Implicit Function Theorem

From the Inverse Function Theorem we derive a theorem that has many applications; it will turn out that it is equivalent to the Inverse Function Theorem.

We consider open sets $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ and a function

$$f : X \times Y \to \mathbb{R}^m.$$

(In the simplest case, $m = n = 1$.)

Let $(a,b) \in X \times Y$. We investigate the question whether there are neighborhoods U of a and V of b and a function $\varphi : U \to V$ with

$$\varphi(a) = b \quad \text{and} \quad (\forall x \in U)\,\big(f(x, \varphi(x)) = f(a,b)\big).$$

If "yes", then such a function φ will be called implicitly defined function.

Example

Let $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,y) = x^2 + y^2$ and $(a,b) \in \mathbb{R}^2$ s.t. $f(a,b) = 1$.

If $-1 < a < 1$, then the equation $f(x, \varphi(x)) = 1 \iff x^2 + \varphi^2(x) = 1$ can be solved for $\varphi(x)$, yielding $\varphi(x) = \sqrt{1 - x^2}$ for $-1 < x < 1$, so $U := \,]-1,1[$ is a neighborhood of a and $V := \,]-1,1[$ is a neighborhood of $b \in \,]-1,1[$. If $a = -1$ or $a = 1$, such a solution function does not exist. $\square$

Let $(a,b) \in X \times Y$. We define the following functions:

$$f_a : Y \subseteq \mathbb{R}^m \to \mathbb{R}^m, \quad f_a(y) = f(a,y) \quad \text{for all } y \in Y,$$

$$f_b : X \subseteq \mathbb{R}^n \to \mathbb{R}^m, \quad f_b(x) = f(x,b) \quad \text{for all } x \in X,$$

$$I_a : Y \longrightarrow \mathbb{R}^{n+m}, \quad I_a(y) = \begin{pmatrix} a \\ y \end{pmatrix} \quad \text{for all } y \in Y,$$

$$I_b : X \longrightarrow \mathbb{R}^{n+m}, \quad I_b(x) = \begin{pmatrix} x \\ b \end{pmatrix} \quad \text{for all } x \in X.$$

<u>Remark 9.47</u>

For all $(a,b) \in X \times Y$,

$$f_a = f \circ I_a, \quad f_b = f \circ I_b.$$

<u>Remark 9.48</u>

For all $(a,b) \in X \times Y$, I_a and I_b are differentiable on X respectively Y. Furthermore, for all $x \in X$, $y \in Y$ we have that

$$J_{I_a}(y) = \begin{pmatrix} O_{n,m} \\ I_m \end{pmatrix} \text{ is a } (m+n) \times m\text{-matrix},$$

$$I_a'(y)(v) = \begin{pmatrix} O_{\mathbb{R}^n} \\ v \end{pmatrix} \quad \text{for all } v \in \mathbb{R}^m,$$

$$J_{I_b}(x) = \begin{pmatrix} I_n \\ O_{m,n} \end{pmatrix} \text{ is a } (m+n) \times n\text{-matrix}$$

$$I_b'(x)(u) = \begin{pmatrix} u \\ O_{\mathbb{R}^m} \end{pmatrix} \quad \text{for all } u \in \mathbb{R}^n.$$

Lemma 1.41

Assume that f is differentiable at (a,b). Then

(i) f_a is differentiable at b and

$$f_a'(b)(v) = f'(a,b)(0_{\mathbb{R}^n}, v) \quad \text{for all } v \in \mathbb{R}^m,$$

$$J_{f_a}(b) = \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(a,b) & \cdots & \frac{\partial f_1}{\partial y_m}(a,b) \\ \vdots & & \\ \frac{\partial f_m}{\partial y_1}(a,b) & \cdots & \frac{\partial f_m}{\partial y_m}(a,b) \end{pmatrix}.$$

(ii) f_b is differentiable at a and

$$f_b'(a)(u) = f'(a,b)(u, 0_{\mathbb{R}^m}) \quad \text{for all } u \in \mathbb{R}^n,$$

$$J_{f_b}(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a,b) & \cdots & \frac{\partial f_1}{\partial x_n}(a,b) \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1}(a,b) & \cdots & \frac{\partial f_m}{\partial x_n}(a,b) \end{pmatrix}.$$

