Optimization Theory

Hao-Che (Howard) Hsu

Optimization Theory[†]

優化理論

優化理論又被稱為「最佳化理論」, 主要應用在處理目標函數 (objective function) 在約束條件下的極大或極小值。因為數學的中文稱法因地制宜，故本文將以英文描寫優化裡論中的細節。

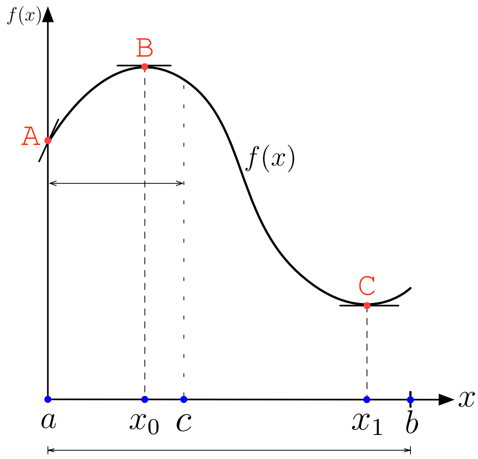
大家在微積分 (Calculus) 或分析 (高微) 中碰到的第一個優化問題，便是求函數極大極小值，這個過程又被稱為求一階條件 (First order condition, FOC)。

定理 1. *Let $f : (a, b) \to \mathbb{R}$ and assume f has a local max (or local min) at $x_0 \in (a, b)$. If f is differentiable at x_0, then $f'(x_0) = 0$.*

Proof. By definition $\lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{x\to x_0^-} \frac{f(x)-f(x_0)}{x-x_0} = \lim_{x\to x_0^+} \frac{f(x)-f(x_0)}{x-x_0}$.
Since $f'(x_0) \geq 0$ and $f'(x_0) \leq 0$, then $f'(x_0) = 0$. ■

極值 (critical value) 發生的地方，就是函數切線斜率為 0 的點。但是這個極大或極小值並不代表此函數在整個定義域 (domain) 中的最大或最小 (global max/min)，我們只能說他是局部 (local) 極值。一階條件的應用，甚至還可以應用在更廣泛的地方。目前，我們的目標都是建立在「函數」上，但是這種優化也可以處理非函數型態[1]。

圖 1: 函數極值的內點解和角解



從圖 1，我們可以看到函數 $f(x)$ 在 a, b 兩點之間有兩個極值，分別為局部極大值 $B(x = x_0)$ 及局部極小值 $C(x = x_1)$，這時的兩個局部極值也是全局的極大、極小值，稱 x_0 為 maximizer、x_1 為 minimizer。同時因為 x_0, x_1 在兩個邊界點 (a,b) 之間，所以又稱這兩點為內點解 (interior solution)。

[†]These notes are written by Hao-Che Hsu(徐晧哲) with XƎLaTeX and serve as a supplementary material for undergraduate Economics Theory. The author thanks Professor Kang Liu for useful comments. Reference can be found at *Sundaram, R., "A First Course in Optimization Theory," Cambridge University Press, Cambridge, 1996*, and *Simon, C. and L. Blume, "Mathematics for Economists," W.E. Norton & Co., New York, 1994*. Materials also have been partially taken from Professor Raymond Deneckere's Ph.D. lecture at University of Wisconsin-Madison. All errors and omissions belong to the author. Comments Welcome. Finally, this manuscript may be printed and reproduced for individual or instructional use, but may not be printed for commercial purposes.

[1]詳細內容請參考後面會提到的「隱函數定理」, Implicit Function Theorem。

但若這時的定義域是在 a, c 之間，則 B 一樣是極大值，但這時的極小值卻是在 A 點。當極值存在的點斜率不為 0 時，我們稱此點 (a) 為角解 (corner solution)。在一般的經濟決策狀況，角點解的情況較少見，最常見的情況為廠商選擇不生產或消費者選擇什麼都不買 (outside option) 及有無限的資源可以讓消費者無限購買，因此我們將以討論內點解為主。內點解可以透過一階條件尋得，但刻劃其為極大或極小，卻須透過二階條件 (Second order condition，SOC) 來判斷。

圖 2: 極大值的充分及必要條件關係圖

$$f''(x^*) < 0 \longrightarrow \text{Maximum} \longrightarrow f''(x^*) \leq 0$$

以極大值為例[2]，其充分條件 (sufficient condition) 為 $f''(x^*) < 0$，而必要條件 (necessary condition) 則為 $f''(x^*) \leq 0$。接下來將介紹多變量 (multivariable) 的一階及二階條件。首先，我們要先區分三種函數形式。第一種為純量函數 (scalar funtion)，他的映射 (mapping) 條件為 $f : \mathbb{R} \to \mathbb{R}$，例如 $f(x) = x^2$。第二種為多變量函數 (multivariate function)，$f : \mathbb{R}^n \to \mathbb{R}$，例如 $f(x_1, x_2, x_3) = \sqrt{x_1^2 + x_2^2 + x_3^2}$。第三種為向量函數 (vector function)，$f : \mathbb{R}^n \to \mathbb{R}^m$，例如衡量二維空間中物體的速度為 $(a_1, a_2, a_3) \to (f_x, f_y)$[3]$= (a_1, \sqrt{a_2 + a_3^2})$，就是 $\mathbb{R}^3 \to \mathbb{R}^2$。假設 x 為一向量，$x = \{x_1, x_2, x_3, ..., x_n\}$，這些變量共同刻劃函數 $F(x_1, ..., x_n)$。例如 F 是一件商品，x 向量則是價格、顏色、大小、廠牌等特徵。將函數進行偏微分[4](降維)，能得到許多有用的性質：

Gradient(梯度) 假設 $F(x_1, ..., x_n)$，我們以 ∇F(The gradient of F) 來描述多變量函數 F 的梯度:

$$\nabla F = DF = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, ..., \frac{\partial F}{\partial x_n} \right) \tag{1}$$

也就是 F 函數的一階條件[5]。另外，∇F 所指的方向，就是 F 函數沿著 x「增長」最快的方向。

Jocobian Matrix(雅可比矩陣)

延續前面的例子，物體在平面上的速度 $(a_1, a_2, a_3) \to (f_x, f_y)$ 是一個從 $\mathbb{R}^3 \to \mathbb{R}^2$ 的向量函數 (以 3 個未知數所組成的 2 條方程式: f_x, f_y)，其一階條件可寫成一個 2×3 的 Jocobian 矩陣[6]:

$$\mathcal{J} = \begin{pmatrix} \dfrac{\partial f_x}{\partial a_1} & \dfrac{\partial f_x}{\partial a_2} & \dfrac{\partial f_x}{\partial a_3} \\ \dfrac{\partial f_y}{\partial a_1} & \dfrac{\partial f_y}{\partial a_2} & \dfrac{\partial f_y}{\partial a_3} \end{pmatrix} \tag{2}$$

Hessian Matrix(黑森矩陣)

假設 $F(x_1, ..., x_n)$，我們以 Hessian 矩陣[7]計算多變量函數 F 的二階條件。

$$\mathcal{H} = D^2F = \begin{pmatrix} \dfrac{\partial^2 F}{\partial x_1^2} & \cdots & \dfrac{\partial^2 F}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 F}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 F}{\partial x_n^2} \end{pmatrix} \tag{3}$$

[2]極小值剛好相反，其必要條件為 $f''(x^*) \geq 0$。

[3]f_x, f_y 為速度在平面兩軸上的分量，每一道分量均由 a_1, a_2, a_3 所描述。

[4]$\frac{\partial F}{\partial x}$: 將函數 F 對 x 偏微分 $\implies$ x「增加」1 單位，F 的變化量。

[5]DF: first derivative of F, D^2F: second derivative of F.

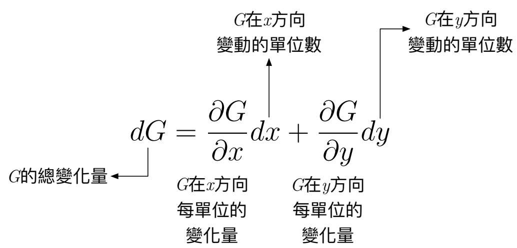
[6]此為 2×3 矩陣。一個 $i \times j$ 矩陣: i為陣中row($\leftrightarrow$)的數量，j為陣中column($\updownarrow$)的數量。

[7]Hessian 矩陣為 $n \times n$ 方陣。根據楊氏定理 (Young‘s Theorem)，二階微分先對 row 微還是對 column 微沒有差。

現在，就要開始介紹優化理論的核心。我們所關心的優化議題的應用其實十分廣泛。在數學中，可以探討以下問題: 當 x, y, z 三點分布在半徑為 1 的球上時，請問 $x+y+z$ 的最大和最小值分別是多少? 在個體經濟學中，我們可問: 現在有兩個商品 x_1 和 x_2 他們的價格分別為 p_1 和 p_2，若某消費者的效用函數為 $U(x_1, x_2) = x_1 x_2$，同時他身上只有帶 I 塊錢，請問這兩種商品他應該分別買多少? 為了能帶入最終 Lagrangian function(拉格朗日函數) 的建立及 Kuhn-Tucker Theorem(庫恩 – 塔克理論)，我們將先介紹全微分 (Total differential of multivariate function) 和隱函數定理 (Implicit function theorem)。

假設現有一函數 $G(x, y) = \mathbf{0}$ 是由 x 和 y 所組成，同時 y 是由 x 組成 $y = y(x)$。這時 G 的變化量可看作兩個部分，其一是 G 因 x 而形成的變化量，其二是 G 因 y 所形成的變化量[8]，而這種分解方法就稱為全微分。

