Lecture 9 (17.4.2019)

(translated and adapted from lecture notes by Martin Klazar)

Theorem 35 ($\partial \Rightarrow$ differential)**.** *Let $U \subset \mathbb{R}^m$ is a neighborhood of a point $\mathbf{a} \in \mathbb{R}^m$. If a function $f : U \to \mathbb{R}$ has all partial derivatives on U and they are continuous at $\mathbf{a}$, then f is differentiable at $\mathbf{a}$.*

Proof. We consider only the case of two variables x and y ($m = 2$). For more variables, the proof is similar (but more technical). We might assume that the point $\mathbf{a} = \mathbf{o}$ and U is a ball $B(\mathbf{o}, \gamma)$ for some $\gamma > 0$. Let $\mathbf{h} = (h_1, h_2) \in U$ (so, $\|bh\| < \gamma$) and $\mathbf{h}' = (h_1, 0)$. Difference $f(\mathbf{h}) - f(\mathbf{o})$ can be expressed as a sum of differences along both coordinate axes:

$$f(\mathbf{h}) - f(\mathbf{o}) = (f(\mathbf{h}) - f(\mathbf{h}')) + (f(\mathbf{h}') - f(\mathbf{o})) \ .$$

Segments $\mathbf{h}'\mathbf{h}$ and $\mathbf{o}\mathbf{h}'$ lie inside U, so f is defined on them, morever, f depends only on variable y on the former and only on variable x on the latter segment. Thus, Lagrange mean value Theorem (for single variable) yields:

$$f(\mathbf{h}) - f(\mathbf{o}) = \frac{\partial f}{\partial y}(\boldsymbol{\zeta_2}) \cdot h_2 + \frac{\partial f}{\partial x}(\boldsymbol{\zeta_1}) \cdot h_1 \ ,$$

where $\boldsymbol{\zeta_1}$ and $\boldsymbol{\zeta_2}$ are internal points of segments $\mathbf{o}\mathbf{h}'$ and $\mathbf{h}'\mathbf{h}$, respectively. In particular, the points $\boldsymbol{\zeta_1}$ and $\boldsymbol{\zeta_2}$ lie inside $B(\mathbf{o}, \|\mathbf{h}\|)$, so by continuity of both partial derivatives at $\mathbf{o}$, we have

$$\frac{\partial f}{\partial y}(\boldsymbol{\zeta_2}) = \frac{\partial f}{\partial y}(\mathbf{o}) + \alpha(\boldsymbol{\zeta_2}) \ \text{ and } \ \frac{\partial f}{\partial x}(\boldsymbol{\zeta_1}) = \frac{\partial f}{\partial x}(\mathbf{o}) + \beta(\boldsymbol{\zeta_1}) \ ,$$

where $\alpha(\mathbf{h})$ i $\beta(\mathbf{h})$ are $o(1)$ as $\mathbf{h} \to \mathbf{o}$ (i.e., for every $\varepsilon > 0$ there is $\delta > 0$, such that $\|\mathbf{h}\| < \delta \Rightarrow |\alpha(\mathbf{h})| < \varepsilon \cdot 1 = \varepsilon$ and the same holds for $\beta(h)$). Thus

$$f(\mathbf{h}) - f(\mathbf{o}) = \frac{\partial f}{\partial y}(\mathbf{o}) \cdot h_2 + \frac{\partial f}{\partial x}(\mathbf{o}) \cdot h_1 + \alpha(\boldsymbol{\zeta_2})h_2 + \beta(\boldsymbol{\zeta_1})h_1 \ .$$

By triangle inequality, and inequalities $0 < \|\boldsymbol{\zeta_1}\|, \|\boldsymbol{\zeta_2}\| < \|\mathbf{h}\|$ and $|h_1|, |h_2| \leq \|\mathbf{h}\|$ it follows that if $\|\mathbf{h}\| < \delta$, then

$$|\alpha(\boldsymbol{\zeta_2})h_2 + \beta(\boldsymbol{\zeta_1})h_1| \leq |\alpha(\boldsymbol{\zeta_2})| \cdot \|\mathbf{h}\| + |\beta(\boldsymbol{\zeta_1})| \cdot \|\mathbf{h}\| \leq 2\varepsilon\|\mathbf{h}\| \ .$$

