 之和等於 1 且大於等於 $n\beta$, 故得 $0 < \beta \leq \frac{1}{n} \leq \frac{1}{2}$ (注意我們有 $n \geq 2$ 之假設). 所以 $0 \leq 1 - 2\beta < 1$, 利用式子 (9.15) 以及數學歸納法得

$$M_{r(k+1)} - m_{r(k+1)} \leq (M_r - m_r)(1 - 2\beta)^k,$$

且因此得證 $(M_{rk} - m_{rk})_{k=1}^{\infty}$ 這個 subsequence 滿足 $\lim_{k\to\infty}(M_{rk} - m_{rk}) = 0$, 故得原 sequence $(M_k - m_k)_{k=1}^{\infty}$ 滿足 $\lim_{k\to\infty}(M_k - m_k) = 0$. 得證本定理. □

再次重申 Theorem 9.6.9 對於一般的 stochastic matrix 未必會成立, 須加上 regular 的條件才一定成立. 這個定理的最重要的結果就是在 Markov process 中不管初始的 probability vector 為何, 經過 regular stochastic matrix 多次的作用之下, 最後會趨近於一個穩定的狀態. 我們將此結果敘述如下.

Corollary 9.6.10. 假設 $A \in \mathscr{M}_{n\times n}$ 為 *regular stochastic matrix* 且 $\mathbf{v} \in \mathbb{R}^n$ 為唯一的 *probability vector* 滿足 $A\mathbf{v} = \mathbf{v}$. 則對於任意 $\mathbb{R}^n$ 的 *probability vector* $\mathbf{w}$ 皆有

$$\lim_{k\to\infty}(A^k\mathbf{w}) = \mathbf{v}.$$

Proof. 令 $\mathbf{w}=\begin{bmatrix}c_1\\ \vdots\\ c_n\end{bmatrix}$. 依 $\mathbf{w}$ 為 probability vector 之假設, 我們知 $c_1+\cdots+c_n=1$. 故由 Theorem 9.6.9 知

$$\lim_{k\to\infty}(A^k\mathbf{w})=(\lim_{k\to\infty}A^k)\mathbf{w}=\begin{bmatrix}| & | & & |\\ \mathbf{v} & \mathbf{v} & \cdots & \mathbf{v}\\ | & | & & |\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ \vdots\\ c_n\end{bmatrix}=c_1\mathbf{v}+c_2\mathbf{v}+\cdots+c_n\mathbf{v}=\mathbf{v}.$$

□

Question 9.11. *Corollary 9.6.10 中* $\mathbf{w}$ *需要假設是 probability vector 才會成立嗎?*

9.7. 結論

從 linear operator 的觀念來看 eigenvector 就是經過 linear operator 運算後仍保持平行的向量. 所以若有一組 basis 是由 eigenvectors 所組成, 則此 linear operator 利用這組 ordered basis 所得的 matrix representation 就會是 diagonal matrix. 這種情況之下便會讓我們很容易掌握這個 linear operator (特別是對其多次重複合成的情況之下). 因此要了解一個 square matrix 是否可以對角化, 是一個重要的課題. 利用 eigenvalue 的 algebraic multiplicity 以及 geometric multiplicity, 可以幫助我們判別一個 square matrix 是否為 diagonalizable. 有一種特別的情形是不必計算就可以知道可以對角化, 就是 symmetric matrix. 事實上 symmetric matrix 不只是 diagonalizable, 其實它還是 orthogonal diagonalizable. 也就是說 symmetric matrix 都會存在一組由 eigenvectors 所組成的 orthogonal basis. 這就是所謂的 Spectral Theorem, 它在數學和物理上有許多的應用. 我們特別利用它, 可將坐標平面和坐標空間的二次方程式, 利用坐標變換 (旋轉和平移) 方法將之變換成標準式, 以便判斷其圖形.