圖 3: 函數變化量分解

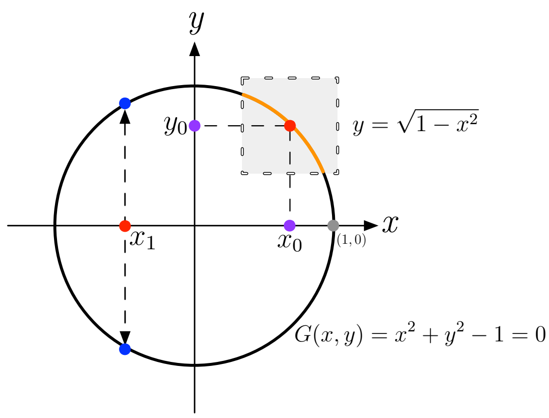


現在給定 $G(x, y) = x^2 - 3xy + y^3 - 7 = 0$，則 G 對 x 的微分為 $\frac{dG}{dx} = \frac{\partial G}{\partial x} + \frac{\partial G}{\partial y} \cdot \frac{dy}{dx} = 2x - 3y + (-3x + 3y^2)\frac{dy}{dx} = 0$，注意，這時 y 依然是 x 的函數 $y = y(x)$。為了找到一般式，假設 $G(x, y(x)) = c$[9] 且是一個 C^1 函數[10]，則 $\frac{\partial G}{\partial x}(x, y(x)) + \frac{\partial G}{\partial y}(x, y(x)) \cdot \frac{dy}{dx}(x) = 0$，經過移項可得

$$y'(x^*) = \frac{dy}{dx} = -\frac{\frac{\partial G}{\partial x}}{\frac{\partial G}{\partial y}} \tag{4}$$

這就是隱函數定理，讓我們僅用原函數 $G = 0$ 對 x, y 的微分，就能刻劃局部函數在 x^* 點的斜率: $y'(x^*)$。這個定理主要是問: 在點 (x_0, y_0) 為 $G(x, y)$ 的解的前提下，在其附近 (往左往右做極微小的移動) 可否將 y 表示成 x 的連續函數? (若可以，我們就能算出該點的函數斜率。)

圖 4: 隱函數



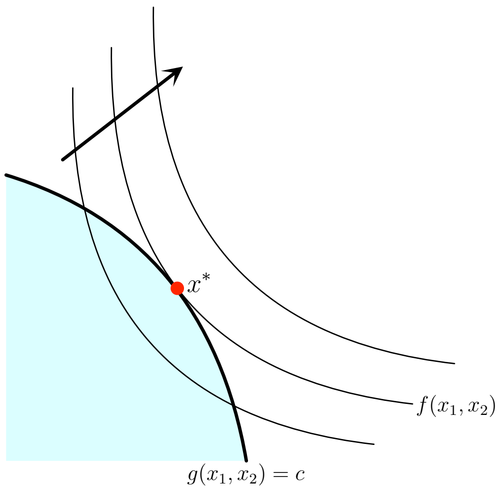
[8] 也可理解為 G 在 x 方向的變化量 (分量) 及 G 在 y 方向的變化量。
[9] c 為任意常數。
[10] C^1 代表此函數的一階微分存在。若是 C^2，則代表函數的二階微分存在。

由圖 4 所示，雖然 $x = x_1$ 對應兩個不同的 y 值，一個在上半圓另一個在下半圓，但我們能分清楚甚至從兩個值中選一個。就如 $G(x, y) = 0$ 不是一個函數，但如果我們侷限在右上角的方格內，此局部圖形卻能表示成 $y = f(x)$。但是如果是點 $(1, 0)$，就算稍微[11]向左向右移動 x，依然會同時得到兩個 y 值，而且這兩個值無從區分。因為在 $(1, 0)$ 附近，無法分辨在上半圓 $(x, +y)$ 或在下半圓 $(x, -y)$ 的兩點，導致在 $(1, 0)$ 這點沒有辦法寫出一個獨特的 x, y 關係式，這結論也可從式 (4) 中得到。

如果要使隱函數斜率 $(\frac{dy}{dx})$ 存在，則分母 $(\frac{\partial G}{\partial y})$ 不能為 0，但 $\frac{\partial G}{\partial y}(1, 0) = 2y|_{(1,0)} = 0$，因此在優化理論中，必須把擁有垂直切線的切點另外拿出來檢查[12]。接著，我們將描述隱函數定理，因為嚴謹的定義較複雜，請自行參考。

定理 2 (**The Implicit Function Theorem**). *Let E be open and $f : E \subset \mathbb{R}^{m+n} \to \mathbb{R}^n$ be a C^1 function s.t. $f(x^*, y^*) = 0$ for some $(x^*, y^*) \in E$ where $x^* \in \mathbb{R}^m$ and $y^* \in \mathbb{R}^n$. Let $Df(x^*, y^*) = \big(Df_x(x^*, y^*),\ Df_y(x^*, y^*)\big)$ and $Df_y(x^*, y^*)$ is invertible. Then $\exists$ a neighborhood $U \in \mathbb{R}^m$ of x^* and a C^1 function $g : U \to \mathbb{R}^n$ s.t. $g(x^*) = y^*$ and $\forall x \in U,\ f\big(x, g(x)\big) \equiv c$ where $\big(x, g(x)\big) \in E$. Then the derivative of g at any $x \in U$ is $Dg(x) = -Df_x(x, y) \cdot Df_y(x, y)^{-1}$.*

圖 5: 優化過程



目標函數的優化必須滿足約束條件 (constraints)。圖 5 的目標函數受到 x_1, x_2 影響，這兩個變數就如想要買的兩種商品的數量或生產時所需的兩種原料。另外，陰影面積就是可行 (feasible) 的區域，這個區域就如身上所帶的錢或生產線上的原料總量。

在極大化目標函數的同時，$f(x_1, x_2)$ 會往右上方向外移動，但不能移到使函數曲線與陰影面積失去交集，因此這時的最優選擇就是在目標函數 $f(x_1, x_2)$ 與約束 $g(x_1, x_2) = c$ 相切的地方 (x^*)。透過隱函數定理，我們可以找到目標函數與約束在 x^* 這點的斜率，又因兩條線相切，故他們在 x^* 的斜率相同:

$$-\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial f}{\partial x_2}(x^*)} = -\frac{\frac{\partial g}{\partial x_1}(x^*)}{\frac{\partial g}{\partial x_2}(x^*)} \tag{5}$$

[11] 此微小，比所有能說出來的微小等級還微小。

[12] 算出來的極值 (內點解) 必須跟擁有垂直切線的切點比較，確保極值沒有發生在這些特殊點上。

接著將等式移項整理，並令這個共同的分數值為 μ

$$-\frac{\frac{\partial f}{\partial x_1}(x^*)}{\frac{\partial g}{\partial x_1}(x^*)} = -\frac{\frac{\partial f}{\partial x_2}(x^*)}{\frac{\partial g}{\partial x_2}(x^*)} = \mu \tag{6}$$

這時將透過目標函數與約束在 x^* 點斜率相同所形成的等式展開，再加上原本的約束就可得到

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x) + \mu\frac{\partial g}{\partial x_1}(x) = 0 \\ \frac{\partial f}{\partial x_2}(x) + \mu\frac{\partial g}{\partial x_2}(x) = 0 \\ g(x_1, x_2) - c = 0 \end{aligned} \tag{7}$$

這時有三個未知數 x_1, x_2, μ 及三條方程式，我們就可解出在約束的條件下，讓目標函數極大的最優點 x_1, x_2 和 μ 分別是多少。這種優化求解 (式 7) 的聯立方程式，可以透過 Lagrangian 函數，將三條式子合併為一條。列式方法如下:

$$\mathcal{L}(x_1, x_2, \mu) = f(x_1, x_2) + \mu\Big(g(x_1, x_2) - c\Big) \tag{8}$$

接著在「特定的條件」下，只要讓全微分 $D\mathcal{L}(\boldsymbol{x}^*, \mu^*) = 0$[13]，也就是計算 $\frac{\partial \mathcal{L}}{\partial x_1}, \frac{\partial \mathcal{L}}{\partial x_2}, \frac{\partial \mathcal{L}}{\partial \mu}$ 並令他們為 0，就能得到式 7 的三條方程式。因此，在面臨「等式約束」條件下求目標函數最大化的優化問題時，我們將 Lagrangian 函數 ($\mathcal{L}$) 定義為: 目標函數 $+\mu_1$(第一條等式約束)$+\mu_2$(第二條等式約束)$+\cdots+\mu_n$(第 n 條等式約束)[14]。若是求最小化，則在目標函數前加上負號，其他列式方法不變: $\mathcal{L} = -f(x_1, x_2) + \mu$(等式約束)。注意，目標函數和約束中間，是以「加號」連接[15]。

透過比較式 7 和式 8 就可以發現 Lagrangian 函數巧妙的把一個三條式子的約束問題 (constrained problem) 變成僅有一條式子的非約束問題 (unconstrained problem)。注意這時的非約束，不是指優化問題中沒有約束 (限制式) 的存在，而是指這個最優點是內點解[16]，不像圖 5 那樣是角解 (邊界點)。也就是說，原本我們需要在約束條件下將目標函數平移來找到切點，但現在因為合成單一條式子，我們可直接令 $\mathcal{L}$ 函數的一階條件為 0，算出最優點。