(iii) $$f'(a,b)(u,v) = f_a'(b)(v) + f_b'(a)(u) \quad \text{for all } u \in \mathbb{R}^n, v \in \mathbb{R}^m;$$

$$J_f(a,b) = \left(J_{f_b}(a) \quad J_{f_a}(b) \right).$$

Proof

Apply the above two remarks and the chain rule. □

We shall denote $f'_b(a)$ with $\frac{\partial f}{\partial x}(a,b)$ and call it the partial derivative of f at (a,b) with respect to x. Similarly, we denote $f'_a(b)$ with $\frac{\partial f}{\partial y}(a,b)$ and call it the partial derivative of f at (a,b) with respect to y.

Theorem 8.50 (Implicit Function Theorem)

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be open sets and $f: X \times Y \to \mathbb{R}^m$ be a continuously differentiable function. Suppose $(a,b) \in X \times Y$ is s.t.

$$f(a,b) = 0 \quad \text{and} \quad \frac{\partial f}{\partial y}(a,b) \text{ is invertible.}$$

Then there are open neighborhoods U of a in X and V of b in Y and a uniquely determined continuously differentiable function $\varphi: U \to V$ s.t.

$$(\forall (x,y) \in U \times V)\left(f(x,y) = 0 \iff y = \varphi(x) \right).$$

Moreover, U and V can be chosen s.t. $\frac{\partial f}{\partial y}(x,y)$ is invertible for all $(x,y) \in U \times V$. In these conditions, for all $x \in U$,

$$\varphi'(x) = -\left(\frac{\partial f}{\partial y}(x,\varphi(x))\right)^{-1} \circ \frac{\partial f}{\partial x}(x,\varphi(x)).$$

Proof

Step 1

Let us define $F: X \times Y \to \mathbb{R}^{n+m}$, $F(x,y) = \begin{pmatrix} x \\ f(x,y) \end{pmatrix}$ for all $x \in X, y \in Y$.

Then, using the Chain Rule, we get that F is continuously differentiable,

$$J_F(x,y) = \begin{pmatrix} I_n & O_{n,m} \\ J_{f_y}(x) & J_{f_x}(y) \end{pmatrix}$$

and for all $(u,v) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$F'(x,y)(u,v) = \begin{pmatrix} u \\ \frac{\partial f}{\partial x}(x,y)(u) + \frac{\partial f}{\partial y}(x,y)(v) \end{pmatrix}.$$

Step 2

Let us prove that $F'(a,b)$ is invertible.

For a given vector $\binom{s}{t} \in \mathbb{R}^n \times \mathbb{R}^m$, we have that

$$F'(a,b)(u,v) = \binom{s}{t} \iff \begin{pmatrix} u \\ \frac{\partial f}{\partial x}(a,b)(u) + \frac{\partial f}{\partial y}(a,b)(v) \end{pmatrix}$$

$$\iff u = s \quad \text{and} \quad t = \frac{\partial f}{\partial x}(a,b)(u) + \frac{\partial f}{\partial y}(a,b)(v)$$

$$\iff u = s \quad \text{and} \quad \frac{\partial f}{\partial y}(a,b)(v) = t - \frac{\partial f}{\partial x}(a,b)(u)$$

$$\iff u = s \quad \text{and} \quad v = \left(\frac{\partial f}{\partial y}(a,b)\right)^{-1}\left(t - \frac{\partial f}{\partial x}(a,b)(s)\right)$$

Thus, for any $\binom{s}{t} \in \mathbb{R}^n \times \mathbb{R}^m$, the equation $F'(a,b)(u,v) = \binom{s}{t}$ has a unique solution (u,v), so $F'(a,b)$ is invertible and for all $(s,t) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$(F'(a,b))^{-1}(s,t) = \begin{pmatrix} s \\ \left(\frac{\partial f}{\partial y}(a,b)\right)^{-1}\left(t - \frac{\partial f}{\partial x}(a,b)(s)\right) \end{pmatrix}$$

$$= \begin{pmatrix} s \\ \left[-\left(\frac{\partial f}{\partial y}(a,b)\right)^{-1} \circ \frac{\partial f}{\partial x}(a,b)\right](s) + \left(\frac{\partial f}{\partial y}(a,b)\right)^{-1}(t) \end{pmatrix}.$$

Step 3

We have got that $F: X \times Y \to \mathbb{R}^{n+m}$ is continuously differentiable and $F'(a,b)$ is invertible. Then we can apply the Inverse Function Theorem to get an open neighborhood $D \subseteq X \times Y$ of (a,b) and an open neighborhood $W \subseteq \mathbb{R}^n \times \mathbb{R}^m$ of $F(a,b) = \begin{pmatrix} a \\ f(a,b) \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix}$

s.t.

