MATEMATICKÝ ÚSTAV SLEZSKÉ UNIVERZITY V OPAVĚ

Učební texty Matematického ústavu

UT 5/2003

Vladimir Averbuch, Michal Málek

Matematická analýza 3-4

Učební texty Matematického ústavu Slezské univerzity v Opavě, Text UT 5/2003

Autor:	Prof. Vladimir Averbuch, DrSc., RNDr. Michal Málek, PhD.
Název:	Matematická analýza 3,4
Určeno:	Učební text pro studenty Slezské univerzity
Vydavatel:	Matematický ústav Slezské univerzity v Opavě
Technický redaktor:	RNDr. Michal Málek, PhD.
Místo a rok vydání:	Opava, 2003

Za odbornou i jazykovou úroveň odpovídá autor.

Předmluva

Tento text vznikl na základě poznámek ke kurzu přednášek z matematické analýzy 3 a 4, které vedl na Slezské univerzitě první autor, a poznámek k cvičení z téhož kurzu, které vedl druhý autor.

Teoretická část je psána anglicky (přednášky probíhaly v tomto jazyce) a cvičení v češtině.

Autoři by chtěli poděkovat Michalovi Miklášovi za skvělou práci při počítačovém zpracování obrázků.

Jako každá publikace i tento text obsahuje tiskové a jiné chyby, ačkoliv bylo věnováno velké úsilí se jich zbavit. Autoři budou velice vděčni za libovolné korektury, které je možno jim zaslat na emailové adresy:

Vladimir.Averbuch@math.slu.cz a Michal.Malek@math.slu.cz

Opava, září 2003 V. Averbuch a M. Málek

Contents

Introduction

Analysis 3-4 is a generalization of usual, one-dimensional analysis to the case of arbirary finite dimensions.

More precisely, instead of functions $\mathbb{R} \to \mathbb{R}$ we consider functions $\mathbb{R}^n \to \mathbb{R}^m$.

In fact elements of finite-dimensional analysis were developed already in 18$^{\text{th}}$ century, but an appropriate frame for this analysis are normed spaces. These spaces were introduced (independently and almost simultaneously (1916–1922)) by A. Bennett, F. Riesz, H. Hahn, S. Banach, N. Wiener. The differentiation operator for mappings between normed spaces was defined in 1925 by M. Fréchet. (More "weak" notion of differentiability was defined early (1913) by R. Gâteaux, but for this kind of differentiability the chain rule is not valid.) This date can be consider as the birthdate of modern analysis.

* * *

Our course contains 12 Chapters. In Chapter 1 we study normed spaces (up to Chapter 5 these spaces are allowed to be infinite-dimensional). In Chapter 2 we consider Fréchet (and Gâteaux) derivative. Chapter 3 is devoted to the most important theorem of analysis, Inverse Function Theorem (which can be equivalently reformulated as Implicit Function Theorem). As a tool for proving this theorem we prove at first so called Contraction Lemma (this is the main tool also in the final Chapter 12). In Chapter 4 we study higher derivatives, up to Taylor formula. In Chapter 5 we give some applications of the theory to *optimization problems*.

Starting from Chapter 6 we restrict ourselfs just by FINITE-dimensional case.

In Chapter 6 we construct Riemann integral in $\mathbb{R}^n$. Chapter 7 is devoted to two important technical results. In Chapter 8 we consider differential forms, which are in fact generalizations of "length element", "area element" and "volume element" of classical "old" analysis. Chapters 9 and 10 are devoted to the crown theorem of the theory, Stokes Theorem, which is a generalization of different known results of "old" analysis (Euler (1771), Green (1828), Ostrogradskij (1834), Stokes (1854)). For this end we define *manifolds* in $\mathbb{R}^n$.

The last two chapters are facultative and written in more compressed style. In Chapter 11 we apply Stokes Theorem for study of analytic complex function, and in Chapter 12 we apply Contraction Lemma for proving of Existence and Uniqueness Theorem for ordinary differential equations.

* * *

Some remarks on notations.

If we write, e.g., $a \overset{A}{<} b$, this means "using A we conclude that $a < b$".

Symbols $\lhd$ and $\rhd$ denote, resp., the beginning and the end of the proof. If we prove some "small" assertion inside the proof of a "great" one, we use symbols $\lhd\!\lhd$ and $\rhd\!\rhd$ for this "small" proof, et-cetera.

"Exerc." over e.g. an equation mark means that to prove this equation is an exercize for the reader.

The reader has to remember that misprints are POSSIBLE and to use ever his common sense.

Chapter 1

Normed spaces

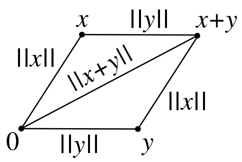
1.1 Norms

Let X be a vector space over $\mathbb{R}$. By a *norm* on (or in) X we mean a function $\|\cdot\| : X \to \mathbb{R}$ with the following properties:

(i) $\forall x \in X\colon \|x\| \geq 0$ (positivity); $\|x\| = 0 \Leftrightarrow x = 0$ (non-degeneracy);

(ii) $\forall x \in X\ \forall t \in \mathbb{R}\colon \|tx\| = |t|\,\|x\|$ (positive homogenity); in particular $\|-x\| = \|x\|$ (symmetry);

(iii) $\forall x, y \in X\colon \|x + y\| \leq \|x\| + \|y\|$ (subadditivity).



If we interpret $\|x\|$ as the LENGTH of the vector x then the property (iii) expresses the *triangle inequality* ($\triangle$-in.).

A *normed space* X is a vector space equipped with a norm. ($X \in$ NS)

Examples.

1. $(\mathbb{R}, |\cdot|)$;

2. $(\mathbb{R}^n, \|\cdot\|_p)$, where $\|\cdot\|_p$ is defined for $1 \leq p < \infty$ by the formula

$$\|x\|_p := \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p} \qquad (x = (x_1, \ldots, x_n)).$$

For $p = 2$ we obtain the usual Euclidean length.

3. $(\mathbb{R}^n, \|\cdot\|_\infty)$, where

$$\|x\|_\infty := \max\{|x_1|, \cdots, |x_n|\}.$$

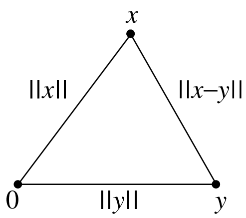
NB $\|x\|_\infty = \lim_{p\to\infty} \|x\|_p$.

