§ 10 Chain Rule

In single variable calculus, we have

Theorem 10.1 (Chain Rule in Single Variable Calculus)

Let $A, B \subseteq \mathbb{R}$ and let $x_0 \in A$.

Suppose that $f: B \to \mathbb{R}$ and $g: A \to \mathbb{R}$ are functions such that

1) $g(A) \subseteq B$ (guarantee $f \circ g: A \to \mathbb{R}$ is well-defined)
2) g is differentiable at $x_0 \in A$
3) f is differentiable at $g(x_0) \in B$

then the composite function $(f \circ g): A \to \mathbb{R}$ defined by $(f \circ g)(x) = f(g(x))$ is differentiable at x_0

and $(f \circ g)'(x_0) = f'(g(x_0)) \cdot g'(x_0)$

Furthermore, suppose that both $f: B \to \mathbb{R}$ and $g: A \to \mathbb{R}$ are differentiable.

Then, $(f \circ g): A \to \mathbb{R}$ is also differentiable.

If we let $u = g(x)$ and $y = f(u)$, then $y = f(g(x))$ is a function of x and

the chain rule can be rephased as $\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$.

Example 10.1

Let $f: B = (0, \infty) \to \mathbb{R}$ defined by $f(x) = \ln x$, and let $g: A = (0, \pi) \to \mathbb{R}$ defined by $g(x) = \sin x$.

Note that $g(A) = (0, 1] \subseteq B$, so $f \circ g: (0, \pi) \to \mathbb{R}$ is well-defined, which is $f \circ g(x) = f(g(x)) = \ln(\sin x)$.

Also, if $x \in A = (0, \pi)$, g is differentiable at x and f is differentiable at $g(x) = \sin x$, so

$(f \circ g)'(x) = f'(g(x))\, g'(x) = \frac{1}{(\sin x)} \cos x = \cot x$.

Theorem 10.2 (Chain Rule)

Let $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^k$ and let $\vec{x}_0 \in A$.

Suppose that $f: B \to \mathbb{R}^m$ and $g: A \to \mathbb{R}^k$ are functions such that

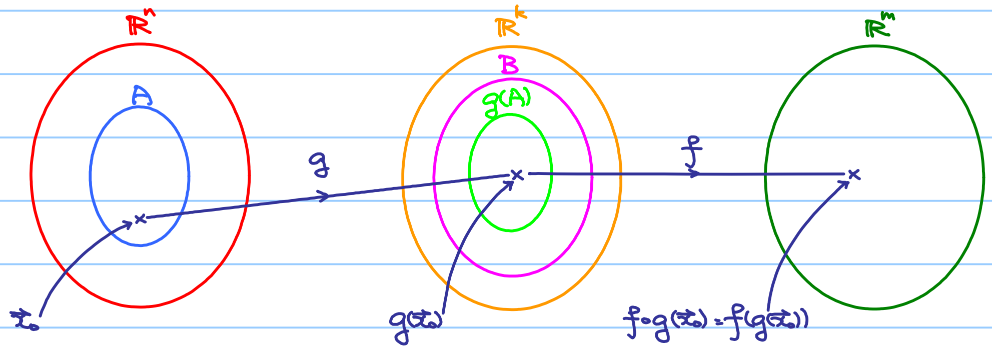
1) $g(A) \subseteq B$

2) g is differentiable at $\vec{x}_0 \in A$, i.e. $Df(g(\vec{x}_0)) \in M_{m \times k}(\mathbb{R})$ is well-defined

3) f is differentiable at $g(\vec{x}_0) \in B$, i.e. $Dg(\vec{x}_0) \in M_{k \times n}(\mathbb{R})$ is well-defined

then the composite function $(f \circ g): A \to \mathbb{R}^m$ defined by $(f \circ g)(\vec{x}) = f(g(\vec{x}))$ is differentiable at $\vec{x}_0$

and $D(f \circ g)(\vec{x}_0) = Df(g(\vec{x}_0)) \cdot Dg(\vec{x}_0) \in M_{m \times n}(\mathbb{R})$.



Remark: If $n = k = m = 1$, then $\vec{x}_0 = x_0 \in \mathbb{R}$

$$D(f \circ g)(\vec{x}_0) = [(f \circ g)'(x_0)] \in M_{1 \times 1}(\mathbb{R})$$

$$Df(g(\vec{x}_0)) = [f'(g(x_0))] \in M_{1 \times 1}(\mathbb{R})$$

$$Dg(\vec{x}_0) = [g'(x_0)] \in M_{1 \times 1}(\mathbb{R})$$

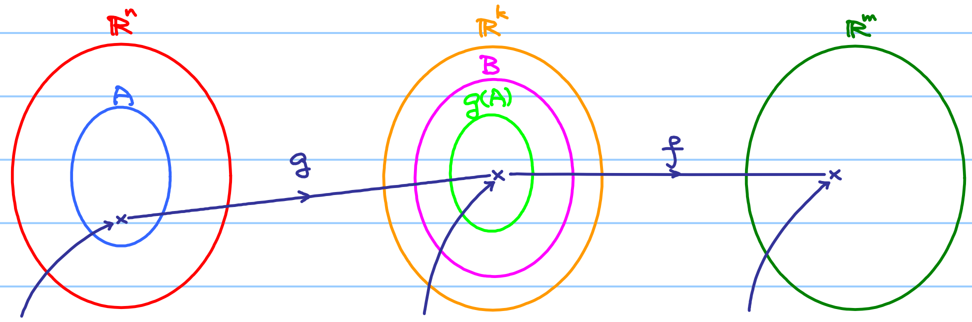
$$D(f \circ g)(\vec{x}_0) = Df(g(\vec{x}_0)) \cdot Dg(\vec{x}_0) \Rightarrow [(f \circ g)'(x_0)] = [f'(g(x_0))] \cdot [g'(x_0)]$$

$$= [f'(g(x_0))\, g'(x_0)] \in M_{1 \times 1}(\mathbb{R})$$

which is the chain rule in single variable calculus.

Furthermore, suppose that both $f: B \to \mathbb{R}^m$ and $g: A \to \mathbb{R}^k$ are differentiable.

Then $(f \circ g): A \to \mathbb{R}^m$ is differentiable.



$\vec{x} = (x_1, x_2, \cdots, x_n)$

$\vec{u} = (u_1, u_2, \cdots, u_k)$

$u_j = u_j(\vec{x}) = u_j(x_1, x_2, \cdots, x_n)$

for $j = 1, 2, \cdots, k$;

$\vec{y} = (y_1, y_2, \cdots, y_m)$.

$y_i = y_i(\vec{u}) = y_i(u_1, u_2, \cdots, u_k)$

for $i = 1, 2, \cdots, m$,

If we write $\vec{y} = (y_1, y_2, \cdots, y_m)$, $\vec{u} = (u_1, u_2, \cdots, u_k)$ and $\vec{x} = (x_1, x_2, \cdots, x_n)$

where $\vec{u} = g(\vec{x})$ and $\vec{y} = f(\vec{u}) = f(g(\vec{x}))$,

then $u_j = u_j(x_1, x_2, \cdots, x_n)$ for $j = 1, 2, \cdots, k$; $y_i = y_i(u_1, u_2, \cdots, u_k)$ for $i = 1, 2, \cdots, m$,

so $y_i = y_i(x_1, x_2, \cdots, x_n) = y_i(u_1(x_1, x_2, \cdots, x_n), u_2(x_1, x_2, \cdots, x_n), \cdots, u_k(x_1, x_2, \cdots, x_n))$.