(1) $F'(x,y)$ is invertible for all $(x,y) \in D$

(2) $F|_D : D \to W$ is bijective and the inverse $F^{-1} : W \to D$ is continuously differentiable with

$$(F^{-1})'(F(x,y)) = (F'(x,y))^{-1} \quad \text{for all } (x,y) \in D.$$

Step 4

Since $(a,b) \in D \subseteq X \times Y$, it is easy to see that there are open sets $U_0 \subseteq X$ and $V \subseteq Y$ s.t. $a \in U_0$, $b \in V$ and $U_0 \times V \subseteq D$.

Let now

$$U := \{x \in U_0 : \text{there exists } y \in V \text{ with } f(x,y) = 0\}.$$

Let us prove firstly that U is open.

We have that $x \in U \iff x \in U_0$ and $(\exists y \in V)\,(f(x,y) = 0)$

$\Longleftrightarrow x \in U_0$ and $(\exists y \in V)\left(F(x,y) = \begin{pmatrix} x \\ 0 \end{pmatrix}\right)$

$\Longleftrightarrow x \in U_0$ and $(\exists y \in V)\left(F(x,y) = I_0(x)\right)$

$\Longleftrightarrow I_0(x) \in F(U_0 \times V) \Longleftrightarrow x \in \underset{\mathbb{R}^m}{I_0^{-1}}(F(U_0\times V))$

Since $F(U_0\times V) = (F^{-1})^{-1}(U_0\times V)$ and F^{-1} is continuous, we get that $F(U_0\times V)$ is open. Since I_0 is continuous, it follows that $U = I_0^{-1}(F(U_0\times V))$ is open.

Step 5

Let us prove that

(*) $\quad x \in U \Longrightarrow$ there exists a unique $y \in V$ s.t. $f(x,y)=0$.

Assume that $y', y'' \in V$, $y' \neq y''$ are such that $f(x,y') = f(x,y'')$.

It follows that $F(x,y') = \begin{pmatrix} x \\ f(x,y') \end{pmatrix} = \begin{pmatrix} x \\ f(x,y'') \end{pmatrix} = F(x,y'')$

But F is injective on D and $U\times V \subseteq U_0 \times V \subseteq D$. Thus, we have got a contradiction. Thus, (*) is satisfied.

As a consequence, there exists a uniquely determined function $\varphi: U \to V$, $\varphi(x) =$ the unique y satisfying $f(x,y)=0$.

Obviously,

$$(\forall (x,y) \in U\times V)\left(f(x,y) = 0 \Longleftrightarrow y = \varphi(x)\right).$$

Step 6

We shall prove that φ is continuously differentiable.

Let us denote with p_2 the projection

$$p_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n+m}, \quad p_2(x,y) = y.$$

Then it is easy to see that

$$\varphi(x) = p_2 \circ F^{-1}(x,0) = \left(p_2 \circ F^{-1} \circ I_{0_{\mathbb{R}^m}}\right)(x)$$

Since p_2, I_0 are linear, it follows that they are continuously differentiable. Moreover, F^{-1} is continuously differentiable on W, by Step 3.

Step 7

Since, by Step 3, $F'(x,y)$ is invertible for all $(x,y) \in U \times V$ and, by Step 1, $J_F(x,y) = \begin{pmatrix} I_n & O_{n,m} \\ \partial f_y(x) & \partial f_x(y) \end{pmatrix}$ is invertible, it follows that $\partial f_x(y)$ is also invertible, so $f'_x(y) = \frac{\partial f}{\partial y}(x,y)$ is also invertible for all $(x,y) \in U \times V$.

