Lecture 6:
Chain rule, Mean Value Theorem, Tangent Plane

Rafikul Alam
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

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

Proof: Use

$$f(\mathbf{x}(t)) = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x}(t) - \mathbf{x}(t_0)) + e(\mathbf{x}) \|\mathbf{x} - \mathbf{a}\|$$

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

Chain rule for partial derivatives

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

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

Proof: Apply Theorem-A.

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

Mean Value Theorem

Theorem: Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$. Then there exists $\mathbf{c} \in [\mathbf{a}, \mathbf{b}]$ such that

$$f(\mathbf{b}) - f(\mathbf{a}) = \nabla f(\mathbf{c}) \bullet (\mathbf{b} - \mathbf{a}),$$

where $[\mathbf{a}, \mathbf{b}] := \{\mathbf{a}(1-t) + t\mathbf{b} : t \in [0, 1]\}$ is the line segment joining $\mathbf{a}$ and $\mathbf{b}$.

Proof: Consider $\phi(t) := f(\mathbf{a}(1-t) + t\mathbf{b})$ for $t \in [0, 1]$ and invoke chain rule.

Total derivative

Let $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then

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

is called total derivative of f.

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

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} f(\mathbf{x}(t)) &= \nabla f(\mathbf{x}) \bullet \mathbf{x}'(t) = (2x, -2y) \bullet (\cos t, -\sin t) \\ &= 2\sin(t)\cos(t) + 2\cos(t)\sin(t) = 2\sin(2t). \end{aligned}$$

Differential

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Then

$$\mathrm{d}f = \sum_{i=1}^{n} \partial_i f(\mathbf{a}) \mathrm{d}x_i$$

is called differential of f at $\mathbf{a}$.

For $n = 1$ this gives familiar differential $\mathrm{d}f = f'(a)dx$.

Question: What is $\mathrm{d}f$? What does it represent?

Well, $\mathrm{d}f = \mathrm{D}f(\mathbf{a})$, the derivative map. The differential $\mathrm{d}f$ is a fancy way of writing the derivative map

$$\mathrm{D}f(\mathbf{a}) : \mathbb{R}^n \to \mathbb{R}, \mathbf{h} \mapsto \nabla f(\mathbf{a}) \bullet \mathbf{h}.$$

Differential = Derivative map

Define coordinate projection $\mathrm{d}x_i : \mathbb{R}^n \to \mathbb{R}$ by $\mathrm{d}x_i(\mathbf{h}) = h_i$. Then

$$\begin{aligned} \mathrm{D}f(\mathbf{a})\mathbf{h} &= \nabla f(\mathbf{a}) \bullet \mathbf{h} = \sum_{j=1}^{n} \partial_j f(\mathbf{a}) h_j \\ &= \sum_{j=1}^{n} \partial_j f(\mathbf{a}) \mathrm{d}x_j(\mathbf{h}) = \mathrm{d}f(\mathbf{h}). \end{aligned}$$

Hence $\mathrm{D}f(\mathbf{a}) = \sum_{i=1}^{n} \partial_i f(\mathbf{a}) \mathrm{d}x_i = \mathrm{d}f.$

Bottomline: The differential $\mathrm{d}f$ at $\mathbf{a}$ denotes $\mathrm{D}f(\mathbf{a})$ when $\mathrm{D}f(\mathbf{a})$ is expressed in terms of the partial derivatives $\partial_j f(\mathbf{a})$.

Differential and increament

Consider an increment $\Delta \mathbf{x} = (\Delta x_1, \ldots, \Delta x_n) \in \mathbb{R}^n$ and define

$$\Delta f(\Delta \mathbf{x}) := f(\mathbf{a} + \Delta \mathbf{x}) - f(\mathbf{a}).$$

Then

$$\Delta f(\Delta \mathbf{x}) = \mathrm{d}f(\Delta \mathbf{x}) + e(\Delta \mathbf{x})\|\Delta \mathbf{x}\| \simeq \mathrm{d}f(\Delta \mathbf{x})$$

gives

$$\Delta f(\Delta \mathbf{x}) \simeq \mathrm{d}f(\Delta \mathbf{x}) = \sum_{j=1}^{n} \partial_j f(\mathbf{a}) \mathrm{d}x_j(\Delta \mathbf{x}) = \sum_{j=1}^{n} \partial_j f(\mathbf{a}) \Delta x_j$$

Increment of f at $\mathbf{a} \simeq$ sum of scaled increments of the components of $\mathbf{a}$.

Level sets

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 curve/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 curve/surface in $\mathbb{R}^n$.

Examples:

- $f(x, y) := x^2 + y^2$. Then $G(f)$ is a paraboloid in $\mathbb{R}^3$ and $S(f, \alpha)$ with $\alpha > 0$ is a circle in $\mathbb{R}^2$.
- $f(x, y) := 4x^2 + y^2$. Then $G(f)$ is an elliptic paraboloid in $\mathbb{R}^3$ and $S(f, \alpha)$ with $\alpha > 0$ is an ellipse in $\mathbb{R}^2$.

Tangent plane to level sets

Equation of a hyperplane in $\mathbb{R}^n$ passing through $\mathbf{a}$ and a normal vector $\mathbf{n}$ is given by $(\mathbf{x} - \mathbf{a}) \bullet \mathbf{n} = 0$.

