Chapter 8

Calculus of Functions of Several Variables

In this chapter we consider functions $f\colon U \to \mathbb{R}$ or $f\colon U \to \mathbb{R}^m$ where $U \subset \mathbb{R}^n$ is an open set. In Proposition 6.17, we collected the main properties of *continuous* functions f. Now we will study differentiation and integration of such functions in more detail

The Norm of a linear Mapping

Proposition 8.1 *Let $T \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ be a linear mapping of euclidean spaces.*
(a) *Then there exists some $C > 0$ such that*

$$\|Tx\| \le C\,\|x\|, \quad \textit{for all } x \in \mathbb{R}^n. \tag{8.1}$$

(b) *T is uniformly continuous on $\mathbb{R}^n$.*

Proof. (a) Using the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ we identify T with its matrix $T = (a_{ij})$, $T\mathrm{e}_j = \sum_{i=1}^m a_{ij}\mathrm{e}_i$. For $x = (x_1, \dots, x_n)$ we have

$$T\,x = \left(\sum_{j=1}^n a_{1j}x_j,\ \dots,\ \sum_{j=1}^n a_{mj}x_j\right);$$

hence by the Cauchy–Schwarz inequality we have

$$\|T\,x\|^2 = \sum_{i=1}^m\sum_{j=1}^n |\,a_{ij}x_j\,|^2 \le \sum_{i=1}^m\sum_{j=1}^n |\,a_{ij}\,|^2 \sum_{j=1}^n |\,x_j\,|^2 = \left(\sum_{i,j} a_{ij}^2\right) \sum_{j=1}^n |\,x_j\,|^2 = C^2\,\|x\|^2,$$

where $C = \sqrt{\sum a_{ij}^2}$. Consequently,

$$\|Tx\| \le C\,\|x\|.$$

(b) Let $\varepsilon > 0$. Put $\delta = \varepsilon / C$ with the above C. Then $\|x - y\| < \delta$ implies

$$\|Tx - Ty\| = \|T(x - y)\| \le C\,\|x - y\| < \varepsilon,$$

which proves (b). ∎

Any linear operator $T \in \mathrm{L}(E, F)$ from a normed vector space E into a normed vector space F such that (8.1) holds for some $C > 0$ is called *bounded*.

Definition 8.1 Let V and W normed vector spaces and $A \in \mathrm{L}(V, W)$. The smallest number C with (8.1) is called the *norm* of the linear map A and is denoted by $\|A\|$.

$$\|A\| = \inf\{C \mid \|Ax\| \leq C\,\|x\| \quad \text{for all } x \in V\}. \tag{8.2}$$

By definition,

$$\|Ax\| \leq \|A\|\;\|x\|\,. \tag{8.3}$$

Let $T \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ be a linear mapping. One can show that

$$\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\| = \sup_{\|x\| \leq 1} \|Tx\|\,.$$

8.1 Partial Derivatives

We consider functions $f \colon U \to \mathbb{R}$ where $U \subset \mathbb{R}^n$ is an open set. We want to find derivatives “one variable at a time.”

Definition 8.2 Let $U \subset \mathbb{R}^n$ be open and $f \colon U \to \mathbb{R}$ a real function. Then f is called *partial differentiable* at $a = (a_1, \dots, a_n) \in U$ with respect to the ith coordinate if the limit

$$D_i f(a) = \lim_{h \to 0} \frac{f(a_1, \dots, a_i + h, \dots, a_n) - f(a_1, \dots, a_n)}{h} \tag{8.4}$$

exists where h is real and sufficiently small (such that $(a_1, \dots, a_i + h, \dots, a_n) \in U$).
$D_i f(x)$ is called the *ith partial derivative of f at a*. We also use the notations

$$D_i f(a) = \frac{\partial f}{\partial x_i}(a) = \frac{\partial f(a)}{\partial x_i} = f_{x_i}(a).$$

It is important that $D_i f(a)$ is the ordinary derivative of a certain function; in fact, if $g(x) = f(a_1, \dots, x, \dots, a_n)$, then $D_i f(a) = g'(a_i)$. That is, $D_i f(a)$ is the slope of the tangent line at $(a, f(a))$ to the curve obtained by intersecting the graph of f with the plane $x_j = a_j$, $j \neq i$. It also means that computation of $D_i f(a)$ is a problem we can already solve.

Example 8.1 (a) $f(x, y) = \sin(xy^2)$. Then $D_1 f(x, y) = y^2 \cos(xy^2)$ and $D_2 f(x, y) = 2xy\cos(xy^2)$.
(b) Consider the radius function $r \colon \mathbb{R}^n \to \mathbb{R}$

$$r(x) = \|x\|_2 = \sqrt{x_1^2 + \dots + x_n^2},$$

$x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Then r is partial differentiable on $\mathbb{R}^n \setminus 0$ with

$$\frac{\partial r}{\partial x_i}(x) = \frac{x_i}{r(x)}, \quad x \neq 0. \tag{8.5}$$

Indeed, the function

$$f(\xi) = \sqrt{x_1^2 + \cdots + \xi^2 + \cdots + x_n^2}$$

is differentiable, where $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ are considered to be constant. Using the chain rule one obtains (with $\xi = x_i$)

$$\frac{\partial r}{\partial x_i}(x) = f'(\xi) = \frac{1}{2} \frac{2\xi}{\sqrt{x_1^2 + \cdots + \xi^2 + \cdots + x_n^2}} = \frac{x_i}{r}.$$

(c) Let $f\colon (0, +\infty) \to \mathbb{R}$ be differentiable. The composition $x \mapsto f(r(x))$ (with the above radius function r) is denoted by $f(r)$, it is partial differentiable on $\mathbb{R}^n \setminus 0$. The chain rule gives

$$\frac{\partial}{\partial x_i} f(r) = f'(r) \frac{\partial r}{\partial x_i} = f'(r) \frac{x_i}{r}.$$

(d) Partial differentiability does not imply continuity. Define

$$f(x, y) = \begin{cases} \frac{xy}{(x^2+y^2)^2} = \frac{xy}{r^4}, & (x, y) \neq (0, 0), \\ 0, & (x, y) = (0, 0). \end{cases}$$

Obviously, f is partial differentiable on $\mathbb{R}^2 \setminus 0$. Indeed,

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{h \to 0} \frac{f(h, 0)}{h} = \lim_{h \to 0} 0 = 0.$$

Since f is symmetric in x and y, $\frac{\partial f}{\partial y}(0, 0) = 0$, too. However, f is not continuous at 0 since $f(\varepsilon, \varepsilon) = 1/(4\varepsilon^2)$ becomes large as ε tends to 0.

Remark 8.1 In the next section we will become acquainted with stronger notion of differentiability which implies continuity. In particular, a *continuously* partial differentiable function is continuous.

Definition 8.3 Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ partial differentiable. The vector

$$\operatorname{grad} f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) \tag{8.6}$$

is called the *gradient* of f at $x \in U$.

Example 8.2 (a) For the radius function $r(x)$ defined in Example 8.1 (b) we have

$$\operatorname{grad} r(x) = \frac{x}{r}.$$

Note that x/r is a unit vector (of the euclidean norm 1) in the direction x. With the notations of Example 8.1 (c),

$$\operatorname{grad} f(r) = f'(r) \frac{x}{r}.$$

(b) Let $f, g\colon U \to \mathbb{R}$ be partial differentiable functions. Then we have the following product rule

$$\operatorname{grad}(fg) = g \operatorname{grad} f + f \operatorname{grad} g. \tag{8.7}$$

This is immediate from the product rule for functions of one variable

$$\frac{\partial}{\partial x_i}(fg) = \frac{\partial f}{\partial x_i} g + f \frac{\partial g}{\partial x_i}.$$

(c) $f(x, y) = x^y$. Then $\operatorname{grad} f(x, y) = ((y-1)x^{y-1}, x^y \log x)$.

Notation. Instead of $\operatorname{grad} f$ one also writes ∇f ("Nabla f"). ∇ is a vector-valued differential operator:

$$\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right).$$

Definition 8.4 Let $U \subset \mathbb{R}^n$. A *vector field* on U is a mapping

$$v = (v_1, \dots, v_n)\colon U \to \mathbb{R}^n. \tag{8.8}$$

To every point $x \in U$ there is associated a vector $v(x) \in \mathbb{R}^n$.
If the vector field v is partial differentiable (i. e. all components v_i are partial differentiable) then

$$\operatorname{div} v = \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i} \tag{8.9}$$

is called the *divergence* of the vector field v.

Formally the divergence of v can be written as a scalar product of ∇ and v

$$\operatorname{div} v = \langle \nabla, v \rangle = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} v_i.$$

The product rule gives the following rule for the divergence. Let $f\colon U \to \mathbb{R}$ a partial differentiable function and

$$v = (v_1, \dots, v_n)\colon U \to \mathbb{R}$$

a partial differentiable vector field, then

$$\frac{\partial}{\partial x_i}(f v_i) = \frac{\partial f}{\partial x_i} \cdot v_i + f \cdot \frac{\partial v_i}{\partial x_i}.$$

Summation over i gives

$$\operatorname{div}(fv) = \langle \operatorname{grad} f, v \rangle + f \operatorname{div} v. \tag{8.10}$$

Using the nabla operator this can be rewritten as

$$\langle \nabla, fv \rangle = \langle \nabla f, v \rangle + f \langle \nabla, v \rangle .$$

Example 8.3 Let $F\colon \mathbb{R}^n \setminus 0 \to \mathbb{R}^n$ be the vector field $F(x) = \frac{x}{r}$, $r = \|x\|$. Since

$$\operatorname{div} x = \sum_{i=1}^{n} \frac{\partial x_i}{\partial x_i} = n \quad \text{and} \quad \langle x, x\rangle = r^2,$$

Example 8.2 gives

$$\operatorname{div} \frac{x}{r} = \left\langle \operatorname{grad} \frac{1}{r}, x \right\rangle + \frac{1}{r} \operatorname{div} x = \left\langle -\frac{x}{r^3}, x \right\rangle + \frac{n}{r} = \frac{n-1}{r}.$$

8.1.1 Higher Derivatives

Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ a partial differentiable function. If all partial derivatives $D_i f\colon U \to \mathbb{R}$ are again partial differentiable, f is called *twice partial differentiable*. We can form the partial derivatives $D_j D_i f$ of the second order.
More general, $f\colon U \to \mathbb{R}$ is said to be $(k+1)$-*times partial differentiable* if it is k-times partial differentiable and all partial derivatives of order k

$$D_{i_k} D_{i_{k-1}} \cdots D_{i_1} f\colon U \to \mathbb{R}$$

are partial differentiable.
A function $f\colon U \to \mathbb{R}$ is said to be *k-times continuously partial differentiable* if it is k-times partial differentiable and all partial derivatives of order less than or equal to k are continuous. The set of all such functions on U is denoted by $\mathrm{C}^k(U)$.
We also use the notation

$$D_j D_i f = \frac{\partial^2 f}{\partial x_j \partial x_i} = f_{x_i x_j}, \;\; D_i D_i f = \frac{\partial^2 f}{\partial x_i^2}, \;\; D_{i_k} \cdots D_{i_1} f = \frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}.$$

Example. Let $f(x, y) = \sin(xy^2)$. One easily sees that

$$f_{yx} = f_{xy} = 2y \cos(xy^2) - y^2 \sin(xy^2) 2xy.$$

Proposition 8.2 (Schwarz's Lemma) *Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ be twice continuously partial differentiable.*
Then for every $a \in U$ and all $i, j = 1, \dots, n$ we have

$$D_j D_i f(a) = D_i D_j f(a). \tag{8.11}$$

Proof. Without loss of generality we assume $n = 2$, $i = 1$, $j = 2$, and $a = 0$; we write (x, y) in place of (x_1, x_2). Since U is open, there is a small square of length $2\delta > 0$ completely contained in U:

$$\{(x, y) \in \mathbb{R}^2 \mid |\,x\,| < \delta, |\,y\,| < \delta\} \subset U.$$

For fixed $y \in U_\delta(0)$ define the function $F\colon (-\delta, \delta) \to \mathbb{R}$ via

$$F(x) = f(x, y) - f(x, 0).$$

By the mean value theorem (Theorem 4.9) there is a ξ with $|\,\xi\,| \le |\,x\,|$ such that

$$F(x) - F(0) = xF'(\xi).$$

But $F'(\xi) = D_1 f(\xi, y) - D_1 f(\xi, 0)$. Applying the mean value theorem to the function $y \mapsto D_1(\xi, y)$, there is an η with $|\,\eta\,| \le |\,y\,|$ and

$$D_1 f(\xi, y) - D_1 f(\xi, 0) = D_2 D_1(\xi, \eta) y.$$

Altogether we have

$$f(x,y) - f(x,0) - f(0,y) + f(0,0) = D_2 D_1 f(\xi, \eta) xy. \tag{8.12}$$

The same arguments but starting with the function $G(y) = f(x,y) - f(0,y)$ show the existence of ξ' and η' with $|\,\xi'\,| \le |\,x\,|$, $|\,\eta'\,| \le |\,y\,|$ and

$$f(x,y) - f(x,0) - f(0,y) + f(0,0) = D_1 D_2 f(\xi', \eta') xy. \tag{8.13}$$

From (8.12) and (8.13) for $xy \neq 0$ it follows that

$$D_2 D_1 f(\xi, \eta) = D_1 D_2(\xi', \eta').$$

If (x, y) approaches $(0,0)$ so do (ξ, η) and (ξ', η'). Since $D_2 D_1 f$ and $D_1 D_2 f$ are both continuous it follows from the above equation

$$D_2 D_1 f(0,0) = D_1 D_2 f(0,0).$$

■

Corollary 8.3 *Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}^n$ be k-times continuously partial differentiable. Then*

$$D_{i_k} \cdots D_{i_1} f = D_{i_{\pi(k)}} \cdots D_{i_{\pi(1)}} f$$

for every permutation π of $1, \ldots, k$.

