LECTURE 10

Higher Order Derivatives and Taylor Expansions

1. Higher Order Derivatives

Since a partial derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ is (wherever it exists) again a function from $\mathbb{R}^n$ to $\mathbb{R}$ it makes sense to talk about partial derivatives of partial derivatives; i.e., higher order partial derivatives.

EXAMPLE 10.1. Compute $\frac{\partial^2 f}{\partial x^2} \equiv \frac{\partial}{\partial x}\frac{\partial f}{\partial x}$, $\frac{\partial^2 f}{\partial x \partial y} \equiv \frac{\partial}{\partial x}\frac{\partial f}{\partial y}$ and $\frac{\partial^2 f}{\partial y \partial x} \equiv \frac{\partial}{\partial y}\frac{\partial f}{\partial x}$ where $f(x,y) = 3x^2y + x^2$.

$$\begin{aligned}
\frac{\partial^2 f}{\partial x^2} &\equiv \frac{\partial}{\partial x}\frac{\partial f}{\partial x} \\
&= \frac{\partial}{\partial x}(6xy + 2x) \\
&= 6y + 2
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial x \partial y} &\equiv \frac{\partial}{\partial x}\frac{\partial f}{\partial y} \\
&= \frac{\partial}{\partial x}\left(3x^2 + 0\right) \\
&= 6x
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial y \partial x} &\equiv \frac{\partial}{\partial y}\frac{\partial f}{\partial x} \\
&= \frac{\partial}{\partial y}(6xy + 2x) \\
&= 6x + 0 \\
&= 6x
\end{aligned}$$

Note that in this example

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

This is in fact a general phenomenon; *the value of a mixed partial derivative does not depend on the order in which the derivatives are taken.* Stated more formally;

THEOREM 10.2. *If $f : \mathbb{R}^n \to \mathbb{R}$ is such that all double partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exist and are continous, then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

2. Taylor's Formula for Functions of Several Variables

Recall that if $f(x)$ is a function of a single variable that is continuous and differentiable up to order $n+1$ then Taylor's theorem says that

$$f(x)=f(a)+f'(a)(x-a)+\frac{1}{2!}f''(a)(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n+R_n(x,a)$$

where the *error term* $R_n(x,a)$ is given by the formula

$$R_n(x,a)=\int_a^x\frac{x-s}{n!}f^{(n+1)}(s)ds$$

and that, moreover, the error term is of order $(x-a)^{n+1}$. Thus, to order $(x-a)^n$ we can approximate the function $f(x)$ by the polynomial function

$$T_n(x)=f(a)+f'(a)(x-a)+\frac{1}{2!}f''(a)(x-a)^2+\cdots+\frac{f^{(n)}(a)}{n!}(x-a)^n$$

There is an analogous theorem for functions of severa variables. However, since its general statement is a bit messy unless we introduce some new notation, we'll simply state the first and second order Taylor formulae

THEOREM 10.3. *Let $f:\mathbb{R}^n\to\mathbb{R}$ have continuous partial derivatives up to order 2. Then we may write*

$$f(\mathbf{x})=f(\mathbf{a})+\nabla f(\mathbf{a})\cdot(\mathbf{x}-\mathbf{a})+R_1(\mathbf{x},\mathbf{a})$$

with the error term $R_1(\mathbf{x},\mathbf{a})$ going to zero faster that a constant times $\|\mathbf{x}-\mathbf{a}\|^2$ as $\mathbf{x}\to\mathbf{a}$.

The first order Taylor polynomial is the function

$$\begin{aligned}T_1(\mathbf{x})&=f(\mathbf{a})+\nabla f(\mathbf{a})\cdot(\mathbf{x}-\mathbf{a})\\&=f(\mathbf{a})+\left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{a}}(x_1-a_1)+\cdots+\left.\frac{\partial f}{\partial x_n}\right|_{\mathbf{a}}(x_n-a_n)\quad.\end{aligned}$$

Note that this function is linear in the coordinates of $\mathbf{x}$. It's graph is thus a flat plane and generalizes the idea of the *best straight line fit to a curve*: it represents the best flat plane approximation to the graph of $f(\mathbf{x})$ near the point $\mathbf{x}_o$.

