Chain rule, Tangents and Higher order derivatives

Department of Mathematics
IIT Guwahati

Chain rule

Theorem-A: Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be differentiable at t_0 and $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} := \mathbf{x}(t_0)$. Then $f \circ \mathbf{x}$ is differentiable at t_0 and

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x})|_{t=t_0} = \nabla f(\mathbf{a}) \bullet \mathbf{x}'(t_0) = \sum_{i=1}^{n} \partial_i f(\mathbf{a}) \frac{\mathrm{d}x_i(t_0)}{\mathrm{d}t}.$$

Chain rule

Theorem-A: Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be differentiable at t_0 and $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} := \mathbf{x}(t_0)$. Then $f \circ \mathbf{x}$ is differentiable at t_0 and

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x})|_{t=t_0} = \nabla f(\mathbf{a}) \bullet \mathbf{x}'(t_0) = \sum_{i=1}^{n} \partial_i f(\mathbf{a}) \frac{\mathrm{d}x_i(t_0)}{\mathrm{d}t}.$$

Proof: Use $\mathbf{h} := \mathbf{x}(t) - \mathbf{x}(t_0) = \mathbf{x}(t) - \mathbf{a}$ in

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet \mathbf{h} + e(\mathbf{h})\|\mathbf{h}\|$$

and the fact that $e(\mathbf{h}) \to 0$ as $\mathbf{h} \to 0$.

Chain rule

Theorem-A: Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ be differentiable at t_0 and $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} := \mathbf{x}(t_0)$. Then $f \circ \mathbf{x}$ is differentiable at t_0 and

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x})|_{t=t_0} = \nabla f(\mathbf{a}) \bullet \mathbf{x}'(t_0) = \sum_{i=1}^{n} \partial_i f(\mathbf{a}) \frac{\mathrm{d}x_i(t_0)}{\mathrm{d}t}.$$

Proof: Use $\mathbf{h} := \mathbf{x}(t) - \mathbf{x}(t_0) = \mathbf{x}(t) - \mathbf{a}$ in

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet \mathbf{h} + e(\mathbf{h})\|\mathbf{h}\|$$

and the fact that $e(\mathbf{h}) \to 0$ as $\mathbf{h} \to 0$.

Remark: $\dfrac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t_0)) = \nabla f(\mathbf{a}) \bullet \mathbf{x}'(t_0) = \displaystyle\sum_{i=1}^{n} \partial_i f(\mathbf{a}) \frac{\mathrm{d}x_i(t_0)}{\mathrm{d}t}$

sometimes referred to as total derivative.

Chain rule for partial derivatives

Theorem-B: If $\mathbf{x} : \mathbb{R}^2 \to \mathbb{R}^n, (u, v) \mapsto (x_1(u, v), \ldots, x_n(u, v))$ has partial derivatives at (a, b) and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{p} := \mathbf{x}(a, b)$ then $F(u, v) := f(\mathbf{x}(u, v))$ has partial derivatives at (a, b) and

$$\begin{aligned} \partial_u F(a, b) &= \nabla f(\mathbf{p}) \bullet \partial_u \mathbf{x}(a, b) = \sum_{j=1}^{n} \frac{\partial_j f(\mathbf{p})}{\partial x_j} \frac{\partial x_j(a, b)}{\partial u}, \\ \partial_v F(a, b) &= \nabla f(\mathbf{p}) \bullet \partial_v \mathbf{x}(a, b) = \sum_{j=1}^{n} \frac{\partial_j f(\mathbf{p})}{\partial x_j} \frac{\partial x_j(a, b)}{\partial v}. \end{aligned}$$

Chain rule for partial derivatives

Theorem-B: If $\mathbf{x} : \mathbb{R}^2 \to \mathbb{R}^n, (u, v) \mapsto (x_1(u, v), \ldots, x_n(u, v))$ has partial derivatives at (a, b) and $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable at $\mathbf{p} := \mathbf{x}(a, b)$ then $F(u, v) := f(\mathbf{x}(u, v))$ has partial derivatives at (a, b) and

$$
\begin{aligned}
\partial_u F(a, b) &= \nabla f(\mathbf{p}) \bullet \partial_u \mathbf{x}(a, b) = \sum_{j=1}^{n} \frac{\partial_j f(\mathbf{p})}{\partial x_j} \frac{\partial x_j(a, b)}{\partial u}, \\
\partial_v F(a, b) &= \nabla f(\mathbf{p}) \bullet \partial_v \mathbf{x}(a, b) = \sum_{j=1}^{n} \frac{\partial_j f(\mathbf{p})}{\partial x_j} \frac{\partial x_j(a, b)}{\partial v}.
\end{aligned}
$$

Proof: Apply Theorem-A.

