Partial Derivatives (偏導數)

$\boxed{\text{differential of } f(\mathbf{x}),\ \mathbf{x}\in\mathbb{R}^n}$ [1] $\Rightarrow$

- $\boxed{\text{tangent plane/linear approx.}}$ [2] ,
- $\boxed{\text{derivative wrt parameter}}$ [3] $\Rightarrow$ $\boxed{\text{Taylor expansion}}$ [6] $\Rightarrow$ $\boxed{\text{quadratic approx.}}$ [7] $\Rightarrow$ $\boxed{\text{find local extrema}}$ [8]
- $\boxed{\text{gradient and directional derivative}}$ [4] .... $\Rightarrow$ $\boxed{\text{Lagrange multipliers}}$ [9]
- $\boxed{\text{chain rule for multi-variable fucntions}}$ [5] ,

---

- Basics: 之前學過的,

  Let $\vec{u}=(u_1,u_2,u_3)\in\mathbb{R}^3$ and $\vec{v}=(v_1,v_2,v_3)\in\mathbb{R}^3$.

  Inner product(內積) $\vec{u}\bullet\vec{v}:=u_1v_1+u_2v_2+u_3v_3$, 讀作 "u dot v";

  Cross product(外積) $\vec{u}\times\vec{v}:=\begin{vmatrix}\hat{i}&\hat{j}&\hat{k}\\u_1&u_2&u_3\\v_1&v_2&v_3\end{vmatrix}$, 讀作 "u cross v", 其中 $\begin{array}{l}\hat{i}=(1,0,0)\\\hat{j}=(0,1,0)\\\hat{k}=(0,0,1)\end{array}$, 也記作 $\begin{array}{l}e_1\\e_2\\e_3\end{array}$ 。

  很明顯地, $(\vec{u}\times\vec{v})\perp\vec{u}$ and $\vec{v}$.

  若 (x_0,y_0,z_0) 是平面$\mathcal{H}$ 上的一點, $\vec{n}$ 是平面$\mathcal{H}$ 的 normal(法向量), 則平面$\mathcal{H}$ 的方程式爲
  $\mathcal{H}:\vec{n}\bullet(x-x_0,y-y_0,z-z_0)=0$, 意思就是: 該平面上的向量必然和該平面的 normal 垂直。

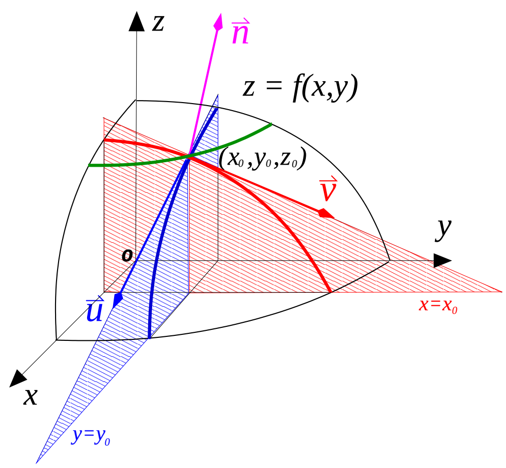


Figure 1: 沿座標平面方向的變化方式

- 假設 (x_0,y_0,z_0) 是曲面 $S:z=f(x,y)$ 上的一個點, 即 $f(x_0,y_0)=z_0$.

  當以 $y=y_0$ 代入 $z=f(x,y)$, 即 $y=y_0$ 藍色平面 與 $z=f(x,y)$ 曲面相交於藍色曲線, 在藍色平面上, 可以求 函數 $f(x,y)$ 在 y 固定成 y_0 時, y_0 附近的變化率: $\frac{\partial f}{\partial x}(x_0,y_0)=f_x(x_0,y_0)=\lim\limits_{\triangle x\to 0}\dfrac{f(x_0+\triangle x,y_0)-f(x_0,y_0)}{\triangle x}$
  $\Longrightarrow$ 切向量 $\vec{u}=(1,0,f_x(x_0,y_0))$.