THEOREM 10.4. *Let $f:\mathbb{R}^n\to\mathbb{R}$ have continuous partial derivatives up to order 3. Then we may write*

$$f(\mathbf{x})=f(\mathbf{a})+\sum_{i=0}^n\frac{\partial f}{\partial x_i}(\mathbf{a})\,(x_i-a_i)+\frac{1}{2}\sum_{i=0}^n\sum_{j=0}^n\frac{\partial^2 f}{\partial x_i\partial x_j}(\mathbf{a})\,(x_i-a_i)\,(x_j-a_j)+R_2(\mathbf{x},\mathbf{a})$$

with the error term $R_2(\mathbf{x},\mathbf{a})$ going to zero faster that a constant times $\|\mathbf{x}-\mathbf{a}\|^3$ as $\mathbf{x}\to\mathbf{a}$.

EXAMPLE 10.5. Compute the second order Taylor formula for the function $f(x,y)=xy+x^2+y^2$ about the point $(1,1)$.

- We have

$$
\begin{aligned}
f(1,1) &= 1+1+1=3 \\
\left.\frac{\partial f}{\partial y}\right|_{(1,1)} &= (y+2x+0)|_{(1,1)} = 3 \\
\left.\frac{\partial f}{\partial y}\right|_{(1,1)} &= (x+0+2y)|_{(1,1)} = 3 \\
\left.\frac{\partial^2 f}{\partial x^2}\right|_{(1,1)} &= (0+2+0)|_{(1,1)} = 2 \\
\left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(1,1)} &= \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{(1,1)} = (1+0+0)|_{(1,1)} = 1 \\
\left.\frac{\partial^2 f}{\partial y^2}\right|_{(1,1)} &= (0+0+2)|_{(1,1)} = 2
\end{aligned}
$$

So

$$
\begin{aligned}
f(x,y) = f(1,1) &+ \left.\frac{\partial f}{\partial y}\right|_{(1,1)} (x-1) + \left.\frac{\partial f}{\partial y}\right|_{(1,1)} (y-1) \\
&+ \frac{1}{2}\left(\left.\frac{\partial^2 f}{\partial x^2}\right|_{(1,1)} (x-1)^2 + \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(1,1)} (x-1)(y-1) \right. \\
&+ \left. \left.\frac{\partial^2 f}{\partial y \partial x}\right|_{(1,1)} (y-1)(x-1) + \left.\frac{\partial^2 f}{\partial y^2}\right|_{(1,1)} (y-1)^2 \right) \\
&+ \mathcal{O}\left(\|(x,y)-(1,1)\|^3 \right) \\
&= 3 + 3(x-1) + 3(y-1) + \frac{1}{2}\left(2(x-1)^2 + 2(x-1)(y-1) + 2(y-1)^2\right) \\
&+ \mathcal{O}\left(\|(x,y)-(1,1)\|^3 \right) \\
&= 3 + 3(x-1) + 3(y-1) + (x-1)^2 + (x-1)(y-1) + (y-1)^2 \\
&+ \mathcal{O}\left(\|(x,y)-(1,1)\|^3 \right)
\end{aligned}
$$

$\square$

Below I present another (equivalent) formula for the second order Taylor expansion.

Let $(\mathbf{x}-\mathbf{a})$ be the n-dimensional column vector with components

$$
(\mathbf{x}-\mathbf{a}) = \begin{pmatrix} x_1 - a_1 \\ x_2 - a_{21} \\ \vdots \\ x_n - a_n \end{pmatrix}
$$

and let $(\mathbf{x}-\mathbf{a})^T$ be the matrix transpose of $(\mathbf{x}-\mathbf{a})$ (an n-dimensional row vector)

$$
(\mathbf{x}-\mathbf{a})^T = (x_1 - a_1, x_2 - a_2, \cdots, x_n - a_n) \quad .
$$

The gradient vector $\nabla f(\mathbf{a}) = Df(\mathbf{a})$, according to the conventions of Section 2.3 is an n-dimensional row vector;

$$
\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{x2}(\mathbf{a}), \cdots, \frac{\partial f}{\partial x_n}(\mathbf{a}) \right) \quad .
$$

Let us now define the **Hessian matrix** at the point $\mathbf{a}$ as the $n \times n$ matrix $\mathbf{H}f(\mathbf{a})$ defined by

$$\mathbf{H}f(\mathbf{a}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{a}) & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{pmatrix} .$$

Then we can write

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \mathbf{H}f(\mathbf{a})\,(\mathbf{x} - \mathbf{a}) + \mathcal{O}\left(\|\mathbf{x} - \mathbf{a}\|^3\right)$$

for the second order Taylor expansion of f about $\mathbf{a}$.