Example: Find $\partial w / \partial u$ and $\partial w / \partial v$ when $w = x^2 + xy$ and $x = u^2 v, y = uv^2$.

Chain rule for derivatives

Fact: If $f : \mathbb{R}^n \to \mathbb{R}^m$ differentiable at $\mathbf{a}$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ differentiable at $\mathbf{c} := f(\mathbf{a})$ then $g \circ f$ differentiable at $\mathbf{a}$ and

$$\mathrm{D}(g \circ f)(\mathbf{a}) = \mathrm{D}g(\mathbf{c})\mathrm{D}f(\mathbf{a}), \;\; J_{g\circ f}(\mathbf{a}) = J_g(\mathbf{c})J_f(\mathbf{a}).$$

Chain rule for derivatives

Fact: If $f : \mathbb{R}^n \to \mathbb{R}^m$ differentiable at $\mathbf{a}$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ differentiable at $\mathbf{c} := f(\mathbf{a})$ then $g \circ f$ differentiable at $\mathbf{a}$ and

$$\mathrm{D}(g \circ f)(\mathbf{a}) = \mathrm{D}g(\mathbf{c})\mathrm{D}f(\mathbf{a}), \;\; J_{g\circ f}(\mathbf{a}) = J_g(\mathbf{c})J_f(\mathbf{a}).$$

Example: Consider $g(x, y) := (xy,\, x^2 - y^2)$ and $f(r, \theta) := (r\cos\theta,\, r\sin\theta)$. Then

$$J_g(x, y) = \begin{bmatrix} y & x \\ 2x & -2y \end{bmatrix}, J_f(r, \theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix},$$

Chain rule for derivatives

Fact: If $f:\mathbb{R}^n \to \mathbb{R}^m$ differentiable at $\mathbf{a}$ and $g:\mathbb{R}^m \to \mathbb{R}^p$ differentiable at $\mathbf{c} := f(\mathbf{a})$ then $g\circ f$ differentiable at $\mathbf{a}$ and

$$\mathrm{D}(g\circ f)(\mathbf{a}) = \mathrm{D}g(\mathbf{c})\mathrm{D}f(\mathbf{a}), \;\; J_{g\circ f}(\mathbf{a}) = J_g(\mathbf{c})J_f(\mathbf{a}).$$

Example: Consider $g(x,y) := (xy,\, x^2-y^2)$ and $f(r,\theta) := (r\cos\theta,\, r\sin\theta)$. Then

$$J_g(x,y) = \begin{bmatrix} y & x \\ 2x & -2y \end{bmatrix}, J_f(r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix},$$

$$J_{g\circ f}(r,\theta) = \begin{bmatrix} r\sin\theta & r\cos\theta \\ 2r\cos\theta & -2r\sin\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}.$$

Graph and level set

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then $G(f) := \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$ is the graph of f. $G(f)$ represents a hyper-surface in $\mathbb{R}^{n+1}$.

Graph and level set

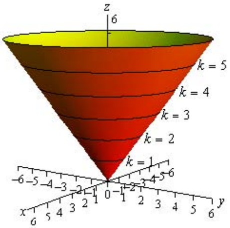
Let $f : \mathbb{R}^n \to \mathbb{R}$. Then $G(f) := \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$ is the graph of f. $G(f)$ represents a hyper-surface in $\mathbb{R}^{n+1}$.

The set $S(f, \alpha) := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = \alpha\}$ is called a level set of f and represents a hyper-surface in $\mathbb{R}^n$.

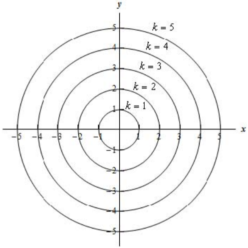
Graph and level set

Let $f : \mathbb{R}^n \to \mathbb{R}$. Then $G(f) := \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^{n+1}$ is the graph of f. $G(f)$ represents a hyper-surface in $\mathbb{R}^{n+1}$.