但是這個極其便捷的方法卻也必須付出代價，那就是一個原本只有兩個變數的目標函數，變成三個變數的 Lagrangian 函數。這個多出來的變數，也就是在 $\mathcal{L}$ 函數中的 μ_1, μ_2 等，被稱為拉格朗日乘數 (Lagrange multiplier)，後面會再做更詳細的介紹，在那之前我們必須注意這個方法成立的必要條件。

Lagrangian 函數並不是任何時候都可以使用，有兩個前提我們必須預先確認。首先，最優點 (極值) 必須先存在，否則算出來的是一個空集合的極值，對一個虛空的東西來說，任何點都是最優點，畢竟沒有任何點存在其中。第二點，必須確定式 7 存在。例如，在只有一條雙變數約束的情況下，約束對兩變數分別的微分 ($\frac{\partial g}{\partial x_1}$ 和 $\frac{\partial g}{\partial x_2}$) 不能同時為 0，否則 μ 會直接消失。現在將先針對第二點說明。

[13] 粗體英文字代表向量 (vector)。這裡的 $\boldsymbol{x} = (x_1, x_2)$。

[14] 這些進入 Lagrangian 的等式約束必須以 (約束 =0) 的形式出現。例如: $x_1 + x_2 = 10$ 要寫成 $x_1 + x_2 - 10 = 0$ 的形式才能放到 Lagrangian 函數中: $\mu(x_1 + x_2 - 10)$。

[15] 在等式約束條件下，其實用加號、減號都可以，但為了配合後面處理「不等式約束」條件，建議全部都用加號。如此整套系統才能固定列式方法，也能避免使用減號在進行最大/最小化、等式/不等式時，被調整方法混淆。

[16] 內點，指的是就算「稍微」偏離最優點，也不會離開可行集 (feasible set)。圖 5 的最優點，只要稍微往右上方向外移動就會離開可行集。

假設現在有三條 4 變數的約束，$\boldsymbol{g}:\mathbb{R}^4\to\mathbb{R}^3$ 如下:

$$\boldsymbol{g}=\Big(g_1(x_1,x_2,x_3,x_4)=c_1,\quad g_2(x_1,x_2,x_3,x_4)=c_2,\quad g_3(x_1,x_2,x_3,x_4)=c_3\Big) \tag{9}$$

這個向量約束 ($\boldsymbol{g}$) 對 4 個變數 (x_1,x_2,x_3,x_4) 微分，會得到一個 3×4 的 Jocobian 矩陣:

$$D\boldsymbol{g}(\boldsymbol{x}^*)=\begin{pmatrix}\frac{\partial g_1}{\partial x_1}(\boldsymbol{x}^*) & \frac{\partial g_1}{\partial x_2}(\boldsymbol{x}^*) & \frac{\partial g_1}{\partial x_3}(\boldsymbol{x}^*) & \frac{\partial g_1}{\partial x_4}(\boldsymbol{x}^*)\\ \frac{\partial g_2}{\partial x_1}(\boldsymbol{x}^*) & \frac{\partial g_2}{\partial x_2}(\boldsymbol{x}^*) & \frac{\partial g_2}{\partial x_3}(\boldsymbol{x}^*) & \frac{\partial g_2}{\partial x_4}(\boldsymbol{x}^*)\\ \frac{\partial g_3}{\partial x_1}(\boldsymbol{x}^*) & \frac{\partial g_3}{\partial x_2}(\boldsymbol{x}^*) & \frac{\partial g_3}{\partial x_3}(\boldsymbol{x}^*) & \frac{\partial g_3}{\partial x_4}(\boldsymbol{x}^*)\end{pmatrix},\quad \boldsymbol{x}^*=(x_1^*,x_2^*,x_3^*,x_4^*) \tag{10}$$

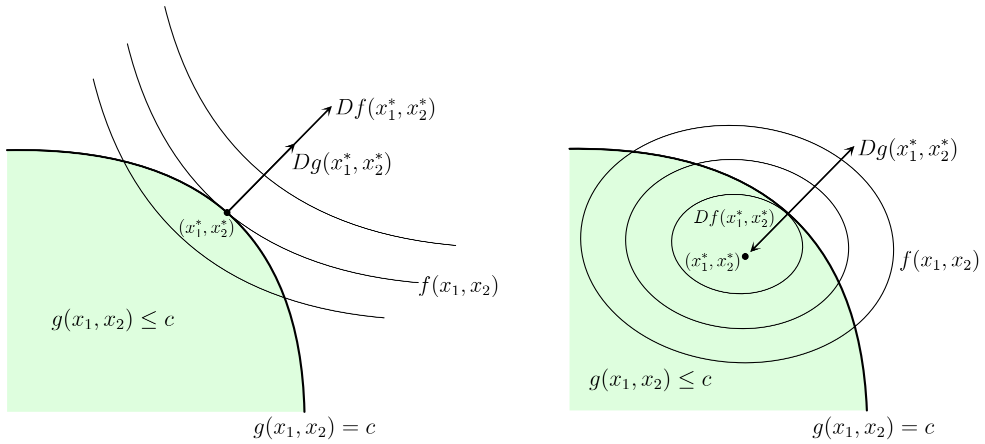
$D\boldsymbol{g}(\boldsymbol{x}^*)$ 矩陣必須是 full rank(滿秩)，才會存在 $\boldsymbol{\mu}^*$ 使 $D\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\mu}^*)=0$，Lagrangian 函數才可使用。如果非滿秩，代表 Jocobian 矩陣少一個 row。以式 (7) 為例，非滿秩等於少了一條等式，因此就不能將剩下的等式以 Lagrangian 函數表現。所以，必須先檢查有哪幾組 $\boldsymbol{x}^*$ 會使 $D\boldsymbol{g}(\boldsymbol{x}^*)$ 不再是滿秩，這幾組 $\boldsymbol{x}^*$ 就是約束向量的 critical point(s)(臨界點)。這些點是 Lagrangian 函數所算不出來的點，因此最後必須再把 Lagrangian 函數算出來的點和這些 critical point(s) 比較，確保正確的最優點不會被忽略。在說明這個結論之前，我們將先解釋什麼是滿秩矩陣。

一個矩陣的 rank(秩)，就是在 row 階梯形矩陣 (row echelon form) 中[17]，非零 row 的數量。下方矩陣的 rank 為 2，因為第三 row 可以和第一、第二 row 進行運算後化為 0。

$$\begin{pmatrix}0&1&2\\1&2&1\\2&7&8\end{pmatrix}\to\begin{pmatrix}1&0&-3\\0&1&2\\0&0&0\end{pmatrix} \tag{11}$$

而滿秩就是把矩陣的 rank 和約束數量進行比較。約束有分為等式約束和不等式約束。但我們真正在意的，是這個約束到底有沒有效 (efficient)，並且將所有有效的約束，統稱為 binding 約束。

圖 6: Binding 約束 與 Nonbinding 約束



在上方圖中，左圖要極大化 f，這時目標函數和約束的梯度是朝向同一方向，最優點就在約束 $g(x_1,x_2)=c$ 上，因此這個等式約束是有效的，也就是 binding 約束。相反的，右圖要極小化 f，現

[17] 若一個 row 階梯形矩陣的每一個 row 的領導係數為 1，同時這些 row 的領到係數也是該 column 中唯一的非零係數，則稱這矩陣為簡化 row 階梯行矩陣 (reduced row echelon form)。式 11 右方的矩陣就是簡化 row 階梯行矩陣。

在最優點在 $g(x_1, x_2) \le c$ 裡面，因此等式約束 $g(x_1, x_2) = c$ 沒有任何效果 (ineffective)，故稱這是 inactive 或 nonbinding 約束。因此，隨時都要注意問題中的約束是否依然是 binding。

判斷矩陣是否為滿秩的方法分成兩種，若矩陣是方陣 $(n \times n)$，則計算其行列式 (determinant) 是否為 0。若方陣行列式為 0，稱為奇異 (singular)；若行列式不為 0，稱為非奇異 (nonsingular)，這時矩陣是可逆 (invertible) 的，同時也是滿秩。若矩陣不是方陣，是一個 $m < n$[18]、row 之間線性獨立的 $m \times n$ 的長方形矩陣，同時 m 等於 binding 約束的數量，那這個矩陣就是滿秩。

回到之前介紹的 Lagrangian 函數，他的必要條件是約束向量的一階微分矩陣 $D\boldsymbol{g}(\boldsymbol{x}^*)$ 必須在滿秩的狀態下才可使用。原因是因為這個 $m \times n (m < n)$ 的滿秩矩陣 (m=binding 約束的數量)，可以分割成左右: $m \times (n - m)$ 和 $m \times m$ 兩個子矩陣。

$$\underset{m\times n}{D\boldsymbol{g}(\boldsymbol{x}^*)} = \Big(\underset{m\times(n-m)}{D_w\boldsymbol{g}(\boldsymbol{x}^*)},\ \underset{m\times m}{D_z\boldsymbol{g}(\boldsymbol{x}^*)}\Big) \tag{12}$$

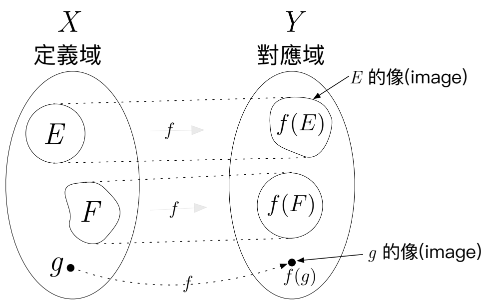
第二個方陣 $D_z\boldsymbol{g}(\boldsymbol{x}^*)$ 是滿秩，表示此矩陣為非奇異，同時也可逆。因此，可以配合第一個矩陣，根據隱函數定理找出 $\boldsymbol{\mu}^*$，讓式 7(透過斜率相等所得到的方程組) 成立。針對約束向量的一階微分矩陣 $D\boldsymbol{g}(\boldsymbol{x}^*)$ 必須滿秩這個必要條件，我們稱之為約束規範 (constraint qualification)。