The chain rule is:

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial y_1}{\partial x_n}(\vec{x}) \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial y_m}{\partial x_n}(\vec{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial y_1}{\partial u_1}(g(\vec{x})) & \cdots & \frac{\partial y_1}{\partial u_k}(g(\vec{x})) \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial u_1}(g(\vec{x})) & \cdots & \frac{\partial y_m}{\partial u_k}(g(\vec{x})) \end{bmatrix} \begin{bmatrix} \frac{\partial u_1}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial u_1}{\partial x_n}(\vec{x}) \\ \vdots & & \vdots \\ \frac{\partial u_k}{\partial x_1}(\vec{x}) & \cdots & \frac{\partial u_k}{\partial x_n}(\vec{x}) \end{bmatrix}$$

$D(f \circ g)(\vec{x}) \in M_{m \times n}(\mathbb{R})$ $\quad Df(g(\vec{x})) \in M_{m \times k}(\mathbb{R})$ $\quad Dg(\vec{x}) \in M_{k \times n}(\mathbb{R})$

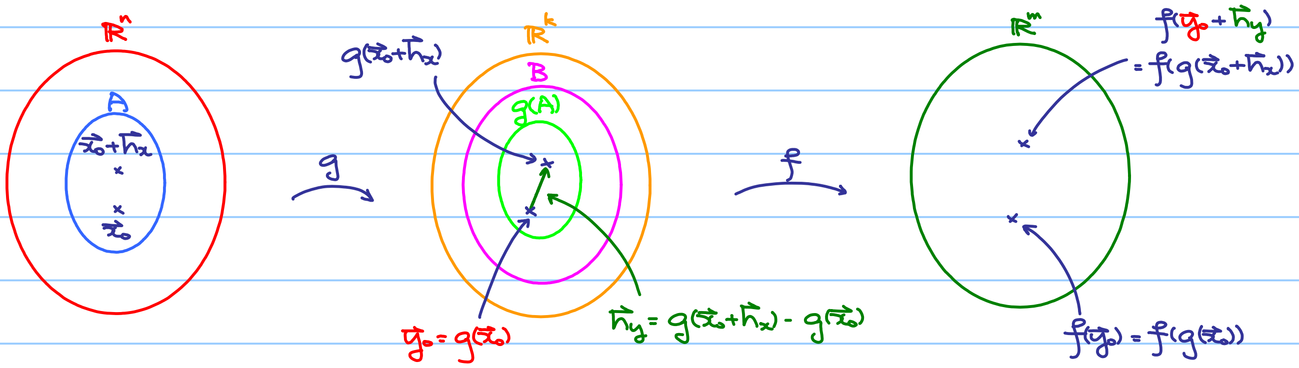
For $1 \le i \le m$, $1 \le j \le n$, $\frac{\partial y_i}{\partial x_j}(\vec{x}) = \sum_{\ell=1}^{k} \frac{\partial y_i}{\partial u_\ell}(g(\vec{x})) \cdot \frac{\partial u_\ell}{\partial x_j}(\vec{x})$

(OR write $\frac{\partial(y_1, \cdots, y_m)}{\partial(x_1, \cdots, x_n)} = \frac{\partial(y_1, \cdots, y_m)}{\partial(u_1, \cdots, u_k)} \frac{\partial(u_1, \cdots, u_k)}{\partial(x_1, \cdots, x_n)}$, so it looks like $\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$ in single variable calculus.)

Idea: Let $\varepsilon_g(\vec{h}_x) = g(\vec{x}_0+\vec{h}_x) - g(\vec{x}_0) - Dg(\vec{x}_0)\cdot\vec{h}_x$

$\varepsilon_f(\vec{h}_y) = f(\vec{y}_0+\vec{h}_y) - f(\vec{y}_0) - Df(\vec{y}_0)\cdot\vec{h}_y$

where $\lim_{\vec{h}_x\to\vec{0}} \frac{|\varepsilon_g(\vec{h}_x)|}{|\vec{h}_x|} = 0$ and $\lim_{\vec{h}_y\to\vec{0}} \frac{|\varepsilon_f(\vec{h}_y)|}{|\vec{h}_y|} = 0$.



Put $\vec{y}_0 = g(\vec{x}_0)$ and $\vec{h}_y = g(\vec{x}_0+\vec{h}_x) - g(\vec{x}_0) = Dg(\vec{x}_0)\,\vec{h}_x + \varepsilon_g(\vec{h}_x)$ we have

$$\begin{aligned}\varepsilon_f(\vec{h}_y) &= f(\vec{y}_0+\vec{h}_y) - f(\vec{y}_0) - Df(\vec{y}_0)\cdot\vec{h}_y \\ &= f(g(\vec{x}_0) + (g(\vec{x}_0+\vec{h}_x) - g(\vec{x}_0))) - f(g(\vec{x}_0)) - Df(g(\vec{x}_0))\cdot(Dg(\vec{x}_0)\cdot\vec{h}_x + \varepsilon_g(\vec{h}_x)) \\ &= f(g(\vec{x}_0+\vec{h})) - f(g(\vec{x}_0)) - [Df(g(\vec{x}_0))\cdot Dg(\vec{x}_0)]\cdot\vec{h}_x - Df(g(\vec{x}_0))\cdot\varepsilon_g(\vec{h}_x)\end{aligned}$$

$$\therefore \underbrace{Df(g(\vec{x}_0))\cdot\varepsilon_g(\vec{h}_x) + \varepsilon_f(\vec{h}_y)}_{\varepsilon_{f\circ g}(\vec{h}_x)} = f(g(\vec{x}_0+\vec{h})) - f(g(\vec{x}_0)) - \underbrace{[Df(g(\vec{x}_0))\cdot Dg(\vec{x}_0)]\cdot\vec{h}_x}_{\text{linear transformation on } \vec{h}_x.}$$

If $\lim_{\vec{h}_x\to\vec{0}} \frac{|\varepsilon_{f\circ g}(\vec{h}_x)|}{|\vec{h}_x|} = 0$, then $D(f\circ g)(\vec{x}_0) = Df(g(\vec{x}_0))\cdot Dg(\vec{x}_0)$

Example 10.2

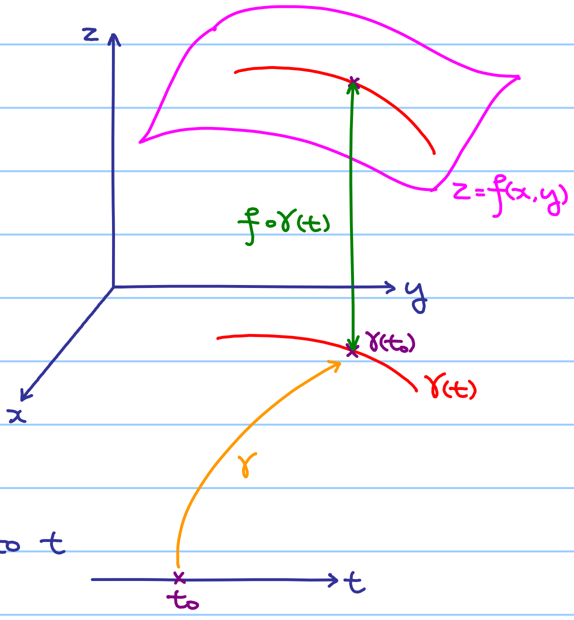
Let $f:\mathbb{R}^2\to\mathbb{R}$ defined by $f(x,y)=x^2y$ and $\gamma:\mathbb{R}\to\mathbb{R}^2$ defined by $\gamma(t)=(x(t),y(t))=(t,t^2)$.

Then $f\circ\gamma:\mathbb{R}\to\mathbb{R}$.

Note γ is a curve in $\mathbb{R}^2$,

i.e. given $t\in\mathbb{R}$, $\gamma(t)=(x(t),y(t))$ gives a point in $\mathbb{R}^2$.