The set $S(f, \alpha) := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = \alpha\}$ is called a level set of f and represents a hyper-surface in $\mathbb{R}^n$.

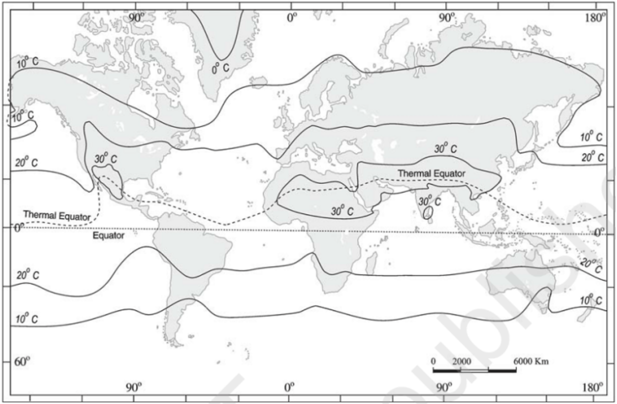


(e) Graph of $f(x, y) := \sqrt{x^2 + y^2}$



(f) Level curve $\sqrt{x^2 + y^2} = k$

Level curve (isothermal contour)



The distribution of surface air temperature in the month of July

Linearization and tangent

Let $L : \mathbb{R}^n \to \mathbb{R}$ be a linear map. Then $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is said to be tangent hyperplane to the hyper-surface $y = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a}))$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{|f(\mathbf{x}) - [f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})]|}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Linearization and tangent

Let $L : \mathbb{R}^n \to \mathbb{R}$ be a linear map. Then $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is said to be tangent hyperplane to the hyper-surface $y = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a}))$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{|f(\mathbf{x}) - [f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})]|}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

The tangent hyperplane $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is called a linearization of f at $\mathbf{a}$ or a linear approximation of f at $\mathbf{a}$.

Linearization and tangent

Let $L : \mathbb{R}^n \to \mathbb{R}$ be a linear map. Then $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is said to be tangent hyperplane to the hyper-surface $y = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a}))$ if

$$\lim_{\mathbf{x} \to \mathbf{a}} \frac{|f(\mathbf{x}) - [f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})]|}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

The tangent hyperplane $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is called a linearization of f at $\mathbf{a}$ or a linear approximation of f at $\mathbf{a}$.

Fact: $y = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a})$ is tangent to the hyper-surface $y = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a})) \Leftrightarrow f$ is differentiable at $\mathbf{a}$.

Linearization and tangent

Let $L : \mathbb{R}^n \to \mathbb{R}$ be a linear map. Then $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is said to be tangent hyperplane to the hyper-surface $y = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a}))$ if

$$\lim_{\mathbf{x}\to\mathbf{a}} \frac{|f(\mathbf{x}) - [f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})]|}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

The tangent hyperplane $y = f(\mathbf{a}) + L(\mathbf{x} - \mathbf{a})$ is called a linearization of f at $\mathbf{a}$ or a linear approximation of f at $\mathbf{a}$.

Fact: $y = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a})$ is tangent to the hyper-surface $y = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a})) \Leftrightarrow f$ is differentiable at $\mathbf{a}$.

Example: $z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ is a tangent plane to the surface $z = f(x, y)$ at $(a, b, f(a, b))$.

Level sets and gradients

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Suppose that $\mathbf{a}$ is point on the hyper-surface $f(\mathbf{x}) = \alpha$.

Level sets and gradients

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Suppose that $\mathbf{a}$ is point on the hyper-surface $f(\mathbf{x}) = \alpha$.

Let $\mathbf{x} : (-\epsilon, \epsilon) \to \mathbb{R}^n$ be a curve on the hyper-surface $f(\mathbf{x}) = \alpha$ passing through $\mathbf{a}$, i.e, $\mathbf{x}(0) = \mathbf{a}$ and $f(\mathbf{x}(t)) = \alpha$ for $t \in (-\epsilon, \epsilon)$.

Level sets and gradients

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Suppose that $\mathbf{a}$ is point on the hyper-surface $f(\mathbf{x}) = \alpha$.