  當以 $x=x_0$ 代入 $z=f(x,y)$, 即 $x=x_0$ 紅色平面 與 $z=f(x,y)$ 曲面相交於紅色曲線, 在紅色平面上, 可以求 函數 $f(x,y)$ 在 x 固定成 x_0 時, x_0 附近的變化率: $\frac{\partial f}{\partial y}(x_0,y_0)=f_y(x_0,y_0)=\lim\limits_{\triangle y\to 0}\dfrac{f(x_0,y_0+\triangle y)-f(x_0,y_0)}{\triangle y}$
  $\Longrightarrow$ 切向量 $\vec{v}=(0,1,f_y(x_0,y_0))$. 所以 垂直於曲面 的 normal $\vec{n}$ 和

$$\vec{u}\times\vec{v}=\begin{vmatrix}\hat{i}&\hat{j}&\hat{k}\\1&0&f_{x(x_0,y_0)}\\0&1&f_{y(x_0,y_0)}\end{vmatrix}=-(f_x(x_0,y_0),f_y(x_0,y_0),-1)$$ 平行。

因此, 很容易就得到 S 在 (x_0, y_0, z_0) 處的切平面 $\mathcal{H}$ 的方程式:

$$\begin{aligned}
\mathcal{H}: &\quad (\vec{u}\times\vec{v})\bullet(x-x_0, y-y_0, z-z_0) = 0\\
\Longleftrightarrow &\quad f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0) - 1(z-z_0) = 0\\
\Longleftrightarrow &\quad z - f(x_0,y_0) = f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)
\end{aligned}$$

還記得所謂的 *linearization (linear approximation)* 嗎? 同樣地, 我們也要用此說明 $f(x,y)$ 的 differential 怎麼得到的。由 $\mathcal{S}$ 在 (x_0, y_0, z_0) 處的切平面 $\mathcal{H}$ 的方程式

$$\mathcal{H}: z = f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$$

得知 $f(x,y)$ 在(x_0,y_0)處的 linearization 爲

$$f(x,y) \approx f(x_0,y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0),$$

$$\begin{array}{lllllll}
\text{即} & f(x,y)-f(x_0,y_0) & \approx f_x(x_0,y_0) & (x-x_0) & +f_y(x_0,y_0) & (y-y_0), \\
\text{即} & \triangle f & \approx f_x(x_0,y_0) & \triangle x & +f_y(x_0,y_0) & \triangle y, \\
 & \downarrow & & \downarrow & & \downarrow \\
\Longrightarrow & df & = f_x(x_0,y_0) & dx & +f_y(x_0,y_0) & dy
\end{array}$$

df —— 函數值的 微變化**量**, 和以前一樣, 叫做 *(total) differential.*

- 多變量函數 $f(x,y,\cdots)$ 的 differential df 是 f在沒有特定變化方式時的微變化量 (所有變量皆可稍微變動)。其 partial derivative 考慮的則是 只有某特定變數在變 (其他不動) 時的變化率, 是具有相對概念的。所以 df 有意義, $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \cdots$ 等 partial derivatives 有意義, 但是 ∂f 沒有意義。且 當 $\begin{cases} x = x(u,v) \\ y = y(u,v)\end{cases}$, $f = f\big(x(u,v), y(u,v)\big)$ 在 只有u變化時的 變化率 (“df 除以 du”) 現在就要寫成 $\cancel{\frac{df}{du}}\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\cancel{\frac{dx}{dt}}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\cancel{\frac{dy}{dt}}\frac{\partial y}{\partial u}$, 因爲 f, x, y 的變量不光是 u。同樣地,f 在 只有v變化時的 變化率 爲 $\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial v}$, 這就是所謂的 *chain rule*。