4. ℓ_2. This is the set of all sequences $x = (x_1, x_2, \ldots)$ of real numbers such that $\sum_{i=1}^\infty x_i^2 < \infty$, with the norm defined so:

$$\|x\|^2 := \sum_{i=1}^\infty x_i^2.$$

5. C([0, 1]). This is the set of all continuous real-valued functions on [0, 1] equipped with the norm

$$\|x\| := \max_{t\in[0,1]} |x(t)| .$$



Second Triangle Inequality. Let $\|\cdot\|$ be a norm in X. Then

$$\forall x, y \in X\,\vdots \quad \big|\|x\| - \|y\|\big| \le \|x - y\| .$$

$\lhd$ Without loss of generality we can assume that $\|x\| > \|y\|$ (since $\|x - y\| = \|y - x\|$). We need to verify that $\|x\| - \|y\| \le \|x - y\|$. But indeed

$$\|x\| \overset{\text{trick}}{=} \|x - y + y\| \overset{\triangle\text{-in.}}{\le} \|x - y\| + \|y\| . \rhd$$

1.2 Balls

Let X be a normed space. Put for $x \in X, r > 0$

$\mathrm{B}_r(x) := \{y \in X \mid \|y - x\| \le r\}$ (the *closed ball* with the center x and radius r);
$\overset{\circ}{\mathrm{B}}_r(x) := \{y \in X \mid \|y - x\| < r\}$ (the *open ball* with the center x and radius r);

For balls with center at 0 we write for short

$$\mathrm{B}_r := \mathrm{B}_r(0), \qquad \overset{\circ}{\mathrm{B}}_r := \overset{\circ}{\mathrm{B}}_r(0).$$

Properties of balls. It is easy to verify (please!) that[1]

1) $\mathrm{B}_r(x) = x + \mathrm{B}_r$; $\overset{\circ}{\mathrm{B}}_r(x) = x + \overset{\circ}{\mathrm{B}}_r$;
2) $\mathrm{B}_r = r\,\mathrm{B}_1$; $\overset{\circ}{\mathrm{B}}_r = r\overset{\circ}{\mathrm{B}}_1$;
3) $\overset{\circ}{\mathrm{B}}_r \subsetneqq \mathrm{B}_r$;
4) if $r_1 < r_2$, then $\mathrm{B}_{r_1} \subsetneqq \overset{\circ}{\mathrm{B}}_{r_2}$;
5) $\overset{\circ}{\mathrm{B}}_r = \bigcup\limits_{(0<)\alpha<r} \overset{\circ}{\mathrm{B}}_\alpha = \bigcup\limits_{\alpha<r} \mathrm{B}_\alpha$;
6) $\mathrm{B}_r = \bigcap\limits_{\alpha>r} \mathrm{B}_\alpha = \bigcap\limits_{\alpha>r} \overset{\circ}{\mathrm{B}}_\alpha$;

NB Here and below we use the following notations:

$$A + B := \{a + b \mid a \in A,\ b \in B\} \quad (A,\ B \subset X),$$
$$TA := \{ta \mid t \in T,\ a \in A\} \quad (T \subset \mathbb{R},\ A \subset X).$$

In particular

$$x + A := \{x\} + A = \{x + a \mid a \in A\} \quad (x \in X),$$
$$tA := \{t\}A = \{ta \mid a \in A\} \quad (t \in \mathbb{R}).$$

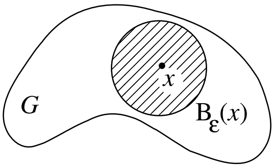
Notation. For balls in $\mathbb{R}$ we write I ("interval") instead of B. For example

$$I_1 = [-1, 1].$$

[1] In 3) and 4) we suppose that our normed space is non-trivial: $X \neq \{0\}$.

1.3 Norm topology

Let X be a normed space. We define the topology τ *generated by the norm* so: a set G is open if for each point $x \in G$ there exists a ball $\mathrm{B}_\varepsilon(x)$ that is contained in G:



$$G \in \tau :\Leftrightarrow \forall x \in G\ \exists \varepsilon > 0 : \mathrm{B}_\varepsilon(x) \subset G.$$

The first assertion of the following theorem means that this definition is correct.

Theorem 1.3.1.

a) *So defined τ is a topology.*

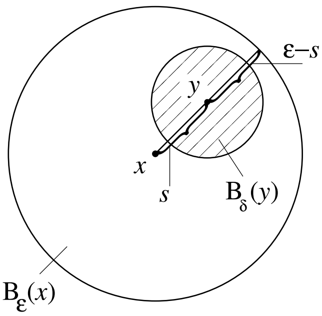
b) *Open balls in X are open sets in this topology, and closed balls are closed sets.*

c) *Both the open balls and the closed balls with the center at x are bases of neighbourhoods of x in this topology.*

$\lhd$ a) $G_\alpha \in \tau \Rightarrow \bigcup_\alpha G_\alpha \in \tau$. Indeed if $x \in \bigcup G_\alpha$ then $x \in G_{\alpha_0}$ for some α_0, hence $\mathrm{B}_\varepsilon(x) \subset G_{\alpha_0}$ for some $\varepsilon > 0$; a fortiori $\mathrm{B}_\varepsilon(x) \subset \bigcup G_\alpha$.

Further, $G_1, G_2 \in \tau \Rightarrow G_1 \cap G_2 \in \tau$. Indeed, if $x \in G_1 \cap G_2$, then $x \in G_1$ and hence $\mathrm{B}_{\varepsilon_1} \subset G_1$ for some $\varepsilon_1 > 0$. Analogously $\mathrm{B}_{\varepsilon_2} \subset G_2$ for some $\varepsilon_2 > 0$. Put $\varepsilon := \min(\varepsilon_1, \varepsilon_2)$. Then $\mathrm{B}_\varepsilon(x) \subset G_1 \cap G_2$.

