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

$$\lambda_1 \bar{x}^2 + \lambda_2 \bar{y}^2 + \lambda_3 \bar{z}^2 + c'\bar{x} + d'\bar{y} + d'\bar{z} + f = 0. \tag{8.8}$$

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一般書籍上的圖形.

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 (8.8) 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 (8.7) 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 *hyperbolic 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 (8.7) 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 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 8.6. 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?

Exercise 8.13. 將以下的三元二次方程式經由坐標變換寫成標準式（中心在原點）。需寫下坐標如何變換。

(1) $3x^2 + 2xy + 2xz + 4yz = 4$.

(2) $2x^2 + 2xy + 2xz + 2yz - x + y + z = 1$.

(3) $3x^2 + 2xz - y^2 + 3z^2 + 2y = 0$.

Exercise 8.14. 套用 Exercise 8.13 的結果，寫下以下三元二次方程式的圖形之相關參數。

(1) $3x^2 + 2xy + 2xz + 4yz = 4$. 說明此曲面之名稱。求中心軸的參數式，以及一組平面方程式使得圖形與平面之截痕為橢圓。

(2) $2x^2 + 2xy + 2xz + 2yz - x + y + z = 1$. 說明此曲面之名稱。求鞍點坐標、以及平面方程式使得圖形與平面之截痕為兩直線。

(3) $3x^2 + 2xz - y^2 + 3z^2 + 2y = 0$. 說明此曲面之名稱。求 2 個頂點之坐標，以及過頂點的切平面方程式。

Exercise 8.15. 閱讀以下附錄：“Second Derivative Test”。

附錄：Second Derivative Test

本單元介紹對稱矩陣可對角化應用在微積分中多變數函數的極值問題。因為本課程並不假設微積分的基本知識，所以置於附錄且微積分的部分我們不去證明，只探討如何用對角化（eigenvalues）來判定是否有極值。

我們先看兩個變數的實函數。設 $f:\mathbb{R}^2 \to \mathbb{R}$ 為在 $p = (a_1, a_2)$ 附近可無限次任意偏微的兩個變數實函數。我們有 $f(x_1, x_2)$ 在 p 附近的 Tayler 展開式：

$$f(x_1,x_2) = f(a_1,a_2) + \frac{\partial f}{\partial x_1}(a_1,a_2)(x_1-a_1) + \frac{\partial f}{\partial x_2}(a_1,a_2)(x_2-a_2) +$$
$$\frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2}(a_1,a_2)(x_1-a_1)^2 + 2\frac{\partial^2 f}{\partial x_2 \partial x_1}(a_1,a_2)(x_1-a_1)(x_2-a_2) + \frac{\partial^2 f}{\partial x_2^2}(a_1,a_2)(x_2-a_2)^2\right) + \cdots$$

其中 $\cdots$ 表示更高次項，也因此當 (x_1,x_2) 很接近 (a_1,a_2) 時，這部分的取值是會遠小於前面一次項與二次項的取值。由於唯有當 p 點為 $f(x_1,x_2)$ 的 critical point 時，$f(x_1,x_2)$ 才有可能在 p 為極值。因此我們僅考慮

$$\frac{\partial f}{\partial x_1}(a_1,a_2)=\frac{\partial f}{\partial x_2}(a_1,a_2)=0$$

的情形。也就是說此時當 (x_1,x_2) 很接近於 (a_1,a_2)，要判斷 $f(x_1,x_2)$ 是否會大於 $f(a_1,a_2)$ 或小於 $f(a_1,a_2)$ 完全取決於

$$\frac{\partial^2 f}{\partial x_1^2}(a_1,a_2)(x_1-a_1)^2+2\frac{\partial^2 f}{\partial x_2\partial x_1}(a_1,a_2)(x_1-a_1)(x_2-a_2)+\frac{\partial^2 f}{\partial x_2^2}(a_1,a_2)(x_2-a_2)^2 \tag{8.9}$$

是正的還是負的。

利用如同探討 quadratic form 的看法，我們可以考慮矩陣

$$H=\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a_1,a_2) & \frac{\partial^2 f}{\partial x_2\partial x_1}(a_1,a_2) \\ \\ \frac{\partial^2 f}{\partial x_2\partial x_1}(a_1,a_2) & \frac{\partial^2 f}{\partial x_2^2}(a_1,a_2) \end{pmatrix}.$$