接下來，將完整的描述 Lagrange 定理，並繼續說明前面所提到的另一個 Lagrangian 函數的使用前提: 最優點 (極值) 必須先存在，最後再討論拉格朗日乘數來總結「等式約束」的優化問題。

定理 3 (The Theorem of Lagrange). *Let E be open and $f : E \subset \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ $(m < n)$ be C^1 functions. Let x^* be a local maximum (minimum) of f on D where $D \subset E$. Suppose then rank of $Dg(x^*)$ equals m. Then $\exists \mu^* \in \mathbb{R}^m$ s.t. $D\mathcal{L}(x^*, \mu^*) = 0$[19] where $\mathcal{L}(x, \mu) : \mathbb{R}^{n+m} \to \mathbb{R}$ is given by $\mathcal{L}(x, \mu) = f(x) + \mu' g(x)$[20].*

大家在大學期間所面臨的優化問題，幾乎不會出現「無解」的情形。我們知道函數 (fuction) 刻劃的是一種數與數之間的關係，因此函數是一種映射 (mapping)，將定義域 (domain) 裡面的數經過處理後，產生對應域 (codomain) 中的數。

圖 7: 函數圖



在上圖中，$f(E)$ 及 $f(F)$ 加上 $f(g)$，又稱為這個 f 函數的值域 (range)，因此值域又是對應域的子集。我們要確定的就是當定義域的數透過函數映射後，對應域中的數是否會存在極大、極小值? 而有個方法能確認這一點，那就是使用極值定理 (The Weierstrass Theorem)。

[18] 若是 $m > n$，則此矩陣必須同時滿足 column 之間線性獨立，同時 n 等於 binding 的約束數量。
[19] This implies $Df(x^*) + \mu^* Dg(x^*) = 0$.
[20] The x and μ can also be considered as vectors.

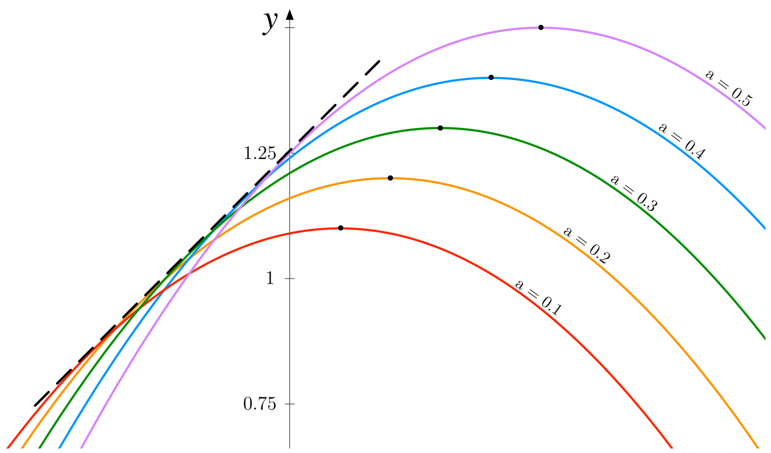
定理 4 (**The Weierstrass Extreme Value Theorem**[21]). *Let $f : X \to \mathbb{R}$ be continuous and $k \subseteq X$ be compact. Then $\exists x_1, x_2 \in k$ s.t. $f(x_1) = \inf_{x\in k} f(x)$ and $f(x_2) = \sup_{x\in k} f(x)$*[22]*. In particular, the* inf *and* sup *are finite, so f is bounded on k.*

Proof. By continuous of f and compactness of k, $f(k) \subseteq \mathbb{R}$ is compact. Therefore, $f(k)$ is closed and bounded. Hence, $\sup f(k), \inf f(x) \in f(k)$. ■

極值定理牽涉到實數分析 (Real Analysis) 中的概念，但定理的內容很簡單，那就是只要確定函數是連續的，同時函數的定義域是緊集 (compact)[23]，例如一個包含 a, b 兩端點的線段 $[a, b]$，那麼在這函數的對應域中，必存在極大和極小值。因此，我們必須先用 Weierstrass 定理確認極值存在，才能開始使用 Lagrangian 函數來找尋那些會產生極值的最優點。當然在大學期間，我們可當作極值都存在，因此可以跳過這個最重要的前提檢查，直接進入 Lagrangian 函數的建構。

在結束等式優化問題前，我們將討論拉格朗日乘數 (Lagrange multiplier)。拉格朗日乘數不僅存在經濟意義，也是包絡定理 (Envelope theorem) 重要的應用結果。因此，我們將先介紹包絡定理。

圖 8: 包絡曲線



包絡定理的核心概念，就是探討當目標函數中的參數 (外生變數[24]) 發生「微小」變化時，目標函數在最優點的函數值會如何改變? 此定理為計算提供了一條捷徑，讓我們在做優化運算時，不用經歷複雜的連鎖律就能求出最優點。上圖 8 的基本目標函數為:

$$y(x; a) = -x^2 + 2xa - a^2 + a + 1 \tag{13}$$

其中 x 為內生變數，a 為外生變數。y 軸右側最下方是 $a = 0.1$ 的線，最上方則是 $a = 0.5$ 的線。包絡定理刻劃的就是當目標函數中 a 發生微小變化時，函數在圖中黑點，也就是最優點 (極值) 的函數值是如何受到 a 的影響而改變? 另外，圖中因外生變數 (a) 改變，使函數平移而勾勒出的虛線，就是包絡曲線[25]。這個定理又分成「無約束條件」及「有 (等式) 約束條件」兩種，但使用方法大同小異。

[21]很多定理皆以 Karl Weierstrass 命名。注意，這裡的 Weierstrass 定理 (極值定理) 不是 Bolzano–Weierstrass 定理。
[22]The supremum(sup) of a set is its least upper bound and the infimum(inf) is its greatest lower bound.
[23]所謂緊集 (compact set)，代表如果這個集合是歐幾里得空間 (R^n) 的子集，那他就是一個閉集合 (closed set)，並且有界 (bounded)。
[24]外生變數 (exogenous variable) 為模型外已給定的參數，非由模型來決定的內生變數 (endogenous variable)。
[25]在幾何學中，某曲線的無窮集 (曲線族) 的包絡線 (Envelope) 是一條跟該曲線族的每條線都有至少一點相切的曲線。

定理 5 (**Unconstrained Envelope Theorem**). *Let $f(\boldsymbol{x};a)$ be a C^1 function of $\boldsymbol{x} \in \mathbb{R}^n$ and a given scalar a. Let $\boldsymbol{x}^*(a) = \arg\max_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x};a)$. Suppose $\boldsymbol{x}^*(a)$ is a C^1 function of a, then*

$$\frac{d}{da}f\Big(\boldsymbol{x}^*(a);a\Big) = \frac{\partial}{\partial a}f\Big(\boldsymbol{x}^*(a);a\Big).$$

當沒有約束條件時，如定理 5 中所述，我們可以將函數對外生變數 a 偏微分，就能直接得到 a 和最優點函數值兩者間變化的關係。如果不使用包絡定理，我們就必須對目標函數進行複雜的隱函數微分來判斷變化關係，這個結論簡化了複雜的計算，同時也能幫助得到很多個體經濟學理論中重要的定理[26]。當目標函數在有約束條件下進行優化時，找最優點函數值和外生變數之間的關係不再是將函數拿來偏微，而是如下方定理 6 所述，先拿目標函數與約束建構 Lagrangian 函數，接著直接將 Lagrangian 函數對 a 偏微分即可。

定理 6 (**Constrained Envelope Theorem**). *Let $f(\boldsymbol{x};a), g_1(\boldsymbol{x},a), ..., g_k(\boldsymbol{x},a)$[27]$: \mathbb{R}^n \times \mathbb{R}^1 \to \mathbb{R}^1$ be C^1 functions of $\boldsymbol{x} \in \mathbb{R}^n$ and a given scalar a. Let $\boldsymbol{x}^*(a) = \arg\max_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x};a)$ on the constraint set $g_i(\boldsymbol{x},a) = 0\ \forall a$ and $i = 1,...,k$. Suppose $\boldsymbol{x}^*(a)$ and $\mu_i(a)$ are C^1 functions of $a\ \forall i = 1,...,k$ and constraint qualification holds, then*

$$\frac{d}{da}f\Big(\boldsymbol{x}^*(a);a\Big) = \frac{\partial}{\partial a}\mathcal{L}\Big(\boldsymbol{x}^*(a),\boldsymbol{\mu}^*(a);a\Big)$$

where $\mathcal{L}(\boldsymbol{x}^*,\boldsymbol{\mu}^*;a) = f(\boldsymbol{x}^*;a) + \sum_{i=1}^{k}\mu_i^*(a)' g_i(\boldsymbol{x}^*;a)$.