Equation of a line in $\mathbb{R}^n$ passing through $\mathbf{a}$ in the direction of $\mathbf{v}$ is given by $\mathbf{x} = \mathbf{a} + t\mathbf{v}$.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$.

Let $\mathbf{x} : (-\delta, \delta) \to S(f, \alpha)$ be differentiable at 0 and $\mathbf{x}(0) = \mathbf{a}$. Then

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

$\Rightarrow \mathbf{x}'(0)$ lies in the hyperplane $(\mathbf{x} - \mathbf{a}) \bullet \nabla f(\mathbf{a}) = 0$.

Tangent plane

Since the velocity $\mathbf{x}'(0)$ is tangent to the curve $\mathbf{x}(t)$ at $\mathbf{a}$, the hyperplane $(\mathbf{x} - \mathbf{a}) \bullet \nabla f(\mathbf{a}) = 0$ is tangent to the level set $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

Hence the line $\mathbf{x} = \mathbf{a} + t\nabla f(\mathbf{a})$ is normal to the level set $f(\mathbf{x}) = \alpha$ at $\mathbf{a}$.

- $(x - a)f_x(a, b) + (y - b)f_y(a, b) = 0$ 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).

Tangent to the graph

Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable at $\mathbf{a} \in \mathbb{R}^n$.

Define $g : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ by $g(\mathbf{x}, z) = f(\mathbf{x}) - z$. Then $S(g, 0) = G(f)$ and $\nabla g = (\nabla f, -1)$.

The tangent plane to $S(g, 0) = G(f)$ at $(\mathbf{a}, f(\mathbf{a}))$ is given by

$$(\mathbf{x} - \mathbf{a}, z - f(\mathbf{a})) \bullet (\nabla f(\mathbf{a}), -1) = 0$$

which gives $z = f(\mathbf{a}) + \nabla f(\mathbf{a}) \bullet (\mathbf{x} - \mathbf{a})$.

- $y = f(a) + f'(a)(x - a)$ is tangent to $y = f(x)$ at $(a, f(a))$.
- $z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$ is tangent to $z = f(x, y)$ at $(a, b, f(a, b))$.

Differentiability of $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$

Definition: Let $U \subset \mathbb{R}^n$ be open. Then $f : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\mathbf{a} \in U$ if there exists a linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ such that

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

The linear map L is called the derivative of f at $\mathbf{a}$ and is denoted by $\mathrm{D}f(\mathbf{a})$, that is, $L = \mathrm{D}f(\mathbf{a})$.

Other notations: $f'(a)$, $\frac{\mathrm{d}f}{\mathrm{d}\mathbf{x}}(a)$.

Characterization of differentiability

Theorem: Consider $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, where $f_i : \mathbb{R}^n \to \mathbb{R}$. Then f is differentiable at $\mathbf{a} \in \mathbb{R}^n \iff f_i$ is differentiable at $\mathbf{a}$ for $i = 1, 2, \ldots, m$. Further

$$\mathrm{D}f(\mathbf{a})\mathbf{h} = (\nabla f_1(\mathbf{a}) \bullet \mathbf{h}, \ldots, \nabla f_m(\mathbf{a}) \bullet \mathbf{h}).$$

Proof: Blackboard.

The matrix of $\mathrm{D}f(\mathbf{a})$ is called the Jacobian matrix of f at $\mathbf{a}$. Jacobian matrix is obtained by writing $\mathbf{x}$ and $f(\mathbf{x})$ as column vectors.

Jacobian matrix of $f : \mathbb{R}^n \to \mathbb{R}^m$

The jacobian matrix $\mathrm{D}f(\mathbf{a}) = \begin{bmatrix} -\nabla f_1(\mathbf{a})^\top - \\ \vdots \\ -\nabla f_m(\mathbf{a})^\top - \end{bmatrix}_{m\times n}$.

- $f(x, y) = (f_1(x, y), f_2(x, y), f_3(x, y))$

$$\mathrm{D}f(a, b) = \begin{bmatrix} \partial_x f_1(a, b) & \partial_y f_1(a, b) \\ \partial_x f_2(a, b) & \partial_y f_2(a, b) \\ \partial_x f_3(a, b) & \partial_y f_3(a, b) \end{bmatrix}$$

- $f(x, y, z) = (f_1(x, y), f_2(x, y))$

$$\mathrm{D}f(a, b, c) = \begin{bmatrix} \partial_x f_1(a, b, c) & \partial_y f_1(a, b, c) & \partial_z f_1(a, b, c) \\ \partial_x f_2(a, b, c) & \partial_y f_2(a, b, c) & \partial_z f_2(a, b, c) \end{bmatrix}$$

Examples

- If $f(x,y) = (xy, e^x y, \sin y)$ then

$$\mathrm{D}f(x,y) = \begin{bmatrix} y & x \\ e^x y & e^x \\ 0 & \cos y \end{bmatrix}$$

- If $f(x,y,z) = (x+y+z, xyz)$ then

$$\mathrm{D}f(x,y,z) = \begin{bmatrix} 1 & 1 & 1 \\ yz & xz & xy \end{bmatrix}$$

*** End ***