H 稱為 $f(x_1,x_2)$ 在點 $p=(a_1,a_2)$ 的 *Hessian matrix.* 我們可以利用矩陣將式子 (1) 寫成

$$\begin{pmatrix} x_1-a_1 & x_2-a_2 \end{pmatrix} H \begin{pmatrix} x_1-a_1 \\ x_2-a_2 \end{pmatrix}. \tag{8.10}$$

由於 H 為對稱矩陣，我們知道存在 orthogonal matrix U 使得 $U^{\mathrm{t}}HU=D$ 其中 D 為對角矩陣其對角線位置分別為 H 的 eigenvalue λ_1,λ_2。考慮變換變數 $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=U^{\mathrm{t}}\begin{pmatrix} x_1-a_1 \\ x_2-a_2 \end{pmatrix}$。由於 $\begin{pmatrix} x_1-a_1 \\ x_2-a_2 \end{pmatrix}=U\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$，我們又可以將式子 (2) 寫成

$$\begin{pmatrix} y_1 & y_2 \end{pmatrix} U^{\mathrm{t}}HU \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}=\begin{pmatrix} y_1 & y_2 \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \tag{8.11}$$

也就是說當 (x_1,x_2) 很接近 (a_1,a_2) 時，$f(x_1,x_2)-f(a_1,a_2)$ 會與 $\lambda_1 y_1^2+\lambda_2 y_2^2$ 同號。注意，利用內積計算 (y_1,y_2) 的長度，我們會有 $\|(y_1,y_2)\|=\|(x_1-a_1,x_2-a_2)\|$。也就是說考慮點 (x_1,x_2) 在 p 附近，就等同於考慮 (y_1,y_2) 在原點 $(0,0)$ 附近。

我們很容易看出當 λ_1,λ_2 皆為正時，若 $(y_1,y_2)\neq(0,0)$ 則 $\lambda_1 y_1^2+\lambda_2 y_2^2$ 也是正的。也就是說此時在 $p=(a_1,a_2)$ 附近的點 (x_1,x_2) 皆滿足 $f(x_1,x_2)>f(a_1,a_2)$。因此知此時 $f(x_1,x_2)$ 在點 $p=(a_1,a_2)$ 有極小值 $f(a_1,a_2)$。同理，當 λ_1,λ_2 皆為負時，$f(x_1,x_2)$ 在點 p 有極大值 $f(a_1,a_2)$。而當 λ_1,λ_2 一正一負時，$f(x_1,x_2)$ 在點 p 沒有極值，事實上 p 會是 $f(x_1,x_2)$ 的 *saddle point*。而當 λ_1,λ_2 其中有一個是 0 時，就無法判定是否有極值了。這很容易由 $f(x_1,x_2)=\lambda x_1^2+rx_2^4$ 這個例子看出，不管 λ,r 為何，$f(x_1,x_2)$ 在原點 $(0,0)$ 的 Hessian matrix 的 eigenvalues 皆為 λ 與 0。而當 λ,r 同號時 $f(x_1,x_2)$ 在 $(0,0)$ 有極值，但當 λ,r 異號時就沒有極值了。

我們可以驗證，上述這種用 Hessian matrix 的 eigenvalues 來判斷一個兩個變數的實函數在其 critical point 是否有極值的方法，與微積分中所介紹（用判別式）的方法其實是等價的。不過這裡所介紹的方法其優點就是很容易推廣到更多變數的情況。

令 $f:\mathbb{R}^n\to\mathbb{R}$ 為定義在 $\mathbb{R}^n$ 中可無限次任意偏微的 n 個變數實函數。對於 $f(x_1,\ldots,x_n)$ 上的一個 critical point $p=(a_1,\ldots,a_n)$ (即 $\dfrac{\partial f}{\partial x_i}(a_1,\ldots,a_n)=0,\ \forall i=1,\ldots,n)$，令 $n\times n$ 矩陣 $H=(h_{ij})$，其 (i,j)-th entry 是

$$h_{ij}=\frac{\partial^2 f}{\partial x_i\partial x_j}(a_1,\ldots,a_n)$$

為 $f(x_1,\ldots,x_n)$ 在點 p 的 Hessian matrix. 由於 H 為對稱矩陣 (因為 f 的假設會使得 $h_{ij}=h_{ji}$)，故 H 為 orthogonal diagonalizable, 也因此存在 $\lambda_1,\ldots,\lambda_n\in\mathbb{R}$ 為 H 的 eigenvalues (包含重根)。利用與兩個變數函數相同的論述，我們有以下的結果。

(1) 若 $\lambda_1,\ldots,\lambda_n$ 皆為正實數，則 $f(x_1,\ldots,x_n)$ 在 $p=(a_1,\ldots,a_n)$ 點有極小值 $f(a_1,\ldots,a_n)$.