Let $\mathbf{x} : (-\epsilon, \epsilon) \to \mathbb{R}^n$ be a curve on the hyper-surface $f(\mathbf{x}) = \alpha$ passing through $\mathbf{a}$, i.e, $\mathbf{x}(0) = \mathbf{a}$ and $f(\mathbf{x}(t)) = \alpha$ for $t \in (-\epsilon, \epsilon)$.

Suppose that $\mathbf{x}(t)$ differentiable at 0. Then

$$\frac{\mathrm{d}f}{\mathrm{d}t}(\mathbf{x}(t))|_{t=0} = \nabla f(\mathbf{a}) \bullet \mathbf{x}'(0) = 0 \Rightarrow \nabla f(\mathbf{a}) \perp \mathbf{x}'(0)$$

Level sets and gradients

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Suppose that $\mathbf{a}$ is point on the hyper-surface $f(\mathbf{x}) = \alpha$.

Let $\mathbf{x} : (-\epsilon, \epsilon) \to \mathbb{R}^n$ be a curve on the hyper-surface $f(\mathbf{x}) = \alpha$ passing through $\mathbf{a}$, i.e, $\mathbf{x}(0) = \mathbf{a}$ and $f(\mathbf{x}(t)) = \alpha$ for $t \in (-\epsilon, \epsilon)$.

Suppose that $\mathbf{x}(t)$ differentiable at 0. Then

$$\frac{\mathrm{d}f}{\mathrm{d}t}(\mathbf{x}(t))|_{t=0} = \nabla f(\mathbf{a}) \bullet \mathbf{x}'(0) = 0 \Rightarrow \nabla f(\mathbf{a}) \perp \mathbf{x}'(0)$$

Since the line $\mathbf{a} + t\,\mathbf{x}'(0)$ is tangent to the curve $\mathbf{x}(t)$ at $\mathbf{a}$, $\nabla f(\mathbf{a})$ is normal to the hyper-surface $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

Tangent and normal to level set

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Then the line $\mathbf{x} = \mathbf{a} + t\nabla f(\mathbf{a})$ is normal to the hyper-surface $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

Tangent and normal to level set

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Then the line $\mathbf{x} = \mathbf{a} + t\nabla f(\mathbf{a})$ is normal to the hyper-surface $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

The hyperplane $(\mathbf{x} - \mathbf{a}) \bullet \nabla f(\mathbf{a}) = 0$ is tangent to the hyper-surface $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

Tangent and normal to level set

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$. Then the line $\mathbf{x} = \mathbf{a} + t\nabla f(\mathbf{a})$ is normal to the hyper-surface $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

The hyperplane $(\mathbf{x} - \mathbf{a}) \bullet \nabla f(\mathbf{a}) = 0$ is tangent to the hyper-surface $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

Note that $(\nabla f(\mathbf{a}), -1)$ is normal to the hyper-surface $F(x, z) := f(\mathbf{x}) - z = 0$ at $(\mathbf{a}, f(\mathbf{a}))$. Hence

$$
\begin{aligned}
& (\mathbf{x} - \mathbf{a}, z - f(\mathbf{a})) \bullet (\nabla f(\mathbf{a}), -1) = 0 \\
\Rightarrow \quad & z = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a})
\end{aligned}
$$

is tangent to the hyper-surface $z = f(\mathbf{x})$ at $(\mathbf{a}, f(\mathbf{a}))$.

Examples:

- $(x-a)f_x(a,b)+(y-b)f_y(a,b)=0$ is the equation of the tangent line to $f(x,y)=\alpha$ at (a,b).

Examples:

- $(x-a)f_x(a,b)+(y-b)f_y(a,b)=0$ is the equation of the tangent line to $f(x,y)=\alpha$ at (a,b).
- $(x-a)f_x(a,b,c)+(y-b)f_y(a,b,c)+(z-c)f_z(a,b,c)=0$ is the equation of the tangent plane to $f(x,y,z)=\alpha$ at (a,b,c).

Examples:

- $(x-a)f_x(a,b)+(y-b)f_y(a,b)=0$ is the equation of the tangent line to $f(x,y)=\alpha$ at (a,b).
- $(x-a)f_x(a,b,c)+(y-b)f_y(a,b,c)+(z-c)f_z(a,b,c)=0$ is the equation of the tangent plane to $f(x,y,z)=\alpha$ at (a,b,c).