Then, the point $\gamma(t)$ is inputted into f and a real number $f\circ\gamma(t)\in\mathbb{R}$ is outputted.



Question: As t changes, how does $f\circ\gamma(t)$ change?

We write $z=f(x,y)$, then $z(t)=f(\gamma(t))=f(x(t),y(t))$,

so $\frac{dz}{dt}$ is the rate of change of z with respect to t

Note $Df\in M_{1\times 2}(\mathbb{R})$, $D\gamma\in M_{2\times 1}(\mathbb{R})$ and $D(f\circ\gamma)\in M_{1\times 1}(\mathbb{R})$.

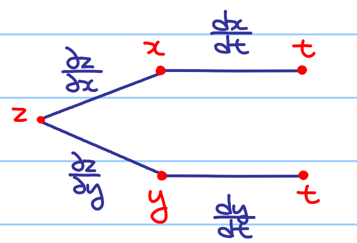
$$Df(\gamma(t))=\begin{bmatrix}\frac{\partial z}{\partial x} & \frac{\partial z}{\partial y}\end{bmatrix}=\begin{bmatrix}2xy & x^2\end{bmatrix},\quad D\gamma(t)=\begin{bmatrix}\frac{dx}{dt}\\ \frac{dy}{dt}\end{bmatrix}=\begin{bmatrix}1\\ 2t\end{bmatrix}$$

$$D(f\circ\gamma)(t)=\begin{bmatrix}\frac{dz}{dt}\end{bmatrix}=\begin{bmatrix}2xy & x^2\end{bmatrix}\begin{bmatrix}1\\ 2t\end{bmatrix}=\begin{bmatrix}2xy+2x^2t\end{bmatrix}=\begin{bmatrix}4t^3\end{bmatrix}$$

Verification: $z=f(\gamma(t))=f(t,t^2)=(t)^2(t^2)=t^4$ and so $\frac{dz}{dt}=4t^3$.

Find it difficult? Try this:

Tree diagram:



$$\frac{dz}{dt}=\frac{\partial z}{\partial x}\frac{dx}{dt}+\frac{\partial z}{\partial y}\frac{dy}{dt}$$

Example 10.3

Let $f:\mathbb{R}^3 \to \mathbb{R}$ defined by $w = f(x,y,z) = x^2+xy+yz$ and

$g:\mathbb{R} \to \mathbb{R}^3$ defined by $g(t) = (x(t), y(t), z(t)) = (t, t^2, t^3)$

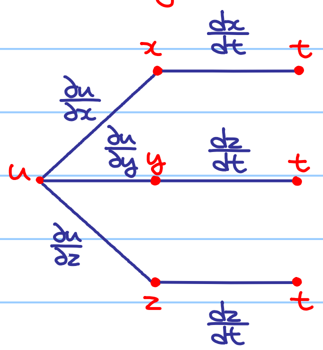
Note $Df \in M_{1\times 3}(\mathbb{R})$, $Dg \in M_{3\times 1}(\mathbb{R})$ and $D(f\circ g) \in M_{1\times 1}(\mathbb{R})$.

$$D(f\circ g)(t) = Df(g(t)) \cdot Dg(t)$$

$$\left[\frac{dw}{dt}\right] = \left[\frac{\partial w}{\partial x} \quad \frac{\partial w}{\partial y} \quad \frac{\partial w}{\partial z}\right] \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix}$$

$$= \left[\frac{\partial w}{\partial x}\frac{dx}{dt} + \frac{\partial w}{\partial y}\frac{dy}{dt} + \frac{\partial w}{\partial z}\frac{dz}{dt}\right]$$

Tree diagram :



$$\therefore \frac{dw}{dt} = (2x+y)(1) + (x+z)(2t) + (y)(3t^2)$$

$$= (2t+t^2)(1) + (t+t^3)(2t) + (t^2)(3t^2)$$

$$= 5t^4 + 3t^2 + 2t$$

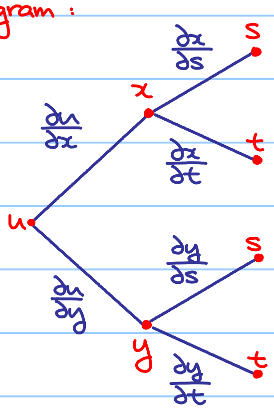
Example 10.4

Let $u = u(x,y)$, $x = x(s,t)$ and $y = y(s,t)$ be differentiable functions.

Therefore $(s,t) \mapsto (x(s,t), y(s,t)) \mapsto u(x(s,t), y(s,t))$

$$\mathbb{R}^2 \longrightarrow \mathbb{R}^2 \longrightarrow \mathbb{R}^1$$

Tree diagram :



Then,

$$\left[\frac{\partial u}{\partial s} \quad \frac{\partial u}{\partial t}\right] = \left[\frac{\partial u}{\partial x} \quad \frac{\partial u}{\partial y}\right] \begin{bmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{bmatrix}$$

$$M_{1\times 2}(\mathbb{R}) \qquad M_{1\times 2}(\mathbb{R}) \qquad M_{2\times 2}(\mathbb{R})$$

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}$$

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t}$$

Compute $\frac{\partial u}{\partial s}$: Sum of paths end at s .

Exercise 10.1

Let $u = x^2 - 2xy + 2y^3$ where $x = s^2 \ln t$ and $y = 2st^3$.

Find $\frac{\partial u}{\partial s}$ and $\frac{\partial u}{\partial t}$.

Ans : $\frac{\partial u}{\partial s} = (2s^2\ln t - 4st^3)(2s\ln t) + (-2s^2\ln t + 24s^2t^6)(2t^3)$

$\frac{\partial u}{\partial t} = (2s^2\ln t - 4st^3)\left(\frac{s^2}{t}\right) + (-2s^2\ln t + 24s^2t^6)(6st^2)$

Example 10.5

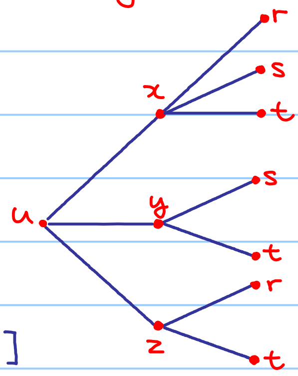
Let $u = u(x,y,z)$, $x = x(r,s,t)$, $y = y(s,t)$ and $z = z(r,t)$ be differentiable functions.

$$\frac{\partial u}{\partial r} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial r}$$

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s}$$

$$\frac{\partial u}{\partial t} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial t}$$

Tree diagram :



$$\begin{bmatrix} \frac{\partial u}{\partial r} & \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \end{bmatrix} \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial s} & \frac{\partial z}{\partial t} \end{bmatrix} \qquad \left(\because \frac{\partial y}{\partial r} = \frac{\partial z}{\partial s} = 0\right)$$

$$= \begin{bmatrix} \frac{\partial u}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial r} & \frac{\partial u}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial s} & \frac{\partial u}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial t} \end{bmatrix}$$

Implicit Differentiation

Example 10.6

Let $\mathcal{C}$ be the circle defined by $x^2+y^2-1=0$. Find $\frac{dy}{dx}$.