(2) 若 $\lambda_1,\ldots,\lambda_n$ 皆為負實數，則 $f(x_1,\ldots,x_n)$ 在 $p=(a_1,\ldots,a_n)$ 點有極大值 $f(a_1,\ldots,a_n)$.

(3) 若 $\lambda_1,\ldots,\lambda_n$ 有正也有負，則 $f(x_1,\ldots,x_n)$ 在 p 點無極值.

(4) 若 $\lambda_1,\ldots,\lambda_n$ 其中有 0，則 $f(x_1,\ldots,x_n)$ 在 p 點無法判定是否有極值.

8.4. Application: Singular Value Decomposition

AI 的運用需要大量的資訊處理，Singular Value Decomposition 是幫助大量資訊處理重要的方法。本節我們利用 Spectral Theorem 說明如何得到一個矩陣的 singular value decomposition. 然後再說明相關定義及其概念。

對於任意矩陣 $A\in\mathsf{M}_{m\times n}(\mathbb{R})$，我們考慮 $A^{\mathrm{t}}A\in\mathsf{M}_{n\times n}(\mathbb{R})$ 這一個對稱矩陣。由 spectral theorem (Theorem 8.2.6)，我們知存在 $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 為 $\mathbb{R}^n$ 的一組 orthonormal basis 且為 $A^{\mathrm{t}}A$ 的 eigenvectors。現假設 $\mathbf{v}_i$ 所對應的 eigenvalue 為 λ_i，亦即 $(A^{\mathrm{t}}A)\mathbf{v}_i=\lambda_i\mathbf{v}_i$。這些 eigenvectors 和 eigenvalues 有許多有趣的性質。

首先利用矩陣運算結合律，我們有

$$\langle A^{\mathrm{t}}(A\mathbf{v}_i),\mathbf{v}_j\rangle=\langle (A^{\mathrm{t}}A)\mathbf{v}_i,\mathbf{v}_j\rangle=\lambda_i\langle\mathbf{v}_i,\mathbf{v}_j\rangle,\ \forall i,j\in\{1,\ldots,n\} \tag{8.12}$$

另一方面，利用矩陣與內積的關係，得

$$\langle A^{\mathrm{t}}(A\mathbf{v}_i),\mathbf{v}_j\rangle=\langle A\mathbf{v}_i,A\mathbf{v}_j\rangle,\ \forall i,j\in\{1,\ldots,n\} \tag{8.13}$$

因此，當 $i=j$ 時，由於 $\langle\mathbf{v}_i,\mathbf{v}_i\rangle=\|\mathbf{v}_i\|^2=1$，式子 (8.12), (8.13) 告訴我們

$$\lambda_i=\langle A\mathbf{v}_i,A\mathbf{v}_i\rangle=\|A\mathbf{v}_i\|^2\geq 0,$$

而且 $\lambda_i=0$ 若且唯若 $A\mathbf{v}_i=\mathbf{0}$。簡單來說：矩陣 $A^{\mathrm{t}}A$ 的 eigenvalues 一定是非負實數。若計算代數重根數，到底 $A^{\mathrm{t}}A$ 有幾個 eigenvalues 會是正的呢？因為 $A^{\mathrm{t}}A$ 可對角化，其 eigenvalue 0 的代數重根數等於其幾何重根數，即 0 的代數重根數為 $\dim(N(A^{\mathrm{t}}A))=n-\operatorname{rank}(A^{\mathrm{t}}A)$。因為 $\operatorname{rank}(A^{\mathrm{t}}A)=\operatorname{rank}(A)$ (參見 Corollary 4.4.5)，所以若 $\operatorname{rank}(A)=r$，則 n 階方陣 $A^{\mathrm{t}}A$ 會有 $n-(n-r)=r$ 個正的 eigenvalue λ_i。一般我們會將 $A^{\mathrm{t}}A$ 的 eigenvectors 所形成的 orthonormal basis $\mathbf{v}_1,\ldots,\mathbf{v}_n$ 重排，使得其對應的 eigenvalues 滿足 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_r>0$ 且 $\lambda_{r+1}=\cdots=\lambda_n=0$。由此設定，我們知道當 $i\leq r$ 時，$A\mathbf{v}_i$ 為非零向量且長度為 $\sqrt{\lambda_i}$；而當 $r<i\leq n$ 時 $A\mathbf{v}_i=\mathbf{0}$。接下來我們對 $A\mathbf{v}_1,\ldots,A\mathbf{v}_r$ 這 r 個向量做進一步探討。