- 若將 $\mathcal{S}: z = f(x,y)$ 裡面的 x, y 參數化, i.e. 沿某曲線行進, 得到的是 S 上面 的曲線。則根據上面得到的 df, 我們可以求出 f 在該曲線上的變化**率** df/dt:

$$\frac{df}{dt} = f_x\frac{dx}{dt} + f_y\frac{dy}{dt} = (f_x, f_y)\bullet(\dot{x},\dot{y}) = \nabla f\bullet(\dot{x},\dot{y}),$$

因此得知, 如果希望函數的變化率高,$(\dot{x},\dot{y})$ 就必須朝 ∇f 的方向走 才會使得 $|\frac{df}{dt}|$ 裡的 inner product 最大。這個方向 $\nabla f \stackrel{\text{def}}{=} (f_x, f_y)$ 就叫做 *gradient*(梯度)。把上面 $\frac{d}{dt}[f\big(x(t),y(t)\big)]$ 式子中的切向量 $(\dot{x},\dot{y})$ 換成單位向量 $\mathbf{u}$, 就是 *directional derivative* $\mathbf{D_u}f = \nabla f\bullet\mathbf{u}$ 。

- 假設 點(x_0, y_0, z_0) 在曲面$\mathcal{S}: z = f(x,y) \in \mathbb{R}^3$ 上 (即 $f(x_0,y_0) = z_0$), 用 $z = z_0$ 砍 $\mathcal{S}$, 就得到 *level curve* $\mathcal{C}: f(x,y) = z_0$, 將此曲線 $\mathcal{C}$ 參數化後 對參數 t 求導: $\frac{d}{dt}f\big(x(t),y(t)\big) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})\bullet(\dot{x},\dot{y}) = \frac{d}{dt}z_0 = 0$, 意即 $(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}) \perp (\dot{x},\dot{y})$, 意即: 在曲線$\mathcal{C}$上 每一點的 $\nabla f = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y})$ 皆垂直於 曲線$\mathcal{C}: f(x,y) = z_0$。

  同一個曲面$\mathcal{S}$, 用另一個方式來看, 曲面 $\mathcal{S}$ 可被視爲 $w = g(x,y,z) := f(x,y) - z \in \mathbb{R}^4$ 用 $w = 0$ 砍後 所得到的 *level surface*。將 $\mathcal{S}$ 上通過 (a,b,c) 的任一曲線參數化後 對參數 t 求導 $\frac{d}{dt}g\big(x(t),y(t),z(t)\big) = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z})\bullet(\dot{x},\dot{y},\dot{z}) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1)\bullet(\dot{x},\dot{y},\dot{z}) = \frac{d}{dt}c = 0$, 意即 $(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}) \perp (\dot{x},\dot{y},\dot{z})$, 意即: 在曲面$\mathcal{S}$上 每一點的 $\nabla g = (\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}, \frac{\partial g}{\partial z}) = \nabla\big(f(x,y)-z\big) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1) = (\nabla f, -1)$ 垂直於曲面 $\mathcal{S}$。(用前面的 Figure 1 來看, $\vec{n} /\!\!/ (\vec{u}\times\vec{v}) = -\nabla\big(f(x,y)-z\big)$ 本來就是垂直於曲面的)

  所以, 三維向量 $(\nabla f(x_0,y_0), -1) = (\frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial y}(x_0,y_0), -1)$ 垂直曲面 $\mathcal{S}: f(x,y) = z$ 於(x_0, y_0, z_0)處,

  投影到 xy 平面 $\Downarrow$

  二維向量 $\nabla f(x_0,y_0) = (\frac{\partial f}{\partial x}(x_0,y_0), \frac{\partial f}{\partial y}(x_0,y_0))$ 垂直曲線 $\mathcal{C}: f(x,y) = z_0$ 於(x_0, y_0)處。