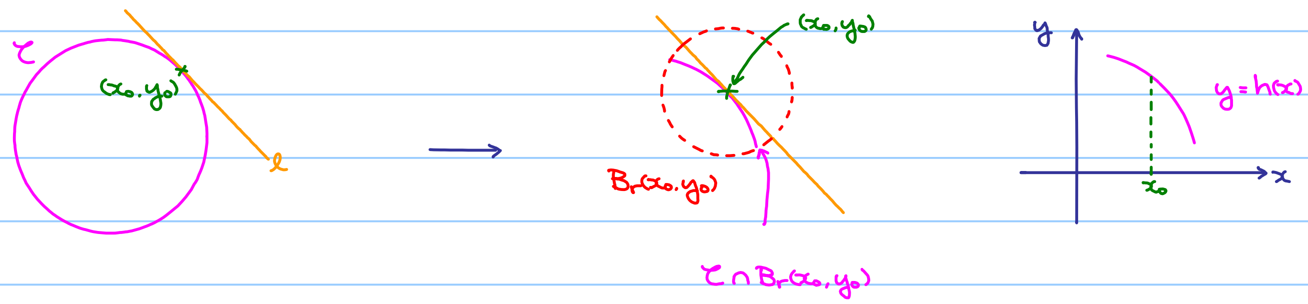
Differentiate both sides with respect to x:

$$2x + 2y\frac{dy}{dx} = 0 \quad \longleftarrow \text{ Why can it be done ?}$$

$$\frac{dy}{dx} = -\frac{x}{y}$$

Let $F(x,y) = x^2+y^2-1$. Then the circle $\mathcal{C}$ is just defined by $F(x,y)=0$.

Suppose that (x_0,y_0) lying on $\mathcal{C}$ such that $y_0 \neq 0$.



If $r>0$ is sufficiently small, $\mathcal{C} \cap B_r(x_0,y_0)$ is a small piece of arc of the circle $\mathcal{C}$, which is the graph of some function $y = h(x)$ and $\frac{dy}{dx}$ means $h'(x)$.

In general, suppose that

1) $F(x,y)$ is differentiable ;

2) $F(x_0,y_0)=0$;

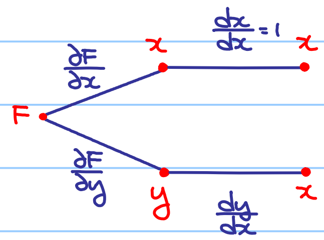
3) there exists $r>0$ such that $\{(x,y) : F(x,y)=0\} \cap B_r(x_0,y_0)$ can be regarded as graph of a differentiable function $y=h(x)$.

Then,

$$F(x,y)=0$$

$$\frac{\partial F}{\partial x}\frac{dx}{dx}+\frac{\partial F}{\partial y}\frac{dy}{dx}=0 \qquad \text{(Chain Rule)}$$

$$\therefore\ \frac{\partial F}{\partial x}\cdot 1+\frac{\partial F}{\partial y}\frac{dy}{dx}=0$$



At (x_0,y_0), $\frac{\partial F}{\partial x}(x_0,y_0)+\frac{\partial F}{\partial y}(x_0,y_0)\cdot \frac{dy}{dx}\Big|_{x=x_0}=0$

if $\frac{\partial F}{\partial y}(x_0,y_0)\neq 0$, then $\frac{dy}{dx}\Big|_{x=x_0}=-\left(\frac{\partial F}{\partial x}(x_0,y_0)\Big/\frac{\partial F}{\partial y}(x_0,y_0)\right)$.

Remark: Implicit function theorem :

$\frac{\partial F}{\partial y}(x_0,y_0)\neq 0$ if and only if condition (3) holds. (Discuss later!)

§ 11 More on Gradient

Gradient and Directional Derivatives

Proposition 11.1

Let $D \subseteq \mathbb{R}^n$ be an open subset, let $\vec{x}_0 \in D$ and let $f: D \to \mathbb{R}$.

If f is differentiable at $\vec{x}_0$, then the directional derivative exists along any nonzero vector $\vec{v}$ and we have $\nabla_{\vec{v}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \frac{\vec{v}}{|\vec{v}|}$.

proof:

f is differentiable at $\vec{x}_0 \Rightarrow$ there exists $\vec{L} \in \mathbb{R}^n$ such that $\lim_{\vec{h} \to \vec{0}} \frac{f(\vec{x}_0 + \vec{h}) - (f(\vec{x}_0) + \vec{L} \cdot \vec{h})}{|\vec{h}|} = 0$

(in fact, $\vec{L} = \nabla f(\vec{x}_0)$.)

In particular, take $\vec{h} = h\hat{v}$, where $\hat{v} = \frac{\vec{v}}{|\vec{v}|}$,

$$\lim_{h \to 0} \frac{f(\vec{x}_0 + h\hat{v}) - (f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot h\hat{v})}{|h\hat{v}|} = 0$$

$$\lim_{h \to 0} \frac{f(\vec{x}_0 + h\hat{v}) - (f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot h\hat{v})}{|h|} = 0$$

$$\Rightarrow \lim_{h \to 0} \frac{f(\vec{x}_0 + h\hat{v}) - (f(\vec{x}_0) + \nabla f(\vec{x}_0) \cdot h\hat{v})}{h} = 0$$

$$\lim_{h \to 0} \frac{f(\vec{x}_0 + h\hat{v}) - f(\vec{x}_0)}{h} = \nabla f(x_0) \cdot \hat{v}, \quad \text{so } \nabla_{\vec{v}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \; \hat{v} = \nabla f(\vec{x}_0) \; \frac{\vec{v}}{|\vec{v}|}$$

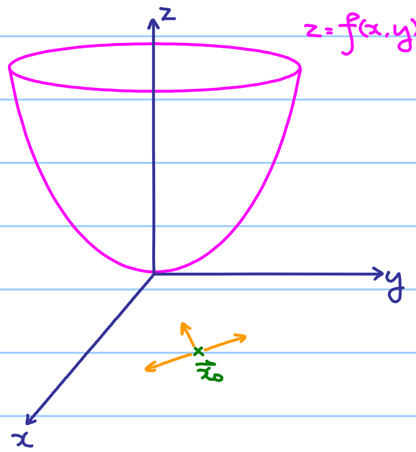
Example 11.1

Let $f(x,y) = xy$ and $\vec{v} = (1,2) \in \mathbb{R}^2$.

Then $\hat{v} = \frac{\vec{v}}{|\vec{v}|} = \frac{1}{\sqrt{5}}(1,2)$ and f is differentiable on $\mathbb{R}^2$, $\nabla f(x,y) = (y,x)$.

$\nabla_{\vec{v}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \frac{\vec{v}}{|\vec{v}|} = \frac{2}{\sqrt{5}} x + \frac{1}{\sqrt{5}} y$ (Compare with example 8.4)

Geometrical Meaning of Gradient



$z = f(x,y) = x^2 + y^2$

Question: If we move away from $\vec{x}_0$, then the value of f changes.

However, which direction shall we go to increase / decrease f most rapidly?

Since $\nabla_{\vec{v}} f(\vec{x}_0)$ is the rate of change of f at $\vec{x}_0$ along $\vec{v}$, our question is equivalent to maximize / minimize $\nabla_{\vec{v}} f(\vec{x}_0)$ among all possible direction $\vec{v}$.

Let $D \subseteq \mathbb{R}^n$ be an open subset, let $\vec{x}_0 \in D$ and let $f: D \to \mathbb{R}$.

