Lecture 10 (24.4.2019)

(translated and adapted from lecture notes by Martin Klazar)

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 $x_{i_1}, \ldots, x_{i_{k-1}}, x_j$* in point $\mathbf{a}$ and we denote its value by

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

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 43 (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 $(f(\mathbf{d}) - f(\mathbf{b}) - f(\mathbf{c}) + f(\mathbf{a}))/h^2$). □

For an open set $U \subset \mathbb{R}^m$ we denote by $\mathcal{C}^k(U)$ the set of functions $f : U \to \mathbb{R}$, such that all their partial derivatives of order up to k (inclusive) (exist and) are continuous on U.

Corollary 44 (Reordering partial derivatives)**.** *For every function $f = f(x_1, x_2, \ldots, x_m)$ from $\mathcal{C}^k(U)$ values of its partial derivatives up to order k do not depend on the order of variables—for $l \le k$ and $\mathbf{a} \in U$ it holds that*

$$\frac{\partial^l f}{\partial x_{i_l} \partial x_{i_{l-1}} \ldots \partial x_{i_1}}(a) = \frac{\partial^l f}{\partial x_{j_l} \partial x_{j_{l-1}} \ldots \partial x_{j_1}}(\mathbf{a}) \ ,$$

whenever $(i_1, \ldots, i_l)$ and $(j_1, \ldots, j_l)$ differ only by permutation of the elements.

Proof. (idea) If a sequence $v = (j_1, \ldots, j_l)$ is a permutation of the sequence $u = (i_1, \ldots, i_l)$, one can turn u into v only by swapping consecutive pairs of elements (in a bubble sort like manner). Then the equality of partial derivatives follows from the previous theorem. □

Since only the multiset of variables matters in case of continuous partial derivatives, we can more briefly write ∂x^2 instead of $\partial_x \partial_x$. For instance, for f from $\mathcal{C}^5(U)$ on U we have

$$\frac{\partial^5 f}{\partial y\, \partial x\, \partial y\, \partial y\, \partial z} = \frac{\partial^5 f}{\partial y^2\, \partial x\, \partial z\, \partial y} = \frac{\partial^5 f}{\partial x\, \partial z\, \partial y^3} = \frac{\partial^5 f}{\partial z\, \partial y^3\, \partial x}\ .$$

Local extrema of multivariate functions

Extrema of the multivariate functions are defined as follows. A function $f:\ U \to \mathbb{R}$, where $U \subset \mathbb{R}^m$ is an open neighborhood of a point $\mathbf{a}$, has in $\mathbf{a}$

- *strict local minimum*, if there exists $\delta > 0$, such that $0 < \|\mathbf{x} - \mathbf{a}\| < \delta \Rightarrow f(\mathbf{x}) > f(\mathbf{a})$,
- *(non-strict) local minimum*, if there exists $\delta > 0$, such that $0 < \|\mathbf{x} - \mathbf{a}\| < \delta \Rightarrow f(\mathbf{x}) \geq f(\mathbf{a})$.

Strict and non-strict local minimum are defined analogously. A function $f:\ M \to \mathbb{R}$, where $M \subset \mathbb{R}^m$, has *maximum on a set* M if $f(\mathbf{a}) \geq f(\mathbf{x})$ for every $\mathbf{x} \in M$. Again, minimum is defined analogously.

Recall facts about extrema of function of a single variable from winter:

1. if $f'(\mathbf{a}) \neq 0$, f does not have a local extremum in $\mathbf{a}$;
2. if $f'(\mathbf{a}) = 0$ and $f''(\mathbf{a}) > 0$, f has a strict local minimum in $\mathbf{a}$ and
3. if $f'(\mathbf{a}) = 0$ and $f''(\mathbf{a}) < 0$, f has a strict local maximum in $\mathbf{a}$.

If $f'(\mathbf{a}) = f''(\mathbf{a}) = 0$, we canot decide whether f has extremum in $\mathbf{a}$ or not without further analysis. If $f'(\mathbf{a}) = 0$ ($\mathbf{a}$ is a "suspicious" point), we cannot, based on the value of the second derivative $f''(\mathbf{a})$ rule out the existence of a local extremum. As we shall see, this is not the case for multivariate functions.

In winter term, it was shown that continuous function has extrema on closed bounded interval. This generalizes to multivariate functions. We say that a set $M \subset \mathbb{R}^m$ is *bounded*, if there exists a real $R > 0$, such that $M \subset B(\overline{0}, R)$. Recall that M is *closed*, if its complement $\mathbb{R}^m \backslash M$ is open. We say that $M \subset \mathbb{R}^m$ is *compact*, when it is closed and bounded.