Proof. The proof is by induction on k using the fact that any permutation can be written as a product of transpositions $(j \leftrightarrow j+1)$. ■

Example 8.4 Let $U \subset \mathbb{R}^3$ be open and let $v\colon U \to \mathbb{R}^3$ be a partial differentiable vector field. One defines a new vector field $\operatorname{curl} v\colon U \to \mathbb{R}^3$, the *curl* of v by

$$\operatorname{curl} v = \left(\frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3}, \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1}, \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \right). \tag{8.14}$$

Formally one can think of $\operatorname{curl} v$ as being the vector product of ∇ and v

$$\operatorname{curl} v = \nabla \times v = \begin{vmatrix} \mathrm{e}_1 & \mathrm{e}_2 & \mathrm{e}_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ v_1 & v_2 & v_3 \end{vmatrix},$$

where e_1, e_2, and e_3 are the unit vectors in $\mathbb{R}^3$. By Proposition 8.2, if $f\colon U \to \mathbb{R}$ has continuous second partial derivatives then

$$\operatorname{curl}\operatorname{grad} f = 0. \tag{8.15}$$

Indeed, the first coordinate of $\operatorname{curl}\operatorname{grad} f$ is by definition

$$\frac{\partial^2 f}{\partial x_2 \partial x_3} - \frac{\partial^2 f}{\partial x_3 \partial x_2} = 0.$$

The other two components are obtained by cyclic permutation of the indices.
We have found: $\operatorname{curl} v = 0$ is a necessary condition for a continuously partial differentiable vector field $v\colon U \to \mathbb{R}^3$ to be the gradient of a function $f\colon U \to \mathbb{R}$.

8.1.2 The Laplacian

Let $U \subset \mathbb{R}^n$ be open and $f\colon U \to \mathbb{R}$ has continuous partial second derivatives. Put

$$\Delta f = \operatorname{div}\operatorname{grad} f = \frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2}, \tag{8.16}$$

and call

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2}$$

the *Laplacian* or *Laplace operator*. The Laplacian plays a fundamental role in mathematical physics. The equation $\Delta f = 0$ is called the *potential equation*; its solution are the *harmonic* functions.
If f depends on an additional time variable t, $f\colon U \times I \to \mathbb{R}$, $(x,t) \mapsto f(x,t)$ one considers the so called *wave equation*

$$\Delta f - \frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} = 0, \tag{8.17}$$

and the so called *heat equation*

$$\Delta f - \frac{1}{k}\frac{\partial f}{\partial t} = 0. \tag{8.18}$$

Example 8.5 (a) Let $f\colon (0,+\infty) \to \mathbb{R}$ be twice continuously differentiable. We want to compute the Laplacian $\Delta f(r)$, $r = \|x\|$, $x \in \mathbb{R}^n \setminus 0$. By Example 8.2 we have

$$\operatorname{grad} f(r) = f'(r)\frac{x}{r},$$

and by the product rule and Example 8.3 we obtain

$$\begin{aligned}\Delta f(r) = \operatorname{div}\operatorname{grad} f(r) &= \left\langle \operatorname{grad} f'(r), \frac{x}{r} \right\rangle + f'(r)\operatorname{div}\frac{x}{r}\\ &= \left\langle f''(r)\frac{x}{r}, \frac{x}{r} \right\rangle + f'(r)\frac{n-1}{r};\end{aligned}$$

thus $\Delta f(r) = f''(r) + \frac{n-1}{r} f'(r)$. In particular,

$$\Delta \frac{1}{r^{n-2}} = 0,$$
$$\Delta \log r = 0, \quad \text{if} \quad n = 2.$$

(b) We show that $F\colon (\mathbb{R}^3 \setminus 0) \times \mathbb{R} \to \mathbb{R}$ given by

$$F(x,t) = \frac{1}{r} \cos(r - ct), \quad r = \|x\|$$

is a solution of the wave equation (8.17)—of course, there are a lot of other solutions. According to the above example we have

$$\Delta F = \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) \frac{\cos(r - ct)}{r}.$$

Since

$$\frac{\partial}{\partial r} \frac{\cos(r - ct)}{r} = -\frac{\sin(r - ct)}{r} - \frac{\cos(r - ct)}{r^2},$$
$$\frac{\partial^2}{\partial r^2} \frac{\cos(r - ct)}{r} = -\frac{\cos(r - ct)}{r} + 2\frac{\sin(r - ct)}{r^2} + 2\frac{\cos(r - ct)}{r^3},$$

we have

$$\Delta F = -\frac{\cos(r - ct)}{r}.$$

On the other hand,

$$\frac{\partial^2}{\partial t^2} \frac{\cos(r - ct)}{r} = -\frac{\cos(r - ct)}{r},$$

and the assertion follows.

8.2 Differentiation

In this section we define (total) differentiability of a function f from $\mathbb{R}^n$ to $\mathbb{R}^m$. Roughly speaking, f is differentiable (at some point) if it can be approximated by a linear mapping. In contrast to partial differentiability we need not to refer to single coordinates. Differentiable functions are continuous. In this section U always denotes an open subset of $\mathbb{R}^n$. The vector space of linear mappings f of a vector space V into a vector space W will be denoted by $\mathrm{L}(V, W)$.

Motivation: If $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable at $x \in \mathbb{R}$ and $f'(x) = a$, then

$$\lim_{h \to 0} \frac{f(x+h) - f(x) - ax}{h} = 0.$$

Definition 8.5 The mapping $f\colon U \to \mathbb{R}^m$ is said to be *differentiable* at a point $x \in U$ if there exist a linear map $A\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0. \tag{8.19}$$

The linear map $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^m)$ is called the *derivative* of f at x and will be denoted by $Df(x)$.

Remarks 8.2 (a) In case $n = m = 1$ this notion coincides with the ordinary differentiability of a function.
(b) If it exists, the linear map A is unique, see [9, 2-1 Theorem]; we refer to the euclidean norm. However switching to other norms does not change the linear map A.

(c) It is often convenient to consider the matrix of $Df(x)$ with respect to the usual bases of $\mathbb{R}^n$ and $\mathbb{R}^m$. This $m \times n$ matrix $A = (a_{ij})$, $i = 1, \dots, m$, $j = 1, \dots, n$ is called the *Jacobian matrix* of f at x, and denoted by $f'(x)$. Using column vectors, $h = (h_1, \dots, h_n)^\top$ the map Ah is then simply given by matrix multiplication

$$A\,h = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}.$$

By Proposition 6.12 the limit

$$\lim_{h\to 0} \frac{1}{\|h\|} \left(f(x+h) - f(x) - Ah\right).$$

exists and is equal to 0 if and only if the limit of every coordinate ($i = 1, \dots, m$) exists and is 0

$$\lim_{h\to 0} \frac{1}{\|h\|} \left(f_i(x+h) - f_i(x) - \sum_{j=1}^{n} a_{ij} h_j\right) = 0, \quad i = 1, \dots, m. \tag{8.20}$$

We see, f is differentiable at x if and only if all f_i, $i = 1, \dots, m$, are differentiable at x. In this case

$$f'(x) = \begin{pmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{pmatrix}.$$

(d) Define a function $\varphi_x \colon U \to \mathbb{R}^m$ (depending on both x and h) by

$$f(x+h) = f(x) + A\,h + \varphi_x(h).$$

Then f is differentiable if and only if

$$\lim_{h\to 0} \frac{\|\varphi_x(h)\|}{\|h\|} = 0,$$

cf. Proposition 4.2.

Example 8.6 Let us consider the function $f \colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = g(x)$ where g is differentiable on $\mathbb{R}$. Then $A = Df(a, b)$ satisfies $A(x, y) = g'(a)x$ (the corresponding 1×2-matrix is $f'(a, b) = (g'(a)\ 0)$). To prove this, note that

$$\frac{|\, f(a+h, b+k) - f(a, b) - A\,(h, k)\,|}{\|(h, k)\|} = \frac{|\, g(a+h) - g(a) - g'(a)h\,|}{\|(h, k)\|}.$$

Since $\|(h,k)\| \geq |\, h \,|$ we continue

$$\leq \left| \frac{g(a+h)-g(a)}{h} - g'(a) \right| .$$

But this term tends to 0 as $h \to 0$ by the definition of $g'(a)$. We conclude that

$$\lim_{(h,k)\to 0} \frac{|\, f(a+h,b+k) - f(a,b) - A\,(h,k)\,|}{\|(h,k)\|} = 0,$$

which proves the claim.
(b) Let $C = (c_{ij}) \in \mathrm{M}(n \times n, \mathbb{R})$ be a symmetric $n \times n$ matrix and define $f\colon \mathbb{R}^n \to \mathbb{R}$ by

$$f(x) = \langle x, Cx \rangle = \sum_{i,j=1}^{n} c_{ij} x_i x_j, \quad x = (x_1, \dots, x_n)^\top \in \mathbb{R}^n.$$

If $a, h \in \mathbb{R}^n$ we have

$$\begin{aligned} f(a+h) &= \langle a+h, C(a+h) \rangle = \langle a, Ca \rangle + \langle h, Ca \rangle + \langle a, Ch \rangle + \langle h, Ch \rangle \\ &= \langle a, Ca \rangle + 2\,\langle Ca, h \rangle + \langle h, Ch \rangle \\ &= f(a) + \langle v, h \rangle + \varphi(h), \end{aligned}$$

where $v = 2Ca$ and $\varphi(h) = \langle h, Ch \rangle$. Since, by the Cauchy–Schwarz inequality,

$$|\, \varphi(h) \,| \leq \|h\|\,\|Ch\| \leq \|h\|\,\|C\|\,\|h\| \leq \|C\|\,\|h\|^2 ,$$

$\lim_{h\to 0} \frac{\varphi(h)}{\|h\|} = 0$. This proves f to be differentiable at $a \in \mathbb{R}^n$ with derivative A, $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R})$, and $Ax = \langle 2Ca, x \rangle$. The Jacobian matrix is a row vector $f'(a) = (2Ca)^\top$.