Thus τ is a topology.



b) Let us prove that $\overset{\circ}{\mathrm{B}}_\varepsilon(x) \in \tau$. Let $y \in \overset{\circ}{\mathrm{B}}_\varepsilon(x)$. Then

$$s := \|y - x\| < \varepsilon.$$

Take any $\delta > 0$ such that

$$\delta < \varepsilon - s. \tag{1}$$

Then

$$z \in \mathrm{B}_\delta(y) \Rightarrow \|z - y\| \le \delta \Rightarrow \|z - x\| \overset{\triangle\text{-in.}}{\le} \underbrace{\|z - y\|}_{\le \delta} + \underbrace{\|y - x\|}_{= s} \le \delta + s \overset{(1)}{<} \varepsilon \ \Rightarrow z \in \overset{\circ}{\mathrm{B}}_\varepsilon(x),$$

which means that $\mathrm{B}_\delta(y) \subset \overset{\circ}{\mathrm{B}}_\varepsilon(x)$. Thus, $\overset{\circ}{\mathrm{B}}_\varepsilon(x) \in \tau$.

That $(\mathrm{B}_\varepsilon(x))^c \in \tau$ can be proved analogously.

c) If U is an open neighbourhood of x in τ, then (by our definition of τ), $\mathrm{B}_\varepsilon(x) \subset U$ for some $\varepsilon > 0$; a fortiori $\overset{\circ}{\mathrm{B}}_\varepsilon(x) \subset U$. But $\overset{\circ}{\mathrm{B}}_\varepsilon(x)$ is an open set (by b)) and contains x (obviously), so $\overset{\circ}{\mathrm{B}}_\varepsilon(x)$ is an open neighbourhood of x, therefore $\mathrm{B}_\varepsilon(x)$ is a (closed by b)) neighbourhood of x. All is proved. $\rhd$

By a *norm topology* we mean the topology generated by a norm.

NB Any norm topology is Hausdorff. (Prove!)

Convergence and continuity

Convergence in a normed space X means convergence in the topology generated by the norm. It follows from the definitions that

$$x_n \xrightarrow[n\to\infty]{} x \in X \Leftrightarrow \|x_n - x\| \xrightarrow[n\to\infty]{} 0.$$

Continuity of a mapping $f : X \to Y$ (where X, Y are normed spaces) means continuity in topologies generated by the norms in X and Y. It follows from the definitions that f is continuous at a point $\hat{x} \in X$ if and only if (iff)

$$\forall \varepsilon > 0 \, \exists \delta > 0 : \quad \|x - \hat{x}\| \le \delta \Rightarrow \|f(x) - f(\hat{x})\| \le \varepsilon \tag{2}$$

(just as in usual analysis, only with $\|\cdot\|$ instead of $|\cdot|$).

For short we write (2) in the form

$$\|x - \hat{x}\| \to 0 \Rightarrow \|f(x) - f(\hat{x})\| \to 0,$$

or

$$\|f(x) - f(\hat{x})\| \xrightarrow[\|x-\hat{x}\|\to 0]{} 0.$$

Theorem 1.3.2. *Let X be a normed space. Then the norm $\|\cdot\| : X \to \mathbb{R}$ is a continuous function*

$\lhd$ Continuity of $\|\cdot\|$ at a point $\hat{x}$ means that

$$\big|\|x\| - \|\hat{x}\|\big| \xrightarrow[\|x\|-\|\hat{x}\|\to 0]{} 0.$$

But the latter relation is true, since, by the Second Triangle Inequality,

$$\big|\|x\| - \|\hat{x}\|\big| \le \|x\| - \|\hat{x}\|. \quad \rhd$$

1.4 Equivalent norms

Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms on a vector space X. We say that the norm $\|\cdot\|_1$ is *stronger* than the norm $\|\cdot\|_2$, and write

$$\|\cdot\|_1 \succ \|\cdot\|_2,$$

if the topology τ_1 generated by $\|\cdot\|_1$ is FINER than the topology τ_2, generated by $\|\cdot\|_2$:

$$\|\cdot\|_1 \succ \|\cdot\|_2 :\Leftrightarrow \tau_1 \supset \tau_2.$$

Theorem 1.4.1. *The following conditions are equivalent (TFAE):*

a) $\|\cdot\|_1 \succ \|\cdot\|_2$;

b) $\exists r_1, r_2 > 0 : \quad \overset{\circ}{\mathrm{B}}{}_{r_1}^{\|\cdot\|_1} \subset \overset{\circ}{\mathrm{B}}{}_{r_2}^{\|\cdot\|_2}$;

c) $\exists r_1, r_2 > 0 : \quad \mathrm{B}_{r_1}^{\|\cdot\|_1} \subset \mathrm{B}_{r_2}^{\|\cdot\|_2}$;

d) $\exists r_1, r_2 > 0 : \quad r_2 \|\cdot\|_1 \ge r_1 \|\cdot\|_2$. *(Which means that $\forall x \in X \,\vdots\, r_1 \|x\|_1 \ge r_2 \|x\|_2$.).*

$\lhd$ 1° (a) $\Rightarrow$ (b):

$$\overset{\circ}{\mathrm{B}}{}_{r_1}^{\|\cdot\|_2} \overset{1.3}{\in} \tau_2 \overset{\tau_1 \supset \tau_2}{\Rightarrow} \overset{\circ}{\mathrm{B}}{}_{1}^{\|\cdot\|_2} \in \tau_1 \overset{\text{def. of } \tau_1}{\Rightarrow} \exists \varepsilon > 0 : \overset{\circ}{\mathrm{B}}{}_{1}^{\|\cdot\|_2} \supset \mathrm{B}_{\varepsilon}^{\|\cdot\|_1} \Rightarrow \overset{\circ}{\mathrm{B}}{}_{1}^{\|\cdot\|_2} \supset \overset{\circ}{\mathrm{B}}{}_{\varepsilon}^{\|\cdot\|_1};$$

thus, we can put $r_1 = \varepsilon, \; r_2 = 1$.