Thus, $\alpha(\boldsymbol{\zeta_2})h_2 + \beta(\boldsymbol{\zeta_1})h_1 = o(\|\mathbf{h}\|)$ for $\mathbf{h} \to \mathbf{o}$. So by definition of the total differential, f is differentiable at $\mathbf{o}$. $\square$

Lagrange Mean Value Theorem can be generalized for functions of several variables as follows.

Theorem 36 (Lagrange Mean Value Theorem for several variables)**.** *Let $U \subset \mathbb{R}^m$ be an open set containing a segment $u = \mathbf{ab}$ with endpoints $\mathbf{a}$ and $\mathbf{b}$ and let $f : U \to \mathbb{R}$ be a function which is continuous at every point of u and differentiable at every internal point of u. Then there exists an internal point ζ of u satisfying*

$$f(\mathbf{b}) - f(\mathbf{a}) = \mathrm{D}f(\zeta)(\mathbf{b} - \mathbf{a}) .$$

In other words, difference of functional values at endpoints of the segment equals value of differential at some internal point of the segment for the vector of the segment.

Proof. Idea: Apply Lagrange Mean Value Theorem of single variable for an auxiliary function $F(t) = f(\mathbf{a} + t(\mathbf{b} - \mathbf{a}))$ and $t \in [0, 1]$. □

We say that an open set $D \subset \mathbb{R}^m$ is *connected*, if every two of its points can be connected by a broken line contained in D. Examples of connected open sets: an open ball in $\mathbb{R}^m$, whole $\mathbb{R}^m$ and $\mathbb{R}^3 \backslash L$, where L is the union of finitely many lines. On the other hand, $B \backslash R$, where B is an open ball $\mathbb{R}^3$ and R a plane intersecting B, is an open set which is not connected.

Corollary 37 ($\partial = 0 \Rightarrow f \equiv \text{const.}$)**.** *If a function f of m variables has zero differential at every point of an open connected set U, then f is constant on U. The same conclusion holds if f has all partial derivatives on U zero.*

Proof. Idea: Consider two points of U and a broken line connecting them. Apply Lagrange Mean Value Theorem for several variables for each segment of the broken line.

□

Calculating partial derivatives and differentials. For two functions $f, g : U \to \mathbb{R}$, defined on a neighborhood $U \subset \mathbb{R}^m$ of a point $\mathbf{a} \in U$ that have a partial derivative with repect to x_i at a point $\mathbf{a}$, formulae for partial derivative their sum, product and quotient are analogous to those for single variable:

$$\begin{aligned}
\partial_i(\alpha f + \beta g)(\mathbf{a}) &= \alpha \partial_i f(\mathbf{a}) + \beta \partial_i g(a) \\
\partial_i(fg)(a) &= g(\mathbf{a}) \partial_i f(\mathbf{a}) + f(\mathbf{a}) \partial_i g(\mathbf{a}) \\
\partial_i(f/g)(\mathbf{a}) &= \frac{g(\mathbf{a}) \partial_i f(\mathbf{a}) - f(\mathbf{a}) \partial_i g(\mathbf{a})}{g(\mathbf{a})^2} \quad (\text{if } g(\mathbf{a}) \neq 0) .
\end{aligned}$$

Similarly, for differentials, we have:

Theorem 38 (Arithmetic of differentials)**.** *Let $U \subset \mathbb{R}^m$ is a neighborhood of $\mathbf{a}$ and $f, g : U \to \mathbb{R}$ are functions differentiable at $\mathbf{a}$.*