Lemma 8.4 *Let $f\colon U \to \mathbb{R}^m$ differentiable at x, then f is continuous at x.*

Proof. Define $\varphi_x(h)$ as in Remarks 8.2 (d) with $Df(x) = A$, then

$$\lim_{h\to 0} \|\varphi_x(h)\| = 0$$

since f is differentiable at x. This gives

$$\lim_{h\to 0} f(x+h) = f(x) + \lim_{h\to 0} Ah + \lim_{h\to 0} \varphi_x(h) = f(x).$$

This shows continuity. ∎

8.2.1 Basic Theorems

Proposition 8.5 *Let $f\colon U \to \mathbb{R}^m$, $f(x) = (f_1(x), \dots, f_m(x))$ be differentiable at $a = (a_1, \dots, a_n) \in U$. Then all partial derivatives $\frac{\partial f_i(a)}{\partial x_j}$, $i = 1, \dots m$, $j = 1, \dots, n$ exist and the Jacobian matrix $f'(a)$ takes the form*

$$f'(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} (a) = \left(\frac{\partial f_i(a)}{\partial x_j} \right)_{\substack{i=1,\dots,m \\ j=1,\dots,n}} . \tag{8.21}$$

Proof. Let $A = (A_{ij}) = f'(a)$ be the Jacobian matrix of f at a. Inserting $h = te_j = (0, \dots, t, \dots, 0)$ into (8.20) (see Remark 8.2 (d)) we have since $\|h\| = |\,t\,|$ and $h_k = t\delta_{kj}$

$$\begin{aligned} 0 &= \lim_{t\to 0} \frac{\|f_i(a + te_j) - f_i(a) - \sum_{k=1}^n A_{ik}h_k\|}{\|te_j\|} \\ &= \lim_{t\to 0} \frac{|\,f_i(a_1, \dots, a_j + t, \dots, a_n) - f_i(a) - tA_{ij}\,|}{|\,t\,|} \\ &= \lim_{t\to 0} \left| \frac{f_i(a_1, \dots, a_j + t, \dots, a_n) - f_i(a)}{t} - A_{ij} \right| \\ &= \left| \frac{\partial f_i(a)}{\partial x_j} - A_{ij} \right|. \end{aligned}$$

Hence $A_{ij} = \dfrac{\partial f_i(a)}{\partial x_j}$. ∎

Example 8.7 (a) Special case $n = 1$; let $f\colon (a, b) \to \mathbb{R}^m$, $f(x) = (f_1(x), \dots, f_m(x))^\top$ be differentiable. Then $f'(x) = (f_1'(x), \dots, f_m'(x))^\top$ is the Jacobian matrix of f at x. It is a column vector. It is also the tangent vector to the curve f at x.
Special case $m = 1$; let $f\colon U \to \mathbb{R}$ be differentiable. Then

$$f'(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right) = \operatorname{grad} f(x).$$

It is a row vector and gives a linear functional on $\mathbb{R}^n$ which linearly associates to each vector $y = (y_1, \dots, y_n)^\top \in \mathbb{R}^n$ a real number

$$Df(x) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \langle \operatorname{grad} f(x), y \rangle = \sum_{j=1}^n f_{x_j}(x) y_j.$$

(b) Let $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ be given by

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

Then

$$f'(x, y, z) = \left(\frac{\partial(f_1, f_2)}{\partial(x, y, z)} \right) = \begin{pmatrix} 3x^2 - 3y^2 & -6xy & 1 \\ yz^2 \cos(xy^2 z) & xz^2 \cos(xy^2 z) & 2xyz \cos(xy^2 z) \end{pmatrix}.$$

Remark 8.3 Note that the existence of all partial derivatives does not imply the existence of $f'(a)$. Recall Example 8.1 (d). There was given a function having partial derivatives at the origin not being continuous at $(0, 0)$, and hence not being differentiable at $(0, 0)$. However, the next proposition shows that the converse is true provided all partial derivatives are continuous.

We say that f is *continuously differentiable* at x if f is differentiable in a neighborhood of x and the assignment $x \mapsto f'(x)$ is continuous.

Proposition 8.6 *Let $f: U \to \mathbb{R}^m$ be continuously partial differentiable at a point $a \in U$. Then f is continuously differentiable at a.*

The proof in case $n = 2$, $m = 1$ is in the appendix to this chapter.

Theorem 8.7 (Chain Rule) *If $f: \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at a point a and $g: \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(a)$, then the composition $g \circ f: \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at a and*

$$D(g \circ f)(a) = Dg(f(a)) \circ Df(a). \tag{8.22}$$

In matrix notation this equation can be written as

$$(g \circ f)'(a) = g'(f(a)) \cdot f'(a). \tag{8.23}$$

Proof. Let $b = f(a)$, $A = Df(a)$, and $B = Dg(f(a))$. If we define functions φ, ψ, and ρ by

$$\varphi(x) = f(x) - f(a) - A(x - a), \tag{8.24}$$
$$\psi(y) = g(y) - g(b) - B(y - b), \tag{8.25}$$
$$\rho(x) = g \circ f(x) - g \circ f(a) - B \circ A(x - a) \tag{8.26}$$

then

$$\lim_{x \to a} \frac{\|\varphi(x)\|}{\|x - a\|} = 0, \quad \lim_{y \to b} \frac{\|\psi(y)\|}{\|y - b\|} = 0 \tag{8.27}$$

and we have to show that

$$\lim_{x \to a} \frac{\|\rho(x)\|}{\|x - a\|} = 0.$$

Inserting (8.24) and (8.25) we find

$$\rho(x) = g(f(x)) - g(f(a)) - BA\,(x - a) = g(f(x)) - g(f(a)) - B(f(x) - f(a) - \varphi(x))$$
$$\rho(x) = [g(f(x)) - g(f(a)) - B(f(x) - f(a))] + B \circ \varphi(x)$$
$$\rho(x) = \psi(f(x)) + B(\varphi(x)).$$

Using $\|Tx\| \le \|T\|\,\|x\|$ (see Proposition 8.1) this shows

$$\frac{\|\rho(x)\|}{\|x - a\|} \le \frac{\|\psi(f(x))\|}{\|x - a\|} + \frac{\|B \circ \varphi(x)\|}{\|x - a\|} \le \frac{\|\psi(y)\|}{\|y - b\|} \cdot \frac{\|f(x) - f(a)\|}{\|x - a\|} + \|B\| \frac{\|\varphi(x)\|}{\|x - a\|}.$$

Inserting (8.24) again into the above equation we continue

$$= \frac{\|\psi(y)\|}{\|y - b\|} \cdot \frac{\|\varphi(x) + A(x - a)\|}{\|a - x\|} + \|B\| \frac{\|\varphi(x)\|}{\|x - a\|}$$
$$\le \frac{\|\psi(y)\|}{\|y - b\|} \left(\frac{\|\varphi(x)\|}{\|a - x\|} + \|A\| \right) + \|B\| \frac{\|\varphi(x)\|}{\|x - a\|}.$$

All terms on the right side tend to 0 as x approaches a. This completes the proof. ∎

Corollary 8.8 *If $k = g \circ f$, $A = f'(a)$, $B = g'(f(a))$, and $C = k'(a)$, then*

$$\left(\frac{\partial(k_1, \dots, k_p)}{\partial(x_1, \dots, x_n)}\right) = \left(\frac{\partial(g_1, \dots, g_p)}{\partial(y_1, \dots, y_m)}\right) \left(\frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)}\right) \tag{8.28}$$

$$\frac{\partial k_r}{\partial x_j}(a) = \sum_{i=1}^{m} \frac{\partial g_r}{\partial y_i}(f(a)) \frac{\partial f_i}{\partial x_j}(a), \quad r = 1, \dots, p,\ j = 1, \dots, n. \tag{8.29}$$

Proof. This is immediate by inserting the Jacobian matrix $f'(a) = \left(\frac{\partial f_i(a)}{\partial x_j}\right)$ (8.21) into (8.23) . ∎

Example 8.8 (a) Let $f(u, v) = uv$, $u = g(x, y) = x^2 + y^2$, $v = h(x, y) = xy$, and $z = f(g(x, y), h(x, y)) = (x^2 + y^2)xy = x^3y + x^2y^3$.

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial u} \cdot \frac{\partial g}{\partial x} + \frac{\partial f}{\partial v} \cdot \frac{\partial u}{\partial x} = v \cdot 2x + u \cdot y = 2x^2y + y(x^2 + y^2)$$
$$\frac{\partial z}{\partial x} = 3x^2y + y^3.$$

(b) Let $f(u, v) = u^v$, $u(t) = v(t) = t$. Then $F(t) = f(u(t), v(t)) = t^t$ and

$$F'(t) = \frac{\partial f}{\partial u} u'(t) + \frac{\partial f}{\partial v} v'(t) = vu^{v-1} \cdot 1 + u^v \log u \cdot 1$$
$$= t \cdot t^{t-1} + \log t t^t = t^t(\log t + 1).$$

8.3 Taylor's Formula

The Jacobian matrix gives an approximation of a function f by a linear function. Taylor's formula generalizes this concept of approximation to higher order. We consider quadratic approximation of f to determine local extrema.

8.3.1 Directional Derivatives

Definition 8.6 Let $f : U \to \mathbb{R}$ be a function, $a \in U$, and $x \in \mathbb{R}^n$ a unit vector, $\|x\| = 1$. The *directional derivative* of f at a in the direction of the unit vector x is the limit

$$(D_x f)(a) = \lim_{t \to 0} \frac{f(a + tx) - f(a)}{t}. \tag{8.30}$$

Note that for $x = \mathrm{e}_j$ we have $D_x f = D_j f = \frac{\partial f}{\partial x_j}$.

Proposition 8.9 *Let $f : U \to \mathbb{R}$ be continuously differentiable. The for every $a \in U$ and every unit vector $x \in \mathbb{R}^n$, $\|x\| = 1$, we have*

$$D_x f(a) = \langle x, \operatorname{grad} f(a) \rangle \tag{8.31}$$

Proof. Define $g\colon \mathbb{R} \to \mathbb{R}^n$ by

$$g(t) = a + tx = (a_1 + tx_1, \dots, a_n + tx_n).$$

For sufficiently small $t \in \mathbb{R}$, say $|\,t\,| \le \varepsilon$, the function

$$h\colon \mathbb{R} \xrightarrow{g} \mathbb{R}^n \xrightarrow{f} \mathbb{R},$$
$$h(t) = f(g(t)) = f(a_1 + tx_1, \dots, a_n + tx_n)$$

is defined. We compute $h'(t)$ using the chain rule.

$$h'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(g(t))\, g_j'(t).$$

Since $g_j'(t) = \frac{\mathrm{d}}{\mathrm{d}t}(a_j + tx_j) = x_j$ and $g(0) = a$, it follows

$$h'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a + tx)\, x_j, \tag{8.32}$$

$$h'(0) = \sum_{j=1}^{n} f_{x_j}(a) x_j = \langle x,\, \operatorname{grad} f(a)\rangle .$$

On the other hand, by definition of the directional derivative

$$h'(0) = \lim_{t\to 0} \frac{h(t) - h(0)}{t} = \lim_{t\to 0} \frac{f(a + tx) - f(a)}{t} = D_x f(a).$$

This completes the proof. ∎

Corollary 8.10 *Let $f\colon U \to \mathbb{R}$ be k-times continuously differentiable, $a \in U$ and $x \in \mathbb{R}^n$ such that the whole segment $a + tx$, $t \in [0, 1]$ is contained in U.*
Then the function $h\colon [0, 1] \to \mathbb{R}$, $h(t) = f(a + tx)$ is k-times continuously differentiable, where

$$h^{(k)}(t) = \sum_{i_1,\dots,i_k=1}^{n} D_{i_k} \cdots D_{i_1} f(a + tx) x_{i_1} \cdots x_{i_k}. \tag{8.33}$$

In particular

$$h^{(k)}(0) = \sum_{i_1,\dots,i_k=1}^{n} D_{i_k} \cdots D_{i_1} f(a) x_{i_1} \cdots x_{i_k}. \tag{8.34}$$

Proof. The proof is by induction on k. For $k = 1$ it is exactly the statement of the Proposition. We demonstrate the step from $k = 1$ to $k = 2$. By (8.32)

$$h''(t) = \sum_{i_1=1}^{n} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f(a + tx)}{\partial x_{i_1}}\right) x_{i_1}$$
$$= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \frac{\partial f}{\partial x_{i_2}\partial x_{i_1}}(a + tx) x_{i_2} x_{i_1}.$$

In the second line we applied the chain rule to $\tilde{h}(t) = f_{x_{i_1}}(a + tx)$. ■

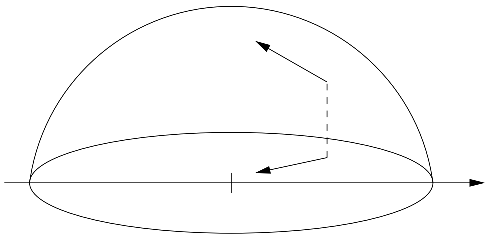
For brevity we use the following notation for the term on the right of (8.34):

$$h^{(k)}(0) = (x\,\nabla)^k f(a) = \sum_{i_1,\dots,i_k=1}^{n} D_{i_k} \cdots D_{i_1} f(a) x_{i_1} \cdots x_{i_k}.$$

Remark 8.4 (Geometric meaning of $\operatorname{grad} f$**)** If $\operatorname{grad} f(x) \neq 0$, the angle θ between v and $\operatorname{grad} f(x)$ is defined and

$$D_v f(x) = \langle v, \operatorname{grad} f(x)\rangle = \|\operatorname{grad} f(x)\| \cos\theta.$$

It follows that $D_v f(x)$ becomes maximal if v and $\operatorname{grad} f(x)$ have the same direction ($\theta = 0$). Hence the vector $\operatorname{grad} f(x)$ points in the direction of maximal slope of f.



For example $f(x, y) = \sqrt{1 - x^2 - y^2}$ has $\operatorname{grad} f(x, y) = \left(\frac{-x}{\sqrt{1-x^2-y^2}}, \frac{-y}{\sqrt{1-x^2-y^2}}\right)$. This corresponds to the picture that the maximal slope on the upper unit half-sphere is in direction to the north pole $(0, 0, 1)$.