2° (b) $\Rightarrow$ (c):

$$\mathrm{B}_1^{\|\cdot\|_1} \overset{1.2}{=} \bigcap_{r > r_1} \overset{\circ}{\mathrm{B}}{}_{r}^{\|\cdot\|_1} \overset{1.2}{=} \bigcap_{\alpha > 1} \alpha \overset{\circ}{\mathrm{B}}{}_{r_1}^{\|\cdot\|_1} \overset{(b)}{\subset} \bigcap_{\alpha > 1} \alpha \overset{\circ}{\mathrm{B}}{}_{r_2}^{\|\cdot\|_2} \overset{1.2}{=} \bigcap_{r > r_2} \overset{\circ}{\mathrm{B}}{}_{r}^{\|\cdot\|_2} \overset{1.2}{=} \mathrm{B}_{r_2}^{\|\cdot\|_2}.$$

3° (c) $\Rightarrow$ (d): Let (c) is true. We need to verify that $\forall x \in X \colon r_2 \|x\|_1 \geq r_1 \|x\|_2$. Without loss of generality $x \neq 0$. We have $\|r_1 x/\|x\|_1\|_1 = r_1 \Rightarrow r_1 x/\|x\|_1 \in \mathrm{B}_{r_1}^{\|\cdot\|_1} \overset{(c)}{\Rightarrow} r_1 x/\|x\|_1 \in \mathrm{B}_{r_2}^{\|\cdot\|_2} \Rightarrow \|r_1 x/\|x\|_1\|_2 \leq r_2 \Rightarrow r_1 \|x\|_2 \geq r_2 \|x\|_1$.
4° (d) $\Rightarrow$ (a): Let (d) is true. We need to verify that $\tau_1 \supset \tau_2$. Let $U \in \tau_2$ and let x be an arbitrary point in U. By the definition of τ_2, for some $\varepsilon > 0$

$$\mathrm{B}_\varepsilon^{\|\cdot\|_2}(x) \subset U. \tag{1}$$

Now, $\|x\|_1 < r_1 \overset{(d)}{\Rightarrow} \|x\|_2 < r_2$, which means that $\mathring{\mathrm{B}}_{r_1}^{\|\cdot\|_1} \subset \mathring{\mathrm{B}}_{r_2}^{\|\cdot\|_2}$. Multiplying by ε/r_2 we obtain $\mathring{\mathrm{B}}_{\varepsilon r_1/r_2}^{\|\cdot\|_1} \subset \mathring{\mathrm{B}}_{\varepsilon}^{\|\cdot\|_2}$. And the translation by x yields (by the property 2 of balls, see 1.2) $\mathring{\mathrm{B}}_{\varepsilon r_1/r_2}^{\|\cdot\|_1}(x) \subset \mathring{\mathrm{B}}_{\varepsilon}^{\|\cdot\|_2}(x) \overset{(1)}{\subset} U$. Thus $U \in \tau_1$. $\rhd$

Example. In ℓ_2

$$\|\cdot\|_2 \succ \|\cdot\|_\infty\,, \quad \|\cdot\|_\infty \not\succ \|\cdot\|_2\,,$$

where

$$\|x\|_\infty := \sup_{i\in\{1,2,\ldots\}} |x_i| \qquad (x = (x_1, x_2, \ldots)).$$

(Prove!)

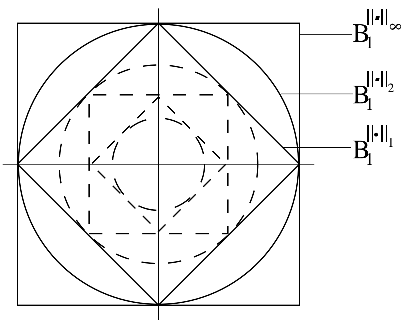
Equivalent norms

We say that two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space X are equivalent and write

$$\|\cdot\|_1 \sim \|\cdot\|_2\,,$$

if each norm is stronger than the other, that is, if they generate one and the same topology:

$$\|\cdot\|_1 \sim \|\cdot\|_2 :\Leftrightarrow (\|\cdot\|_1 \succ \|\cdot\|_2\,,\ \|\cdot\|_2 \succ \|\cdot\|_1) \Leftrightarrow \tau_1 = \tau_2.$$



Theorem 1.4.2. *In* $\mathbb{R}^n$

$$\|\cdot\|_1 \sim \|\cdot\|_2 \sim \|\cdot\|_\infty$$

$\lhd$ 1° $\|x\|_1 \geq \|x\|_2 \geq \|x\|_\infty$ (prove!), hence (by Theorem 1.4.1.) $\|\cdot\|_1 \succ \|\cdot\|_2 \succ \|\cdot\|_\infty$.
2° $n\,\|x\|_\infty \geq \sqrt{n}\,\|x\|_2 \geq \|x\|_1$ (prove!), hence $\|\cdot\|_\infty \succ \|\cdot\|_2 \succ \|\cdot\|_1$. $\rhd$
NB In fact ALL norms in $\mathbb{R}^n$ are equivalent (see 1.9).

1.5 Bounded sets

A set A in a normed space X is called *bounded* if A is contained in the ball B_r for some $r > 0$.
NB A set A is bounded iff the number set $\{\|x\| \mid x \in A\} \subset \mathbb{R}$ is bounded.

Example. Each ball (closed or open), with any center, is bounded. (Prove!)

Theorem 1.5.1. *For equivalent norms the set of all bounded sets is one and the same, that is, if $\|\cdot\|_1 \sim \|\cdot\|_2$, then*

$$A \text{ is bounded in } \|\cdot\|_1 \Leftrightarrow A \text{ is bounded in } \|\cdot\|_2 .$$

$\lhd$ This follows from equivalence (a) $\Leftrightarrow$ (d) in Theorem 1.4.1. $\rhd$
Remark. The theorem suggests that boundedness can be expressed in terms of the TOPOLOGY τ generated by the norm. And indeed A is bounded iff for each neighbourhood U of zero in τ there exists $\delta > 0$ such that $\delta A \subset U$.