回到式子 (8.12), (8.13)，當 $i \neq j$ 時，由於 $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$，我們得 $\langle A\mathbf{v}_i, A\mathbf{v}_j \rangle = 0$。也就是 $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ 是兩兩互相垂直的向量。現令 $\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i$，則 $\mathbf{u}_1, \ldots, \mathbf{u}_r$ 為 $\mathbb{R}^m$ 中長度為 1 且兩兩互相垂直的向量，此時我們又可以利用 Gram-Schmidt 將之擴展程 $\mathbb{R}^m$ 的一組 normal basis $\mathbf{u}_1, \ldots, \mathbf{u}_r, \mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$。

通常我們令 $\sigma_i = \sqrt{\lambda_i}$，也因此 $\mathbf{u}_i$ 會滿足 $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$。注意 $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$ 不只對 $1 \leq i \leq r$ 時成立，當 $r+1 \leq i \leq n$ 時，也會成立，因為此時 $\sigma_i = 0$ 且 $A\mathbf{v}_i = \mathbf{0}$。所以若令

$$V = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} \in \mathsf{M}_{n\times n}(\mathbb{R}), \quad U = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \\ | & | & & | \end{bmatrix} \in \mathsf{M}_{m\times m}(\mathbb{R})$$

以及 $\Sigma = [s_{ij}] \in \mathsf{M}_{m\times n}(\mathbb{R})$，其中 $s_{ij} = \begin{cases} \sigma_i, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$ 亦即

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & \mathbf{0} \\ & & \sigma_r & \\ \hline & \mathbf{0} & & \mathbf{0} \end{array}\right] \in \mathsf{M}_{m\times n}(\mathbb{R}).$$

則依矩陣乘法定義，我們有 $AV = U\Sigma$。注意 U, V 為 orthogonal matrix，而 Σ 為 (i,i)-th entry 外皆為 0 的 $m \times n$ matrix。因為 $V^{-1} = V^{\mathrm{t}}$，我們可以將 $AV = U\Sigma$ 寫成

$$A = U\Sigma V^{\mathrm{t}}.$$

這就是矩陣 A 的 *singular value decomposition*，以下簡稱為 svd. 之後我們會討論若限定 σ_i 為遞減，則這些 σ_i 是唯一的，我們稱之為 A 的 *singular value*, 而 $\mathbf{v}_i$、$\mathbf{u}_i$ 分別稱為 A 的 *right singular vector*、*left singular vector*。注意 singular vectors 和 eigenvectors 一樣，並不唯一。

當 A 為方陣時，不管 A 是否可對角化，$A = U\Sigma V^{\mathrm{t}}$ 中的 Σ 為對角矩陣，或許會讓人以為 svd 是對角化的推廣。不過這樣的說法，不太正確，因為若 A 可對角化，其對角化後的對角矩陣 D 有可能和 Σ 不同。事實上當 $\mathbf{w} \neq \mathbf{0}$ 且 $A\mathbf{w} = \lambda\mathbf{w}$，我們立即知 $\mathbf{w}$ 為 A 的 eigenvector 且 λ 為其 eigenvalue；但對任意 $\mathbf{v}$ 我們都可寫成 $A\mathbf{v} = \sigma\mathbf{u}$ 的形式，不能由單一情況就稱 σ 為 singular value，$\mathbf{v}, \mathbf{u}$ 為 singular vector。必需整體 $\mathbf{v}_1, \ldots, \mathbf{v}_n$ 和 $\mathbf{u}_1, \ldots, \mathbf{u}_m$ 皆為 orthonormal basis 才可確認 singular value 和 singular vector；或許說 svd 是正交對角化的推廣比較合適（留在習題討論）。以下我們看一些簡單的例子。