8.3.2 Taylor's Formula

Theorem 8.11 *Let $f: U \to \mathbb{R}$ be $(k+1)$-times continuously differentiable, $a \in U$, $x \in \mathbb{R}^n$ such that $a + tx \in U$ for all $t \in [0, 1]$. Then there exists $\theta \in [0, 1]$ such that*

$$f(a + x) = \sum_{m=0}^{k} \frac{1}{m!}(x\,\nabla)^m f(a) + \frac{1}{(k+1)!}(x\,\nabla)^{k+1} f(a + \theta x) \tag{8.35}$$

$$\begin{aligned} f(a + x) &= f(a) + \sum_{i=1}^{n} x_i f_{x_i}(a) + \frac{1}{2!}\sum_{i,j=1}^{n} x_i x_j f_{x_i x_j}(a) + \cdots + \\ &\quad + \frac{1}{(k+1)!} \sum_{i_1,\dots i_{k+1}} x_{i_1} \cdots x_{i_{k+1}} f_{x_{i_1} \cdots x_{i_{k+1}}}(a + \theta x). \end{aligned}$$

Proof. Consider the function $h: [0, 1] \to \mathbb{R}$, $h(t) = f(a + tx)$. By Corollary 8.10, h is a $(k+1)$-times continuously differentiable. By Taylor's theorem for functions in one variable (Theorem 4.15 with $x = 1$ and $a = 0$ therein) Thm 4.15, we have

$$g(1) = \sum_{m=0}^{k} \frac{g^{(m)}(0)}{m!} + \frac{g^{(k+1)}(\theta)}{(k+1)!}.$$

By Corollary 8.10 for $m = 1, \dots, k$ we have

$$\frac{h^{(m)}(0)}{m!} = \frac{1}{m!}(x\,\nabla)^m f(a).$$

and

$$\frac{g^{(k+1)}(\theta)}{(k+1)!} = \frac{1}{(k+1)!}(x\,\nabla)^{k+1} f(a+\theta x);$$

the assertion follows. ■

It is often convenient to substitute $x := x + a$. Then the Taylor expansion reads

$$f(x) = \sum_{m=0}^{k} \frac{1}{m!}((x-a)\,\nabla)^m f(a) + \frac{1}{(k+1)!}((x-a)\,\nabla)^{(k+1)} f(a+\theta(x-a))$$

$$\begin{aligned} f(x) = f(a) &+ \sum_{i=1}^{n}(x_i - a_i) f_{x_i}(a) + \frac{1}{2!}\sum_{i,j=1}^{n}(x_i-a_i)(x_j-a_j) f_{x_i x_j}(a) + \cdots + \\ &+ \frac{1}{(k+1)!}\sum_{i_1,\dots i_{k+1}} (x_{i_1} - a_{i_1})\cdots(x_{i_{k+1}} - a_{i_{k+1}}) f_{x_{i_1}\cdots x_{i_{k+1}}}(a+\theta(x-a)). \end{aligned}$$

Let us abbreviate the remainder term by

$$R_{k+1}(a,x) = \frac{1}{(k+1)!}\sum_{i_1,\dots i_{k+1}} x_{i_1}\cdots x_{i_{k+1}} f_{x_{i_1}\cdots x_{i_{k+1}}}(a+\theta x).$$

We write the Taylor formula for the case $n = 2$, $k = 3$:

$$\begin{aligned} f(a+x, b+y) &= f(a,b) + (f_x(a,b)x + f_y(a,b)y) + \\ &+ \frac{1}{2!}\left(f_{xx}(a,b)x^2 + 2f_{xy}(a,b)\,xy + f_{yy}f(a,b)y^2\right) + \\ &+ \frac{1}{3!}\left(f_{xxx}(a,b)x^3 + 3f_{xxy}(a,b)x^2y + 3f_{xyy}(a,b)xy^2 + f_{yyy}f(a,b)y^3)\right) + R_4(a,x). \end{aligned}$$

Corollary 8.12 (Mean Value Theorem) *Let $f\colon U \to \mathbb{R}$ be continuously differentiable, $a \in U$, $x \in \mathbb{R}^n$ such that $a + tx \in U$ for all $t \in [0,1]$. Then there exists $\theta \in [0,1]$ such that*

$$f(a+x) - f(a) = \sum_{j=1}^{n} D_j f(a+\theta x)\, x_j = \langle\, \mathrm{grad}\, f(a+\theta x), x\rangle\,. \tag{8.36}$$

Proof. This is the special case $k = 0$. ■

Corollary 8.13 *Let $f\colon U \to \mathbb{R}$ be k times continuously differentiable, $a \in U$, $x \in \mathbb{R}^n$ such that $a + tx \in U$ for all $t \in [0,1]$. Then there exists $\varphi\colon U \to \mathbb{R}$ such that*

$$f(a+x) = \sum_{m=0}^{k} \frac{1}{m!}(x\,\nabla)^m f(a) + \varphi(x), \tag{8.37}$$

where

$$\lim_{x\to 0} \frac{\varphi(x)}{\|x\|^k} = 0.$$

In case $k = 1$ this is exactly the characterization of differentiability from Remark 8.2 (d).

Remarks 8.5 (a) With the above notations let

$$P_m(x) = \frac{(x\,\nabla)}{m!} f(a).$$

Then P_m is a homogeneous polynomial of degree m in the set of variables $x = (x_1, \dots, x_n)$ and we have

$$f(a + x) = \sum_{m=0}^{k} P_m(x) + \varphi(x),$$

where

$$\lim_{x \to 0} \frac{\|\varphi(x)\|}{\|x\|^k} = 0.$$

Let us consider in more detail the cases $m = 0, 1, 2$.
Case $m = 0$.

$$P_0(x) = \frac{D^0 f(a)}{0!} x^0 = f(a).$$

P_0 is the constant polynomial with value $f(a)$.
Case $m = 1$. We have

$$P_1(x) = \sum_{j=1}^{n} D_j f(a) x_j = \langle\, \operatorname{grad} f(a), x \rangle$$

Using Corollary 8.13 the first order approximation of a continuously differentiable function is

$$f(a + x) = f(a) + \langle\, \operatorname{grad} f(a), x \rangle + \varphi(x), \quad \lim_{x \to 0} \frac{\varphi(x)}{\|x\|} = 0. \tag{8.38}$$

The linearization of f at a is

$$\begin{aligned} \ell(x) &= f(a) + \langle\, \operatorname{grad} f(a), x - a \rangle\,, \\ x_{n+1} &= f(a) + \sum_{i=1}^{n} f_{x_i}(a)(x_i - a_i). \end{aligned}$$

The graph of this linear function $x_{n+1} = \ell(x)$ is the hyperplane through $(a, f(a))$ which touches the graph of f.
Case $m = 2$.

$$P_2(x) = \frac{1}{2} \sum_{i,j=1}^{n} D_i D_j f(a) x_i x_j.$$

Hence, $P_2(x)$ is a quadratic form with the corresponding matrix $\left(\frac{1}{2} D_i D_j(a)\right)$.

Definition 8.7 Let $f \colon U \to \mathbb{R}$ be twice continuously differentiable. The *Hessian matrix* of f at $a \in U$ is the $n \times n$-matrix

$$(\operatorname{Hess} f)(a) = \left(D_i D_j f(a)\right)_{i,j=1}^{n}. \tag{8.39}$$

The Hessian matrix is symmetric.

As a special case of Corollary 8.13 ($m = 2$) we have for $f \in C^2(U)$

$$f(a+x) = f(a) + \langle \operatorname{grad} f(a), x\rangle + \frac{1}{2}\langle x, \operatorname{Hess} f(a)x\rangle + \varphi(x), \tag{8.40}$$

where

$$\lim_{x\to 0} \frac{\varphi(x)}{\|x\|^2} = 0.$$

Example 8.9 (a) We compute the Taylor expansion of $f(x,y) = \cos x \sin y$ at $(0,0)$ to the third order. We have

$$\begin{aligned}
f_x &= -\sin x \sin y, & f_y &= \cos x \cos y, & & \\
f_x(0,0) &= 0, & f_y(0,0) &= 1, & & \\
f_{xx} &= -\cos x \sin y, & f_{yy} &= -\cos x \sin y, & f_{xy} &= -\sin x \cos y \\
f_{xx}(0,0) &= 0, & f_{yy}(0,0) &= 0, & f_{xy}(0,0) &= 0. \\
f_{xxy} &= -\cos x \cos y, & f_{yyy} &= -\cos x \cos y, & & \\
f_{xxy}(0,0) &= -1, & f_{yyy}(0,0) &= -1, & f_{xyy}(0,0) &= f_{xxx}(0,0) = 0.
\end{aligned}$$

Inserting this gives

$$f(x,y) = y + \frac{1}{3!}\left(-3x^2y - y^3\right) + R_4(x,y;0).$$

The same result can be obtained by multiplying the Taylor series for $\cos x$ and $\sin y$:

$$\left(1 - \frac{x^2}{2} + \frac{x^4}{4!} \mp \cdots\right)\left(y - \frac{y^3}{3!} \pm \cdots\right) = y - \frac{1}{2}x^2y - \frac{y^3}{6} + \cdots$$

(b) The Taylor series of $f(x,y) = e^{xy^2}$ at $(0,0)$ is

$$e^{xy^2} = \sum_{n=0}^{\infty} \frac{(xy^2)^n}{n!} = 1 + xy^2 + \frac{1}{2}x^2y^4 + \cdots;$$

it converges all over $\mathbb{R}^2$ to f.

8.4 Extrema of Functions of Several Variables

Definition 8.8 Let $f\colon U \to \mathbb{R}$ be a function. The point $x \in U$ is called *local maximum (minimum)* of f if there exists a neighborhood $V \subset U$ of x such that

$$f(x) \ge f(y) \quad (f(x) \le f(y)) \quad \text{for all } y \in V.$$

A *local extremum* is either a local maximum or a local minimum.

Proposition 8.14 *Let $f\colon U \to \mathbb{R}$ be partial differentiable. If f has a local extremum at $x \in U$ then* $\operatorname{grad} f(x) = 0$.

Proof. For $i = 1, \dots, n$ consider the function

$$g_i(t) = f(x + te_i).$$

This is a differentiable function of one variable, defined on a certain interval $(-\varepsilon, \varepsilon)$. If f has an extremum at x, then g_i has an extremum at $t = 0$. By Proposition 4.7

$$g_i'(0) = 0.$$

Since $g_i'(0) = \lim_{t\to 0} \frac{f(x+te_i)-f(x)}{t} = f_{x_i}(x)$ and i was arbitrary, it follows that

$$\operatorname{grad} f(x) = (D_i f(x), \dots, D_n f(x)) = 0.$$

■

Example 8.10 Let $f(x, y) = \sqrt{x^2 + y^2}$ and $U = \{(x, y) \mid x^2 + y^2 < 1\}$. Then $\operatorname{grad} f(x, y) = (-x/r, -y/r) = 0$ if $x = y = 0$. If f has an extremum in U then at the origin.

To obtain a sufficient criterion for the existence of local extrema we have to consider the Hessian matrix. Before, we need some facts from Linear Algebra.

Definition 8.9 Let A be a real, symmetric $n \times n$-matrix ($a_{ij} = a_{ji}$). The corresponding quadratic form

$$Q(x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = \langle x, Ax \rangle$$

is called

positive definite	if $Q(x) > 0$ for all $x \neq 0$,
negative definite	if $Q(x) < 0$ for all $x \neq 0$,
indefinite	if $Q(x) > 0$, $Q(y) < 0$ for some x, y,
positive semidefinite	if $Q(x) \geq 0$ for all x,
negative semidefinite	if $Q(x) \leq 0$ for all x.

Also, we say that the corresponding matrix A has the above properties.

Example 8.11 Let $n = 2$, $Q(x) = Q(x_1, x_2)$. Then $Q_1(x) = 3x_1^2 + 7x_2^2$ is positive definite, $Q_2(x) = -x_1^2 - 2x_2^2$ is negative definite, $Q_3(x) = x_1^2 - 2x_2^2$ is indefinite, $Q_4(x) = x_1^2$ is positive semidefinite, and $Q_5(x) = -x_2^2$ is negative semidefinite.

Proposition 8.15 (Sylvester) *Let A be a real symmetric $n \times n$-matrix and $Q(x) = \langle x, Ax \rangle$ the corresponding quadratic form. For $k = 1, \cdots, n$ let*

$$A_k = \begin{pmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \cdots & a_{kk} \end{pmatrix}, \quad D_k = \det A_k.$$

Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of A. Then

(a) Q *is positive definite if and only if* $\lambda_1 > 0,\ \lambda_2 > 0, \dots, \lambda_n > 0$. *This is the case if and only if* $D_1 > 0,\ D_2 > 0,\ \dots,\ D_n > 0$.
(b) $Q(x)$ *is negative definite if and only if* $\lambda_1 < 0,\ \lambda_2 < 0,\ \dots, \lambda_n < 0$. *This is the case if and only if* $(-1)^k D_k > 0$ *for all* $k = 1, \dots, n$.
(c) $Q(x)$ *is indefinite if and only if,* A *has both positive and negative eigenvalues.*

Proposition 8.16 *Let* $f\colon U \to \mathbb{R}$ *be twice continuously differentiable and let* $\operatorname{grad} f(a) = 0$ *at some point* $a \in U$.