By applying the Chain Rule to the equality

$$f(x, \varphi(x)) = 0 \qquad \text{for all } x \in U,$$

it follows that for all $x \in U$,

$$\frac{\partial f}{\partial x}(x, \varphi(x)) + \frac{\partial f}{\partial y}(x, \varphi(x)) \circ \varphi'(x) = 0,$$

so

$$\varphi'(x) = -\left(\frac{\partial f}{\partial y}(x, \varphi(x))\right)^{-1} \circ \frac{\partial f}{\partial x}(x, \varphi(x)).$$

$\square$

Corollary 9.51

Let $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}$ be open sets and $f: X \times Y \to \mathbb{R}$ be continuously differentiable. Suppose that $(a_1, \dots, a_n, a_{n+1}) \in X \times Y$ is such that

$$f(a_1, \dots, a_n, a_{n+1}) = 0 \quad \text{and} \quad \frac{\partial f}{\partial x_{n+1}}(a_1 \dots, a_n, a_{n+1}) \neq 0.$$

Then there are open neighborhoods U of $(a_1 \dots a_n)$ in X and V of a_{n+1} in Y and a uniquely determined continuously differentiable function $\varphi: U \to V$ s.t.

$$(\forall (x_1 \dots x_n, x_{n+1}) \in U \times V) \left(f(x_1 \dots x_n, x_{n+1}) = 0 \iff x_{n+1} = \varphi(x_1 \dots x_n) \right)$$

Moreover, $\frac{\partial f}{\partial x_{n+1}}(x_1 \dots x_n, x_{n+1}) \neq 0$ for all $(x_1 \dots x_n, x_{n+1}) \in U \times V$

and for all $(x_1, \dots x_n) \in U$,

$$\varphi'(x_1 \dots x_n) = -\frac{1}{\frac{\partial f}{\partial x_{n+1}}(x_1 \dots x_n, \varphi(x_1 \dots x_n))} \cdot \left(\frac{\partial f}{\partial x_1}(x_1 \dots x_n, \varphi(x_1 \dots x_n)) \dots \frac{\partial f}{\partial x_n}(x_1 \dots x_n, \varphi(x_1 \dots x_n)) \right)$$

Proof

Apply the Implicit Function Theorem with $m = 1$. □

Example 9.52

Consider the function

$$f: \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}, \quad f(x, y, z) = x^2 + y^2 + z^2 - 1.$$

Here $n = 2$ and $m = 1$.

We have that $f(x_0,y_0,z_0)=0 \Leftrightarrow x_0^2+y_0^2+z_0^2=1 \Leftrightarrow$
$\Leftrightarrow (x_0,y_0,z_0)$ lies on the surface of a sphere centred at the origin $(0,0,0)$ and with radius 1. On the other hand,

$$\frac{\partial f}{\partial z}(x_0,y_0,z_0)=2z_0\neq 0 \quad \text{whenever } z_0\neq 0.$$

Thus, this partial derivative does not vanish on the surface of the sphere except on $z_0=0$ and $x_0^2+y_0^2=1$. Hence, for $(x_0,y_0,z_0)\in\mathbb{R}^3$ s.t. $x_0^2+y_0^2+z_0^2=1$ and $z\neq 0$ we can apply the Implicit Function Theorem (or its Corollary 3.51) to get a uniquely determined function φ defined on a sufficiently small neighborhood U of (x_0,y_0) s.t.

$$f(x,y,\varphi(x,y))=0 \quad \text{for all } (x,y)\in U.$$

Moreover,

$$\frac{\partial \varphi}{\partial x}(x,y)=-\frac{\frac{\partial f}{\partial x}(x,y,\varphi(x,y))}{\frac{\partial f}{\partial z}(x,y,\varphi(x,y))}=-\frac{2x}{2\varphi(x,y)}$$

$$\frac{\partial f}{\partial y}(x,y)=-\frac{\frac{\partial f}{\partial y}(x,y,\varphi(x,y))}{\frac{\partial f}{\partial z}(x,y,\varphi(x,y))}=-\frac{2y}{2\varphi(x,y)}.$$

It is easy to see that the function

$$z=\varphi(x,y)=\sqrt{1-x^2-y^2}$$

satisfies the requirements if $z_0>0$, and the function

$$z=\varphi(x,y)=-\sqrt{1-x^2-y^2}$$

satisfies the requirements if $z_0<0$.