一般來說, 如果 $\mathbf{r}=(x_1,\cdots,,x_n)\in\mathbb{R}^n$, $f(\mathbf{r})\in\mathbb{R}$ 的 gradient $\nabla f(\mathbf{r})=(\frac{\partial f}{\partial x_1},\cdots,\frac{\partial f}{\partial x_n})\in\mathbb{R}^n$, 並且

$$\text{函數值的微變化量 } df=\nabla f\bullet d\mathbf{r}$$

若 t 是 $\mathbf{r}$ 的變量, $\xrightarrow{\text{除以 } dt}$ 則變化率 $\left\{\begin{array}{ll}\frac{df}{dt}=\nabla f\bullet\frac{d\mathbf{r}}{dt}, & \text{if } \mathbf{r}=\mathbf{r}(t),\\ \frac{\partial f}{\partial t}=\nabla f\bullet\frac{\partial\mathbf{r}}{\partial t}, & \text{if } \mathbf{r}=\mathbf{r}(t,u,v,\cdots).\end{array}\right\}$, 即一般的 chain rule。

$\frac{d\clubsuit}{d\spadesuit}$, $\frac{\partial\clubsuit}{\partial\spadesuit}$ 兩個看似不同的符號 只是用來區分 單變量/多變量 函數的導數罷了, 本質上還是 differential 和 differential 間的比值。

- 令 $\left\{\begin{array}{l}x(t)=a+ht,\\ y(t)=b+kt,\end{array}\right.$ 爲 $\underset{(t=0)}{(a,b)}$ 和 $\underset{(t=1)}{(a+h,b+k)}$ 連線 的參數式, 則$g(t):=f\big(x(t),y(t)\big)$ 僅有一個變量 t, 我們就可以求 $g(t)$ 在 $t=0$ 處的 Taylor expansion:

$$\begin{array}{rcl}g(t)&=&g(0)+g'(0)(t-0)+\frac{g''(0)}{2!}(t-0)^2+\frac{g'''(0)}{3!}(t-0)^3+\cdots\\ g(1)&=&g(0)+g'(0)+\frac{g''(0)}{2!}+\frac{g'''(0)}{3!}+\cdots\end{array}$$

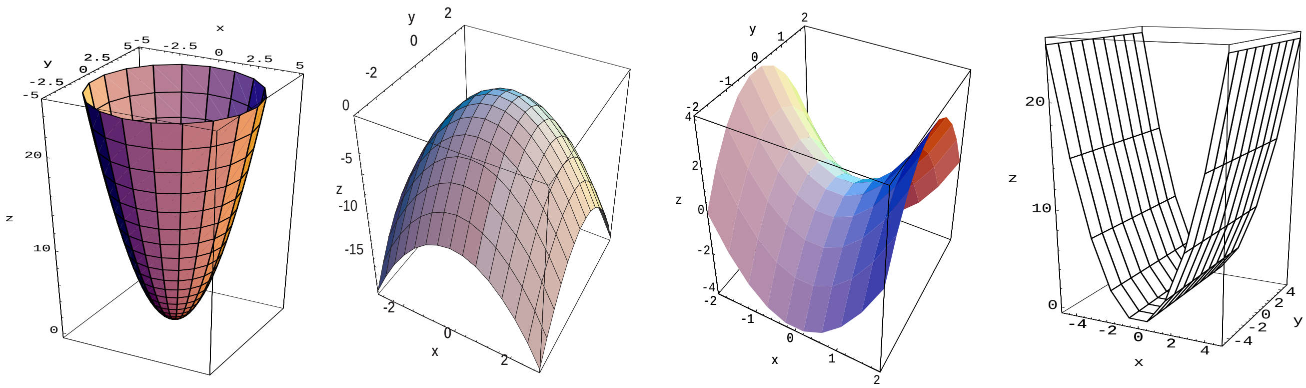
$$\begin{array}{l}g(0)=f\big(x(0),y(0)\big)=f(a,b)\\ g(1)=f\big(x(1),y(1)\big)=f(a+h,b+k)\end{array},$$