Example 8.4.1. (1) 考慮 $A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$。注意 A_1 只有一個 eigenvalue 0 但其幾何重根數為 1 小於代數重根數 2，所以 A_1 不能對角化。然而 $A_1^{\mathrm{t}} A_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ 可正交對角化，其中 $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$、$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 為 eigenvectors 且其 eigenvalue 分別為 1、0。故得 $V = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$、$\sigma_1 = 1$, $\sigma_2 = 0$ 且 $\mathbf{u}_1 = A_1\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$。由於 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 與 $\mathbf{u}_1$ 形成 $\mathbb{R}^2$ 的一組 orthonormal basis，所以我們可以選 $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$，得 $A_1 = U\Sigma V^{\mathrm{t}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ 為 A_1 的 singular value decomposition. 注意當初 $\mathbf{v}_1$ 可以選 $\mathbf{v}_1=\begin{bmatrix}0\\-1\end{bmatrix}$，不過此時 $\mathbf{u}_1$ 就應為 $\begin{bmatrix}0&1\\0&0\end{bmatrix}\begin{bmatrix}0\\-1\end{bmatrix}=\begin{bmatrix}-1\\0\end{bmatrix}$，故 $\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}0&1\\-1&0\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}0&-1\\1&0\end{bmatrix}$ 也是 A_1 的 svd. 另一方面，$\mathbf{v}_2,\mathbf{u}_2$ 差個負號也可以，而且它們互不影響，所以

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

也都是 A_1 的 svd.

(2) 考慮 $A_2=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$。因為 A_2 的 characteristic polynomial 為 t^2+1 不可分解，所以 A_2 不能對角化。然而 A_2 本身是 orthogonal matrix，$A_2^{\mathrm{t}}A_2=\begin{bmatrix}1&0\\0&1\end{bmatrix}$。由於任意 $\mathbb{R}^2$ 中的非零向量都是 I_2 的 eigenvector 且其 eigenvalue 為 1，所以 $\sigma_1=\sigma_2=1$，亦即 $\Sigma=I_2$。而且任意二階的 orthogonal matrix 都可以當成 V。例如 $\mathbf{v}_1=\begin{bmatrix}0\\1\end{bmatrix}$、$\mathbf{v}_2=\begin{bmatrix}1\\0\end{bmatrix}$ 則 $\mathbf{u}_1=A_2\mathbf{v}_1=\begin{bmatrix}-1\\0\end{bmatrix}$ 且 $\mathbf{u}_2=A_2\mathbf{v}_2=\begin{bmatrix}0\\1\end{bmatrix}$。得 $A_2=U\Sigma V^{\mathrm{t}}=\begin{bmatrix}-1&0\\0&1\end{bmatrix}\begin{bmatrix}1&0\\0&1\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}$ 為 A_2 的 svd. 我們也可選 $V=I_2$ 或 $V=A_2$ 本身，所以 $A_2=A_2I_2I_2$ 或 $A_2=A_2^2I_2A_2^{\mathrm{t}}$ 也都是 A_2 的 svd. 一般來說，若 $Q\in\mathsf{M}_{n\times n}(\mathbb{R})$ 為 orthogonal matrix，任選另一個 orthogonal matrix $V\in\mathsf{M}_{n\times n}(\mathbb{R})$，則 $U=QV$, $\Sigma=I_n$，亦即 $Q=(QV)I_nV^{\mathrm{t}}$ 就會是 Q 的 svd.

(3) 考慮 $A_3=\begin{bmatrix}0&1\\0&1\end{bmatrix}$。因為 A_3 有兩個相異的 eigenvalue 1,0，所以可以對角化成

$$A_3=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}1&1\\1&0\end{bmatrix}^{-1}=\begin{bmatrix}1&1\\1&0\end{bmatrix}\begin{bmatrix}1&0\\0&0\end{bmatrix}\begin{bmatrix}0&1\\1&-1\end{bmatrix}.$$