現在我們將透過包絡定理得到拉格朗日乘數。假設目標函數 $f(x,y)$ 和單一約束條件 $g(x,y,) = a$ 都是由 x,y 兩個內生變數構成，其最優點 (極值) 為 $x^*(a)$ 和 $y^*(a)$，而在最優點的函數值為 $f\Big(x^*(a),y^*(a)\Big)$。我們可建構 Lagrangian 函數 $\mathcal{L}(x,y,\mu) = f(x,y) - \mu\Big(g(x,y)-a\Big)$，根據定理 6，解出最優點 $x^*(a),y^*(a),\mu^*(a)$ 後，函數值在最優點所受到 a 的影響為:

$$\begin{aligned}
\frac{d}{da}f\Big(x^*(a),y^*(a)\Big) &= \frac{\partial}{\partial a}\mathcal{L}\Big(x^*(a),y^*(a),\mu^*(a)\Big) \\
&= \frac{\partial}{\partial a}\Big[f\big(x^*(a),y^*(a)\big) - \mu^*(a)\Big(g\big(x^*(a),y^*(a)\big)-a\Big)\Big] \\
&= \mu^*(a)
\end{aligned} \tag{14}$$

我們稱這個 μ 為拉格朗日乘數。注意，式 14 我們是將 Lagrangian 函數對 a 進行偏微分而不是全微分，因此不須將每個變數都用隱函數展開再微分，可直接看成 $\frac{d}{da}\Big[f(x^*,y^*) - \mu^*\cdot g(x^*,y^*) - \boldsymbol{\mu^* a}\Big]$，微分後除了粗體那項，其餘都是 0。另外注意，除了 x^* 和 y^*，μ^* 也會受到 a 的影響。式 14 的結論也說明: 當約束條件中的 a 增加 θ 個單位，最優點的函數值則會增加 $\theta\mu^*$ 個單位。

接著，我們將透過一個簡單的例子驗證式 14 的結論。現有一函數 $f(x_1,x_2) = x_1^2 x_2$ 及約束條件 $2x_1^2 + x_2^2 = 3$，已知最優點發生在 $(x_1,x_2,\mu) = (1,1,0.5)$，而在這點的函數值 $f(1,1) = 1$。現在如果將約束調整為 $2x_1^2 + x_2^2 = 3.3$，請問在新最優點的函數值是多少? 依照正規的作法，我們必

[26] Hotelling's Lemma、Shephard's Lemma、Roy's Identity 都是應用包絡定理而得到的重要經濟理論。
[27] This is a $n \times 1$ matrix.

須透過 Lagrangian 函數找出新的最優點 $(x_1, x_2) = (\sqrt{1.1}, \sqrt{1.1})$ 才能算出在新最優點的函數值是 $f(\sqrt{1.1}, \sqrt{1.1}) = \sqrt{1.1^3} \approx 1.153$。但如果透過拉格朗日乘數，整個過程將會簡便許多。我們知道在約束調整前，達到最優點的拉格朗日乘數 $\mu^* = 0.5$，調整後約束的 a 增加 0.3，最優點的函數值則會增加 $\theta\mu^* = 0.3 \times 0.5 = 0.15$ 個單位，也就是從 1 變成 1.15。透過泰勒展開 (Taylor expansion)，也可以看出拉格朗日乘數是包絡定理的結果:

$$f\Big(x_1^*(\underbrace{3+0.3}_{=a+h}), x_2^*(3+0.3)\Big) = \underbrace{f\Big(\underbrace{x_1^*(3)}_{1}, \underbrace{x_2^*(3)}_{1}\Big)}_{1} \cdot (\underbrace{0.3}_{h})^0 + \underbrace{f'\Big(x_1^*(3), x_2^*(3)\Big)}_{=\mu^*=0.5} \cdot (0.3)^1$$
$$= 1 + 0.5 \times 0.3 = 1.15$$

最優點的函數值 $f\Big(x^*(3.3), y^*(3.3)\Big) = 1.15$，同時乘數 $\mu^* = \dfrac{d}{da} f\Big(x^*(a), y^*(a)\Big)\Big|_{a=0.3} = 0.5$。從經濟的角度上來看，我們可以看作生產某商品需要 x_1, x_2 兩種原料，同時每一商品能賺 $x_1^2 x_2$ 元，現在兩種原料在 $2x_1^2 + x_2^2$ 下總共只有 3 個單位，請問兩種原料分別要用多少，才能將利潤極大? 這時，拉格朗日乘數 (μ^*) 就能說明當原料變成 3.3 單位時，利潤能增加多少。換個角度來想，我們也可將拉格朗日乘數看成廠商為了獲得額外 1 單位原料所願意付的價格，這個價格又稱為內部價值 (internal value) 或「 影子價格 (shadow price) 」。

不等式約束 (inequality constraint) 的優化問題，會直接建立在之前所提到的觀念上，因此將比前面所討論的內容複雜，但重要性卻不在其之下。我們將從錐形 (cones) 開始討論，最後帶入不等式約束優化的核心: 庫恩-塔克理論 (Kuhn-Tucker Theorem)。

定義 1 (**Cone**). *Let $a^1, ..., a^m \in \mathbb{R}^n$. The cone with vertex at the origin $\mathbf{0} \in \mathbb{R}^n$ that is generated by the vectors $a^1, ..., a^m$ is the set*

$$\mathcal{K}(a^1, ..., a^m) = \big\{x \in \mathbb{R}^n | x = \rho a^i, \rho \geq 0, i = 1, ..., m\big\}.$$

錐形，是以原點 (origin) 為錐點在 n 維空間中所形成的錐體。圖 9 左側為平面上的錐形 $\mathcal{K}\Big((-1,3),(2,4)\Big) = \big\{x \in \mathbb{R}^2 | x = \rho(-1,3) \text{ or } x = \rho(2,4), \rho \geq 0\big\}$ 是以原點 $(0,0)$ 至 $(-1,3)$ 和 $(2,4)$ 兩點所形成的方向，將點 $(-1,3)$ 和 $(2,4)$ 依比例任意放大、縮小，所形成的 V 形線。如果將錐形平移就會得到右側的圖，又稱為仿射錐 (affine cone)。

這時，如果將錐形上的任意兩點相加，則會得到一個可能依舊在錐上或在錐形內部陰影面積上的點。如圖 10 所示，若將錐上的 $(-1,3)$ 和 $(2,4)$ 兩點相加，則會得到在陰影中的點:$(1,7)$。若取錐形上的同一點相加如 $(-1,3)$，則會得到一個在錐上的點 $(-2,6)$。

定義 2 (**Convex Cone**). *Let $\mathcal{K}$ be a cone in $\mathbb{R}^n$. The set*

$$\mathcal{CK} = \big\{x \in \mathbb{R}^n | x = y + z \ \forall y, z \in \mathcal{K}\big\}$$

is the convex cone generated by the cone $\mathcal{K}$.

根據定義 1，$(-2,6)$ 不屬於錐形 V 線上面的點，因此該點不屬於圖 10 中的錐形。因此，我們將所有錐形上任意兩點相加在 V 線內陰影區所形成的點、加上 V 線本身，定義成一個廣義的錐體，稱為「 凸錐 (Convex cone) 」。所以定義 2 所描繪的凸錐，就是 V 線 (錐形) 加上陰影區。

接下來，我們將介紹法卡斯引理 (Farkas's Lemma)。此引理主要描述兩向量相加後，合成向量與原兩向量所形成的凸錐之間的關係。

圖 9: 錐形和仿射錐

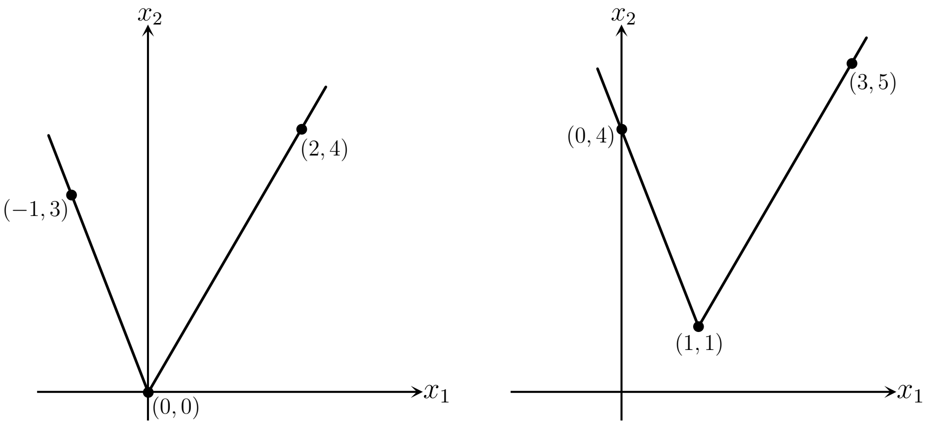


圖 10: 凸錐

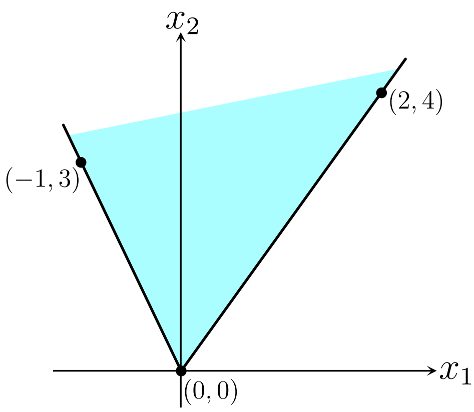
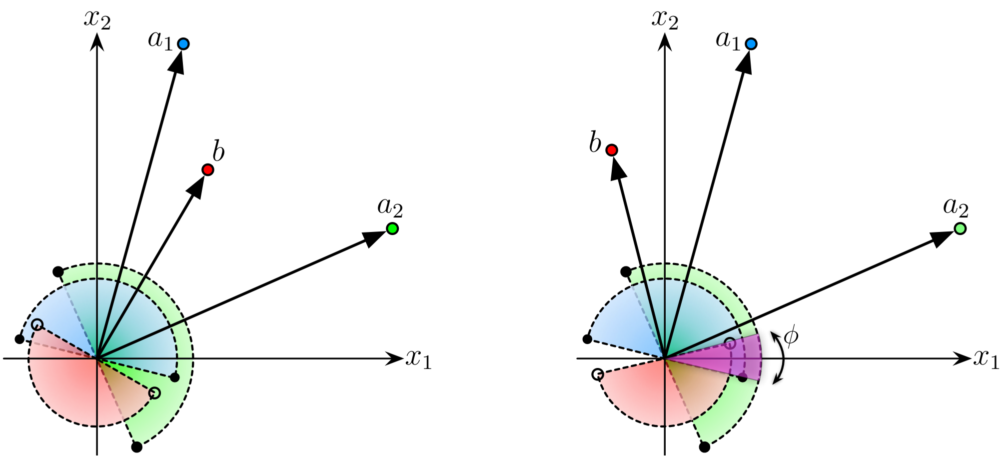


