Lecture 7: Higher order partial derivatives, Hessian, Maxima and Minima

Rafikul Alam
Department of Mathematics
IIT Guwahati

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 $\partial_i f : \mathbb{R}^n \to \mathbb{R}$ 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: $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}), \quad f_{x_i x_j}(\mathbf{a}).$

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.

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

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

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

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 $\mathbf{a}$.

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

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

Proof: Apply EMVT to $\phi(t) := f(\mathbf{a}+t\mathbf{h})$ for $t \in [0,1]$ and use chain rule.

Local extremum

Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be continuous, where U is open. Then f has a

$$\left(\begin{array}{c} \text{local minimum at } \mathbf{a} \\ \text{local maximum at } \mathbf{a} \end{array} \right) \text{ if } \left(\begin{array}{c} f(\mathbf{x}) \geq f(\mathbf{a}) \\ f(\mathbf{x}) \leq f(\mathbf{a}) \end{array} \right)$$

for $\mathbf{x} \in B(\mathbf{a}, \epsilon)$ for some $\epsilon > 0$.

Necessary condition for local extremum: Suppose that f has a local extremum at $\mathbf{a}$. If $\nabla f(\mathbf{a})$ exists then $\nabla f(\mathbf{a}) = \mathbf{0}$.

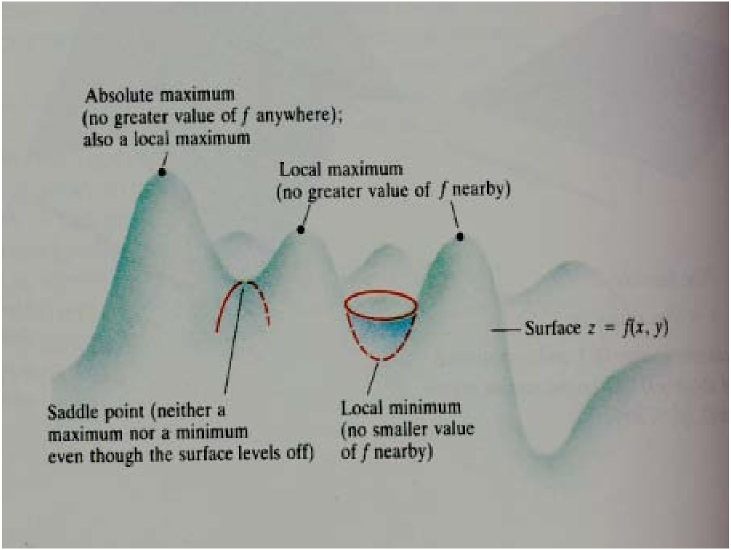

Figure: Local extremum of $z = f(x, y)$

Sufficient condition for extremum

Theorem: Let $f : U \subset \mathbb{R}^2 \to \mathbb{R}$ be C^2 and $\mathbf{p} \in U$. Suppose that $f_x(\mathbf{p}) = 0 = f_y(\mathbf{p})$ and

$$\det(H_f(\mathbf{p})) = f_{xx}(\mathbf{p}) f_{yy}(\mathbf{p}) - f_{xy}^2(\mathbf{p}) > 0.$$

- If $f_{xx}(\mathbf{p}) > 0$ then f has a local minimum at $\mathbf{p}$.
- If $f_{xx}(\mathbf{p}) < 0$ then f has a local maximum at $\mathbf{p}$.

Proof: Use EMVT and

$$f_{xx}(\mathbf{p})\mathbf{h}^\top H_f(\mathbf{p}))\mathbf{h} = (\nabla f(\mathbf{p}) \bullet \mathbf{h})^2 + k^2 \det(H_f(\mathbf{p})) > 0$$

to conclude that $\mathbf{h}^\top H_f(\mathbf{p}))\mathbf{h} > 0$ or < 0.

*** End ***