Suppose that f is differentiable $\vec{x}_0$. If $\vec{v} \in \mathbb{R}^n$ and $|\vec{v}| = 1$, i.e $\hat{v} = \vec{v}$, then $\nabla_{\vec{v}} f(\vec{x}_0) = \nabla f(\vec{x}_0) \cdot \vec{v}$.

$$|\nabla_{\vec{v}} f(\vec{x}_0)| = |\nabla f(\vec{x}_0) \cdot \vec{v}|$$

$$= |\nabla f(\vec{x}_0)|\,|\vec{v}| \cdot |\cos\theta| \quad \text{, where } \theta \text{ is the angle between } \nabla f(\vec{x}_0) \text{ and } \vec{v}$$

$$\overset{(*)}{\leq} |\nabla f(\vec{x}_0)| \cdot |\vec{v}|$$

$$= |\nabla f(\vec{x}_0)| \qquad (\because |\vec{v}| = 1)$$

$$\therefore \; -|\nabla f(\vec{x}_0)| \leq \nabla_{\vec{v}} f(\vec{x}_0) \leq |\nabla f(\vec{x}_0)|$$

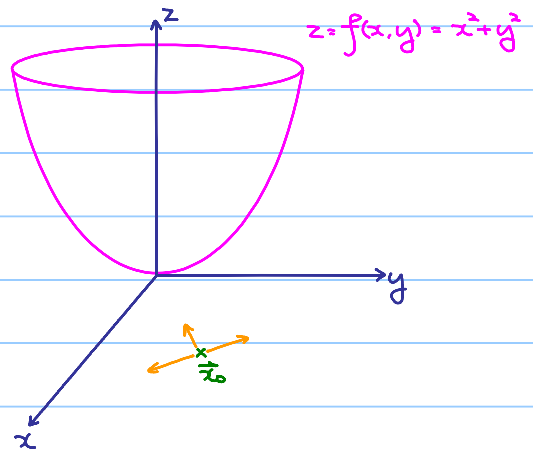
Furthermore, equality of (*) holds if and only if $\theta = 0$ or π, i.e. $\vec{v} \parallel \nabla f(\vec{x}_0)$

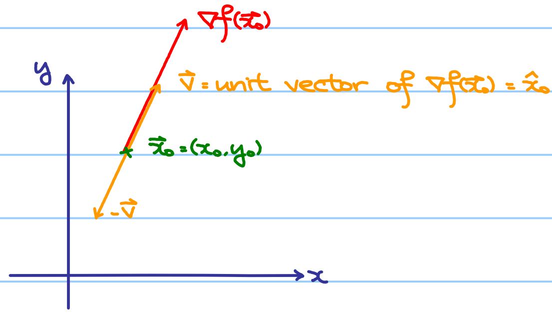
Therefore, if f changes most rapidly if we move along the direction of $\nabla f(\vec{x}_0)$ or $-\nabla f(\vec{x}_0)$.

Example 11.2

Let $f(x,y) = x^2 + y^2$. Then, $\nabla f(x,y) = (2x, 2y)$.

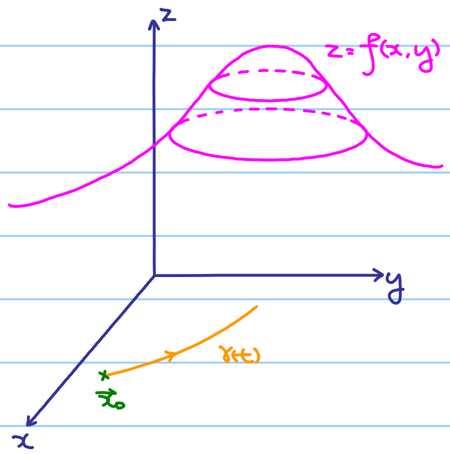
$\therefore$ For any $\vec{x}_0 = (x_0, y_0)$, $\nabla f(\vec{x}_0) = 2\vec{x}_0$





f changes most rapidly if we move along the direction of $\nabla f(\vec{x}_0)$ or $-\nabla f(\vec{x}_0)$.

Remark.



Idea: If we start from $\vec{x}_0$ and move along the curve $\gamma(t)$ such that $\gamma(0) = \vec{x}_0$ and $\gamma'(t) = \nabla f(\gamma(t))$, i.e. always moves along the gradient direction.

Eventually, we arrive extreme points of f!

(Gradient flow, useful in optiminization)

Gradient and Level Set

Let $D \subseteq \mathbb{R}^n$ be an open subset, let $\vec{x}_0 \in D$ and let $f: D \to \mathbb{R}$ be a smooth function.

Suppose that $L_c(f) = \{\vec{x} \in D : f(\vec{x}) = c\}$ and $\nabla f(\vec{x}) \neq \vec{0}$ for all points $\vec{x} \in L_c(f)$ ($\Rightarrow L_c(f)$ is a smooth $(n-1)$-dimensional manifold in $\mathbb{R}^n$) and $\vec{x}_0 \in L_c(f)$ i.e. $f(\vec{x}_0) = c$.

Let $\gamma : (-\varepsilon, \varepsilon) \to D$ be a differentiable curve such that γ lies on $L_c(f)$, $\gamma(0) = \vec{x}_0$ and $\gamma'(0)$ is nonzero. Then $\gamma'(0)$ is a vector tangent to $L_c(f)$ at $\vec{x}_0$.

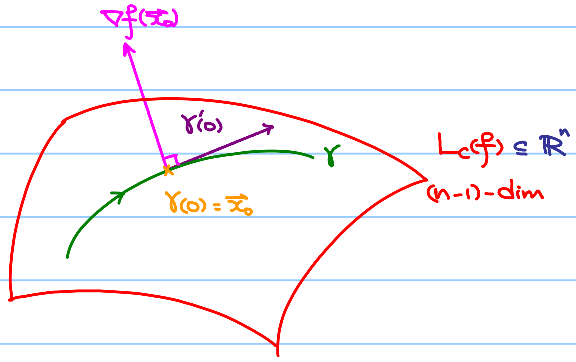
Furthermore, since γ lies on $L_c(f)$, we have $f(\gamma(t)) = c$ for all $t \in (-\varepsilon, \varepsilon)$.

Then, $\frac{d}{dt} f(\gamma(t)) = \nabla f(\gamma(t)) \cdot \gamma'(t) = 0$ (Chain Rule).

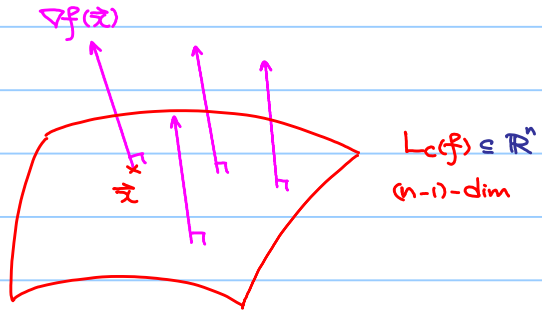
In particular, put $t = 0$, we have

$$\nabla f(\gamma(0)) \cdot \gamma'(0) = \nabla f(\vec{x}_0) \cdot \gamma'(0) = 0$$

Therefore, $\nabla f(\vec{x}_0)$ is orthogonal to every vector tangent to $L_c(f)$ at $\vec{x}_0$, and $\nabla f(\vec{x}_0)$ gives a normal of $L_c(f)$ at $\vec{x}_0$.



For every point $\vec{x} \in L_c(f)$, we draw $\nabla f(\vec{x})$, then we will see this:

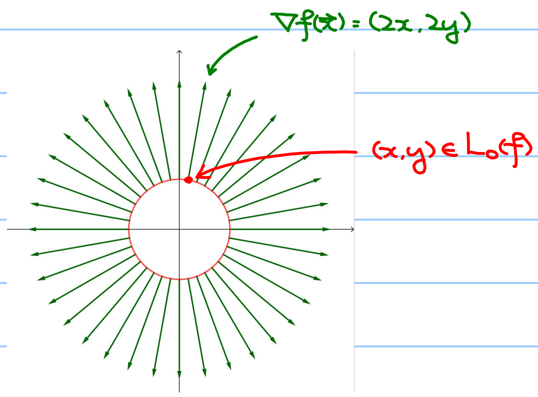


Example 11.3

Let $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = x^2 + y^2 - 1$.