Example 9.53

Let $f: \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$f(x,y,z) = \begin{pmatrix} 3x^2+xy-z-3 \\ 2xz+y^3+xy \end{pmatrix} \quad \text{for all } (x,y,z) \in \mathbb{R}^3.$$

Thus, $f = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}$, where
$$f_1(x,y,z) = 3x^2+xy-z-3$$
$$f_2(x,y,z) = 2xz+y^3+xy.$$

Here $n=1$ and $m=2$.

We have $f(1,0,0)=0$. To given $z \in \mathbb{R}$ from a neighborhood of 0 we seek $(x,y) \in \mathbb{R}^2$ s.t. $f(x,y,z)$. To this end we must test, whether the matrix

$$\begin{pmatrix} \frac{\partial f_1}{\partial x}(x,y,z) & \frac{\partial f_1}{\partial y}(x,y,z) \\ \frac{\partial f_2}{\partial x}(x,y,z) & \frac{\partial f_2}{\partial y}(x,y,z \end{pmatrix} = \begin{pmatrix} 6x+y & x \\ 2z+y & 3y^2+x \end{pmatrix}$$

is invertible at $(x,y,z)=(1,0,0)$. At this point, the determinant of this matrix is $\begin{vmatrix} 6 & 1 \\ 0 & 1 \end{vmatrix} = 6 \neq 0$, hence the matrix is invertible. Consequently, we can apply the Implicit Function Theorem to get a neighborhood $U =]-\delta,\delta[$ of 0 in $\mathbb{R}$ and a neighborhood V of $(1,0)$ in $\mathbb{R}^2$ and a uniquely determined function $\varphi:]-\delta,\delta[\to V$ s.t. $f(\varphi_1(z), \varphi_2(z), z) = 0$, where φ_1, φ_2 are the components of φ. Moreover, φ is continuously differentiable and for all $z \in V$, if $x = \varphi_1(z)$ and $y = \varphi_2(z)$,

$$J\varphi(z) = -\begin{pmatrix} 6x+y & x \\ 2z+y & 3y^2+x \end{pmatrix}^{-1} \cdot \begin{pmatrix} \frac{\partial f_1}{\partial z}(x,y,z) \\ \frac{\partial f_2}{\partial z}(x,y,z \end{pmatrix}$$

$$= \frac{-1}{(6x+y)(3y^2+x)-x(2z+y)} \cdot \begin{pmatrix} 3y^2+x & -x \\ -(2z+y) & 6x+y \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 2x \end{pmatrix}$$

$$= -\frac{1}{(6x+y)(3y^2+x)-x(2z+y)} \cdot \begin{pmatrix} -3y^2-x-2x^2 \\ 2z+y+12x^2+2xy \end{pmatrix}.$$

Thus,

$$J\varphi(z) = \frac{-1}{(6\varphi_1(z)+\varphi_2(z))(3\varphi_2^2(z)+\varphi_1(z))-\varphi_1(z)(2z+\varphi_2(z))} \cdot \begin{pmatrix} -3\varphi_2^2(z)-\varphi_1(z)-2\varphi_1^2(z) \\ 2z+\varphi_2(z)+12\varphi_1^2(z)+2\varphi_1(z)\varphi_2(z) \end{pmatrix}$$

Since $\varphi(0) = (1,0)$, so $\varphi_1(0)=1$, $\varphi_2(0)=0$, we get that

$$J\varphi(0) = -\frac{1}{6} \cdot \begin{pmatrix} -3 \\ 12 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -2 \end{pmatrix}.$$

Thus, for all $h \in \mathbb{R}$,

$$\varphi'(0)(h) = J\varphi(0) \cdot h = \begin{pmatrix} \frac{1}{2} \\ -2 \end{pmatrix} \cdot h = \begin{pmatrix} \frac{1}{2}h \\ -2h \end{pmatrix}.$$

$\square$

Higher order partial derivatives

Higher order partial derivatives are defined similarly to higher order derivatives of real functions.

Let $D \subseteq \mathbb{R}^n$ be an open set and $f: D \to \mathbb{R}$ be a real-valued function. Let $i,j=1,\dots,n$ and $a \in D$. If the i-th partial derivative exists on D, $\frac{\partial f}{\partial x_i} : D \to \mathbb{R}$, we define the <u>ij-th second order partial derivative of f at $a \in D$</u> as the j-th partial derivative of $\frac{\partial f}{\partial x_i}$ at a, provided this exists. Thus, the ij-th partial derivative of f at a is $\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(a)$. This is usually abbreviated $\frac{\partial^2 f}{\partial x_j \partial x_i}(a)$ or $D_j D_i f(a)$.