$$\begin{array}{rcll}g'(t)&=&\frac{d}{dt}(f)=\frac{dx}{dt}\frac{\partial f}{\partial x}+\frac{dy}{dt}\frac{\partial f}{\partial y}=h\frac{\partial f}{\partial x}+k\frac{\partial f}{\partial y}=(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})f, & \big(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\text{算子對}f\text{作用1次}\big)\\ g''(t)&=&\frac{d}{dt}(\frac{df}{dt})=(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})(\frac{df}{dt})=(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})f & \\ &=&(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^2f, & \big(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\text{算子對}f\text{作用2次}\big)\\ & & (\text{即 } h^2\frac{\partial}{\partial x}\frac{\partial}{\partial x}f+hk\frac{\partial}{\partial x}\frac{\partial}{\partial y}f+hk\frac{\partial}{\partial y}\frac{\partial}{\partial x}f+k^2\frac{\partial}{\partial x}\frac{\partial}{\partial x}f) & \\ &\vdots& & \\ g^{(i)}(t)&=&(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^if, & \big(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y}\text{算子對}f\text{作用 }i\text{次}\big)\end{array}$$

$$\begin{array}{rl}f(a+h,b+k)&=f(a,b)+\dfrac{(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^1f|_{(a,b)}}{1!}+\dfrac{(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^2f|_{(a,b)}}{2!}+\cdots+\dfrac{(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^nf|_{(a,b)}}{n!}+\cdots\\ &=\displaystyle\sum_{n=0}^{\infty}\dfrac{(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial y})^nf|_{(a,b)}}{n!}\end{array}$$

$$\text{Let }\left\{\begin{array}{l}x=a+h,\\ y=b+k,\end{array}\right.,\Longrightarrow f(x,y)=\sum_{n=0}^{\infty}\frac{\Big((x-a)\frac{\partial}{\partial x}+(y-b)\frac{\partial}{\partial y}\Big)^nf\,\Big|_{(a,b)}}{n!}.$$

$$\left(\begin{array}{l}\text{In general, for } \mathbf{x}\in\mathbb{R}^n\text{, Taylor expansion of } f(\mathbf{x}) \text{ about } \mathbf{y}\in\mathbb{R}^n \text{ is}\\ \begin{array}{rl}f(\mathbf{x})&=f(\mathbf{y})+((\mathbf{x}-\mathbf{y})\bullet\nabla)f\,|_{\mathbf{y}}+((\mathbf{x}-\mathbf{y})\bullet\nabla)^2f\,|_{\mathbf{y}}/2!+\cdots\\ &=\displaystyle\sum_{n=0}^{\infty}\frac{((\mathbf{x}-\mathbf{y})\bullet\nabla)^nf\,\Big|_{\mathbf{y}}}{n!},\\ &\text{where } \nabla:=(\frac{\partial}{\partial x_1},\cdots,\frac{\partial}{\partial x_n}).\end{array}\end{array}\right)$$



(1) (2) (3) (4)

Figure 1: critical point 附近 二次近似的 可能情況

- **用 二次曲面 來近似** $z = f(x, y)$

假設 $f(x, y)$ 是 至少二階可微, 有了 $f(x, y)$ 在 (a, b) 處的 Taylor expansion,

$$
\begin{aligned}
f(x,y) &= \underbrace{\underbrace{f(a,b) + \begin{pmatrix} f_x(a,b)(x-a) \\ +f_y(a,b)(y-b) \end{pmatrix}}_{\text{linear approximation } \ell(x,y)} + \frac{1}{2}\overbrace{\begin{pmatrix} f_{xx}(a,b)(x-a)^2 \\ +f_{xy}(a,b)(x-a)(y-b) \\ +f_{yx}(a,b)(x-a)(y-b) \\ +f_{yy}(a,b)(y-b)^2 \end{pmatrix}}^{\widetilde{q}(x,y)}}_{\text{quadratic approximation } q(x,y)} + \underbrace{\boxed{\text{次數大於等於 3 的項}}}_{\mathcal{O}(3)} \\
&= q(x,y) + \mathcal{O}(3)
\end{aligned}
$$