(a) *If* $\operatorname{Hess} f(a)$ *is positive definite, then* f *has a local minimum at* a.
(b) *If* $\operatorname{Hess} f(a)$ *is negative definite, then* f *has a local maximum at* a.
(c) *If* $\operatorname{Hess} f(a)$ *is indefinite, then* f *has not a local extremum at* a.

Proof. By (8.40) and since $\operatorname{grad} f(a) = 0$,

$$f(a+x) = f(a) + \frac{1}{2}\langle x, Ax\rangle + \varphi(x), \tag{8.41}$$

where $A = \operatorname{Hess} f(a)$

$$\lim_{x\to 0}\frac{\varphi(x)}{\|x\|^2} = 0,$$

see Corollary 8.13. (a) Let A be positive definite. Since the unit sphere $\mathrm{S} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ is compact and the map $x \mapsto \langle x, Ax\rangle$ is continuous, the function attains its minimum, say α, on S,

$$\alpha = \min\{\langle x, Ax\rangle \mid x \in \mathrm{S}\} > 0.$$

We show that

$$\langle x, Ax\rangle \geq \alpha \|x\|^2, \quad \text{for all } x \in U. \tag{8.42}$$

This is obvious in case $x = 0$. If x is nonzero, $y = x/\|x\| \in \mathrm{S}$ and therefore

$$\alpha \leq \langle y, Ay\rangle = \left\langle \frac{x}{\|x\|}, \frac{Ax}{\|x\|}\right\rangle = \frac{1}{\|x\|^2}\langle x, Ax\rangle,$$

and (8.42) follows.
Now choose $\delta > 0$ such that

$$|\,\varphi(x)\,| \leq \frac{\alpha}{4}\,\|x\|^2, \quad \text{if } \|x\| < \delta.$$

From (8.41) and (8.42) it follows

$$f(a+x) \geq f(a) + \frac{\alpha}{4}\|x\|^2,$$

hence

$$f(a+x) > f(a), \quad \text{if} \quad 0 < \|x\| < \delta,$$

and f has a local minimum at a.
(b) If $A = \operatorname{Hess} f(a)$ is negative definite, consider $-f$ in place of f and apply (a).
(c) Let $A = \operatorname{Hess} f(a)$ indefinite. We have to show that in every neighborhood of a there exist x' and x'' such that $f(x'') < f(a) < f(x')$. Since A is indefinite, there is a vector $x \in \mathbb{R}^n \setminus 0$ such that $\langle x, Ax \rangle = \alpha > 0$. Then for small t we have

$$f(a + tx) = f(a) + \frac{1}{2} \langle tx, Atx \rangle + \varphi(tx) = f(a) + \frac{\alpha}{2} t^2 + \varphi(tx).$$

If t is small enough, $|\,\varphi(tx)\,| \leq \frac{\alpha}{4} t^2$, hence

$$f(a + tx) > f(a), \quad \text{if} \quad 0 < |\,t\,| < \delta.$$

Similarly, if $y \in \mathbb{R}^n \setminus 0$ satisfies $\langle y, Ay \rangle < 0$, for sufficiently small t we have $f(a+ty) < f(a)$. ■

Example 8.12 $z = f(x, y) = 4x^2 - y^2$, $U = \{(x, y) \mid x^2 + y^2 < 1\}$ (a hyperbolic paraboloid). We find

$$D_1 f = 8x = 0, \quad D_2 f = -2y = 0 \quad \text{implies} \quad x = y = 0.$$

Further,

$$\begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = \begin{vmatrix} 8 & 0 \\ 0 & -2 \end{vmatrix} = -16 < 0.$$

The form is indefinite; the function has not an extremum at the origin $(0, 0)$.
We want to determine the global extrema on the closed unit disc $\overline{U} = \{(x, y) \mid x^2 + y^2 \leq 1\}$. Since there are no local extrema in the inner region, the global extrema are attained on the boundary $x^2 + y^2 = 1$. Solving for $y^2 = 1 - x^2$ we find $g(x) = 4x^2 - (1 - x^2) = 5x^2 - 1$ which has local minima at $x = 0$, $y = \pm 1$. Solving for $x^2 = 1 - y^2$, $h(y) = 4 - 5y^2$ has local maxima at $y = 0$, $x = \pm 1$. Finally, the global maxima and minima are $f(\pm 1, 0) = 4$ and $f(0, \pm 1) = -1$, respectively.

Remark 8.6 To compute the *global* extrema of a function $f \in \mathrm{C}^2(K)$ where K is a compact subset of $\mathbb{R}^n$ we have

(a) to compute the local extrema on the interior of K;

(b) to compute the global extrema on the boundary ∂K;

If f has no partial derivatives at some points one has to consider these points separately. Later we will give a method to solve (b).

Example 8.13 Find the local and global extrema of $f(x, y) = x^2 y$ on $G = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

We have $f \in C^\infty(G)$. Since $\operatorname{grad} f = (f_x, f_y) = (2xy, x^2)$ local extrema can appear only on the y-axis $x = 0$, y is arbitrary. The Hessian matrix is

$$\operatorname{Hess} f(0,y) = \begin{pmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{pmatrix}\Bigg|_{x=0} = \begin{pmatrix} 2y & 2x \\ 2x & 0 \end{pmatrix}\Bigg|_{x=0} = \begin{pmatrix} 2y & 0 \\ 0 & 0 \end{pmatrix}.$$

This matrix is positive semidefinite in case $y > 0$, negative semidefinite in case $y < 0$ and 0 at $(0,0)$. Hence, the above criterion gives *no answer*. We have to apply the definition directly. In case $y > 0$ we have $f(x,y) = x^2y \geq 0$ for all x. In particular $f(x,y) \geq f(0,y) = 0$. Hence $(0,y)$ is a local minimum. Similarly, in case $y < 0$, $f(x,y) \leq f(0,y) = 0$ for all x. Hence, f has a local maximum at $(0,y)$, $y < 0$. However f takes both positive and negative values in a neighborhood of $(0,0)$, for example $f(\varepsilon, \varepsilon) = \varepsilon^3$ and $f(\varepsilon, -\varepsilon) = -\varepsilon^3$. Thus $(0,0)$ is not a local extremum.
We have to consider the boundary $x^2 + y^2 = 1$. Inserting $x^2 = 1 - y^2$ we obtain

$$g(y) = f(x,y)|_{x^2+y^2=1} = x^2y\big|_{x^2+y^2=1} = (1-y^2)y = y - y^3, \quad |\,y\,| \leq 1.$$

We compute the local extrema of the boundary $x^2 + y^2 = 1$ (note, that the circle has no boundary, such that the local extrema are actually the global extrema).

$$g'(y) = 1 - 3y^2 \overset{!}{=} 0, \quad |\,y\,| = \frac{1}{\sqrt{3}}.$$

Since $g''(1/\sqrt{3}) < 0$ and $g''(-1/\sqrt{3}) > 0$, g attains its maximum $\frac{2}{3\sqrt{3}}$ at $y = 1/\sqrt{3}$. Since this is greater than the local maximum of f at $(0,y)$, $y > 0$, f attains its global maximum at the two points

$$M_{1,2} = \left(\pm\sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}\right),$$

where $f(M_{1,2}) = x^2y = \frac{2}{3\sqrt{3}}$. g attains its minimum $-\frac{2}{3\sqrt{3}}$ at $y = -1/\sqrt{3}$. Since this is less than the local minimum of f at $(0,y)$, $y < 0$, f attains its global minimum at the two points

$$m_{1,2} = \left(\pm\sqrt{\frac{2}{3}}, -\frac{1}{\sqrt{3}}\right),$$

where $f(m_{1,2}) = x^2y = -\frac{2}{3\sqrt{3}}$.
The arithmetic-geometric mean inequality shows the same result for $x, y > 0$:

$$\frac{1}{3} \geq \frac{x^2+y^2}{3} = \frac{\frac{x^2}{2} + x^2 2 + y^2}{3} \geq \left(\frac{x^2}{2}\frac{x^2}{2}y^2\right)^{\frac{1}{3}} \implies x^2y \leq \frac{2}{3\sqrt{3}}.$$

In Homework 25.5 there is an example where $\operatorname{Hess} f(0,0)$ is positive semidefinite; however, $(0,0)$ is not a local minimum.

8.5 The Inverse Mapping Theorem

Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ is differentiable on an open set $U \subset \mathbb{R}$, containing $a \in U$, and $f'(a) \neq 0$. If $f'(a) > 0$, then there is an open interval $V \subset U$ containing a such that $f'(x) > 0$ for all $x \in V$. Thus f is strictly increasing on V and therefore injective with an inverse function g defined on some open interval W containing $f(a)$. Moreover g is differentiable (see Proposition 4.5) and $g'(y) = 1/f'(x)$ if $f(x) = y$. An analogous result in higher dimensions is more involved but the result is very important.

Theorem 8.17 (Inverse Mapping Theorem) *Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable on an open set U containing a, and $\det f'(a) \neq 0$. Then there is an open set $V \subset U$ containing a and an open set W containing $f(a)$ such that $f : V \to W$ has a continuous inverse $g\colon W \to V$ which is differentiable and for all $y \in W$. For $y = f(x)$ we have*

$$g'(y) = (f'(x))^{-1}, \quad Dg(y) = (Df(x))^{-1}. \tag{8.43}$$

For the proof see [8, 9.24 Theorem] or [9, 2-11].

Corollary 8.18 *Let $U \subset \mathbb{R}^n$ be open, $f : U \to \mathbb{R}^n$ continuously differentiable and $\det f'(x) \neq 0$ for all $x \in U$. Then $f(U)$ is open in $\mathbb{R}^n$*

Definition 8.10 Let $f\colon \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable at x,

$$f(x) = f(x_1, \dots, x_n) = (f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)).$$

then we call

$$\frac{\partial(f_1, \dots, f_n)}{\partial(x_1, \dots, x_n)}(x) = \det f'(x) = \det \left(\frac{\partial f_i(x)}{\partial x_j}\right)_{i,j=1,\dots n} \tag{8.44}$$

the *Jacobian* of f at x.

Remarks 8.7 (a) One main part is to show that there is an open set $V \subset U$ which is mapped onto an *open* set W. This is not true for *continuous* mappings. For example $\sin x$ maps the open interval $(0, 2\pi)$ onto the closed set $[-1, 1]$. Note that $\sin x$ does not satisfy the assumptions of the corollary since $\sin'(\pi/2) = \cos(\pi/2) = 0$.
(b) From linear algebra it is well known that the Jacobian matrix $f'(x)$ is invertible if and only if its determinant $\det f'(x)$ is non-zero.
(c) Let us reformulate the statement of the theorem. Suppose

$$\begin{aligned} y_1 &= f_1(x_1, \dots, x_n), \\ y_2 &= f_2(x_1, \dots, x_n), \\ &\vdots \\ y_n &= f_n(x_1, \dots, x_n) \end{aligned}$$

is a system of n equations in n variables $x_1, \dots, x_n$; $y_1, \dots, y_n$ are given in a neighborhood W of $f(a)$. Under the assumptions of the theorem, there exists a unique solution $x = g(y)$ of this system of equations

$$\begin{aligned} x_1 &= g_1(y_1, \dots, y_n), \\ x_2 &= g_2(y_1, \dots, y_n), \\ &\vdots \\ x_n &= g_n(y_1, \dots, y_n) \end{aligned}$$

in a certain neighborhood $(x_1, \dots, x_n) \in U$ of a. Note that the theorem states the existence of such a solution. It doesn't provide an explicit formula.
(d) It should be noted that the inverse function g may exist even if $\det f'(x) = 0$. For example $f\colon \mathbb{R} \to \mathbb{R}$, defined by $f(x) = x^3$ has $f'(0) = 0$; however $g(y) = \sqrt[3]{x}$ is inverse to $f(x)$. One thing is certain if $\det f'(a) = 0$ then g cannot be differentiable at $f(a)$. If g were differentiable at $f(a)$, the chain rule applied to $g(f(x)) = x$ would give

$$g'(f(a)) \cdot f'(a) = \mathrm{id}$$

and consequently

$$\det g'(f(a)) \, \det f'(a) = \det \mathrm{id} \, = 1$$

contradicting $\det f'(a) = 0$.
(e) Note that the theorem states that under the given assumptions f is *locally* invertible. There is no information about the existence of an inverse function g to f on a fixed open set. See Example 8.14 (a) below.