圖 11: 法卡斯引理



引理 1 (Farkas's Lemma). *Given vectors $a^1, ..., a^m, b \in \mathbb{R}^n$ with $b \neq \mathbf{0}$, one and only one of the following two statements is true.*

(i) $\exists \lambda_1, ..., \lambda_m \in \mathbb{R}$, all nonnegative and not all zero, such that

$$b = \lambda_1 a^1 + \cdots + \lambda_m a^m.$$

(ii) $\exists x \in \mathbb{R}^n$, such that

$$a^1 \cdot x \geq 0, ..., a^m \cdot x \geq 0 \text{ and } b \cdot x < 0.$$

法卡斯引理保證兩向量 a_1、a_2 經過一非負的倍數調整長度後，他們的合成向量必在 a_1 和 a_2 所形成的凸錐之中。因此，如果某一向量不在凸錐之內，則其必不是 a_1、a_2 經調整後的合成向量。透過內積 (inner product) 的定義:

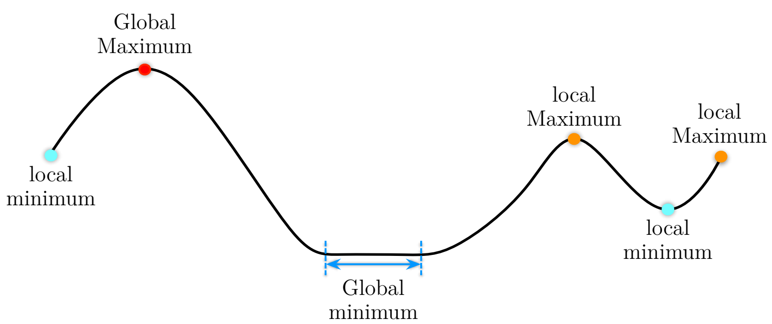
$$0 \leq \cos\theta = \frac{\overrightarrow{a_1} \cdot \overrightarrow{a_2}}{|\overrightarrow{a_1}||\overrightarrow{a_2}|} \leq 1, \text{ where } -90° \leq \theta \leq 90° \tag{15}$$

兩向量 $\overrightarrow{a_1}$、$\overrightarrow{a_2}$ 的內積若介於 0 於 1 間，則其夾角為銳角；若其內積為負數，則為鈍角；若內積為 1，則是直角。如圖 11 左側所示，向量 $\overrightarrow{x}$ 若與 $\overrightarrow{a_1}$ 為銳角，其必在以 $\overrightarrow{a_1}$ 為中心，向左、向右各 90° 的範圍內，也就是最靠近 x_1 軸的兩個實心黑點所經過的扇形角度。若 $\overrightarrow{x}$ 與 $\overrightarrow{b}$ 為鈍角，則其範圍為兩空心圓所經過的扇形角度。

我們根據引理 (ii)，在圖 11 中將 $\overrightarrow{x}$ 的潛在範圍標示出來。其中，最大的扇形[28]表示 $\overrightarrow{x}$ 在 $\overrightarrow{a_2}$ 的銳角範圍內，中等的扇形表示 $\overrightarrow{x}$ 在 $\overrightarrow{a_1}$ 的銳角範圍內，最小的扇形表示 $\overrightarrow{x}$ 在 $\overrightarrow{b}$ 的鈍角範圍內。法卡斯引理二分了所有可能的情況，只要符合 (i) 就必不符合 (ii)，現在我們用一個簡單的方法說明二分法存在。引理 (ii) 說明: 只要向量 b 不是 $\overrightarrow{a_1}$、$\overrightarrow{a_2}$ 的合成，則 $\overrightarrow{x}$ 同時會和 $\overrightarrow{a_1}$、$\overrightarrow{a_2}$ 成銳角關係，並和 $\overrightarrow{b}$ 成鈍角關係，也就是說，若 $\overrightarrow{x}$ 在 $\overrightarrow{a_1}$、$\overrightarrow{a_2}$ 所形成的凸錐外，則 $\overrightarrow{x}$ 必須出現在三個半圓交疊 (圖 11 右側 ϕ 角度) 的範圍內。但是在引理 (i) 的情況下，三個半圓不可能重疊，因此就不存在 ϕ 角度讓 $\overrightarrow{x}$ 滿足 (ii) 的條件；同理，只要 $\overrightarrow{x}$ 在三個半圓的交疊內，則 $\overrightarrow{b}$ 就不可能在 $\overrightarrow{a_1}$、$\overrightarrow{a_2}$ 所形成的凸錐內。

接下來，我們正式介紹庫恩-塔克定理 (Kuhn-Tucker Theorem)。這個理論是由 Harold Kuhn 和 Albert Tucker 在 1951 年所提出來的理論，但是後來發現在 1939 年時，William Karush 就已在他的碩士論文中提出這個結果，因此現在又有人稱這個定理為 Karush-Kuhn-Tucker Theorem(KKT 定理)。庫恩-塔克定理主要描述一個可行集 (feasible set $\mathcal{D}$) 中的點 x^*，要成為局部最大值 (local maximum) 的必要條件，這些條件又被稱為最適性必要條件 (necessary optimality condition)。

圖 12: 全域 (global) 與局部 (local) 極值

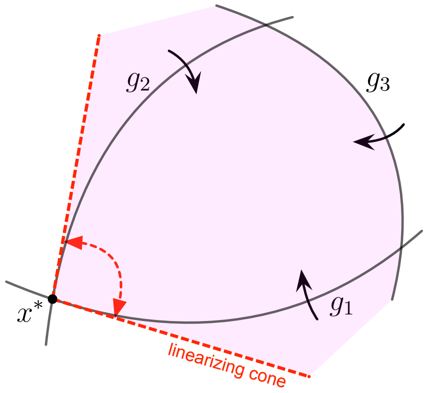


而其之所以被稱為必要條件，是因為只要局部極值 (最優點) 存在，就一定會滿足這些條件。也就是說，我們可以使用庫恩-塔克定理來找到局部極值。但若要使用這個定理，也有一些先決條件必須達成。所以在某些狀況下，雖然最優點存在，但卻無法使用庫恩-塔克求解。這個使用前提，就是要確認在「可行方向[29](feasible direction)」下，目標函數不會無止盡的增長。現在，把這個可行方向以 z 表示，

[28] 扇形的最寬角度為 180° 的半圓。
[29] 可行方向是一個方向向量 (feasible direction vector)。

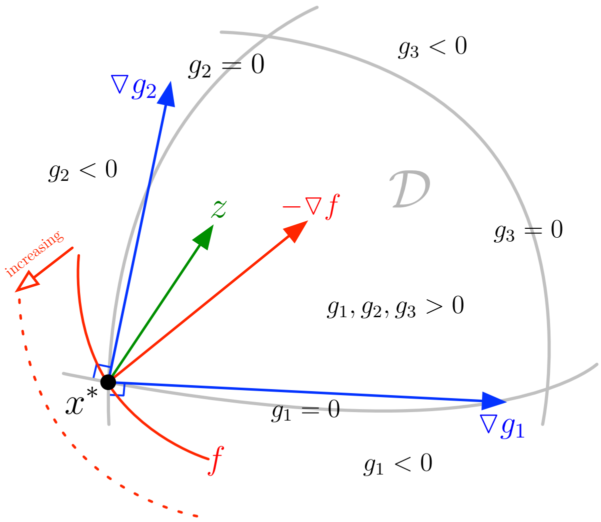
在下圖 13 中 g_1, g_2, g_3 分別為三個不等式約束，假設 x^* 為局部極值，這時從 x^* 出發沿其中兩個約束 g_1, g_2 做切線 (形成圖中的兩條紅虛線)，這兩條切線所圍成的範圍稱為線性化錐形 (linearizing cone)。

圖 13: 線性化錐形



而 z 必會出現在線性化錐形的範圍中 (圖 13 陰影區)。接下來，我們要定義兩個 z 的性質。前面在討論 Lagrangian 時都用 $g_i(x^*)$ 來表示等式約束，現在同時出現「等式」及「不等式」約束時，我們將用 $h_j(x^*)$ 表示等式約束，用 $g_i(x^*)$ 表示不等式約束。

圖 14: 庫恩-塔克定理示意圖



定義 3 (可行方向性質一:z_1). *Let* $g_i(x^*) \geq 0\ \forall i = 1, ...n$ *and* $h_j(x^*) = 0\ \forall j = 1, ..., p$ *denote inequality and equality constraints. Then* $z_1(x^*) = \Big\{z : z'\nabla g_i(x^*) \geq 0\ \forall i \in I(x^*), z'\nabla h_j(x^*) = 0, j = 1, ..., p\Big\}$[30] *where* $I(x^*) = \{i \in \{1, ...m\} | g_i(x) = 0\}$.

