Lecture 11 (15.5.2019)

(translated and slightly adapted from lecture notes by Martin Klazar)

Implicit functions. As we know from linear algebra, system of n linear equations with n variables $a_{i,1}y_1 + a_{i,2}y_2 + \ldots + a_{i,n}y_n + b_i = 0$, $i = 1, 2, \ldots, n$, where $a_{i,j} \in \mathbb{R}$ are given constants and $\det(a_{i,j})_{i,j=1}^n \neq 0$, has for each choice of n constants b_i unique solution $y_1, y_2, \ldots, y_n$. Moreover, this solution y_j is a homogenous linear functions of b_i, that is: $y_j(b_1, b_2, \ldots, b_n) = c_{j,1}b_1 + c_{j,2}b_2 + \ldots + c_{j,n}b_n$, $j = 1, 2, \ldots, n$, for some constants $c_{j,i} \in \mathbb{R}$ (this follows from Crammer's rule).

We now generalize this result to the situation when the linear functions are replaced by general functions. We will consider a system of n equations with $m + n$ variables

$$\begin{aligned} F_1(x_1, \ldots, x_m, y_1, \ldots, y_n) &= 0 \\ F_2(x_1, \ldots, x_m, y_1, \ldots, y_n) &= 0 \\ &\vdots \\ F_n(x_1, \ldots, x_m, y_1, \ldots, y_n) &= 0\,, \end{aligned}$$

where F_i are real functions defined on some neighborhood of a point $(\bar{x}_0, \bar{y}_0)$ in $\mathbb{R}^{m+n}$, where $\bar{x}_0 \in \mathbb{R}^m$ and $\bar{y}_0 \in \mathbb{R}^n$, is a solution of the system, that is $F_1(\mathbf{x}_0, \mathbf{y}_0) = F_2(\mathbf{x}_0, \mathbf{y}_0) = \ldots = F_n(\mathbf{x}_0, \mathbf{y}_0) = 0$. We shall see that under certain conditions it is possible to express variables $y_1, y_2, \ldots, y_n$ as functions $y_i = f_i(x_1, x_2, \ldots, x_m)$ of variables $x_1, x_2, \ldots, x_m$ on some neighborhood of x_0. Even in simplest cases we cannot expect to have necessarily a solution, not to speak of a unique one. Consider example the following single equation

$$F(x, y) = x^2 + y^2 - 1 = 0.$$

For $|x| > 1$ there is no y with $f(x, y) = 0$. For $|x_0| < 1$, we have in a sufficiently small open interval containing x_0 two solutions

$$f(x) = \sqrt{1 - x^2} \text{ and } g(x) = -\sqrt{1 - x^2}.$$

This is better, but we have *two* values in each point, contradicting the definition of a function. To achieve uniqueness, we have to restrict not only the values of x, but *also the values of* y to an interval $(y_0 - \Delta, y_0 + \Delta)$ (where $F(x_0, y_0) = 0$). That is, if we have a particular solution (x_0, y_0) we have a "window"

$$(x_0 - \delta, x_0 + \delta) \times (y_0 - \Delta, y_0 + \Delta)$$

through which we see a unique solution.

But in our example there is also the case $(x_0, y_0) = (1, 0)$, where there is a unique solution, but no suitable window as above, since in every neighborhood of $(1, 0)$, there are no solutions for any value x slightly bigger and two solutions for value x slightly smaller.

Theorem 48 (Implicit function Theorem.). *Let $F(\mathbf{x}, y)$ be a function of $n+1$ variables defined in a neighbourhood of a point $(\mathbf{x}_0, y_0)$. Let F have continuous partial derivatives up to the order $p \geq 1$ and let*

$$F(\mathbf{x}_0, y_0) = 0 \text{ and } \left|\frac{\partial F}{\partial y}(\mathbf{x}_0, y_0)\right| \neq 0.$$

Then there exist $\delta > 0$ and $\Delta > 0$ such that for every $\mathbf{x}$ with $||\mathbf{x} - \mathbf{x}_0|| < \delta$ there exists precisely one y with $|y - y_0| < \Delta$ such that

$$F(\mathbf{x}, y) = 0.$$

Furthermore, if we write $y = f(\mathbf{x})$ for this unique solution y, then the function

$$f : B(\mathbf{x}, \delta) \to \mathbb{R}$$

has continuous partial derivatives up to the order p. Moreover,

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = -\frac{\frac{\partial F}{\partial x_i}(\mathbf{x}, f(\mathbf{x}))}{\frac{\partial F}{\partial y}(\mathbf{x}, f(\mathbf{x}))}$$

for every $i = 1, \dots n$.

We will not prove this theorem, however, we show how to derive the formula for partial derivatives of the implicit function f, assuming they exist.