Let $f(x,y,z)=x^2+2xy-y^2+z^2$. Then $f(1,-1,3)=7$ and $\nabla f(1,-1,3)=(0,4,6)$.

Examples:

- $(x-a)f_x(a,b)+(y-b)f_y(a,b)=0$ is the equation of the tangent line to $f(x,y)=\alpha$ at (a,b).
- $(x-a)f_x(a,b,c)+(y-b)f_y(a,b,c)+(z-c)f_z(a,b,c)=0$ is the equation of the tangent plane to $f(x,y,z)=\alpha$ at (a,b,c).

Let $f(x,y,z)=x^2+2xy-y^2+z^2$. Then $f(1,-1,3)=7$ and $\nabla f(1,-1,3)=(0,4,6)$.

The tangent plane to $f(x,y,z)=7$ at $(1,-1,3)$ is given by

$$0\times(x-1)+4\times(y+1)+6\times(z-3)=0, \quad \text{i.e. }, 2y+3z=7.$$

Examples:

- $(x-a)f_x(a,b)+(y-b)f_y(a,b)=0$ is the equation of the tangent line to $f(x,y)=\alpha$ at (a,b).
- $(x-a)f_x(a,b,c)+(y-b)f_y(a,b,c)+(z-c)f_z(a,b,c)=0$ is the equation of the tangent plane to $f(x,y,z)=\alpha$ at (a,b,c).

Let $f(x,y,z)=x^2+2xy-y^2+z^2$. Then $f(1,-1,3)=7$ and $\nabla f(1,-1,3)=(0,4,6)$.

The tangent plane to $f(x,y,z)=7$ at $(1,-1,3)$ is given by

$$0\times(x-1)+4\times(y+1)+6\times(z-3)=0, \quad \text{i.e. }, 2y+3z=7.$$

The normal line to $f(x,y,z)=7$ at $(1,-1,3)$ is given by

$$(x,y,z)=(1,-1,3)+t(0,4,6) \text{ for } t\in\mathbb{R}.$$

Continuous partial derivatives

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that $\partial_i f(\mathbf{x})$ exists for $\mathbf{x} \in U$ and $i = 1 : n$. Then each $\partial_i f$ defines a function on U.

If $\partial_i f : U \to \mathbb{R}, \mathbf{x} \mapsto \partial_i f(\mathbf{x})$ is continuous for $i = 1 : n$ then f is said to be continuously differentiable (in short, C^1).

Continuous partial derivatives

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$. Suppose that $\partial_i f(\mathbf{x})$ exists for $\mathbf{x} \in U$ and $i = 1 : n$. Then each $\partial_i f$ defines a function on U.

If $\partial_i f : U \to \mathbb{R}, \mathbf{x} \mapsto \partial_i f(\mathbf{x})$ is continuous for $i = 1 : n$ then f is said to be continuously differentiable (in short, C^1).

Fact: f is $C^1 \iff \nabla f : U \subset \mathbb{R}^n \to \mathbb{R}^n, \mathbf{x} \mapsto \nabla f(\mathbf{x})$ is continuous.

Recall: f is $C^1 \Rightarrow f$ is differentiable $\not\Rightarrow f$ is C^1.

Examples:

- Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x^2 + e^{xy} + y^2$. Then f is C^1.
- Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(0, 0) = 0$ and $f(x, y) := (x^2 + y^2)\sin(1/(x^2 + y^2))$ if $(x, y) \neq (0, 0)$. Then f is differentiable but NOT C^1.

Higher order partial derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable so that $\partial_j f : \mathbb{R}^n \to \mathbb{R}$ for $j = 1 : n$.

If the partial derivatives of $\partial_j f$ exist at $\mathbf{a} \in \mathbb{R}^n$ for $j = 1 : n$, that is, $\partial_i \partial_j f(\mathbf{a})$ exists for $i, j = 1, 2, \ldots, n$, then f is said to have second order partial derivatives at $\mathbf{a}$.

Higher order partial derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable so that $\partial_i f : \mathbb{R}^n \to \mathbb{R}$ for $j = 1 : n$.