然而，$A_3^{\mathrm{t}}A_3=\begin{bmatrix}0&0\\0&2\end{bmatrix}$ 有互相垂直的 eigenvector $\mathbf{v}_1=\begin{bmatrix}0\\1\end{bmatrix}$, $\mathbf{v}_2=\begin{bmatrix}1\\0\end{bmatrix}$ 其 eigenvalue 分別為 2, 0，故得 $V=\begin{bmatrix}0&1\\1&0\end{bmatrix}$、$\sigma_1=\sqrt{2}$, $\sigma_2=0$ 且 $\mathbf{u}_1=\frac{1}{\sigma_1}A_3\mathbf{v}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$。另取 $\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\end{bmatrix}$ 與 $\mathbf{u}_1$ 形成 $\mathbb{R}^2$ 的一組 orthonormal basis，所以我們可以選 $U=\frac{1}{\sqrt{2}}\begin{bmatrix}1&-1\\1&1\end{bmatrix}$ 得

$$A_3=U\Sigma V^{\mathrm{t}}=\begin{bmatrix}\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{2}&0\\0&0\end{bmatrix}\begin{bmatrix}0&1\\1&0\end{bmatrix}$$

為 A_3 的 svd. 注意 A_3 的對角化和 svd 有差異。

(4) 考慮 $A_4=\begin{bmatrix}1&2\\2&1\end{bmatrix}$。因為 A_4 為對稱矩陣，有互相垂直的 eigenvector $\mathbf{v}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$ 其 eigenvalue 分別為 3, -1，所以可以正交對角化成

$$A_4=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}3&0\\0&-1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}^{\mathrm{t}}=\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}3&0\\0&-1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}.$$

然而，$A_4^{\mathrm{t}}A_4=\begin{bmatrix}5&4\\4&5\end{bmatrix}$，和 A_4 有相同的 eigenvector $\mathbf{v}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$ 但其 eigenvalue 分別為 9, 1，故得 $V=\frac{1}{\sqrt{2}}\begin{bmatrix}1&1\\1&-1\end{bmatrix}$、$\sigma_1=\sqrt{9}=3,\ \sigma_2=1$ 且

$$\mathbf{u}_1=\frac{1}{\sigma_1}A_4\mathbf{v}_1=\frac{1}{3}\begin{bmatrix}\frac{3}{\sqrt{2}}\\\frac{3}{\sqrt{2}}\end{bmatrix}=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix},\quad \mathbf{u}_2=\frac{1}{\sigma_2}A_4\mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\end{bmatrix}.$$

故取 $U=\frac{1}{\sqrt{2}}\begin{bmatrix}1&-1\\1&1\end{bmatrix}$ 得

$$A_4=U\Sigma V^{\mathrm{t}}=\begin{bmatrix}\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}3&0\\0&1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}.$$

為 A_4 的 svd. 注意 A_4 的對角化和 svd 的差異就在 eigenvalue 為負的情況及其對應的 eigenvector。對稱矩陣的 svd 其實和其正交對角化關係密切（留在習題討論）。

(5) 最後看一個不是方陣的例子。考慮 $B=\begin{bmatrix}1&1\\1&-1\\1&1\end{bmatrix}$，當然無法對角化。然而 $B^{\mathrm{t}}B=\begin{bmatrix}3&1\\1&3\end{bmatrix}$ 有互相垂直的 eigenvector $\mathbf{v}_1=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix}$, $\mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\end{bmatrix}$ 且其對應的 eigenvalue 分別為 4, 2。故得 $V=\frac{1}{\sqrt{2}}\begin{bmatrix}1&1\\1&-1\end{bmatrix}$、$\sigma_1=2,\ \sigma_2=\sqrt{2}$ 且

$$\mathbf{u}_1=\frac{1}{\sigma_1}B\mathbf{v}_1=\frac{1}{2}\begin{bmatrix}\sqrt{2}\\0\\\sqrt{2}\end{bmatrix},\quad \mathbf{u}_2=\frac{1}{\sigma_2}B\mathbf{v}_2=\frac{1}{\sqrt{2}}\begin{bmatrix}0\\\sqrt{2}\\0\end{bmatrix}=\begin{bmatrix}0\\1\\0\end{bmatrix}.$$