Since we have

$$0 \equiv F(\mathbf{x}, f(\mathbf{x}));$$

taking a derivative of both sides (using the Chain Rule) we obtain.

$$0 = \frac{\partial F}{\partial x_i}(\mathbf{x}, f(\mathbf{x})) + \frac{\partial F}{\partial y}(\mathbf{x}, f(\mathbf{x})) \cdot \frac{\partial f}{\partial x_i}(\mathbf{x}).$$

.

From this, we can express $\frac{\partial f}{\partial x_i}(\mathbf{x})$. Differentiating further, we obtain inductively linear equations from which we can compute the values of all the derivatives guaranteed by the theorem.

For more than a system of several functions, we can apply the previous theorem inductively, eliminating variables one by one.

Theorem 49 (Implicit functions). *Let*

$$F = (F_1, F_2, \dots, F_n) : W \to \mathbb{R}^n$$

be a mapping defined on a neighborhood $W \subset \mathbb{R}^{m+n}$ of a point $(\mathbf{x}_0, \mathbf{y}_0)$, where $\mathbf{x}_0 \in \mathbb{R}^m$ and $\mathbf{y}_0 \in \mathbb{R}^n$, satisfying the following conditions:

1. *$F_i = F_i(\mathbf{x}, \mathbf{y}) \in \mathcal{C}^1(W)$ pro $1 \leq i \leq n$.*
2. *$F_i(\mathbf{x}_0, \mathbf{y}_0) = 0$ pro $1 \leq i \leq n$.*

3. $\det\begin{pmatrix} \frac{\partial F_1}{\partial y_1} & \frac{\partial F_1}{\partial y_2} & \cdots & \frac{\partial F_1}{\partial y_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial F_n}{\partial y_1} & \frac{\partial F_n}{\partial y_2} & \cdots & \frac{\partial F_n}{\partial y_n} \end{pmatrix}(\mathbf{x}_0, \mathbf{y}_0)) \neq 0.$

Then there exist neighborhoods $U \subset \mathbb{R}^m$ and $V \subset \mathbb{R}^n$ of $\mathbf{x}_0$ a $\mathbf{y}_0$ such that $U \times V \subset W$ and for every $\mathbf{x} \in U$ there exists exactly one $\mathbf{y} \in V$ satisfying $F_i(\mathbf{x}, \mathbf{y}) = 0$ for $1 \leq i \leq n$. In other words, there exist a mapping $f = (f_1, f_2, \ldots, f_n) : U \to V$ such that

$$\forall (\mathbf{x}, \mathbf{y}) \in U \times V : \ F(\mathbf{x}, \mathbf{y}) = \overline{0} \iff \mathbf{y} = f(\mathbf{x}) \ .$$

Moreover f_i is $\mathcal{C}^1(U)$ for every $i = 1, \ldots n$.

Constrained extrema. From Implicit functions theorem one can derive a necessary condition for local extrema on sets defined by a system of equations.

Let $U \subset \mathbb{R}^m$ be an open set and let

$$f, F_1, \ldots, F_n : \ U \to \mathbb{R}$$

be functions from $\mathcal{C}^1(U)$, where $n < m$. We wish to find extrema of f on a set

$$H = \{\mathbf{x} \in U \mid F_1(\mathbf{x}) = F_2(\mathbf{x}) = \cdots = F_n(\mathbf{x}) = 0\} \ .$$

Such a set usually does not have any internal points. Example of such a set is a unit sphere in $\mathbb{R}^m$:

$$\{\mathbf{x} \in \mathbb{R}^m \mid x_1^2 + x_2^2 + \ldots + x_m^2 - 1 = 0\} \ .$$

Theorem 50 (Lagrange multipliers)**.** *Let $U \subset \mathbb{R}^m$ be an open set,*

$$f, F_1, \ldots, F_n : \ U \to \mathbb{R}$$

be functions from $\mathcal{C}^1(U)$, where $n < m$ and let

$$H = \{\mathbf{x} \in U \mid F_1(\mathbf{x}) = F_2(\mathbf{x}) = \cdots = F_n(\mathbf{x}) = 0\} \ .$$

Let $\mathbf{a} \in H$. If $\nabla F_1(\mathbf{a}), \ldots, \nabla F_n(\mathbf{a})$ are linearly independent and $\nabla f(\mathbf{a})$ is not their linear combination, then f does not have a local extremum with respect to H in $\mathbf{a}$.

Equivalently: if $\nabla F_1(\mathbf{a}), \ldots, \nabla F_n(\mathbf{a})$ are linearly independent and f has local extremum in $\mathbf{a}$ with respect to H , then there exist reals $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, called Lagrange multipliers*, such that*

$$\nabla f(\mathbf{a}) - \sum_{i=1}^{n} \lambda_i \nabla F_i(\mathbf{a}) = \overline{0} \ .$$

that is,

$$\frac{\partial f}{\partial x_j}(\mathbf{a}) - \lambda_1 \frac{\partial F_1}{\partial x_j}(\mathbf{a}) - \cdots - \lambda_n \frac{\partial F_n}{\partial x_j}(\mathbf{a}) = 0$$

for every $1 \leq j \leq m$.