If $i=j$, then we use the notation $\frac{\partial^2 f}{\partial x_i^2}(a)$.

Let now $k=1,\dots,n$. Assume that the ij-th second order partial derivative exists on D, $\frac{\partial^2 f}{\partial x_j \partial x_i} : D \to \mathbb{R}$. We define the <u>ijk-th third order partial derivative of f at a</u> as the k-th partial derivative of $\frac{\partial^2 f}{\partial x_j \partial x_i}$ at a, provided that this exists; it is denoted with $\frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}$ and if it exists for all $a \in D$, we have a function $\frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i} : D \to \mathbb{R}$.

And so on for still higher order partial derivatives.

When still higher order partial derivatives are in question, certain obvious abbreviations are used. For example

$$\frac{\partial^4 f}{\partial x\, \partial y^2\, \partial z} \quad \text{means} \quad \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right)\right)\right).$$

The large number of possible higher order partial derivatives of a function of several variables is much reduced by the circumstance that the order of performing the partial differentiation is usually irrelevant. The simplest case of this is the equation

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

repeated applications of which yields

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

Thus, the following theorem holds

Theorem 9.54 (H.A. Schwarz)

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$. Assume that f is twice continuously partial differentiable. Then for all $a \in D$ and all $i, j = 1, \dots, n$

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

For a proof of this theorem see Forster, Analysis II, Satz 1, p. 40 or Hofmann, Lecture Notes on Analysis II, Chapter 4: Higher derivatives, p. 2-4.

f is called twice partial differentiable at a if all second order partial derivatives at a exist. By induction, for $k \in \mathbb{N}$, we say that f is k-times partial differentiable at a if all partial derivatives of order k at a exist. f is called k-times continuously partial differentiable when f is k-times partial differentiable and the partial derivatives of order $\leq k$ are continuous.

Definition 9.55

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ be twice continuously differentiable. The Hessian matrix of f at $a \in D$, denoted by $(\operatorname{Hess} f)(a)$, is the matrix

$$(\operatorname{Hess} f)(a) = \left(\frac{\partial^2 f}{\partial x_j \, \partial x_i} \right)_{i,j=1,\dots,n}$$

As a consequence of Theorem 9.54, the Hessian matrix is symmetric.

Taylor's Theorem

In the sequel, we use the following abbreviations.

For $p = (p_1, \dots, p_n) \in \mathbb{N}^n$, define

$$|p| := p_1 + \dots + p_n \quad \text{(length of } p)$$

$$p! := p_1! \cdot p_2! \cdot \ldots \cdot p_n!$$

For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ define

$$x^p := x_1^{p_1} \cdot x_2^{p_2} \cdot \ldots \cdot x_n^{p_n}.$$

If f is $|p|$-times continuously partial differentiable, set

$$\partial^p f(x) := \frac{\partial^{|p|} f}{\partial x_1^{p_1} \, \partial x_2^{p_2} \cdots \partial x_n^{p_n}}(x) \quad \text{for every } x \in D.$$

Theorem 3.56 (Taylor's Theorem)

Let $D \subseteq \mathbb{R}^n$ be open and let $f : D \to \mathbb{R}$ be $(k+1)$-times continuously partial differentiable. Assume that $x, a \in D$ are such that $[a, x] \subseteq D$. Then there exists $c \in [a, x]$ s.t.

$$f(x) = \sum_{|p| \leq k} \frac{\partial^p f(a)}{p!} \cdot (x-a)^p + \sum_{|p|=k+1} \frac{\partial^p f(c)}{p!} \cdot (x-a)^p.$$

For a proof of this theorem, see Forster, Analysis II, Satz 2, p. 56-57.

As a corollary, we get the following very useful variant of Taylor's Theorem.

Corollary 3.57

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ be k-times continuously differentiable. Assume that $x \in D$ and $\delta > 0$ are such that $U_\delta(x) \subseteq D$.

Then for all $h \in \mathbb{R}^n$ with $\|h\| < \delta$,

$$f(x+h) = \sum_{|p| \leq k} \frac{\partial^p f(x)}{p!} \cdot h^p + r(h),$$

where r is a function with $r(0) = 0$ and $\lim_{h \to 0} \frac{r(h)}{\|h\|^k} = 0$.