由於 U 必須由 $\mathbb{R}^3$ 上的 orthonormal basis 所組成，所以我們利用 Gram-Schmidt 找到 $\mathbf{u}_3=\frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}$，得 $U=\frac{1}{\sqrt{2}}\begin{bmatrix}1&0&1\\0&\sqrt{2}&0\\1&0&-1\end{bmatrix}$。因此

$$B=U\Sigma V^{\mathrm{t}}=\begin{bmatrix}\frac{1}{\sqrt{2}}&0&\frac{1}{\sqrt{2}}\\0&1&0\\\frac{1}{\sqrt{2}}&0&-\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}2&0\\0&\sqrt{2}\\0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}.$$

為 B 的 svd。注意：因為 B 為 3×2 矩陣，所以 U 應為 3×3、Σ 應為 3×2 且 V 應為 2×2 矩陣。 $\sharp$

Exercise 8.16. 考慮矩陣 $B'=\begin{bmatrix}1&1&1\\1&-1&1\end{bmatrix}$, $C=\begin{bmatrix}1&1&1\\1&0&-1\\1&-2&1\\1&1&1\end{bmatrix}$.

(1) 試分別寫下 B',C 的一組 svd.

(2) 比較 B' 以及 Example 8.4.1(5) 的 B 的關係，並說明它們 svd 的關係。

(3) 試寫下 $C'=\begin{bmatrix}1&1&1&1\\1&0&-2&1\\1&-1&1&1\end{bmatrix}$ 的一組 svd.

Exercise 8.17. 假設 $Q \in \mathsf{M}_{n\times n}(\mathbb{R})$ 是一個 orthogonal matrix.

(1) 試說明 Q 的任一組 svd，$Q = U\Sigma V^{\mathrm{t}}$ 中 Σ 為何？

(2) 說明對任意 orthogonal matrix $V \in \mathsf{M}_{n\times n}(\mathbb{R})$，皆可找到 $U \in \mathsf{M}_{n\times n}(\mathbb{R})$ 為 orthogonal matrix 使得 $Q = U\Sigma V^{\mathrm{t}}$ 為 Q 的一組 svd.

（參見 Example 8.4.1 (2)）

Exercise 8.18. 假設 $S \in \mathsf{M}_{n\times n}(\mathbb{R})$ 是一個 symmetric matrix，且將 S 正交對角化成 $Q^{\mathrm{t}}SQ = D$，其中 Q 為 orthogonal matrix, D 為 diagonal matrix，且 S 的 eigenvalues $\lambda_1, \dots, \lambda_n$ 依 $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ 方式排列，使得 D 的 (i,i)-th entry 為 λ_i。

(1) 試說明 S 的 singular value σ_i 與 eigenvalue λ_i 之間的關係。

(2) 若考慮 S 的一組 svd，$S = U\Sigma V^{\mathrm{t}}$ 其中 $V = Q$，試依 σ_i 與 λ_i 的關係，說明 U 的 i-th column $\mathbf{u}_i$ 與 V 的 i-th column $\mathbf{v}_i$ 之間的關係。

（參見 Example 8.4.1 (4)）

Exercise 8.19. 設 $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$ 皆為非零 column vector，且令 $A = \mathbf{u}\mathbf{v}^{\mathrm{t}} \in \mathsf{M}_{n\times n}(\mathbb{R})$（回顧：此為 outer product, $\operatorname{rank}(A) = 1$）。

(1) 說明 $A^{\mathrm{t}}A$ 有唯一的非零 eigenvalue λ_1 且其對應的 eigenvector 為 $\mathbf{v}$，並依此寫下 A 的唯一非零 singular value σ_1。

(2) 說明若 $A = U\Sigma V^{\mathrm{t}}$ 為 A 的一組 svd，則 U, V 的 1-st column $\mathbf{u}_1, \mathbf{v}_1$ 和 $\mathbf{u}, \mathbf{v}$ 的關係。

（參見 Example 8.4.1 (1) $A_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}$；(3) $A_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}$.）

Exercise 8.20. 令 $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ 且 σ_1 為 A 最大的 singular value.

(1) 證明對任意 $\mathbf{v} \in \mathbb{R}^n$ 皆滿足 $\|A\mathbf{v}\| \leq \sigma_1 \|\mathbf{v}\|$。(Hint: 將 $\mathbf{v}$ 寫成 orthogonal matrix V 的 column vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ 的線性組合)

(2) 假設 A 為方陣（即 $m = n$），證明 A 的 eigenvalue λ 皆滿足 $|\lambda| \leq \sigma_1$。