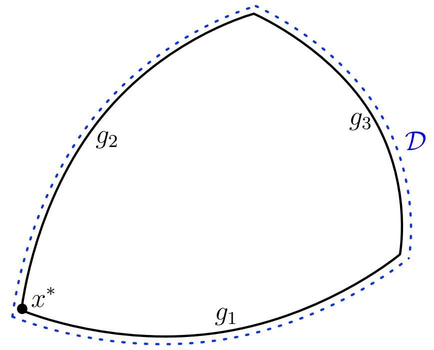
定義 4 (可行方向性質二:z_2). *Let* $f(x^*)$ *denotes the objective function. Then* $z_2(x^*) = \Big\{z \in \mathbb{R}^n \mid z'\nabla f(x^*) > 0\Big\}$.

我們假設一共有 n 條不等式約束 (g_i)、p 條等式約束 (h_i)，同時在這 n 條不等式約束中，一共有 m 條 binding 約束[31](請看圖 6)，另外 f、g_1、g_2、g_3 都是連續可微的函數。上方的圖 14 就是庫

[30]兩向量相乘時，前面的向量要轉置 (transpose)。
[31]因此，$n \geq m$。

恩-塔克定理的示意圖，此優化問題沒有等式約束。此圖是在三個不等式約束: $g_1(x) \geq 0$、$g_2(x) \geq 0$、$g_3(x) \geq 0$ 的限制下，求目標函數 $f(x)$ 的最大值。這三個約束所圍成的面積[32]是一個閉集合 (closed set)[33]，但為了使最優點皆為內點 (interior point) 而非邊界點 (boundary point)，我們取包含這三條約束所圍成面積的最小開集 (open set)[34]，並稱為「可行集 (feasible set)，$\mathcal{D}$」。

圖 15: 可行集



定義 5 (可行集). *Let $\mathcal{D}$ be a feasible (open) set. $\mathcal{D} = \Big\{x \in \mathbb{R}^n \mid g(x) \geq 0,\ h(x) = 0\Big\} \cap \mathcal{U}$ where $\mathcal{U}$ is an open set.*

前面透過線性化錐形，給可行方向 z 加上最寬鬆的限制。可行方向除了要在線性化錐形中之外，其方向也指向可行集 $\mathcal{D}$。性質一、二為較嚴格的條件，稍後即說明。圖 14 中的三個約束，只有 g_1 和 g_2 是 binding 約束，g_3 和 x^* 毫無關係，因此為 nonbinding 約束。圖底部的線代表 $g_1 = 0$，線內 (可行集) 區域是 $g_1 > 0$，線外區域是 $g_1 < 0$。由此可見約束 $g_1 \geq 0$ 包含圖底的 $g_1 = 0$ 線與可行集，而可行集則符合 $g_1, g_2, g_3 > 0$。

在很前面的時候，我們在式 1 有提到「梯度」就是函數增長最快的方向。假設在圖 14 中，x^* 是最優點 (極大值)，則兩個 binding 約束 (g_1, g_2) 在 x^* 的梯度就是分別與兩條約束垂直的線 ∇g_1 和 ∇g_2。因為這兩個不等式約束都是 ≥ 0，因此其成長方向為指向 > 0 的地方，也就是指向可行集 $\mathcal{D}$。接下來，我們要找目標函數 f 的梯度。

已知 x^* 是可行集中的極大值，因此如果 f 函數向右上、往可行集內移動，其值會越來越小: $f(x^*) > f(x)\ \forall x \in \mathcal{D}$。所以 ∇f 會指向左下方，遠離可行集。找到目標函數與約束的梯度後，我們就可回到可行方向的兩個性質 z_1 和 z_2。

性質一 (z_1) 說可行方向必在約束梯度 (∇g_i) 所形成的凸錐中。性質二 (z_2) 表示可行方向一定和目標函數梯度呈銳角關係。使用庫恩-塔克定理的先決條件就是: 在最優點 x^* 的可行方向 z 須滿足性質一，但必須違背性質二:

$$z_1(x^*) \cap z_2(x^*) = \phi \tag{16}$$

[32]為了示意，我們都假設是在二維平面上進行討論。

[33]因為約束是可微函數，其圍成的面積包含了所有的極限點 (limit point)w^*，且 $f'(w^*) = 0$，因此這個面積是一個閉集合。

[34]所謂開集合，代表在集合 G 中的任意點 x，我們都可以 x 點為中心，畫一個半徑為 r、同時也至少包含集合 G 中另一個非 x 點的開球 (open ball)，同時這個球也被包含於 G 集合內。

也就是說，z 必須要在約束所形成的凸錐內，同時這個方向不能讓目標函數增長，也就是要和目標函數梯度呈反方向 (如圖 14 所示: z 和 ∇f 呈鈍角關係、和 $-\nabla f$ 呈銳角關係)。注意，如果 $z_1(x^*) \cap z_2(x^*) \neq \phi$(兩性質皆滿足)，這不代表局部極值不存在，而是極值有可能存在，但我們不能使用庫恩-塔克定理求解，必須尋找其他方法。

圖中的 $-\nabla f$ 是 f 梯度的反向方向向量。這時我們可以看到，$-\nabla f$ 在 ∇g_1 和 ∇g_2 所形成的凸錐 (convex cone) 中。根據法卡斯引理 (i)，這時必存在以下關係:

$$-\nabla f(x^*) = \lambda_1 \nabla g_1(x^*) + \lambda_2 \nabla g_2(x^*) \tag{17}$$

式 16 就是庫恩-塔克定理 (i) 的結論。

定理 7 (Kuhn-Tucker Theorem). *Suppose* $x^* \in \mathcal{D} = \left\{x \in \mathbb{R}^n \mid g(x) \geq 0,\ h(x) = 0\right\} \cap \mathcal{U}_{open}$ *is a solution to the optimization problem with* m *inequality and* p *equality constraints and* $z_1(x^*) \cap z_2(x^*) = \phi$*. And suppose* $rank\left[Dg_{effective}(x^*),\ Dh(x^*)\right] = k$ *where* k *is the number of effective constraint at* x^**. Then* $\exists$ *vectors* $\lambda^* \in \mathbb{R}^m$ *and* $\mu^* \in \mathbb{R}^p$ *such that the following holds:*

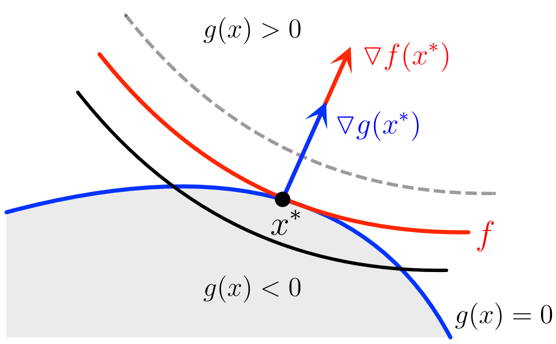
(i) $\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j^* \nabla h_j(x^*) = 0.$

(ii) $\lambda_i^* g_i(x^*) = 0\ \forall i = 1, ..., m$ *(complementary slackness).*

(iii) $\lambda^* \geq 0.$

定理 7 中的 (iii) 是法卡斯引理 (i) 的條件 (λ 為非負值)。定理 7 中的 (iii) 則是法卡斯引理 (i) 的結論。由此看出，庫恩-塔克定理是由法卡斯引理所得到的。其中，$rank\left[Dg_{effective}(x^*),\ Dh(x^*)\right] = k$ 為前面提到的滿秩條件，其中 k 是有效約束的數量[35]。這等式也就是說: 約束規範 (constraint qualification) 必須成立。接下來，我們要討論定理 7(ii) 出現的互補寬鬆條件 (complementary slackness)。為了讓公式簡化，以下我們將用 $\boldsymbol{x}$ 向量來代表在二維平面上的一組座標 (x, y)。

圖 16: binding 優化問題



以圖 16 的優化問題為例，在 binding[36]約束 $g(\boldsymbol{x}) \geq 0$ 下最大化 f。在此情形下，函數 f 的極值 $f(\boldsymbol{x}^*)$ 會發生在定義域 (圖 16 灰色區域) 外；但因為受到約束 $g(\boldsymbol{x}) = 0$ 的作用，極值將發生在定義

[35] 有效約束 = 等式約束 + binding 的不等式約束。
[36] 當一個不等式 $g(\boldsymbol{x})$ 為 binding 時，最優點 $\boldsymbol{x}^*$ 會發生在約束 $g(\boldsymbol{x}) = 0$ 上。換句話說，這時 $g(\boldsymbol{x}) = 0$ 才會發生作用。

域，也就是灰色區域的邊界上。同時，因限制式 g 在定義域外是 $g(\boldsymbol{x}) > 0$ 以及目標函數的「潛在」最大值在定義域外，造成目標函數和約束的梯度都指向右上方 (向外)。雖然兩梯度長度不同，但因為方向「相同[37]」，兩者長度呈等比例 λ 放大、縮小的關係，故可寫成:

$$\nabla f(\boldsymbol{x}^*) = \lambda \nabla g(\boldsymbol{x}^*),\ \lambda \geq 0 \ \text{ 或 } \ \frac{\nabla f(\boldsymbol{x}^*)}{\nabla g(\boldsymbol{x}^*)} = \lambda \tag{18}$$