1.6 Product

Let $X_1, \ldots, X_n$ be normed spaces. The vector space $X_1 \times \cdots \times X_n$ can be equipped with the norm

$$\|(x_1, \ldots, x_n)\| := \| \underbrace{(\|x_1\|_{X_1}, \ldots, \|x_n\|_{X_n})}_{\in\mathbb{R}^n} \|_p, \tag{1}$$

where $1 \le p \le \infty$, and $\|\cdot\|_p$ is the norm in $\mathbb{R}^n$ defined in 1.1. Just as in Theorem 1.4.2., it can be verified that for $p = 1, 2$ and ∞ we obtain equivalent norms. (In fact, the norms (1) are equivalent for ALL $p \in [1, \infty]$, since all norms IN $\mathbb{R}^n$ are equivalent, see 1.9.)
Remark. The norm $\|\cdot\|_p$ in $\mathbb{R}^n$ is a special case of this construction ($\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$).
NB The topology generated by the norms (1) coincides with the product topology in $X_1 \times \cdots \times X_n$, each X_i being supplied with the topology generated by the norm.
Criterion. Let $X_1, \ldots, X_n$ be normed spaces. Then

$$(x_1, \ldots, x_n) \to (\hat{x}_1, \ldots, \hat{x}_n) \in X_1 \times \cdots \times X_n \Leftrightarrow \forall i : \|x_i - \hat{x}_i\| \to 0.$$

$\lhd$ This follows at once from the definitions. $\rhd$

Theorem 1.6.1. *For each normed space the algebraical operations*

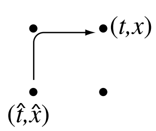
$$\cdot : X \times \mathbb{R} \to X,\ (x, t) \mapsto tx \quad \textit{(multiplication)}$$

and

$$+ : X \times X \to X,\ (x, y) \mapsto x + y \quad \textit{(addition)}$$

are continuous.
$\lhd$ 1°

$$0 \le \|tx - \hat{t}\hat{x}\| \overset{\triangle\text{-in}}{\le} \|tx - t\hat{x}\| + \|t\hat{x} - \hat{t}\hat{x}\| = \underbrace{|t|}_{\xrightarrow[t\to\hat{t}]{}|\hat{t}|} \|x - \hat{x}\| + |t - \hat{t}|\,\|\hat{x}\| \xrightarrow[\substack{\|x-\hat{x}\|\to 0\\ |t-\hat{t}|\to 0}]{} 0,$$



hence $\|tx - \hat{t}\hat{x}\| \to 0$ as $(x, t) \mapsto (\hat{x}, \hat{t})$. Thus, the multiplication is continuous.
2°

$$0 \le \|(x + y) - (\hat{x} + \hat{y})\| \overset{\triangle\text{-in.}}{\le} \|x - \hat{x}\| + \|y - \hat{y}\| \xrightarrow[\substack{\|x-\hat{x}\|\to 0\\ \|y-\hat{y}\|\to 0}]{} 0,$$

hence $\|(x + y) - (\hat{x} + \hat{y})\| \to 0$ as $(x, y) \to (\hat{x}, \hat{y})$. Thus, the addition is continuous. $\rhd$

1.7 Natural topology in $\mathbb{R}^n$

The topology generated by the equivalent norms $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ in $\mathbb{R}^n$ (see 1.4) is called the *natural* topology.
NB We ever consider $\mathbb{R}^n$ with the natural topology.

Theorem 1.7.1. *The natural topology τ_{nat} in $\mathbb{R}^n$ coincides with the product topology τ_{prod} (that is, the topology of the product $\mathbb{R} \times \cdots \times \mathbb{R}$ (n times), where $\mathbb{R}$ is equipped with its standard topology).*

In particular the natural topology in $\mathbb{R}$ is its standard topology.
$\lhd$ For short consider the case $n = 2$.
0° Since the neighbourhoods of x are translations by x of the neighbourhoods of 0 (by Property 1 of balls, see 1.2), it is sufficient to verify that each neighbourhood of 0 in τ_{nat} contains a neighbourhood of 0 in τ_{prod} and vice versa. Below the notation $U \in \mathrm{Nb}_x$ means that U is a neighbourhood of x.
1° Let $U \in \mathrm{Nb}_0(\tau_{\mathrm{prod}})$. Then by the definition of the product topology there exist $\varepsilon_1 > 0$, $\varepsilon_2 > 0$ such that

$$U \supset I_{\varepsilon_1} \times I_{\varepsilon_2} \underset{\varepsilon:=\min(\varepsilon_1,\varepsilon_2)}{\supset} \underbrace{I_\varepsilon \times I_\varepsilon}_{=\{(x,y)\,|\,|x|\le\varepsilon,\ |y|\le\varepsilon\}} = \mathrm{B}_\varepsilon^{\|\cdot\|_\infty} \in \mathrm{Nb}(\tau_{\mathrm{nat}}). \quad \text{O.K.}$$

2° Let $U \in \mathrm{Nb}_0(\tau_{\mathrm{nat}})$. Then (since τ_{nat} is generated by $\|\cdot\|_\infty$) there exists $\varepsilon > 0$ such that $U \supset \mathrm{B}_\varepsilon^{\|\cdot\|_\infty} = I_\varepsilon \times I_\varepsilon \in \mathrm{Nb}_0(\tau_{\mathrm{prod}})$. O.K. $\rhd$

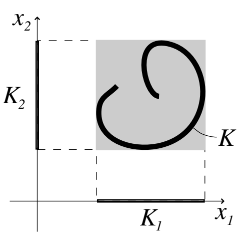
1.8 Bounded sets and compact sets in $\mathbb{R}^n$

A set $A \subset \mathbb{R}^n$ is called *bounded* if it is bounded in one of the norms $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ (then, by Theorem 1.5.1., it it bounded also in the two others; in essence it is boundedness with respect to the natural topology, see Remark in 1.5).

Theorem 1.8.1. *A set $K \subset \mathbb{R}^n$ is compact (in the natural topology) iff it is bounded and closed.*
$\lhd$ For simplicity of notations consider the case $n = 2$.
0° We need the following important theorem of general topology:

> **Tichonov Theorem.** *The product $\prod_i X_i$ of (arbitrary many) topologicalspa-ces (equipped with the product topology) is a compact space iff each X_i is a compact space.*