(i) $\alpha f + \beta g$ is differentiable at $\mathbf{a}$ and

$$\mathrm{D}(\alpha f + \beta g)(\mathbf{a}) = \alpha \mathrm{D}f(\mathbf{a}) + \beta \mathrm{D}g(\mathbf{a}) .$$

for any $\alpha, \beta \in \mathbb{R}$, *(ii) fg is differentiable at $\mathbf{a}$ and*

$$\mathrm{D}(fg)(\mathbf{a}) = g(\mathbf{a})\mathrm{D}f(\mathbf{a}) + f(\mathbf{a})\mathrm{D}g(\mathbf{a}) \ .$$

(iii) If $g(\mathbf{a}) \neq 0$, f/g is differentiable at $\mathbf{a}$ and

$$\mathrm{D}(f/g)(\mathbf{a}) = \frac{1}{g(\mathbf{a})^2}\Big(g(\mathbf{a})\mathrm{D}f(\mathbf{a}) - f(\mathbf{a})\mathrm{D}g(\mathbf{a})\Big) \ .$$

Proof. Follows from Theorem 32 and formulae for partial derivatives. □

The formula for linear combination can be easily generalized for vector valued functions $f, g: \ U \to \mathbb{R}^n$.

Next, we generalize a formula for derivative of a composed function to a composition of multivariable mappings. We use $\circ$ for denoting composition, where $(g \circ f)(\mathbf{x}) = g(f(\mathbf{x}))$.

Theorem 39 (Differential of a composed mapping)**.** *Let*

$$f: \ U \to V, \ g: \ V \to \mathbb{R}^k$$

are two mappings where $U \subset \mathbb{R}^m$ is a neighborhood of $\mathbf{a}$ and $V \subset \mathbb{R}^n$ is a neighborhood of $\mathbf{b} = f(\mathbf{a})$. If the mapping f is differentiable at $\mathbf{a}$ and g is differentialble at $\mathbf{b}$, the composed mapping

$$g \circ f = g(f): \ U \to \mathbb{R}^k$$

is differentiable at $\mathbf{a}$ and the total differential is a composition of differentials of f and g:

$$\mathrm{D}(g \circ f)(\mathbf{a}) = \mathrm{D}g(\mathbf{b}) \circ \mathrm{D}f(\mathbf{a}) \ .$$

Since composition of linear mappings corresponds to multiplication of matrices, total differential of a composed mapping corresponds to a product of the Jacobi matrices.

Partial derivatives of higher orders. If the $f: \ U \to \mathbb{R}$ function defined on a neighborhood $U \subset \mathbb{R}^m$ of a point $\mathbf{a}$ has a partial derivative $F = \partial f x_i$ in each point U and this function $F: \ U \to \mathbb{R}$ has at $\mathbf{a}$ the partial derivative $\partial F x_j(\mathbf{a})$, we say that f has a partial derivative at the point $\mathbf{a}$ of the second order with respect to the variables x_i and x_j and we denote it

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$$

or shortly by $\partial_i \partial_j f(\mathbf{a})$.

Similarly, we define higher order partial derivatives: if $f = f(x_1, x_2, \ ldots, x_m)$ has partial derivative $(i_1, i_2, \ldots, i_{k-1}, j \in \{1, 2, \ldots, m\})$

$$F = \frac{\partial^{k-1} f}{\partial x_{i_{k-1}} \partial x_{i_{k-2}} \ldots \partial x_{i_1}}(x)$$

at every point $x\ inU$ and we say that f has *partial derivative of order k with respect to the variables* $x_{i_1}, \dots, x_{i_{k-1}}, x_j$ in point $\mathbf{a}$ and we denote its value by

$$\frac{\partial^k f}{\partial x_j \partial x_{i_{k-1}} \dots \partial x_{i_1}}(\mathbf{a}) \ .$$

In general, order of variables in higher order derivatives matters. You can verify yourself that $f:\ \mathbb{R}^2 \to \mathbb{R}$,

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & \text{pro } x^2+y^2 \neq 0 \\ 0 & \text{pro } x^2+y^2 = 0 \ , \end{cases}$$

has different *mixed* (i.e., with respect to two different variables) second order partial derivatives in the origin.