Example 8.14 (a) Let $x = r\cos\varphi$ and $y = r\sin\varphi$ be the polar coordinates in $\mathbb{R}^2$. More precisely, let

$$f(r, \varphi) = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\varphi \\ r\sin\varphi \end{pmatrix}, \quad f\colon \mathbb{R}^2 \to \mathbb{R}^2.$$

The Jacobian is

$$\frac{\partial(x, y)}{\partial(r, \varphi)} = \begin{vmatrix} x_r & x_\varphi \\ y_r & y_\varphi \end{vmatrix} = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = r.$$

Let $f(r_0, \varphi_0) = (x_0, y_0) \neq (0, 0)$, then $r_0 \neq 0$ and the Jacobian of f at (r_0, φ_0) is non-zero. Since all partial derivatives of f with respect to r and φ exist and they are continuous on $\mathbb{R}^2$, the assumptions of the theorem are satisfied. Hence, in a neighborhood U of (x_0, y_0) there exists a continuously differentiable inverse function $r = r(x, y)$, $\varphi = \varphi(x, y)$. In this case, the function can be given explicitly, $r = \sqrt{x^2 + y^2}$, $\varphi = \arg(x, y)$. We want to compute the Jacobian matrix of the inverse function. Since the inverse matrix

$$\begin{pmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{pmatrix}^{-1} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix}$$

we obtain by the theorem

$$g'(x,y) = \left(\frac{\partial(r,\varphi)}{\partial(x,y)}\right) = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix};$$

in particular, the second row gives the partial derivatives of the argument function with respect to x and y

$$\frac{\partial \arg(x,y)}{\partial x} = \frac{-y}{x^2+y^2}, \qquad \frac{\partial \arg(x,y)}{\partial y} = \frac{x}{x^2+y^2}.$$

Note that we have not determined the explicit form of the argument function which is not unique since $f(r,\varphi+2k\pi) = f(r,\varphi)$, for all $k \in \mathbb{Z}$. However, the gradient takes always the above form. Note that $\det f'(r,\varphi) \neq 0$ for all $r \neq 0$ is not sufficient for f to be injective on $\mathbb{R}^2 \setminus \{(0,0)\}$.

(b) Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by $(u,v) = f(x,y)$ where

$$u(x,y) = \sin x - \cos y, \quad v(x,y) = -\cos x + \sin y.$$

Since

$$\frac{\partial(u,v)}{\partial(x,y)} = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \begin{vmatrix} \cos x & \sin y \\ \sin x & \cos y \end{vmatrix} = \cos x \cos y - \sin x \sin y = \cos(x+y)$$

f is locally invertible at $(x_0,y_0) = (\frac{\pi}{4}, -\frac{\pi}{4})$ since the Jacobian at (x_0,y_0) is $\cos 0 = 1 \neq 0$. Since $f(\frac{\pi}{4}, -\frac{\pi}{4}) = (0, -\sqrt{2})$, the inverse function $g(u,v) = (x,y)$ is defined in a neighborhood of $(0,-\sqrt{2})$ and the Jacobian matrix of g at $(0,-\sqrt{2})$ is

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Note that at point $(\frac{\pi}{4}, \frac{\pi}{4})$ the Jacobian of f vanishes. There is indeed no neighborhood of $(\frac{\pi}{4}, \frac{\pi}{4})$ where f is injective since for all $t \in \mathbb{R}$

$$f\left(\frac{\pi}{4}+t, \frac{\pi}{4}-t\right) = f\left(\frac{\pi}{4}, \frac{\pi}{4}\right) = (0,0).$$

8.6 The Implicit Function Theorem

Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x,y) = x^2+y^2-1$. If we choose (a,b) with $a \neq -1, 1$, there are open intervals A and B containig a and b with the following property: if $x \in A$, there is a unique $y \in B$ with $f(x,y) = 0$. We can therefore define a function $g\colon A \to B$ by the condition $g(x) \in B$ and $f(x, g(x)) = 0$. If $b > 0$ then $g(x) = \sqrt{1-x^2}$; if $b < 0$ then $g(x) = -\sqrt{1-x^2}$. Both functions g are differentiable. These functions are said to be defined *implicitly* by the equation $f(x,y) = 0$.

More generally we may ask the following: If $f\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ and $f(a_1,\dots,a_n,b) = 0$ when can we find for each $(x_1,\dots,x_n)$ near $(a_1,\dots,a_n)$ a unique y near b such that $f(x_1,\dots,x_n,y) = 0$? The answer is provided by the following theorem.

Theorem 8.19 *Suppose that $f\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $f = f(x, y)$, is continuously differentiable in an open set containing $(a, b) \in \mathbb{R}^{n+m}$ and $f(a, b) = 0$. Let $M(x, y)$ be the $m \times m$-matrix*

$$\big(D_{n+j} f^i(x, y)\big) = \left(\frac{\partial(f_1, \dots, f_m)}{\partial(y_1, \dots, y_m)}(x, y)\right) = \left(\frac{\partial f_i(x, y)}{\partial y_j}\right), \quad i, j = 1, \dots, m. \tag{8.45}$$

If $\det M(a, b) \neq 0$ *there is an open set $A \subset \mathbb{R}^n$ containing a and an open set $B \subset \mathbb{R}^m$ containing b with the following properties: There exists a unique continuously differentiable function $g\colon A \to B$ such that*

(a) $g(a) = b$,
(b) $f(x, g(x)) = 0$ *for all* $x \in A$.

The Jacobian matrix $g'(x)$ is given by

$$\left(\frac{\partial(g_1, \dots, g_m)}{\partial(x_1, \dots, x_n)}(x)\right) = -M(x, g(x))^{-1} \cdot \left(\frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)}(x, g(x))\right) \tag{8.46}$$

$$\frac{\partial(g_k(x))}{\partial x_j} = -\sum_{l=1}^{n} (M(x, g(x))^{-1})_{kl} \cdot \frac{\partial f_l(x, g(x))}{\partial x_j}, \quad k = 1, \dots, m,\ j = 1, \dots, n.$$

Proof. Define $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ by $F(x, y) = (x, f(x, y))$. Let $M = M(a, b)$. Then

$$F'(a, b) = \begin{pmatrix} \mathbb{1}_n & 0_{n,m} \\ 0_{m,n} & M \end{pmatrix} \implies \det F'(a, b) = \det M \neq 0.$$

By the inverse mapping theorem Theorem 8.17 there exists an open set $W \subset \mathbb{R}^n \times \mathbb{R}^m$ containing $F(a, b) = (a, 0)$ and an open set $V \subset \mathbb{R}^n \times \mathbb{R}^m$ containing (a, b) which may be of the form $A \times B$ such that $F\colon A \times B \to W$ has a differentiable inverse $h\colon W \to A \times B$. Clearly, h is of the form $h(x, y) = (x, k(x, y))$ for some differentiable function k (since F is of this form). Let $p_2\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$, $p_2(x, y) = y$ be the projection onto the second argument. Then $p_2 \circ F = f$ and therefore

$$\begin{aligned} f(x,\, k(x, y)) &= f \circ h(x, y) = (p_2 \circ F) \circ h(x, y) \\ &= p_2 \circ (F \circ h)(x, y) = p_2(x, y) = y. \end{aligned}$$

Thus $f(x,\, k(x, 0)) = 0$; in other words we can define $g(x) = k(x, 0)$.
Since g is differentiable, it is easy to find the Jacobian matrix. In fact, since $f_i(x, g(x)) = 0$, $i = 1, \dots n$, taking the partial derivative $\frac{\partial f}{\partial x_j}$ on both sides gives by the chain rule

$$0 = \frac{\partial f_i(x, g(x))}{\partial x_j} + \sum_{k=1}^{m} \frac{\partial f_i(x, g(x))}{\partial y_k} \cdot \frac{\partial g_k(x)}{\partial x_j}$$

$$0 = \frac{\partial f_i(x, g(x))}{\partial x_j} + M(x, g(x)) \cdot \left(\frac{\partial g_k(x)}{\partial x_j}\right) \quad \Big| - \frac{\partial f_i(x, g(x))}{\partial x_j}$$

$$-\left(\frac{\partial f_i(x, g(x))}{\partial x_j}\right) = M(x, g(x)) \cdot \left(\frac{\partial g_k(x)}{\partial x_j}\right).$$

Since $\det M(a,b) \neq 0$, $\det M(x,y) \neq 0$ in a small neighborhood of (a,b). Hence $M(x,g(x))$ is invertible and we can multiply the preceding equation from the left by $M(x,g(x))^{-1}$ which gives (8.46). ∎

Remarks 8.8 (a) The theorem gives a sufficient condition for "locally" solving the system of equations

$$\begin{aligned} 0 &= f_1(x_1,\dots,x_n,y_1,\dots y_m),\\ &\vdots\\ 0 &= f_m(x_1,\dots,x_n,y_1,\dots,y_m) \end{aligned}$$

with given $x_1,\dots,x_n$ for $y_1,\dots,y_m$.
(b) We rewrite the statement in case $n=m=1$: If $f(x,y)$ is continously differentiable on an open set $G\subset\mathbb{R}^2$ which contains (a,b) and $f(a,b)=0$. If $f_y(a,b)\neq 0$ then there exist $\delta,\varepsilon>0$ such that the following holds: for every $x\in U_\delta(a)$ there exists a unique $y=g(x)\in U_\varepsilon(b)$ with $f(x,y)=0$. We have $g(a)=b$; the function $y=g(x)$ is continuously differentiable with

$$g'(x) = -\frac{f_x(x,g(x))}{f_y(x,g(x))}.$$

Be careful, note $f_x(x,g(x)) \neq \frac{\partial f}{\partial x} g(x,g(x))$.

Example 8.15 (a) Let $f(x,y)=\sin(x+y)+\mathrm{e}^{xy}-1$. Note that $f(0,0)=0$. Since

$$f_y(0,0) = \cos(x+y) + x\mathrm{e}^{xy}\big|_{(0,0)} = \cos 0 + 0 = 1 \neq 0$$

$f(x,y)=0$ can uniquely be solved for $y=g(x)$ in a neighborhood of $x=0, y=0$. Further

$$f_x(0,0) = \cos(x+y) + y\mathrm{e}^{xy}\big|_{(0,0)} = 1.$$

By Remark 8.8 (b)

$$g'(x) = -\left.\frac{f_x(x,y)}{f_y(x,y)}\right|_{y=g(x)} = \frac{\cos(x+g(x)) + g(x)\,\mathrm{e}^{xg(x)}}{\cos(x+g(x)) + x\,\mathrm{e}^{xg(x)}}.$$

In particular $g'(0)=-1$.
Remark. Differentiating the equation $f_x + f_y g' = 0$ we obtain

$$\begin{aligned} 0 &= f_{xx} + f_{xy}g' + (f_{yx}+f_{yy}g')g' + f_y g''\\ g'' &= -\frac{1}{f_y}\left(f_{xx} + 2f_{xy}g' + f_{yy}(g')^2\right)\\ g'' &\underset{g'=-\frac{f_x}{f_y}}{=} \frac{-f_{xx}f_y^2 + 2f_{xy}f_xf_y - f_{yy}f_x^2}{f_y^3}. \end{aligned}$$

Since

$$\begin{aligned} f_{xx}(0,0) &= -\sin(x+y) + y^2 e^{xy}\big|_{(0,0)} = 0, \\ f_{yy}(0,0) &= -\sin(x+y) + x^2 e^{xy}\big|_{(0,0)} = 0, \\ f_{xy}(0,0) &= -\sin(x+y) + e^{xy}(1+xy)\big|_{(0,0)} = 1, \end{aligned}$$

we obtain $g''(0) = 2$. Therefore the Taylor expansion of $g(x)$ around 0 reads

$$g(x) = x + x^2 + r_3(x).$$

(b)

$$\begin{aligned} f_1(x,y,z) &= x^2 + 4y^2 + 9z^2 - 1 && = 0, \quad \text{ellipsoid}, \\ f_2(x,y,z) &= x + y + z && = 0, \quad \text{plane} \end{aligned}$$

Find $y = y(x)$ and $z = z(x)$ around $x = 0$.
Inserting $x = 0$ into the two equations we find

$$4y^2 + 9z^2 = 1, \quad y + z = 0, \quad \Longrightarrow 13y^2 = 1 \Longrightarrow |\, y \,| = \frac{1}{\sqrt{13}}.$$

We choose $y = 1/\sqrt{13}$. We have

$$\frac{\partial(f_1, f_2)}{\partial(y, z)} = \begin{vmatrix} \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \end{vmatrix} = \begin{vmatrix} 8y & 18z \\ 1 & 1 \end{vmatrix} = 8y - 18z = 26y \neq 0.$$