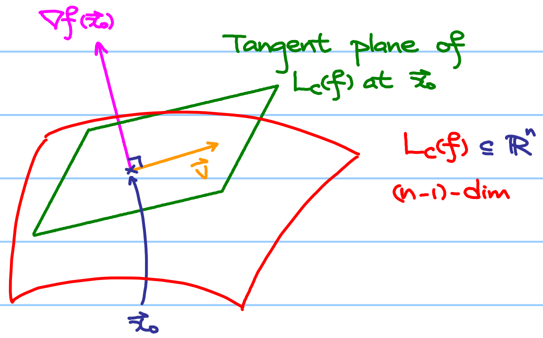
$L_0(f) = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 - 1 = 0\}$ which is the unit circle centered at the origin.

$\nabla f(\vec{x}) = \left(\frac{\partial f}{\partial x}(\vec{x}), \frac{\partial f}{\partial y}(\vec{x})\right) = (2x, 2y)$



Summary

Fix $\vec{x}_0 \in L_c(f)$ (i.e $f(\vec{x}_0) = c$), assume $\nabla f(\vec{x}_0)$ is nonzero,



1) if we move along any direction on the tangent plane of $L_c(f)$ at $\vec{x}_0$, the value of f does not change ;

2) if we move along the normal direction of the tangent plane of $L_c(f)$ at $\vec{x}_0$, the value of f changes mostly rapidly

§ 12 Taylor's Theorem

Taylor's Polynomial for Two variables

Let $D \subseteq \mathbb{R}^2$ be an open set, and let $f: D \to \mathbb{R}$ be a function such that partial derivatives of all order exist, i.e. $\frac{\partial^k f}{\partial x^{k_1} \partial y^{k_2}}$ exists for any positive integer k with integers $k_1, k_2 \geq 0$ and $k_1 + k_2 = k$.

Let $\vec{a} = (a_1, a_2) \in D$.

Idea: Can we approximate $f(x,y)$ around $\vec{a}$ by a polynomial $P_n(x,y)$ of degree n such that f and P_n agree up to n-th order at $\vec{a}$, i.e.

$$\frac{\partial^k f}{\partial x^{k_1} \partial y^{k_2}}(a_1, a_2) = \frac{\partial^k P_n}{\partial x^{k_1} \partial y^{k_2}}(a_1, a_2) \quad , \text{for } k_1 + k_2 = k \ , \ 0 \leq k \leq n \ ?$$

Let
$$\begin{aligned} P_n(x,y) &= C_{0,0} \\ &+ C_{1,0}(x-a_1) + C_{0,1}(y-a_2) \\ &+ C_{2,0}(x-a_1)^2 + C_{1,1}(x-a_1)(y-a_2) + C_{0,2}(y-a_2)^2 + \cdots \\ &+ C_{n,0}(x-a_1)^n + C_{n-1,1}(x-a_1)^{n-1}(y-a_2) + \cdots + C_{0,n}(y-a_2)^n \\ &= \sum_{k=0}^{n} \sum_{k_1+k_2=k} C_{k_1,k_2}(x-a_1)^{k_1}(y-a_2)^{k_2} \end{aligned}$$

Determine C_{k_1,k_2}'s :

- $P_n(a_1,a_2) = f(a_1,a_2) \Rightarrow C_{0,0} = f(a_1,a_2)$
- $\frac{\partial P_n}{\partial x}(a_1,a_2) = \frac{\partial f}{\partial x}(a_1,a_2) \Rightarrow C_{1,0} = \frac{\partial f}{\partial x}(a_1,a_2)$ $\qquad \frac{\partial P_n}{\partial y}(a_1,a_2) = \frac{\partial f}{\partial y}(a_1,a_2) \Rightarrow C_{0,1} = \frac{\partial f}{\partial y}(a_1,a_2)$
- $\frac{\partial^2 P_n}{\partial x^2}(a_1,a_2) = \frac{\partial^2 f}{\partial x^2}(a_1,a_2) \Rightarrow C_{2,0} = \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(a_1,a_2)$

  $\frac{\partial^2 P_n}{\partial y^2}(a_1,a_2) = \frac{\partial^2 f}{\partial y^2}(a_1,a_2) \Rightarrow C_{0,2} = \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(a_1,a_2)$

  $\frac{\partial^2 P_n}{\partial x \partial y}(a_1,a_2) = \frac{\partial^2 f}{\partial x \partial y}(a_1,a_2) \Rightarrow C_{1,1} = \frac{\partial^2 f}{\partial x \partial y}(a_1,a_2)$
- In general, $C_{k_1,k_2} = \frac{1}{k_1!\,k_2!}\frac{\partial^k f}{\partial x^{k_1}\partial y^{k_2}}(a_1,a_2)$ with $k = k_1 + k_2$.

Definition 12.1

Let D be an open subset in $\mathbb{R}^2$, let $\vec{a} = (a_1, a_2) \in D$.
and let $f: D \to \mathbb{R}$ be a function such that partial derivatives of f exist up to n-th order at $\vec{a}$. Then,

$$P_n(x,y) = \sum_{k=0}^{n} \sum_{k_1+k_2=k} \left(\frac{1}{k_1!\,k_2!} \frac{\partial^k f}{\partial x^{k_1} \partial y^{k_2}}(a_1,a_2) \right)(x-a_1)^{k_1}(y-a_2)^{k_2}$$

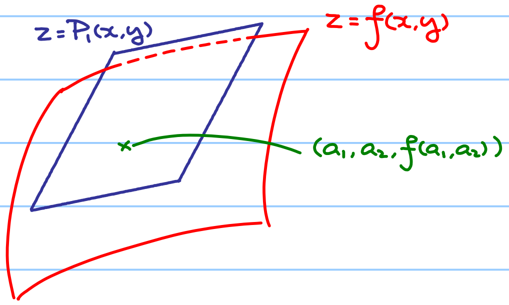
is said to be the Taylor polynomial of degree n generated by $f(x,y)$ at $\vec{a}$.

In particular, $z = P_1(x,y) = f(a_1,a_2) + \left(\frac{\partial f}{\partial x}(a_1,a_2)\right)(x-a_1) + \left(\frac{\partial f}{\partial y}(a_1,a_2)\right)(y-a_2)$

gives the tangent plane of f at $\vec{a}$.

Also, $P_1(x,y)$ is exactly the linearization of f at $\vec{a} = (a_1,a_2)$

(see example 9.3)



$$P_2(x,y) = f(a_1,a_2) + \left(\frac{\partial f}{\partial x}(a_1,a_2)\right)(x-a_1) + \left(\frac{\partial f}{\partial y}(a_1,a_2)\right)(y-a_2)$$
$$+ \left(\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(a_1,a_2)\right)(x-a_1)^2 + \left(\frac{\partial^2 f}{\partial x \partial y}(a_1,a_2)\right)(x-a_1)(y-a_2) + \left(\frac{1}{2}\frac{\partial^2 f}{\partial y^2}(a_1,a_2)\right)(y-a_2)^2$$

Example 12.1

Let $f(x,y) = e^{x^2+y}$.

Let $P_n(x,y)$ be the Taylor polynomial of degree n generated by $f(x,y)$ at $(0,0)$.