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \ \text{ a } \ \frac{\partial^2 f}{\partial y \partial x}(0,0) = -1 \ .$$

However, the order does not matter if the partial derivatives are continuous.

Theorem 40 (Usually $\partial_x\partial_y f = \partial_y \partial_x f$)**.** *Let $f:\ U \to \mathbb{R}$ be a function with second order partial derivatives $\partial_j\partial_i f$ a $\partial_i\partial_j f$, $i \neq j$ on a neighborhood $U \subset \mathbb{R}^m$ of a point* $\mathbf{a}$ *which are continuous in* $\mathbf{a}$. *Then*

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

Proof. We prove the statement for $m = 2$, for $m > 2$, the proof would be analogous but more tedious. Without loss of generality, we may assume that $\mathbf{a} = \mathbf{o} = (0,0)$. By continuity of the partial derivatives in the origin, it is enough to find for arbitrarily small $h > 0$ two points σ, τ in the square $[0,h]^2$ satisfying $\partial_x\partial_y f(\sigma) = \partial_y\partial_x f(\tau)$. Then, for $h \to 0^+$, $\sigma, \tau \to \mathbf{o}$ and from a limit argument and continuity of the partial derivatives we get that $\partial_x\partial_y f(\mathbf{o}) = \partial_y\partial_x f(\mathbf{o})$.

Given h, we find σ and τ as follows. We denote the corners of the square $\mathbf{a} = (0,0)$, $\mathbf{b} = (0,h)$, $\mathbf{c} = (h,0)$, $\mathbf{d} = (h,h)$ and we consider a value $f(\mathbf{d}) - f(\mathbf{b}) - f(\mathbf{c}) + f(\mathbf{a})$. It can be expressed in two different ways:

$$\begin{aligned} f(\mathbf{d}) - f(\mathbf{b}) - f(\mathbf{c}) + f(\mathbf{a}) &= (f(\mathbf{d}) - f(\mathbf{b})) - (f(\mathbf{c}) - f(\mathbf{a})) = \psi(h) - \psi(0) \\ &= (f(\mathbf{d}) - f(\mathbf{c})) - (f(\mathbf{b}) - f(\mathbf{a})) = \phi(h) - \phi(0) \ , \end{aligned}$$

where

$$\psi(t) = f(h,t) - f(0,t) \ \text{ and } \ \phi(t) = f(t,h) - f(t,0) \ .$$

We have that $\psi'(t) = \partial_y f(h,t) - \partial_y f(0,t)$ and $\phi'(t) = \partial_x f(t,h) - \partial_x f(t,0)$. Lagrange mean value theorem gives two expresions

$$\begin{aligned} f(\mathbf{d}) - f(\mathbf{b}) - f(\mathbf{c}) + f(\mathbf{a}) &= \psi'(t_0)h = (\partial_y f(h,t_0) - \partial_y f(0,t_0))h \\ &= \phi'(s_0)h = (\partial_x f(s_0,h) - \partial_x f(s_0,0))h \ , \end{aligned}$$

where $0 < s_0, t_0 < h$ are intermediate points. Applying the theorem once more on differences of partial derivatives of f, we obtain the following

$$f(\mathbf{d}) - f(\mathbf{b}) - f(\mathbf{c}) + f(\mathbf{a}) = \partial_x \partial_y f(s_1, t_0) h^2 = \partial_y \partial_x f(s_0, t_1) h^2,\ \ s_1, t_1 \in (0, h)\ .$$

Points $\sigma = (s_1, t_0)$ and $\tau = (s_0, t_1)$ belong to $[0, h]^2$ and we have $\partial_x \partial_y f(\sigma) = \partial_y \partial_x f(\tau)$ (since both sides equal to $(f(\mathbf{d}) - f(\mathbf{b}) - f(\mathbf{c}) + f(\mathbf{a}))/h^2$). $\square$