Hence $f = 0$ is uniquely solvable for $y = y(x)$, $z = z(x)$ in a certain neighborhood of $x = 0$, $y = 1/\sqrt{13}$. Geometrically, $(x, y(x), z(x))$ is a curve in $\mathbb{R}^3$, the intersection of the ellipsoid $f_1 = 0$ with the plane $f_2 = 0$.
We compute y' and z'. Differentiation of $f_i(x, y(x), z(x)) = 0$, $i = 1, 2$ gives

$$\begin{aligned} 2x + 8y\, y' + 18z\, z' &= 0, \\ 1 + y' + z' &= 0 \end{aligned}$$

Inserting $y' = -1 - z'$ into the first equation gives $2x - 8y - 8yz' + 18zz' = 0$; hence

$$z' = \frac{4y - x}{9z - 4y}, \quad y' = \frac{-9z + x}{9z - 4y}.$$

In particular, at $(0, \frac{1}{\sqrt{13}}, -\frac{1}{\sqrt{13}})$ the tangent vector to the curve $(x, y(x), z(x))$ is $(1, y'(0), z'(0)) = (1, -9/13, 9/13)$.
(c) Let $\gamma(t) = (x(t), y(t))$ be a differentiable curve $\gamma \in C^2([0,1])$ in $\mathbb{R}^2$. Suppose in a neighborhood of $t = 0$ the curve describes a function $y = g(x)$. Find the Taylor polynomial of degree 2 of g at $x_0 = x(0)$.
Inserting the curve into the equation $y = g(x)$ we have $y(t) = g(x(t))$. Differentiation gives

$$\dot{y} = g'\,\dot{x}, \qquad \ddot{y} = g''\dot{x}^2 + g'\ddot{x}$$

Thus

$$g'(x) = \frac{\dot{y}}{\dot{x}}, \qquad g''(x) = \frac{\ddot{y} - g'\ddot{x}}{\dot{x}^2} = \frac{\ddot{y}\dot{x} - \ddot{x}\dot{y}}{\dot{x}^3}$$

Now we have the Taylor ploynomial of g at x_0

$$T_2(g)(x) = x_0 + g'(x_0)(x - x_0) + \frac{g''(x_0)}{2}(x - x_0)^2.$$

8.7 Level Set, Normal Vector, and Lagrange Multiplier

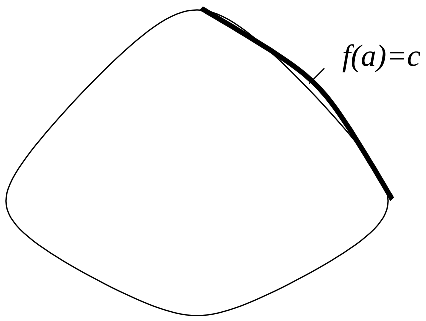
8.7.1 Level Sets

As an application of the implicit function theorem we give another geometric interpretation of the gradient $\operatorname{grad} f$ to be the normal vector to the level set. As usual, $U \subset \mathbb{R}^n$ is open.

Definition 8.11 Let $f\colon U \to \mathbb{R}$ be differentiable. For $c \in \mathbb{R}$ define the *level set* $U_c = \{x \in U \mid f(x) = c\}$.

This set may be empty, may consist of a single point (in case of local extrema) or, in the "generic" case, U_c it is a $(n-1)$-dimensional surface. $\{U_c \mid c \in \mathbb{R}\}$ is family of non-intersecting subsets of U which cover U.

Proposition 8.20 *Let $f\colon U \to \mathbb{R}$ be continuously differentiable, $f(a) = c$, and* $\operatorname{grad} f(a) \neq 0$. *Then the tangent hyperplane to the level set U_c at a is given by*



$$0 = \sum_{i=1}^{n} f_{x_i}(a)(x_i - a_i), \quad \textit{or} \tag{8.47}$$

$$0 = \langle \operatorname{grad} f(a), x - a \rangle . \tag{8.48}$$

Note that the second equation means that the tangent hyperplane is orthogonal to $\operatorname{grad} f(a)$. A vector which is orthogonal to the tangent hyperplane is called a *normal* vector to the given surface.

Proof. Consider U_c in a neighborhood of a to be the graph of a function which is implicitly given by $f(x) = c$. Since $\operatorname{grad} f(a) \neq 0$ at least one coordinate of the gradient $\operatorname{grad} f(a)$ is nonzero. Without loss of generality we may assume that $f_{x_n}(a) \neq 0$. By the implicit function theorem we may solve $f(x) = c$ in a neighborhood of $x = a$ locally for x_n, say $x_n = g(x_1, \dots, x_{n-1})$.

Define the tangent hyperplane to be the graph of the linearization of g at $(a_1, \dots, a_{n-1}, a_n)$. By Remark 8.5 8.5 with $\tilde{a} = (a_1, \dots, a_{n-1})$, $\tilde{x} = (x_1, \dots, x_{n-1})$, the hyperplane is given by

$$x_n = g(\tilde{a}) + \langle \operatorname{grad} g(\tilde{a}), \tilde{x} \rangle . \tag{8.49}$$

Since $f(\tilde{a}, g(\tilde{a})) = c$, by the implicit function theorem

$$\frac{\partial g(\tilde{a})}{\partial x_j} = -\frac{f_{x_j}(a)}{f_{x_n}(a)}, \quad j = 1, \dots, n-1.$$

Inserting this into (8.49) we have

$$x_n - a_n = -\frac{1}{f_{x_n}(a)} \sum_{j=1}^{n-1} f_{x_j}(a)(x_j - a_j).$$

Multiplication by $-f_{x_n}(a)$ gives

$$-f_{x_n}(a)(x_n - a_n) = \sum_{j=1}^{n-1} f_{x_j}(a)(x_j - a_j) \Longrightarrow 0 = \langle \operatorname{grad} f(a), x - a \rangle .$$

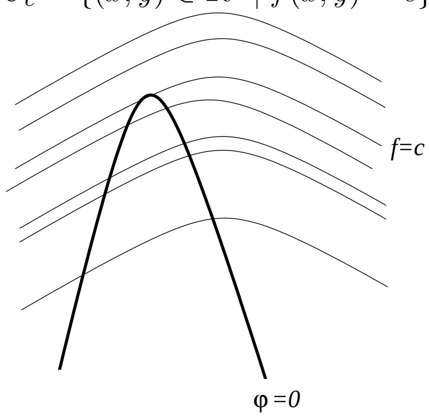
■

8.7.2 Lagrange Multiplier

This is a method to find local extrema of a function under certain constraints. Consider the following problem: Find local extremma of a function $f(x, y)$ of two variables where x and y are not independent from each other but satisfy the constraint

$$\varphi(x, y) = 0.$$

Suppose further that f and φ are continuously differentiable. Note that the level sets $U_c = \{(x, y) \in \mathbb{R}^2 \mid f(x, y) = c\}$ form a family of non-intersecting curves in the plane.



We have to find the curve $f(x, y) = c$ intersecting the constraint curve $\varphi(x, y) = 0$ where c is as large or as small as possible. Usually $f = c$ intersects $\varphi = 0$ if c monotonically changes. However if c is maximal, the curve $f = c$ *touches* the graph $\varphi = 0$. In other words, the tangent lines coincide. This means that the defining normal vectors to the tangent lines are scalar multiples of each other.

Theorem 8.21 *Let $f, \varphi\colon U \to \mathbb{R}$, $U \subset \mathbb{R}^n$ is open, be continuously differentiable and f has a local extrema at $a \in U$ under the constraint $\varphi(x) = 0$. Suppose that* $\operatorname{grad} \varphi(a) \neq 0$. *Then there exists a real number λ such that*

$$\operatorname{grad} f(a) = \lambda \operatorname{grad} \varphi(a).$$

This number λ is called *Lagrange multiplier.*
Proof. The idea is to solve the constraint $\varphi(x) = 0$ for one variable and to consider the "free" extremum problem with one variable less. Suppose without loss of generality that $\varphi_{x_n}(a) \neq 0$. By the implicit function theorm we can solve $\varphi(x) = 0$ for $x_n = g(x_1, \dots, x_{n-1})$ in a neighborhood of $x = a$. Differentiating $\varphi(\tilde{x}, g(\tilde{x})) = 0$ and inserting $a = (\tilde{a}, a_n)$ as before we have

$$\varphi_{x_j}(a) + \varphi_{x_n}(a) g_{x_j}(\tilde{a}) = 0, \quad j = 1, \dots, n-1. \tag{8.50}$$

Since $h(\tilde{x}) = f(\tilde{x}, g(\tilde{x}))$ has a local extremum at $\tilde{a}$ all partial derivatives of h vanish at $\tilde{a}$:

$$f_{x_j}(a) + f_{x_n}(a) g_{x_j}(\tilde{a}) = 0, \quad j = 1, \dots, n-1. \tag{8.51}$$

Setting $\lambda = f_{x_n}(a)/\varphi_{x_n}(a)$ and comparing (8.50) and (8.51) we find

$$f_{x_j}(a) = \lambda \varphi_{x_j}(a), \quad j = 1, \dots, n-1.$$

Since by definition, $f_{x_n}(a) = \lambda \varphi_{x_n}(a)$ we finally obtain $\operatorname{grad} f(a) = \lambda \operatorname{grad} \varphi(a)$ which completes the proof. ∎

Example 8.16 (a) Let $A = (a_{ij})$ be a real symmetric $n \times n$-matrix, and define $f(x) = \langle x, Ax \rangle = \sum_{i,j} a_{ij} x_i x_j$. We aks for the local extrema of f on the unit sphere $\mathrm{S}^{n-1} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$.
This constraint can be written as $\varphi(x) = \|x\|^2 - 1 = \sum_{i=1}^n x_i^2 - 1 = 0$. Suppose that f attains a local minimum at $a \in \mathrm{S}^{n-1}$. By Example 8.6 (b)

$$\operatorname{grad} f(a) = 2A(a).$$

On the other hand

$$\operatorname{grad} \varphi(a) = (2x_1, \dots, 2x_n)|_{x=a} = 2a.$$

By Theorem 8.21 there exists a real number λ_1 such that

$$\operatorname{grad} f(a) = 2A(a) = \lambda_1 \operatorname{grad} \varphi(a) = 2a,$$

Hence $A(a) = \lambda_1 a$; that is, λ is an eigenvalue of A and a the corresponding eigenvector. In particular, A has a real eigenvalue. Since S^{n-1} has no boundary, the global minimum is also a local one. We find: if $f(a) = \langle a, A(a) \rangle = \langle a, \lambda a \rangle = \lambda$ is the global minimum, λ is the smallest eigenvalue.
(b) Let a be the point of a hypersurface $M = \{x \mid \varphi(x) = 0\}$ with minimal distance to a given point $b \notin M$. Then the line through a and b is orthogonal to M.
Indeed, the function $f(x) = \|x - b\|^2$ attains its minimum under the condition $\varphi(x) = 0$ at a. By the Theorem, there is a real number λ such that

$$\operatorname{grad} f(a) = 2(a - b) = \lambda \operatorname{grad} \varphi(a).$$

The assertion follows since by Proposition 8.20 $\operatorname{grad} \varphi(a)$ is orthogonal to M at a and $b - a$, the direction of the line through a and b is a multiple of it.

Theorem 8.22 (Lagrange's Multiplier Rule) *Let $f, \varphi_i \colon U \to \mathbb{R}$, $i = 1, \dots, m$, $m < n$, be continuously differentiable functions. Let $M = \{x \in U \mid \varphi_1(x) = \cdots = \varphi_m(x) = 0\}$ and suppose that f has a local extrema at a under the constraints $a \in M$. Suppose further that the Jacobian $m \times n$-matrix $\varphi'(a)$ has rank m.*
Then there exist real numbers $\lambda_1, \dots, \lambda_m$ such that

$$\operatorname{grad} f(a) = \operatorname{grad}(\lambda_1\varphi_1 + \cdots + \lambda_m\varphi_m)(a) = 0.$$

Note that the rank condition ensures that there is a choice of m variables out of $x_1, \dots, x_n$ such that the Jacobian of $\varphi_1, \dots, \varphi_m$ with respect to this set of variable is nonzero at a.

8.8 Integrals depending on Parameters

Problem: Define $I(y) = \int_a^b f(x,y)\,\mathrm{d}x$; what are the relations between properties of $f(x,y)$ and of $I(y)$ for example with respect to continuity and differentiability.

8.8.1 Continuity of $I(y)$

Proposition 8.23 *Let $f(x,y)$ be continuous on the rectangle $R = [a,b] \times [c,d]$.*
Then $I(y) = \int_a^b f(x,y)\,\mathrm{d}x$ is continuous on $[c,d]$.