- $f(0,0) = 1$

- $\frac{\partial f}{\partial x} = 2xe^{x^2+y} \Rightarrow \frac{\partial f}{\partial x}(0,0) = 0$ $\qquad$ $\frac{\partial f}{\partial y} = e^{x^2+y} \Rightarrow \frac{\partial f}{\partial y}(0,0) = 1$

- $\frac{\partial^2 f}{\partial x^2} = (2+4x^2)e^{x^2+y} \Rightarrow \frac{\partial^2 f}{\partial x^2}(0,0) = 2$ $\qquad$ $\frac{\partial^2 f}{\partial x \partial y} = 2xe^{x^2+y} \Rightarrow \frac{\partial^2 f}{\partial x \partial y}(0,0) = 0$ $\qquad$ $\frac{\partial^2 f}{\partial y^2} = e^{x^2+y} \Rightarrow \frac{\partial^2 f}{\partial y^2}(0,0) = 1$

- $\frac{\partial^3 f}{\partial x^3} = (12x+8x^3)e^{x^2+y} \Rightarrow \frac{\partial^3 f}{\partial x^3}(0,0) = 0$ $\qquad$ $\frac{\partial^3 f}{\partial x^2 \partial y} = (2+4x^2)e^{x^2+y} \Rightarrow \frac{\partial^3 f}{\partial x^2 \partial y}(0,0) = 0$

$\frac{\partial^3 f}{\partial x \partial y^2} = (2+4x^2)e^{x^2+y} \Rightarrow \frac{\partial^3 f}{\partial x \partial y^2}(0,0) = 2$ $\qquad$ $\frac{\partial^3 f}{\partial y^3} = e^{x^2+y} \Rightarrow \frac{\partial^3 f}{\partial y^3}(0,0) = 1$

$P_0(x,y) = f(0,0) = 1$

$P_1(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)\,x + \frac{\partial f}{\partial y}(0,0)\,y = 1 + y$

$P_2(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)\,x + \frac{\partial f}{\partial y}(0,0)\,y + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0)x^2 + \frac{\partial^2 f}{\partial x \partial y}(0,0)\,xy + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(0,0)\,y^2$

$= 1 + y + x^2 + \frac{1}{2}y^2$

$P_3(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)\,x + \frac{\partial f}{\partial y}(0,0)\,y + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(0,0)x^2 + \frac{\partial^2 f}{\partial x \partial y}(0,0)\,xy + \frac{1}{2}\frac{\partial^2 f}{\partial y^2}(0,0)\,y^2$

$+ \frac{1}{6}\frac{\partial^3 f}{\partial x^3}(0,0)\,x^3 + \frac{1}{2}\frac{\partial^3 f}{\partial x^2 \partial y}(0,0)\,x^2y + \frac{1}{2}\frac{\partial^3 f}{\partial x \partial y^2}(0,0)\,xy^2 + \frac{1}{6}\frac{\partial^3 f}{\partial y^3}(0,0)\,y^3$

$= 1 + y + x^2 + \frac{1}{2}y^2 + x^2y + \frac{1}{6}y^3$

Frequently used Taylor series :

1) $\dfrac{1}{1-x} = 1 + x + x^2 + \cdots + x^n + \cdots = \sum_{n=0}^{\infty} x^n \quad , \quad |x| < 1$

2) $\dfrac{1}{1+x} = 1 - x + x^2 - \cdots + (-x)^n + \cdots = \sum_{n=0}^{\infty} (-x)^n \quad , \quad |x| < 1$

3) $e^x = 1 + x + \dfrac{x^2}{2!} + \cdots + \dfrac{x^n}{n!} + \cdots = \sum_{n=0}^{\infty} \dfrac{x^n}{n!} \quad , \quad x \in \mathbb{R}$

4) $\sin x = x - \dfrac{x^3}{3!} + \dfrac{x^5}{5!} - \cdots + (-1)^n \dfrac{x^{2n+1}}{(2n+1)!} + \cdots = \sum_{n=0}^{\infty} \dfrac{(-1)^n x^{2n+1}}{(2n+1)!} \quad , \quad x \in \mathbb{R}$

5) $\cos x = 1 - \dfrac{x^2}{2!} + \dfrac{x^4}{4!} - \cdots + (-1)^n \dfrac{x^{2n}}{(2n)!} + \cdots = \sum_{n=0}^{\infty} \dfrac{(-1)^n x^{2n}}{(2n)!} \quad , \quad x \in \mathbb{R}$

6) $\ln(1+x) = x - \dfrac{x^2}{2} + \dfrac{x^3}{3} - \cdots + (-1)^{n+1} \dfrac{x^n}{n} + \cdots = \sum_{n=1}^{\infty} \dfrac{(-1)^{n+1} x^n}{n} \quad , \quad -1 < x \le 1$

Example 12.1 (Cont.)

$$e^{x^2+y} = 1 + (x^2+y) + \frac{1}{2!}(x^2+y)^2 + \frac{1}{3!}(x^2+y)^3 + \cdots$$

$$= 1 + (x^2+y) + \frac{1}{2}(x^4 + 2x^2y + y^2) + \frac{1}{6}(x^6 + 3x^4y + 3x^2y^2 + y^3) + \cdots$$

$$= \underbrace{1 + y + x^2 + \frac{1}{2}y^2 + x^2y + \frac{1}{6}y^3}_{P_3(x,y)} + \underbrace{\cdots}_{\text{higher order terms}}$$

OR :

$$e^{x^2+y} = e^{x^2} \cdot e^{y}$$

$$= (1 + x^2 + \cdots)(1 + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots)$$

$$= \underbrace{1 + y + x^2 + \frac{1}{2}y^2 + x^2y + \frac{1}{6}y^3}_{P_3(x,y)} + \underbrace{\cdots}_{\text{higher order terms}}$$

Example 12.2

$$\sin(e^y + x - 1) = \sin\left((1 + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots) + x - 1\right)$$

$$= \sin(x + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots)$$

$$= (x + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots) - \frac{1}{3!}(x + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots)^3 + \frac{1}{5!}(x + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots)^5 - \cdots$$

$$= (x + y + \frac{1}{2}y^2 + \frac{1}{6}y^3 + \cdots) - \frac{1}{6}(x^3 + 3x^2y + 3xy^2 + y^3 + \cdots) + \cdots$$

$$= \underbrace{x + y + \frac{1}{2}y^2 - \frac{1}{6}x^3 - \frac{1}{2}x^2y - \frac{1}{2}xy^2}_{P_3(x,y)} + \underbrace{\cdots}_{\text{higher order terms}}$$

Example 12.3

$\ln(1+x+y) = \ln(1+(x+y))$

$$= (x+y) - \frac{(x+y)^2}{2} + \frac{(x+y)^3}{3} + \cdots$$

$$= \underbrace{x+y - \tfrac{1}{2}(x^2+2xy+y^2) + \tfrac{1}{3}(x^3+3x^2y+3xy^2+y^3)}_{P_3(x,y)} + \underbrace{\cdots}_{\text{higher order terms}}$$

In particular, there is a very compact way to write down $P_2(x,y)$ which is very useful later!

We have

$$P_2(x,y) = f(a_1,a_2) + \left(\frac{\partial f}{\partial x}(a_1,a_2)\right)(x-a_1) + \left(\frac{\partial f}{\partial y}(a_1,a_2)\right)(y-a_2)$$

$$+ \left(\frac{1}{2}\frac{\partial^2 f}{\partial x^2}(a_1,a_2)\right)(x-a_1)^2 + \left(\frac{\partial^2 f}{\partial x \partial y}(a_1,a_2)\right)(x-a_1)(y-a_2) + \left(\frac{1}{2}\frac{\partial^2 f}{\partial y^2}(a_1,a_2)\right)(y-a_2)^2$$

$$= f(a_1,a_2) + \left[\frac{\partial f}{\partial x}(a_1,a_2) \quad \frac{\partial f}{\partial y}(a_1,a_2)\right]\begin{bmatrix} x-a_1 \\ y-a_2 \end{bmatrix} + \frac{1}{2}[x-a_1 \quad y-a_2]\begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(a_1,a_2) & \frac{\partial^2 f}{\partial x \partial y}(a_1,a_2) \\ \frac{\partial^2 f}{\partial y \partial x}(a_1,a_2) & \frac{\partial^2 f}{\partial y^2}(a_1,a_2) \end{bmatrix}\begin{bmatrix} x-a_1 \\ y-a_2 \end{bmatrix}$$

$$\left(\because \frac{\partial^2 f}{\partial x \partial y}(a_1,a_2) = \frac{\partial^2 f}{\partial y \partial x}(a_1,a_2)\right)$$

$$= f(\vec{a}) + \nabla f(\vec{a})(\vec{x}-\vec{a}) + \frac{1}{2}(\vec{x}-\vec{a})^T H(\vec{a})(\vec{x}-\vec{a})$$

where $\vec{x}-\vec{a} = \begin{bmatrix} x-a_1 \\ y-a_2 \end{bmatrix}$, $\nabla f(\vec{x}) = \left[\frac{\partial f}{\partial x} \quad \frac{\partial f}{\partial y}\right]$, $H(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$ called Hessian matrix.