If the partial derivatives of $\partial_j f$ exist at $\mathbf{a} \in \mathbb{R}^n$ for $j = 1 : n$, that is, $\partial_i \partial_j f(\mathbf{a})$ exists for $i, j = 1, 2, \ldots, n$, then f is said to have second order partial derivatives at $\mathbf{a}$.

Other notations: $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}), \quad f_{x_i x_j}(\mathbf{a})$.

Higher order partial derivatives

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable so that $\partial_j f : \mathbb{R}^n \to \mathbb{R}$ for $j = 1 : n$.

If the partial derivatives of $\partial_j f$ exist at $\mathbf{a} \in \mathbb{R}^n$ for $j = 1 : n$, that is, $\partial_i \partial_j f(\mathbf{a})$ exists for $i, j = 1, 2, \ldots, n$, then f is said to have second order partial derivatives at $\mathbf{a}$.

Other notations: $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}), \quad f_{x_i x_j}(\mathbf{a})$.

f is said to be C^2 (twice continuously differentiable) if $\partial_i \partial_j f(\mathbf{x})$ exists for $\mathbf{x} \in \mathbb{R}^n$ and $\partial_i \partial_j f : \mathbb{R}^n \to \mathbb{R}$ is continuous for $i, j = 1, 2, \ldots, n$.

- p-th order partial derivatives of f are defined similarly.

Mixed partial derivatives

Fact: $f : \mathbb{R}^n \to \mathbb{R}$ is $C^2 \Rightarrow \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ has second order partial derivatives. Then $\partial_i \partial_j f(\mathbf{x})$ for $i \neq j$ is called mixed partial derivative of order 2.

Mixed partial derivatives

Fact: $f : \mathbb{R}^n \to \mathbb{R}$ is $C^2 \Rightarrow \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ has second order partial derivatives. Then $\partial_i \partial_j f(\mathbf{x})$ for $i \neq j$ is called mixed partial derivative of order 2.

Example: Consider $f(x, y) := x^2 + xy^2 + y^3$. Then $f_x = 2x + y^2 \Rightarrow f_{xy} = 2y$ and $f_y = 2xy + 3y^2 \Rightarrow f_{yx} = 2y$ showing that $f_{xy} = f_{yx}$.

Mixed partial derivatives

Fact: $f : \mathbb{R}^n \to \mathbb{R}$ is $C^2 \Rightarrow \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable.

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ has second order partial derivatives. Then $\partial_i \partial_j f(\mathbf{x})$ for $i \neq j$ is called mixed partial derivative of order 2.

Example: Consider $f(x, y) := x^2 + xy^2 + y^3$. Then $f_x = 2x + y^2 \Rightarrow f_{xy} = 2y$ and $f_y = 2xy + 3y^2 \Rightarrow f_{yx} = 2y$ showing that $f_{xy} = f_{yx}$.

Question: Is $\partial_i \partial_j f(\mathbf{x}) = \partial_j \partial_i f(\mathbf{x})$?

Unequal mixed partial derivatives

Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(0,0) = 0$ and

$$f(x,y) := xy\frac{x^2 - y^2}{x^2 + y^2} \text{ if } (x,y) \neq (0,0)$$

Then

Unequal mixed partial derivatives

Consider $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(0,0) = 0$ and

$$f(x,y) := xy\frac{x^2 - y^2}{x^2 + y^2} \text{ if } (x,y) \neq (0,0)$$

Then

$$\partial_x f(0,y) = -y \Rightarrow \partial_y \partial_x f(0,0) = -1$$

and

$$\partial_y f(x,0) = x \Rightarrow \partial_x \partial_y f(0,0) = 1.$$

This shows that

$$\partial_x \partial_y f(0,0) \neq \partial_y \partial_x f(0,0).$$

Equality of mixed partial derivatives

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$. Suppose that $\partial_i \partial_j f$ is continuous at $\mathbf{a}$ for $i, j = 1, 2, \ldots, n$. Then for all $i \neq j$,

$$\partial_i \partial_j f(\mathbf{a}) = \partial_j \partial_i f(\mathbf{a}).$$

In particular, if f is C^2 then $\partial_i \partial_j f(\mathbf{a}) = \partial_j \partial_i f(\mathbf{a})$. $\blacksquare$