1° Let K be compact (in $\mathbb{R}^2$). Then K is closed (as a compact set in a Hausdorff topological space). Now the projections K_1 and K_2 of K onto the axes x_1 and x_2 are compact (in $\mathbb{R}$) as the images of a compact set by continuous mappings. Hence, as is known from one-dimensional analysis, K_1 and K_2 are bounded. So there exists $a > 0$ such that $K_i \subset I_a$, $i = 1, 2, \ldots$ whence it follows that



$$K_1 \times K_2 \subset I_a \times I_a = \mathrm{B}_a^{\|\cdot\|_\infty}.$$

Thus $K_1 \times K_2$ is bounded. A fortiori $K \subset K_1 \times K_2$ is bounded.
2° Vice versa, let K be a closed bounded set in $\mathbb{R}^2$. Then $K \subset \mathrm{B}_a^{\|\cdot\|_\infty}$ for some $a > 0$. But $\mathrm{B}_a^{\|\cdot\|_\infty} = I_a \times I_a$ is compact by Tichonov Theorem, hence K is compact as a closed subset of a compact set. $\rhd$

1.9 Uniqueness of the norm topology in $\mathbb{R}^n$

Up to equivalence there exists just one norm in $\mathbb{R}^n$ — all norms in $\mathbb{R}^n$generate one and the same topology:

Theorem 1.9.1. *All the norms in $\mathbb{R}^n$ are equivalent.*
$\lhd$ Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$. Show that $\|\cdot\| \sim \|\cdot\|_1$.
1° $\|\cdot\|_1 \succ \|\cdot\|$: Consider the canonical basis $\{\mathrm{e}_1, \dots, \mathrm{e}_n\}$ of $\mathbb{R}^n$, $(\mathrm{e}_i = (0, \dots, 0, \underset{i}{1}, 0, \dots, 0))$. For any point $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ it holds

$$\begin{aligned} \|x\| = \|x_1\mathrm{e}_1 + \cdots + x_n\mathrm{e}_n\| &\overset{\triangle\text{-in.}}{\leq} \|x_1\mathrm{e}_1\| + \cdots + \|x_n\mathrm{e}_n\| \\ &= |x_1|\,\|\mathrm{e}_1\| + \cdots + |x_n|\,\|\mathrm{e}_n\| \\ &\underset{M:=\max\{\|\mathrm{e}_1\|,\cdots,\|\mathrm{e}_n\|\}}{\leq} M(|x_1| + \cdots + |x_n|) = M\,\|x\|_1\,. \end{aligned}$$

Hence, by Theorem 1.4.1., $\|\cdot\|_1 \succ \|\cdot\|$.
2° $\|\cdot\| \succ \|\cdot\|_1$: Consider the unit sphere S in the norm $\|\cdot\|_1$:

$$S := \{x \in \mathbb{R}^n \mid \|x\|_1 = 1\}.$$

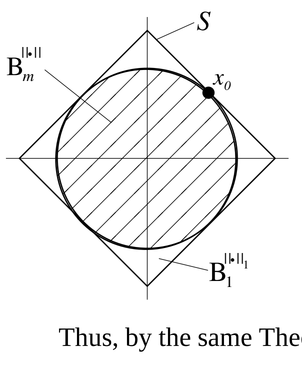
This set is compact (in the natural topology). Indeed, S is obviously bounded, and S is closed as the pre-image of the closed set $\{1\} \subset \mathbb{R}$ by the continuous mapping $\|\cdot\|_1$ (see Theorem 1.3.2.).

Now we claim that $\|\cdot\|$ is a continuous function on $\mathbb{R}^n$ (with the natural topology). Indeed, $\|\cdot\|$ is continuous with respect to the topology τ generated by $\|\cdot\|$ (once again by Theorem 1.3.2.) and τ_{nat} is FINER than τ by 1°.

We conclude that $\|\cdot\|$ attains its MINIMAL value m on S, that is,

$$\begin{aligned} &\|x\|_1 = 1 \Rightarrow \|x\| \geq m, \\ &\|x_0\| = m \quad \text{for some } x_0 \text{ with } \|x_0\|_1 = 1. \end{aligned} \tag{1}$$

This value m must be greater than 0, since otherwise $x_0 = 0$ and $\|x_0\|_1 = 0$. It follows from (1) that



$$\|x\|_1 > 1 \Rightarrow \|x\| = \underbrace{\|x\|_1}_{>1} \underbrace{\left\| \frac{x}{\|x\|_1} \right\|}_{\geq m} > m.$$

Hence

$$\|x\| \leq m \Rightarrow \|x\|_1 \leq 1,$$

that is,

$$B_m^{\|\cdot\|} \subset B_1^{\|\cdot\|_1}\,.$$

Thus, by the same Theorem 1.4.1., $\|\cdot\| \succ \|\cdot\|_1$. $\rhd$

1.10 Linear mappings

For a linear mapping l we usually write lx or $l \cdot x$ instead of $l(x)$:

$$lx \equiv l \cdot x \equiv l(x).$$

The set of all linear mappings from a vector space X into a vector space Y is a vector space with respect to operations

$$(l+m)x := lx + mx, \quad (tl)x := t(lx) \qquad (t \in \mathbb{R}),$$

and we denote this vector space by

$$\mathrm{L}(X,Y).$$

The vector subspace of all CONTINUOUS linear mappings, in the case where X and Y are normed spaces, we denote by

$$\mathscr{L}(X,Y).$$

Theorem 1.10.1. *Let X and Y be normed spaces and let $l \in \mathrm{L}(X,Y)$. Then l is continuous iff l is continuous at* 0.
◁ "Only if": obvious.

"If": Let l be continuous at 0, that is, $\|h\| \to 0 \Rightarrow \|lh\| \to 0$. Then for an arbitrary $x \in X$ it holds

$$\|\underbrace{l(x+h)}_{=lx+lh} - lx\| = \|lh\| \xrightarrow[\|h\|\to 0]{} 0,$$

which means that l is continuous at x. ▷

Operator norm

Let X, Y be normed spaces. We define the *norm of a mapping* $l \in \mathrm{L}(X,Y)$ as

$$\|l\| := \sup_{\|x\|\le 1} \|lx\|.$$

Very often one says "operators" for linear mappings, that is why this norm is usually named *operator norm*. (Below we will see that this is really a norm.)

Example. For any $k \in \mathbb{R}$ the linear mapping $\mathbb{R} \to \mathbb{R},\ x \mapsto kx$ has the norm $|k|$.