Taylor's Theorem for Two Variables

Theorem 12.1 (Taylor's Theorem for Two Variables)

Let D be an open convex set in $\mathbb{R}^2$, let $\vec{a} = (a_1,a_2) \in D$

and let $f: D \to \mathbb{R}$ be a C^{n+1} function on D. Then,

$$f(\vec{x}) = f(x,y) = \underbrace{\sum_{k=0}^{n} \sum_{k_1+k_2=k} \left(\frac{1}{k_1!\,k_2!}\frac{\partial^k f}{\partial x^{k_1} \partial x^{k_2}}(a_1,a_2)\right)(x-a_1)^{k_1}(y-a_2)^{k_2}}_{P_n(x,y)} + R_n(\vec{x}) \quad \text{for } \vec{x} \in D$$

where $R_n(\vec{x}) = \sum_{k_1+k_2=n+1} \left(\frac{1}{k_1!\,k_2!}\frac{\partial^k f}{\partial x^{k_1} \partial x^{k_2}}(\vec{c})\right)(x-a_1)^{k_1}(y-a_2)^{k_2}$ for some $\vec{c} = t\vec{a} + (1-t)\vec{x}$, $t \in (0,1)$.

Example 12.4

Let $f: \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = \cos(x+2y)$

The Taylor polynomial of degree 2 generated by f at $\vec{0} = (0,0)$ is $P_2(x,y) = 1 - \frac{1}{2}x^2 - 2xy - 2y^2$

Therefore, $f(0.2, 0.1)$ can be approximated by $P_2(0.2, 0.1) = 1 - \frac{1}{2}(0.2)^2 - 2(0.2)(0.1) - 2(0.1)^2 = 0.92$

However, how good is the approximation ?

The absolute error

$= |R_2(0.2, 0.1)|$

$= \left| \frac{1}{3!}\frac{\partial^3 f}{\partial x^3}(\vec{c})(0.2-0)^3 + \frac{1}{2!1!}\frac{\partial^3 f}{\partial x^2 \partial y}(\vec{c})(0.2-0)^2(0.1-0) + \frac{1}{1!2!}\frac{\partial^3 f}{\partial x \partial y^2}(\vec{c})(0.2-0)(0.1-0)^2 + \frac{1}{3!}\frac{\partial^3 f}{\partial y^3}(\vec{c})(0.1-0)^3 \right|$

$\le \left| \frac{1}{3!}\frac{\partial^3 f}{\partial x^3}(\vec{c})(0.2-0)^3 \right| + \left| \frac{1}{2!1!}\frac{\partial^3 f}{\partial x^2 \partial y}(\vec{c})(0.2-0)^2(0.1-0) \right| + \left| \frac{1}{1!2!}\frac{\partial^3 f}{\partial x \partial y^2}(\vec{c})(0.2-0)(0.1-0)^2 \right| + \left| \frac{1}{3!}\frac{\partial^3 f}{\partial y^3}(\vec{c})(0.1-0)^3 \right|$

$\le \frac{1}{6}(0.2)^3 + \frac{1}{2}(2)(0.2)^2(0.1) + \frac{1}{2}(4)(0.2)(0.1)^2 + \frac{1}{6}(8)(0.1)^3$ for some

$= \frac{4}{375} \approx 0.010667$ $\vec{c} = t(0,0) + (1-t)(0.2, 0.1)$ $t \in (0,1)$

$= (1-t)(0.2, 0.1)$

Exercise : Show that $\left|\frac{\partial^3 f}{\partial x^3}(\vec{c})\right| \le 1$, $\left|\frac{\partial^3 f}{\partial x^2 \partial y}(\vec{c})\right| \le 2$, $\left|\frac{\partial^3 f}{\partial x \partial y^2}(\vec{c})\right| \le 4$, $\left|\frac{\partial^3 f}{\partial y^3}(\vec{c})\right| \le 8$

Taylor's Theorem

Definition 12.2

Let D be an open subset in $\mathbb{R}^m$, let $\vec{a} = (a_1, a_2, \cdots, a_m) \in D$.

and let $f: D \to \mathbb{R}$ be a function such that partial derivatives of f exist up to n-th order at $\vec{a}$. Then,

$$P_n(\vec{x}) = \sum_{k=0}^{n} \sum_{k_1+k_2+\cdots+k_m=k} \left(\frac{1}{k_1!k_2!\cdots k_m!} \frac{\partial^k f}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_m^{k_m}}(\vec{a}) \; (x_1-a_1)^{k_1}(x_2-a_2)^{k_2}\cdots(x_m-a_m)^{k_m} \right)$$

is said to be the Taylor polynomial of degree n generated by $f(\vec{x})$ at $\vec{a}$.

Exercise 12.1

Define $H(\vec{x}) \in M_{n\times n}(\mathbb{R})$ such that $[H(\vec{x})]_{ij} = \frac{\partial^2 f}{\partial x_i \partial y_j}$

Show that $P_2(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a})\,(\vec{x}-\vec{a}) + \frac{1}{2}(\vec{x}-\vec{a})^T H(\vec{a})(\vec{x}-\vec{a})$

where $\vec{x}-\vec{a} = \begin{bmatrix} x_1-a_1 \\ x_2-a_2 \\ \vdots \\ x_n-a_n \end{bmatrix}$, $\nabla f(\vec{x}) = [\frac{\partial f}{\partial x_1} \;\; \frac{\partial f}{\partial x_2} \;\; \cdots \;\; \frac{\partial f}{\partial x_n}]$.

Theorem 12.2 (Taylor's Theorem)

Let D be an open convex set in $\mathbb{R}^m$, let $\vec{a} = (a_1, a_2, \cdots, a_m) \in D$.

and let $f: D \to \mathbb{R}$ be a C^{n+1} function on D. Then,

$$f(\vec{x}) = \underbrace{\sum_{k=0}^{n} \sum_{k_1+k_2+\cdots+k_m=k} \left(\frac{1}{k_1!\,k_2!\cdots k_m!} \frac{\partial^k f}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_m^{k_m}}(\vec{a}) \; (x_1-a_1)^{k_1}(x_2-a_2)^{k_2}\cdots(x_m-a_m)^{k_m} \right)}_{= P_n(\vec{x})} + R_n(\vec{x}) \quad \text{for } \vec{x} \in D$$

where $R_n(\vec{x}) = \displaystyle\sum_{k_1+k_2+\cdots+k_m=n+1} \left(\frac{1}{k_1!\,k_2!\cdots k_m!} \frac{\partial^k f}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_m^{k_m}}(\vec{c}) \right) (x_1-a_1)^{k_1}(x_2-a_2)^{k_2}\cdots(x_m-a_m)^{k_m}$

for some $\vec{c} = t\vec{a} + (1-t)\vec{x}$, $t \in (0,1)$.