$q(x, y)$ 稱爲 $f(x, y)$ 在 (a, b) 附近的 quadratic approximation(二次函數近似)。

令 $A := f_{xx}(a, b)$, $B := f_{xy}(a, b)$, $C := f_{yy}(a, b)$ 節省符號, 因爲 有無常數項、一次項、乘 $\frac{1}{2}$ 皆不影響形狀, 所以看 $\widetilde{q}(x, y)$ 的形狀 就等於是在看 $q(x, y)$ 形狀 : 在不失一般性的情況下, 假設 (a, b) 即 $(0, 0)$ 以節省符號。於是 二次齊次式 $\widetilde{q}(x, y) = Ax^2 + 2Bxy + Cy^2$, 根據配方法 (completing the square) 就可以寫成

$$
\begin{aligned}
z = \widetilde{q}(x,y) &= Ax^2 + 2Bxy + Cy^2 \\
&= A\left((x + \tfrac{B}{A}y)^2 - (\boldsymbol{B^2 - AC})\,(\tfrac{y}{A})^2\right)
\end{aligned}
$$

$$
\begin{array}{lll}
B^2 - AC < 0, \text{ 即 } \begin{vmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{vmatrix}_{(a,b)} > 0, \Rightarrow z = A\left((x + *y)^2 + \oplus y^2\right), & \text{— paraboloid: } \begin{cases} A > 0 \text{ 開口向上}^{(1)}, \\ A < 0 \text{ 開口向下}^{(2)}, \end{cases} \\
\qquad\qquad > 0, \qquad\qquad\qquad < 0, \Rightarrow z = A\left((x + *y)^2 + \ominus y^2\right), & \text{— saddle}^{(3)}, \\
\qquad\qquad = 0, \qquad\qquad\qquad = 0, \Rightarrow z = A(x + *y)^2, & \text{— parabolic cylinder}^{(4)},
\end{array}
$$

這些是 $\widetilde{q}$、即 q 的形狀, 也就是 f 在 (a, b) 點附近 的形狀。(複習 “`3D.pdf`”)

一個光滑函數 $f(x, y)$ 的極值 應該是會發生在 “任何方向導數均爲 0” 的地方 (*critical point*)、就說是在 (a, b) 處吧, 意即 $\mathbf{D_u} f(a, b) = \nabla f(a, b) \bullet \mathbf{u} = 0 \, \forall \mathbf{u} \in \mathbb{R}^2, \Rightarrow \nabla f(a, b) = (f_x(a, b), f_y(a, b)) = (0, 0)$, 於是,
如果 $\nabla f(a, b) = (0, 0)$, $f(x, y)$ 的極值 可由以上 (1)(2)(3)(4) 點得知 :

(1): local minimum 發生在 (a, b) 處 (即最低點),
(2): local maximum 發生在 (a, b) 處 (即最高點),
(3): no extremum 發生在 (a, b) 處 (即鞍點),
(4): inconlusive (溝底/頂).

當 (x,y) 很靠近 f 的 critical point (a,b) 時, 在 case(1)(2)(3) 的情況下, $z=f(x,y)=q(x,y)+\mathcal{O}(3)$ 和 $z=q(x,y)$ 形狀差不多, 但 case(4) 時 $z=q(x,y)$ 爲 parabolic cylinder, $f(x,y)=q(x,y)+\mathcal{O}(3)$ 有可能受到高次項 $\mathcal{O}(3)$ 的影響 而使得 cylinder 兩端稍稍向上或向下彎。故在 case(4) 下 若想得知 $z=f(x,y)$ 的形狀, 就必須另外從 $\mathcal{O}(3)$ 的部分得到更多的訊息。