Local extrema

Definition 9.58

Let $D \subseteq \mathbb{R}^n$ be open, $f: D \to \mathbb{R}$ be differentiable and $a \in D$.

If $f'(a) = 0$, then a is called a critical point of f.

Proposition 9.59

Let $D \subseteq \mathbb{R}^n$ be open and $f: D \to \mathbb{R}$ be differentiable. If f attains a local extreme value at a, then $f'(a) = 0$. That is, a is a critical point of f.

For a proof, see Alber, Analysis II, Theorem 5.2, p. 97.

As in the case of real functions, the second order partial derivatives can be used to formulate a sufficient criterion for an extreme value. To this end, we recall the following result from Linear Algebra.

Proposition 9.60

Let $A = (a_{ij})_{i,j=1,\dots,n}$ be a symmetric real $(n \times n)$-matrix.

Then A is positive definite iff for all $k = 1, \dots, n$

$$\det \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \\ a_{k1} & \cdots & a_{kk} \end{pmatrix} > 0.$$

Proposition 1.61

Let $D \subseteq \mathbb{R}^n$ be open and $f: D \to \mathbb{R}$ be twice continuously partial differentiable. Assume that a is a critical point of f. Then

(i) if $(\text{Hess } f)(a)$ is positive definite, then f has a local minimum at a;

(ii) if $(\text{Hess } f)(a)$ is negative definite, then f has a local maximum at a

(iii) if $(\text{Hess } f)(a)$ is indefinite, then f does not have an extreme value at a.

Example 1.62

Let $f: \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = 6xy - 3y^2 - 2x^3$.

f is twice continuously partial differentiable and for all $(x,y) \in \mathbb{R}^2$

$$\partial f(x,y) = \left(\frac{\partial f}{\partial x}(x,y) \quad \frac{\partial f}{\partial y}(x,y) \right) = \left(6y - 6x^2 \quad 6x - 6y \right)$$

and

$$\text{Hess } f(x,y) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial y \partial x}(x,y) \\ \frac{\partial^2 f}{\partial x \partial y}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{pmatrix} = \begin{pmatrix} -12x & 6 \\ 6 & -6 \end{pmatrix}.$$

We have that $(x,y) \in \mathbb{R}^2$ is a critical point of f iff

$$f'(x,y) = 0 \iff \partial f(x,y) = (0 \quad 0) \iff \text{grad } f(x,y) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

$$\iff \begin{cases} 6y - 6x^2 = 0 \\ 6x - 6y = 0 \end{cases} \iff (x,y) = (0,0) \text{ or } (x,y) = (1,1).$$

Thus, f has two critical points $(x,y) = (0,0)$ and $(x,y) = (1,1)$.

To determine wether these critical points are extremal points, we use Proposition 8.61.

Let us consider the critical point $(x,y)=(0,0)$. Then

$$(\mathrm{Hess}\, f)(0,0) = \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix}.$$ We shall prove that this matrix is indefinite.

For $h_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we get that

$$h_1^T (\mathrm{Hess}\, f)(0,0)\, h_1 = (1 \;\; 1) \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 6 > 0,$$

and for $h_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, we get that

$$h_2^T (\mathrm{Hess}\, f)(0,0)\, h_2 = (0 \;\; 1) \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -6 < 0.$$

Therefore, by Proposition 8.61 (iii), f does not attain an extreme value at $(0,0)$.

Consider now the second critical point $(x,y)=(1,1)$. Then

$$(\mathrm{Hess}\, f)(1,1) = \begin{pmatrix} -12 & 6 \\ 6 & -6 \end{pmatrix}, \text{ so } \quad -(\mathrm{Hess}\, f)(1,1) = \begin{pmatrix} 12 & -6 \\ -6 & 6 \end{pmatrix}.$$

Since $12 > 0$ and $\det \begin{pmatrix} 12 & -6 \\ -6 & 6 \end{pmatrix} = 72 - 36 = 36 > 0$, by Proposition 8.60 we get that $-(\mathrm{Hess}\, f)(1,1)$ is positive definite. That is, $(\mathrm{Hess}\, f)(1,1)$ is negative definite. Applying Proposition 8.61 (ii), it follows that f has a local maximum at a. □