Equality of mixed partial derivatives

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{a} \in \mathbb{R}^n$. Suppose that $\partial_i \partial_j f$ is continuous at $\mathbf{a}$ for $i, j = 1, 2, \ldots, n$. Then for all $i \neq j$,

$$\partial_i \partial_j f(\mathbf{a}) = \partial_j \partial_i f(\mathbf{a}).$$

In particular, if f is C^2 then $\partial_i \partial_j f(\mathbf{a}) = \partial_j \partial_i f(\mathbf{a})$. ■

Suppose that $f(x, y)$ has second order partial derivatives at $\mathbf{p} := (a, b)$. Then the matrix

$$H_f(\mathbf{p}) := \begin{bmatrix} \partial_x \partial_x f(\mathbf{p}) & \partial_y \partial_x f(\mathbf{p}) \\ \partial_x \partial_y f(\mathbf{p}) & \partial_y \partial_y f(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} f_{xx}(\mathbf{p}) & f_{xy}(\mathbf{p}) \\ f_{yx}(\mathbf{p}) & f_{yy}(\mathbf{p}) \end{bmatrix}$$

is called the Hessian of f at $\mathbf{p}$.

Hessian

Fact: Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 and $\mathbf{a} \in \mathbb{R}^n$. Then the Hessian

$$H_f(\mathbf{a}) := \begin{bmatrix} \partial_1\partial_1 f(\mathbf{a}) & \cdots & \partial_n\partial_1 f(\mathbf{a}) \\ \vdots & \cdots & \vdots \\ \partial_1\partial_n f(\mathbf{a}) & \cdots & \partial_n\partial_n f(\mathbf{a}) \end{bmatrix}$$

is symmetric. Also $H_f(\mathbf{a}) = J_{\nabla f}(\mathbf{a}) =$ Jacobian of ∇f at $\mathbf{a}$.

Hessian

Fact: Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is C^2 and $\mathbf{a} \in \mathbb{R}^n$. Then the Hessian

$$H_f(\mathbf{a}) := \begin{bmatrix} \partial_1\partial_1 f(\mathbf{a}) & \cdots & \partial_n\partial_1 f(\mathbf{a}) \\ \vdots & \cdots & \vdots \\ \partial_1\partial_n f(\mathbf{a}) & \cdots & \partial_n\partial_n f(\mathbf{a}) \end{bmatrix}$$

is symmetric. Also $H_f(\mathbf{a}) = J_{\nabla f}(\mathbf{a}) =$ Jacobian of ∇f at $\mathbf{a}$.

Example: Consider $f(x, y) = x^2 - 2xy + 2y^2$. Then

$$H_f(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}.$$

Extended Mean Value Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be C^2 and $\mathbf{a} \in \mathbb{R}^n$. Then there exists $0 < \theta < 1$ such that

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet \mathbf{h} + \frac{1}{2}\mathbf{h}^\top H_f(\mathbf{a}+\theta\mathbf{h})\mathbf{h}, \\ &= f(\mathbf{a}) + \sum_{i=1}^{n} \partial_i f(\mathbf{a}) h_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \partial_i\partial_j f(\mathbf{a}+\theta\mathbf{h}) h_i h_j. \end{aligned}$$

Extended Mean Value Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}$ be C^2 and $\mathbf{a} \in \mathbb{R}^n$. Then there exists $0 < \theta < 1$ such that

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet \mathbf{h} + \frac{1}{2}\mathbf{h}^\top H_f(\mathbf{a}+\theta\mathbf{h})\mathbf{h}, \\ &= f(\mathbf{a}) + \sum_{i=1}^{n} \partial_i f(\mathbf{a}) h_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \partial_i\partial_j f(\mathbf{a}+\theta\mathbf{h}) h_i h_j. \end{aligned}$$

Proof: Define $\phi(t) := f(\mathbf{a}+t\mathbf{h})$ for $t \in [0,1]$. Then ϕ is twice continuously differentiable. By chain rule

$$\phi'(t) = \nabla f(\mathbf{a}+t\mathbf{h}) \bullet \mathbf{h} \text{ and } \phi''(t) = \mathbf{h}^\top H_f(\mathbf{a}+t\mathbf{h})\mathbf{h}.\blacksquare$$

*** End***