- **Lagrange method**

我們知道 ∇f 提供的是 $z=f(x,y)$ 變化 (增大/減小) 最快的方向。如果對其中的 (x,y) 有所限制, 並想知道 $f(x,y)$ 的極值, 則我們希望找出 這些限制條件的 增大/減小方向 在何處會和 ∇f 的方向一致, 那麼 $f(x,y)$ 的極值就可以從這些地方求值、比較後而得到。

以虛擬的曲線參數式來解釋:

– 假設曲線 $\mathcal{C}: g(x,y)=0$ 的參數式爲 $\mathcal{C}: (x(t),y(t))$, 則 $f(x,y)\Big|_{\mathcal{C}}=f(x(t),y(t))$ 的極值, 一定發生在 $\frac{df}{dt}=\nabla f\bullet(\dot{x},\dot{y})=0$ 的地方, 即 $\nabla f\perp(\dot{x},\dot{y})$, 即 $\nabla f /\!/ \nabla g$, 即 $\nabla f=\lambda\nabla g$ for some constant λ.
Find local extrema of $z=f(x,y)$ under the constraint $g(x,y)=0$: a local extremum occures at which $\begin{cases}\nabla f=\lambda\nabla g\\ g=0\end{cases}$ for some $\lambda\in\mathbb{R}$.

– $\begin{matrix}\text{曲面 } g(x,y,z)=0\\ \text{曲面 } h(x,y,z)=0\end{matrix}$ 交出 曲線$\mathcal{C}$。假設曲線$\mathcal{C}$的參數式爲 $\mathcal{C}: (x(t),y(t),z(t))$, 則曲線$\mathcal{C}$ 的切向量 $(\dot{x},\dot{y},\dot{z})$ $/\!/ (\nabla g\times\nabla h)=\begin{vmatrix}\hat{i}&\hat{j}&\hat{k}\\ g_x&g_y&g_z\\ h_x&h_y&h_z\end{vmatrix}$, 則 $f(x,y,z)\Big|_{\mathcal{C}}=f(x(t),y(t),z(t))$ 的極值, 一定發生在 $\frac{df}{dt}=\nabla f\bullet(\dot{x},\dot{y},\dot{z})=0$ 的地方, 即 $\nabla f\bullet(\nabla g\times\nabla h)=\begin{vmatrix}f_x&f_y&f_z\\ g_x&g_y&g_z\\ h_x&h_y&h_z\end{vmatrix}=0$, 即 $\nabla f,\nabla g,\nabla h$ 線性相關, ∇f 可以寫成 ∇g 和 ∇h 的線性組合, 即 $\nabla f=\lambda\nabla g+\mu\nabla h$ for some $\lambda,\mu\in\mathbb{R}$.
Find local extrema of $w=f(x,y,z)$ under constraints $\begin{cases}g(x,y,z)=0\\ h(x,y,z)=0\end{cases}$: a local extremum occures at which $\begin{cases}\nabla f=\lambda\nabla g+\mu\nabla h\\ g=0\\ h=0\end{cases}$ for some $\lambda,\mu\in\mathbb{R}$.

In general, a local extremum of $f(\mathbf{x}),\mathbf{x}\in\mathbb{R}^n$ under constraints $\begin{cases}g_1(\mathbf{x})=0\\ \vdots\\ g_{n-1}(\mathbf{x})=0\end{cases}$ occures at which

$$\begin{cases}\nabla f(\mathbf{x}) = \lambda_1\nabla g_1(x)+\cdots+\lambda_{n-1}\nabla g_{n-1}(\mathbf{x}) & \text{for some } \lambda_1,\cdots,\lambda_{n-1}\in\mathbb{R}.\\ g_1(\mathbf{x}) = 0 & (\nabla f \text{ is a linear combination of } \nabla g_1,\cdots,\nabla g_{n-1})\\ \vdots & \\ g_{n-1}(\mathbf{x}) = 0 & \end{cases}$$