Basic Inequality (BI). $\forall l \in \mathrm{L}(X,Y)\ \forall x \in X \vdots\ \boxed{\|lx\| \le \|l\|\,\|x\|}$.
◁ If $x = 0$ then our inequality is trivially true. If $x \neq 0$ then

$$\|lx\| = \left\| l\left(\|x\|\,\frac{x}{\|x\|}\right)\right\| = \left\|\|x\|\, l\frac{x}{\|x\|}\right\| = \|x\| \underbrace{\left\| l\frac{x}{\|x\|}\right\|}_{\underset{\|x/\|x\|\|=1}{\le} \|l\|} \le \|l\|\,\|x\|. \ ▷$$

Criteria of continuity. *Let $l \in \mathrm{L}(X,Y)$. The following conditions are equivalent:*

a) *l is continuous;*
b) *the image $l\,\mathrm{B}_1$ of the unit ball in X is bounded in Y;*
c) $\exists k > 0\ \forall x \in X \vdots\ \|lx\| \le k\,\|x\|$ *(the norm of lx admits an estimation linear in $\|x\|$);*
d) $\|l\| < \infty$ *(the operator norm is finite).*

◁ (a)⇒(b): Since l is continuous at 0, there exists $\delta > 0 : l\,\mathrm{B}^X_\delta \subset \mathrm{B}^Y_1$. Multiplying by δ^{-1} we obtain (by linearity of l) $l\,\mathrm{B}^X_1 \subset \mathrm{B}^Y_{\delta^{-1}}$, which just means that the image $l\,\mathrm{B}^X_1$ is bounded.

(b)⇒(c): If $l\,\mathrm{B}_1^X \subset \mathrm{B}_k^Y$, then (without loss of generality $x \neq 0$)

$$\|lx\| \overset{\substack{\text{wlog}\\ x\neq 0}}{=} \|\, l \underbrace{\overbrace{\frac{x}{\|x\|}}^{\in \mathrm{B}_k^Y}}_{\in \mathrm{B}_1^X} \|\, \|x\| \le k\, \|x\|\,.$$

(c)⇒(d): If $\|lx\| \le k\,\|x\|$ for all x, then $\sup_{\|x\|\le 1} \underbrace{\|lx\|}_{\le k\,\underbrace{\|x\|}_{\le 1}} \le k$, that is, $\|l\| \le k$.

(d)⇒(a): If $\|l\| < \infty$ then $0 \le \|lx\| \overset{\text{BI}}{\le} \|l\|\,\|x\| \xrightarrow[\|x\|\to 0]{} 0$, hence $\|lx\| \xrightarrow[\|x\|\to 0]{} 0$, that is, l is continuous at 0. But then, by Theorem 1.10.1., l is continuous everywhere. $\rhd$

Remark. $\|l\| = \inf\{k > 0 \mid \forall x \in X \colon\ \|lx\| \le k\,\|x\|\}$.

Theorem 1.10.2. *The mapping $\mathscr{L}(X,Y) \to \mathbb{R},\ l \mapsto \|l\|$ is a norm.*
$\lhd$ That $\|l\| \ge 0$ is obvious. If $\|l\| = 0$ then $\|lx\| = 0$ for all x with $\|x\| \le 1$ and hence, by linearity of l, for all x, which means that $l = 0$. Further

$$\|tl\| = \sup_{\|x\|\le 1} \|(tl)x\| = \sup_{\|x\|\le 1} \|t(lx)\| = |t| \sup_{\|x\|\le 1} \|lx\| = |t|\,\|l\|\,.$$

At last

$$\|l_1 + l_2\| = \sup_{\|x\|\le 1} \|(l_1 + l_2)x\| = \sup_{\|x\|\le 1} \|l_1 x + l_2 x\| \le \sup_{\|x\|\le 1} \underbrace{\|l_1 x\|}_{\le \|l_1\|\,\underbrace{\|x\|}_{\le 1}} + \underbrace{\|l_2 x\|}_{\le \|l_2\|\,\underbrace{\|x\|}_{\le 1}}$$
$$\le \|l_1\| + \|l_2\|\,.\ \rhd$$

NB We EVER consider $\mathscr{L}(X,Y)$ as a normed space with this norm.

The case $X = \mathbb{R}^n$

Theorem 1.10.3. *Any linear mapping from $\mathbb{R}^n$ (with the natural topology) into a normed space Y is continuous:*

$$\mathrm{L}(\mathbb{R}^n, Y) = \mathscr{L}(\mathbb{R}^n, Y).$$

$\lhd$ Let $l \in \mathrm{L}(\mathbb{R}^n, Y)$. Each element $x = (x_1, \dots, x_n)$ of $\mathbb{R}^n$ can be written as $x_1 \mathrm{e}_1 + \dots + x_n \mathrm{e}_n$, where $\{\mathrm{e}_1, \dots, \mathrm{e}_n\}$ is the canonical basis, so if we put $l\mathrm{e}_i =: a_i$, then, by linearity of l,

$$lx = x_1 a_1 + \dots + x_n a_n. \tag{1}$$

In view of Theorem 1.10.1. it is sufficient to verify that l is continuous at 0, that is, that $x \to 0 \Rightarrow lx \to 0$. But $x \to 0$ means that all $x_i \to 0$ (since the natural topology coincides with the PRODUCT topology), whence it follows (by continuity of algebraical operations in a normed space, see Theorem 1.6.1.) that

$$x_1 a_1 + \dots + x_n a_n \to 0.$$

Thus, by (1), $lx \to 0$.

(ANOTHER PROOF:
$\|lx\| = \|l(x_1\mathrm{e}_1 + \cdots)\| \le |x_1|\|\underbrace{l\mathrm{e}_1}_{a_i}\| + \cdots \le \underbrace{\max\|a_i\|}_{=:k}\underbrace{(|x_1| + \cdots)}_{=\|x\|_1} \le k\,\|x\|_1$,
so l is continuous by Criterium (c) of continuity.) $\triangleright$

Evaluation at a point

Let X, Y be vector spaces, and let h be a FIXED element of X. The *evaluation at* h (or *delta-function* at h) is the mapping

$$\mathrm{ev}_h \equiv \delta_h : \mathrm{L}(X, Y) \to Y,\ l \mapsto lh.$$

This mapping is (obviously) LINEAR.