若 $g(\boldsymbol{x}) \geq 0$ 是 nonbinding 約束，則最優點 $\boldsymbol{x}^*$ 將在約束邊界 $g(\boldsymbol{x}) = 0$ 之內、圖 16 灰色區域內，不受約束的影響。因為 $\boldsymbol{x}^*$ 是目標函數 f 的最優點，所以根據一階條件可得到: $\frac{\partial f}{\partial x}(\boldsymbol{x}^*) = 0$。若把問題用庫恩-塔克定理列式，則會得到 $\mathcal{L}(\boldsymbol{x}^*, \lambda) = f(\boldsymbol{x}^*) + \lambda\big[g(\boldsymbol{x}^*)\big]$。所以當約束是 nonbinding 時，其一階條件為:

$$\frac{\partial \mathcal{L}}{\partial x}(\boldsymbol{x}^*, \lambda) = \underbrace{\frac{\partial f}{\partial x}(\boldsymbol{x}^*)}_{=0} + \lambda \underbrace{\frac{\partial g}{\partial x}(\boldsymbol{x}^*)}_{\neq 0} = 0$$

在優化問題中，一個 nonbinding 的約束式等同沒有效果，所以令 $\lambda = 0$ 來忽略約束式 g，並讓上述等式成立。最後，綜合這兩種情況：當約束為 binding 時 $g(\boldsymbol{x}) = 0$，$\lambda \geq 0$；當約束為 nonbinding 時 $g(\boldsymbol{x}) \geq 0$，$\lambda = 0$。將兩種情況寫在一起，即可得到互補寬鬆條件:

$$\lambda g(\boldsymbol{x}) = 0 \tag{19}$$

這邊所討論的庫恩-塔克定理，可以用在單純只有等式[38]或不等式約束的問題，也可以用於包含混合約束的情況。

接下來，我們將提供處理優化問題的標準流程，其中 $\boldsymbol{x}$ 是座標向量:

1. 使用 Weierstrass 定理檢查定義域是否為緊集 (compact set)、目標函數是否連續 (continuous)，證明極值存在
2. 檢查目標函數、所有的約束是否為 C^1 函數 (函數 1 階導數連續)
3. 透過雅可比矩陣 (Jocobian matrix)，檢查約束規範 (constraint qualification) 是否滿足
4. 將 binding 的不等式約束移項，調整成 $g(\boldsymbol{x}) \boldsymbol{\geq 0}$ 的形式
5. 若求極大值: Max $f(\boldsymbol{x})$，使用庫恩-塔克列式:
   $\mathcal{L} =$ 目標函數 $+\ \mu$(等式約束) $+\ \lambda$(binding 的不等式約束)
6. 算一階條件 (F.O.C)
7. 加上互補寬鬆條件 (complementary slackness): $\lambda g(\boldsymbol{x}) = 0$
8. 不等式約束乘數 (multiplier) 非負: $\lambda \geq 0$

[37] 因為方向相同，故此比例乘數 λ 為正值。
[38] 如果優化問題只有等式約束，則變成前面討論過的 Lagrange 定理。

9. 列出等式約束: $h(\boldsymbol{x}) = 0$

10. 列出不等式約束: $g(\boldsymbol{x}) \geq 0$

11. 將 6-10 列出來的所有式子聯立，解出最優點 (臨界點, critical point)

12. 驗證那些使約束規範失效的點 (與臨界點比較)，檢查那些點是否才是真正的最優點 (極值)。另外，要記得檢查所有變數、乘數皆為 0 的特殊情況：$x = y = ... = 0;\ \lambda_1 = \lambda_2 = ... = \lambda_n = 0.$

13. 若求極小值: $\min\ f(\boldsymbol{x})$，則當成求負目標函數的極大值: $\text{Max}\ -f(\boldsymbol{x})$，使用庫恩-塔克列式:

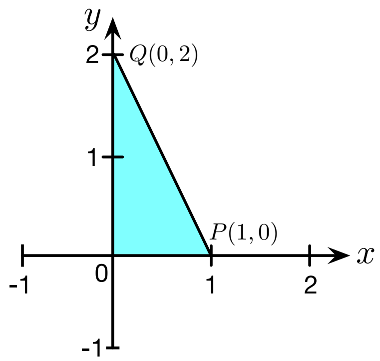
$$\mathcal{L} = -(\text{目標函數}) + \mu(\text{等式約束}) + \lambda(\text{binding 的不等式約束})$$

並執行 6-12。最後要記得，所求出的函數極值，是 $-f$ 的值。

♦ 大學期間所遇到的基礎優化問題，可直接從 4 開始，並可跳過 12。若優化問題沒有不等式約束，則可跳過 7,8,10。

$\boxed{Example}$ Find the maximizer of $f(x, y) = x^2 + y^2$, subject to the constraints $2x + y \leq 2, x \geq 0, y \geq 0$.[39]

$\boxed{Solution}$ From the three constraints: $2x + y \leq 2, x \geq 0$ and $y \geq 0$, we can conclude that the domain ($Area OPQ$ in the graph below) of this function is closed and bounded.



By the Weierstrass Theorem, compactness indicates the existence of an extreme value. Then we check the rank of the Jocobian of the constraints function:

$$\begin{pmatrix} 2-2x-y \\ x \\ y \end{pmatrix} \xrightarrow{\text{Jocobian Matrix}} \begin{pmatrix} -2 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{\text{Row Operation}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

where the row echelon form indicates that the Jocobian matrix has rank two. Since there are only two variables, so at most two of the three constraints can be binding at any time. Hence, the constraint qualification holds at any solution candidate. Next, we will form the Lagrangian equation.

[39] This example is taken from *Mathematics for Economists.*

$$\mathcal{L} = x^2 + y^2 + \lambda_1(2 - 2x - y) + \lambda_2(x) + \lambda_3(y)$$

The first order conditions are the following:

$$\frac{\partial \mathcal{L}}{\partial x} = 2x - 2\lambda_1 + \lambda_2 = 0 \tag{20}$$
$$\frac{\partial \mathcal{L}}{\partial y} = 2y - \lambda_1 + \lambda_3 = 0 \tag{21}$$
$$\lambda_1(2 - 2x - y) = 0 \tag{22}$$
$$\lambda_2(x) = 0 \tag{23}$$
$$\lambda_3(y) = 0 \tag{24}$$
$$\lambda_1 \geq 0 \tag{25}$$
$$\lambda_2 \geq 0 \tag{26}$$
$$\lambda_3 \geq 0 \tag{27}$$
$$2 - 2x - y \geq 0 \tag{28}$$
$$x \geq 0 \tag{29}$$
$$y \geq 0 \tag{30}$$

Then we will solve for the critical point base on the ten equations, from (20) to (30).

$\boxed{\text{CASE I}}$ Let $x = 0$.

Then from (20): $\lambda_2 = 2\lambda_1$, from (23): $\lambda_2 \neq 0$ [40] $\rightarrow \lambda_1 \neq 0$. Then from (22): $2 - 2x - y = 0 \rightarrow y = 2 \neq 0 \rightarrow (24) : \lambda_3 = 0$. Then (21): $\lambda_1 = 2y = 4$. And finally, $\lambda_2 = 2\lambda_1 = 8$. Here we have the critical point: $(0, 2)$ with $\boldsymbol{f(0,2) = 4}$.

$\boxed{\text{CASE II}}$ Let $y = 0$.

Then from (24): $\lambda_3 \neq 0$. From (21): $\lambda_1 = \lambda_3 \neq 0$, so from (22): $2 - 2x - y = 0 \rightarrow x = 1$. Then from (23): $\lambda_2 = 0$ and from (20): $\lambda_1 = 1$. Finally, from (21) again: $\lambda_3 = 1$. The critical point is $(1, 0)$ and $\boldsymbol{f(1,0) = 1}$.

$\boxed{\text{CASE III}}$ Let $x = 0$ and $y = 0$.

This is the *trivial solution*. In this case, we neglect the complementary slackness requirements such that $x = y = \lambda_1 = \lambda_2 = \lambda_3 = 0$. The critical point is $(0, 0)$ with $\boldsymbol{f(0,0) = 0}$.

$\boxed{\text{CASE IV}}$ Let $x \neq 0$ and $y \neq 0$.

Then from (23), (24): $\lambda_2 = \lambda_3 = 0$. From (20), (21): $x = 2y$. Then from (21): since $\lambda_1 = 2y$ and $y \neq 0 \rightarrow \lambda_1 \neq 0$. So from (22): $2 - 2x - y = 0 \rightarrow 2 = 5y \rightarrow y = \frac{2}{5}$. Then $x = 2y = \frac{4}{5}$ and $\lambda_1 = \frac{4}{5}$. The critical point is $\left(\frac{4}{5}, \frac{2}{5}\right)$ and $\boldsymbol{f\left(\frac{4}{5}, \frac{2}{5}\right) = 0.8}$.

[40] The complementary slackness condition states that only one of either constraint or the multiplier can be 0.

Then we compare the function value for these four cases and conclude that the global maximum of the function occurs at the critical point $(0, 2)$ with the function value of 4.

References

Simon, Carl P. and Lawrence E. Blume. 1994. *Mathematics for Economists.* W. W. Norton & Company.

Sundaram, Rangarajan K. 1996. *A First Course in Optimization Theory.* Cambridge University Press.