Theorem 45 (Extrema on compact)**.** *Let $M \subset \mathbb{R}^m$ be a nonempty compact set and $f:\ M \to \mathbb{R}$ a continuous function on M. Then f attains minimum and maximum on M.*

For instance the unit sphere $S = \{\mathbf{x} \in \mathbb{R}^n |\ ||\mathbf{x}|| = 1\}$ in $\mathbb{R}^n$, is a compact set and thus every continuous function $f:\ S \to \mathbb{R}$ attains minimum and maximum on S.

We fist introduce some notation and recall some facts from linear algebra. Let $A = (a_{i,j}) \in \mathbb{R}^{n \times n}$ be a real symmetric matrix ($a_{i,j} = a_{j,i}$) of size $n \times n$. A quadratic form corresponding to this matrix is a function of n variables defined as

$$P_A(x_1, x_2, \dots, x_n) = \mathbf{x}A\mathbf{x}^T = \sum_{i,j=1}^{n} a_{i,j}x_ix_j : \mathbb{R}^n \to \mathbb{R} ,$$

where $\mathbf{x}$ is a row vector $(x_1, x_2, \dots, x_n)$ and $\mathbf{x}^T$ is the corresponding column vector.

A matrix A is

- *pozitive (negative) definite*, if $P_A(\mathbf{x}) > 0$ ($P_A(\mathbf{x}) < 0$) for every $\mathbf{x} \in \mathbb{R}^n \setminus \{\overline{0}\}$;
- *pozitive (negative) semidefinite*, if $P_A(\mathbf{x}) \geq 0$ ($P_A(\mathbf{x}) \leq 0$) for every $\mathbf{x} \in \mathbb{R}^n$ and
- *indefinite*, if it is none of the previous, that is, there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $P_A(\mathbf{x}) > 0$ and $P_A(\mathbf{y}) < 0$.

Hessian matrix $H_f(\mathbf{a})$ of a function f in a point $\mathbf{a}$, where $U \subset \mathbb{R}^m$ is an open neighborhood of $\mathbf{a}$ and $f : U \to \mathbb{R}$ is a function with all partial derivatives of second order on U, is a metrix recording values of these derivatives in $\mathbf{a}$:

$$H_f(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) \right)_{i,j=1}^{m} .$$

By theorem that $\partial_x \partial_y = \partial_y \partial_x$, if $f \in \mathcal{C}^2(U)$ its Hessian matrix is symmetric.

Theorem 46 (Necessary condition for local extremum.)**.** *Let $f : U \to \mathbb{R}$, where $U \subset \mathbb{R}^m$ is an open neighborhood of* $\mathbf{a}$*. If $\nabla f(\mathbf{a}) \neq \overline{0}$, then f does not have local extremum in* $\mathbf{a}$.

Proof. For $i = 1, \dots m$, define auxiliary functions of a single variable $g_i(h) = f(\mathbf{a} + h\bar{e}_i)$. Note that $g_i'(0) = \partial f x_i(\mathbf{a})$. By results from winter term, it follows that if $\partial f x_i(\mathbf{a}) \neq 0$, g_i does not have an extremum in 0. Moreover, if g_i does not have an extremum in 0, f does not have an extremum in $\mathbf{a}$. □

Theorem 47 (Sufficient conditions for local extrema)**.** *Let $f \in \mathcal{C}^2(U)$, where $U \subset \mathbb{R}^m$ is an open neighborhood of* $\mathbf{a}$.

1. *If $\nabla f(\mathbf{a}) = \overline{0}$ and $H_f(\mathbf{a})$ is positive (negative) definite, then f has local minimum (maximum) in* $\mathbf{a}$.
2. *If $\nabla f(\mathbf{a}) = \overline{0}$ and $H_f(\mathbf{a})$ is indefinite, f does not have local extremum in* $\mathbf{a}$.

Sylvester kriterion from linear algebra gives the following way to recognize definiteness of a symmetric matrix: if all subdeterminants $d_m = \det(a_{i,j})_{i,j=1}^m$, $1 \leq m \leq n$, are non-zero, then, if all of them are positive, the matrix A is pozitive definite, if $(-1)^m d_m > 0$, $1 \leq m \leq n$, then A is negative definite, and the matrix is indefinite otherwise. (If some of the determinants are zero, we don't know.)