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。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：

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

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

$$\frac{\partial^2 f}{\partial x_1^2}(a_1,a_2)(x_1-a_1)^2+2\frac{\partial^2 f}{\partial x_2\partial x_1}(a_1,a_2)(x_1-a_1)(x_2-a_2)+\frac{\partial^2 f}{\partial x_2^2}(a_1,a_2)(x_2-a_2)^2 \tag{1}$$

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

$$\begin{pmatrix} x_1-a_1 & x_2-a_2\end{pmatrix} H \begin{pmatrix} x_1-a_1\\ x_2-a_2\end{pmatrix}. \tag{2}$$

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

$$\begin{pmatrix} y_1 & y_2\end{pmatrix} U^{\mathrm{t}}HU\begin{pmatrix} y_1\\ y_2\end{pmatrix}=\begin{pmatrix} y_1 & y_2\end{pmatrix}\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}\begin{pmatrix} y_1\\ y_2\end{pmatrix}. \tag{3}$$

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.