Proof. Let $\varepsilon > 0$. Since f is continuous on the compact set R, f is uniformly continuous on R (see Proposition 6.23). Hence, there is a $\delta > 0$ such that $|\, x - x' \,| < \delta$ and $|\, y - y' \,| < \delta$ and $(x,y), (x',y') \in R$ imply

$$|\, f(x,y) - f(x',y') \,| < \varepsilon.$$

Therefore, $|\, y - y_0 \,| < \delta$ and $y, y_0 \in [c,d]$ imply

$$|\, I(y) - I(y_0) \,| = \left| \int_a^b (f(x,y) - f(x,y_0))\,\mathrm{d}x \right| \le \varepsilon(b-a).$$

This shows continuity of $I(y)$ at y_0. ∎

For example, $I(y) = \int_0^1 \arctan\frac{x}{y}\,\mathrm{d}x$ is continuous for $y > 0$.

8.8.2 Differentiation of Integrals

Proposition 8.24 *Let $f(x,y)$ be defined on $R = [a,b] \times [c,d]$ and continuous as a function of x for every fixed y. Suppose that $f_y(x,y)$ exists for all $(x,y) \in R$ and is continuous as a function of the two variables x and y.*
Then $I(y)$ is differentiable and

$$I'(y) = \frac{\mathrm{d}}{\mathrm{d}y}\int_a^b f(x,y)\,\mathrm{d}x = \int_a^b f_y(x,y)\,\mathrm{d}x.$$

Proof. This prove was not carried out in the lecture.
Since $f_y(x, y)$ is continuous, it is uniformly continuous on R. Hence there exists $\delta > 0$ such that $|x' - x''| < \delta$ and $|y' - y''| < \delta$ imply $|f_y(x', y') - f_y(x'', y'')| < \varepsilon$. We have for $|h| < \delta$

$$\begin{aligned}\left| \frac{I(y_0 + h) - I(y_0)}{h} - \int_a^b f_y(x, y_0)\,\mathrm{d}x \right| &\le \left| \int_a^b \left(\frac{f(x, y_0 + h) - f(x, y_0)}{h} - f_y(x, y_0) \right) \mathrm{d}x \right| \\ &\underset{\text{Mean value theorem}}{\le} \int_a^b |f_y(x, y_0 + \theta h) - f_y(x, y_0)|\,\mathrm{d}x < \varepsilon(b - a)\end{aligned}$$

for some $\theta \in (0, 1)$. Hence,

$$\lim_{h\to 0} \frac{I(y_0 + h) - I(y_0)}{h} = I'(y_0) = \int_a^b f_y(x, y_0)\,\mathrm{d}x.$$

■

In case of variable integration limits we have the following theorem.

Proposition 8.25 *Let $f(x, y)$ be as in Proposition 8.24. Let $\alpha(y)$ and $\beta(y)$ be differentiable on $[c, d]$, and suppose that $\alpha([c, d])$ and $\beta([c, d])$ are contained in $[a, b]$.*
Let $I(y) = \int_{\alpha(y)}^{\beta(y)} f(x, y)\,\mathrm{d}x$. Then $I(y)$ is differentiable and

$$I'(y) = \int_{\alpha(y)}^{\beta(y)} f_y(x, y)\,\mathrm{d}x \; + \beta'(y) f(\beta(y), y) \; - \alpha'(y) f(\alpha(y), y). \tag{8.52}$$

Proof. Let $F(y, u, v) = \int_u^v f(x, y)\,\mathrm{d}x$; then $I(y) = F(y, \alpha(y), \beta(y))$. The fundamental theorem of calculus yields

$$\begin{aligned}\frac{\partial F}{\partial v}(y, u, v) &= \frac{\partial f}{\partial v} \int_u^v f(x, y)\,\mathrm{d}x = f(v, y), \\ \frac{\partial F}{\partial u}(y, u, v) &= \frac{\partial f}{\partial u} \left(- \int_v^u f(x, y)\,\mathrm{d}x \right) = -f(u, y).\end{aligned} \tag{8.53}$$

By the chain rule, the previous proposition and (8.53) we have

$$\begin{aligned}I'(y) &= \frac{\partial F}{\partial y}(y, \alpha(y), \beta(y)) + \frac{\partial F}{\partial u}(y, \alpha(y), \beta(y))\, \alpha'(y) + \frac{\partial F}{\partial v}(y, \alpha(y), \beta(y))\, \beta'(y) \\ &= \int_{\alpha(y)}^{\beta(y)} f_y(x, y)\,\mathrm{d}x + \alpha'(y)(-f(\alpha(y), y)) + \beta'(y) f(\beta(y), y).\end{aligned}$$

■

8.8.3 Improper Integrals with Parameters

Suppose that the improper integral $\int_a^\infty f(x,y)\,\mathrm{d}x$ exists for $y \in [c,d]$.

Definition 8.12 We say that the improper integral $\int_a^\infty f(x,y)\,\mathrm{d}x$ *converges uniformly* with respect to y on $[c,d]$ if for every $\varepsilon > 0$ there is an $A_0 > 0$ such that $A > A_0$ implies

$$\left| I(y) - \int_a^A f(x,y)\,\mathrm{d}x \right| \equiv \left| \int_A^\infty f(x,y)\,\mathrm{d}x \right| < \varepsilon$$

for all $y \in [c,d]$.

Note that the Cauchy and Weierstraß criteria (see Proposition 7.1 and Theorem 7.3) for uniform convergence of series of functions also hold for improper parametric integrals. For example the theorem of Weierstraß now reads as follows.

Proposition 8.26 *Suppose that $\int_a^A f(x,y)\,\mathrm{d}x$ exists for all $A \ge a$ and $y \in [c,d]$. Suppose further that $|\,f(x,y)\,| \le \varphi(x)$ for all $x \ge a$ and $\int_a^\infty \varphi(x)\,\mathrm{d}x$ converges.*
Then $\int_a^\infty f(x,y)\,\mathrm{d}x$ converges uniformly with respect to $y \in [c,d]$.

The following example was not carried out in the lecture.

Example 8.17 $I(y) = \int_1^\infty \mathrm{e}^{-xy} x^y y^2\,\mathrm{d}x$ converges uniformly on $[2,4]$ since

$$|\,f(x,y)\,| = \left|\,\mathrm{e}^{-xy} x^y y^2\,\right| \le \mathrm{e}^{-2x} x^4 4^2 = \varphi(x).$$

and $\int_1^\infty \mathrm{e}^{-2x} x^4 4^2\,\mathrm{d}x < \infty$ converges.

If we add the assumption of *uniform convergence* then the preceding theorems remain true for improper integrals.

Proposition 8.27 *Let $f(x,y)$ be continuous on $\{(x,y) \in \mathbb{R}^2 \mid a \le x < \infty,\ c \le y \le d\}$. Suppose that $I(y) = \int_a^\infty f(x,y)\,\mathrm{d}x$ converges uniformly with respect to $y \in [c,d]$.*
Then $I(y)$ is continuous on $[c,d]$.

Proof. This proof was not carried out in the lecture. Let $\varepsilon > 0$. Since the improper integral converges uniformly, there exists $A_0 > 0$ such that for all $A \ge A_0$ we have

$$\left| \int_A^\infty f(x,y)\,\mathrm{d}x \right| < \varepsilon$$

for all $y \in [c,d]$. Let $A \ge A_0$ be fixed. On $\{(x,y) \in \mathbb{R}^2 \mid a \le x \le A,\ c \le y \le d\}$ $f(x,y)$ is uniformly continuous; hence there is a $\delta > 0$ such that $|\,x' - x''\,| < \delta$ and $|\,y' - y''\,| < \delta$ implies

$$|\,f(x',y') - f(x'',y'')\,| < \frac{\varepsilon}{A-a}.$$

Therefore,

$$\int_a^A |\,f(x,y) - f(x,y_0)\,|\,\mathrm{d}x < \frac{\varepsilon}{A-a}(A-a) = \varepsilon, \quad \text{for} \quad |\,y - y_0\,| < \delta.$$

Finally,

$$|I(y) - I(y_0)| = \left(\int_a^A + \int_A^\infty\right) |f(x,y) - f(x,y_0)| \le 2\varepsilon \quad \text{for} \quad |y - y_0| < \delta.$$

■

We skip the proof of the following proposition.

Proposition 8.28 *Let $f_y(x,y)$ be continuous on $\{(x,y) \mid a \le x < \infty,\ c \le y \le d\}$, $f(x,y)$ continuous with respect to x for all fixed $y \in [ccarryout, d]$.*
Suppose that for all $y \in [c,d]$ the integral $I(y) = \int_a^\infty f(x,y)\,\mathrm{d}x$ exists and the integral $\int_a^\infty f_y(x,y)\,\mathrm{d}x$ converges uniformly with respect to $y \in [c,d]$.
Then $I(y)$ is differentiable and $I'(y) = \int_a^\infty f_y(x,y)\,\mathrm{d}x$.

The following examples were not carried out in the lecture.

Example 8.18 (a) $I(y) = \int_3^4 \frac{\sin(xy)}{x}\,\mathrm{d}x$ is differentiable by Proposition 8.24 since $f_y(x,y) = \frac{\cos(xy)}{x}\,x = \cos(xy)$ is continuous. Hence

$$I'(y) = \int_3^4 \cos(xy)\,\mathrm{d}x = \left.\frac{\sin(xy)}{y}\right|_3^4 = \frac{\sin 4y}{y} - \frac{\sin 3y}{y}.$$

(b) $I(y) = \int_{\log y}^{\sin y} \mathrm{e}^{x^2 y}\,\mathrm{d}x$ is differentiable with

$$I'(y) = \int_{\log y}^{\sin y} x^2 \mathrm{e}^{x^2 y}\,\mathrm{d}x + \cos y \mathrm{e}^{y \sin^2 y} - \frac{1}{y}\mathrm{e}^{y(\log y)^2}.$$

(c) $I(y) = \int_0^\infty \mathrm{e}^{-x^2}\cos(2yx)\,\mathrm{d}x$. $f(x,y) = \mathrm{e}^{-x^2}\cos(2yx)$, $f_y(x,y) = -2x\sin(2yx)\,\mathrm{e}^{-x^2}$ converges uniformly with respect to y since

$$|f_y(x,y)| \le 2x\mathrm{e}^{-x^2} \le K\mathrm{e}^{-x^2/2}.$$

Hence,

$$I'(y) = -\int_0^\infty 2x\sin(2yx)\,\mathrm{e}^{-x^2}\,\mathrm{d}x.$$

Partial integration with $u = \sin(2yx)$, $v' = -\mathrm{e}^{-x^2}2x$ gives $u' = 2y\cos(2yx)$, $v = \mathrm{e}^{-x^2}$ and

$$\int_0^A -\mathrm{e}^{-x^2}2x\,\sin(2yx)\,\mathrm{d}x = \sin 2yA\,\mathrm{e}^{-A^2} - \int_0^A 2y\cos(2yx)\,\mathrm{e}^{-x^2}\,\mathrm{d}x.$$

As $A \to \infty$ the first summand on the right tends to 0; thus $I(y)$ satisfies the ordinary differential equation

$$I'(y) = -2yI(y).$$

ODE: $y' = -2xy$; $\mathrm{d}y = -2xy\,\mathrm{d}x$; $\mathrm{d}y/y = -2x\,\mathrm{d}x$. Integration yields $\log y = -x^2 + c$; $y = c'\mathrm{e}^{-x^2}$.

The general solution is $I(y) = C\mathrm{e}^{-y^2}$. We determine the constant C. Insert $y = 0$. Since $I(0) = \int_0^\infty \mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}/2$, we find

$$I(y) = \frac{\sqrt{\pi}}{2}\mathrm{e}^{-y^2}.$$

(d) The Gamma function is in $\mathrm{C}^\infty(\mathbb{R}_+)$.

Bibliography

[1] Glen E. Bredon, *Topology and geometry*, Graduate Texts in Mathematics, no. 139, Springer-Verlag, New York, 1997.

[2] R. Courant, *Differential and integral calculus I–II*, Wiley Classics Library, John Wiley & Sons, New York etc., 1988.

[3] J. Dieudonné, *Treatise on analysis. Volume I – IX*, Pure and Applied Mathematics, Academic Press, Boston, 1993.

[4] O. Forster, *Analysis 1 – 3 (in German)*, Vieweg Studium: Grundkurs Mathematik, Vieweg, Braunschweig, 2001.

[5] K. Königsberger, *Analysis 1 (English)*, Springer-Verlag, Berlin, Heidelberg, New York, 1990.

[6] S. Lang, *Undergraduate Analysis*, second ed., Undergraduate texts in mathematics, Springer, New York-Heidelberg, 1989.

[7] J. Marsden and A. Weinstein, *Calculus. I, II, III*, Undergraduate Texts in Mathematics, Springer-Verlag, New York etc., 1985.

[8] W. Rudin, *Principles of mathematical analysis*, third ed., International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976.

[9] M. Spivak, *Calculus on manifolds*, W. A. Benjamin, New York, Amsterdam, 1965.

[10] ———, *Calculus*, Publish or Perish, Inc., Berkeley, California, 1980.