Theorem 1.10.4. *Let X, Y be normed spaces. Then for each $h \in X$ the evaluation at h is* CONTINUOUS*:*

$$\mathrm{ev}_h \in \mathscr{L}(\mathscr{L}(X, Y), Y).$$

$\triangleleft$ $\|\mathrm{ev}_h\| = \sup\limits_{\|l\|\le 1} \underbrace{\|\overbrace{\mathrm{ev}_h\, l}^{lh}\|}_{\overset{\mathrm{BI}}{\le} \underbrace{\|l\|}_{\le 1}\|h\|} \le \|h\|$, hence by Criteria (d) of continuity, ev_h is continuous. $\triangleright$

"Lemma" from Functional Analysis

Mappings into $\mathbb{R}$ are usually called *functionals* in the case where the "first" space is infinite-dimensional. (The name "Functional Analysis" originates from this word.) The vector space of all linear (resp., continuous linear) functionals on a given vector space (resp., normed space) X we shall denote by X' (resp., X^*):

$$X' := \mathrm{L}(X, \mathbb{R}), \qquad X^* := \mathscr{L}(X, \mathbb{R}).$$

Later we at least two times shall use the following:

Theorem 1.10.5. ("Lemma" from Functional Analysis) *Let X be a normed space. Then for each vector $x \in X$ there exists a functional $l \in X^*$ of the unit norm such that its value at x is just the norm of x:*

$$\|l\| = 1, \qquad lx = \|x\|.$$

$\triangleleft$ We give the proof for $X = \mathbb{R}^n$ only. If $x = 0$ then we can take as l ANY functional of the norm 1. Let $x \ne 0$. Put (below $\|\cdot\|$ denotes $\|\cdot\|_2$)

$$\mathrm{e} := \frac{x}{\|x\|} \qquad \text{and} \qquad ly := \mathrm{e}\cdot y \quad (y \in \mathbb{R}^n)$$

(where the latter point means scalar product). It is clear that $\|\mathrm{e}\| = 1$, so

$$\|l\| = \sup_{\|y\|\le 1} \underbrace{|\mathrm{e}\cdot y|}_{\underset{\substack{\text{prop.}\\ \text{of scal.}\\ \text{prod.}}}{\le} \underbrace{\|\mathrm{e}\|}_{=1}\underbrace{\|y\|}_{\le 1}} \le 1, \qquad |l\mathrm{e}| = |\underbrace{\mathrm{e}\cdot\mathrm{e}}_{=1}| = 1.$$

Since e belongs to the unit ball, over which we take the supremum, we conclude that $\|l\| = 1$.

At last

$$lx = \mathrm{e}\cdot x = \frac{x}{\|x\|}\cdot x = \frac{x\cdot x}{\|x\|} = \frac{\|x\|^2}{\|x\|} = \|x\|. \ \triangleright$$

Chapter 2

First derivative

2.1 Fréchet and Gâteaux derivatives

The classic definition of the derivative

$$f'(x) := \lim_{\substack{h \to 0 \\ (h \neq 0)}} \frac{f(x+h) - f(x)}{h}$$

can be written in the form (below we drop for short "$h \neq 0$")

$$\frac{r(h)}{h} \xrightarrow[h \to 0]{} 0,$$

where

$$r(h) := f(x+h) - f(x) - f'(x)h.$$

So we can reformulate the definition as follows: a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at a point x if there exists a number l $(= f'(x))$ such that f admits the representation

$$\forall h \in X \,\vdots \quad f(x+h) = f(x) + lh + r(h),$$

where r is a mapping $\mathbb{R} \to \mathbb{R}$, that satisfies the conditions $r(0) = 0$ and

$$\frac{r(h)}{h} \xrightarrow[h \to 0]{} 0. \tag{1}$$

Such a mapping we call *small*.

A key point to generalize this definition is the idea that l can be considered as a (continuous) LINEAR MAPPING $\mathbb{R} \to \mathbb{R}$:

$$l : \mathbb{R} \to \mathbb{R}, \quad h \mapsto lh$$

(we identify a number l with the linear function with the (slope) coefficient l). This leads to the following definition:

A mapping $f : X \to Y$ between normed spaces X and Y is *differentiable* (in a given sense) at a point $x \in X$ (notation: $f \in \mathrm{Dif}(x)$) if there exists a continuous linear mapping $l : X \to Y$ such that f admits the representation

$$\forall h \in X \,\vdots \quad f(x+h) = f(x) + lh + r(h), \tag{2}$$

where r is a mapping $X \to Y$, that is SMALL (in this sense). There are two basic kinds of smallness for mappings between normed spaces:

A mapping $r : X \to Y$ is *Fréchet-small* (*F-small*) if $r(0) = 0$ and

$$\frac{\|r(h)\|}{\|h\|} \xrightarrow[\|h\| \to 0]{} 0; \tag{3}$$

r is *Gâteaux-small* (*G-small*) if $r(0) = 0$ and

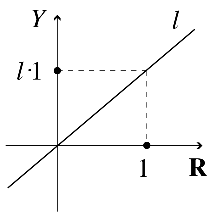
$$\forall h \in X \,\vdots\quad \frac{r(th)}{t} \xrightarrow[t \to 0]{} 0 \qquad (t \in \mathbb{R}). \tag{4}$$

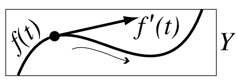
NB For $X = Y = \mathbb{R}$ both (3) and (4) are equivalent to (1) (verify!).

Accordingly we speak about *F-differentiability* and *G-differentiability*. Very often we drop the symbol "F", so "differentiability" means ever "Fréchet differentiability" and "$f \in \mathrm{Dif}(x)$" means "$f \in F\text{-}\mathrm{Dif}(x)$".

Remark. Both our differentiabilities do not depend on the choice of EQUIVALENT norms in X and Y. (Verify!)

The mapping l in the representation (2) is called the *derivative* of f at x and it is denoted by $f'(x)$.





Examples.

1. If $X = \mathbb{R}$ ("time") then we can identify a linear mapping $l : \mathbb{R} \to Y$ with the element $l \cdot 1$ of Y and it is easy to see that F-differentiability is equivalent to G-differentiability, and

$$f'(t) \cdot 1 = \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t} \qquad \text{(the limit in } Y).$$

Below we shall denote the last limit by $\dot f(t)$ and call $\dot f(t)$ (which is a vector in Y) the *usual derivative* of f at t (it is the velocity of a point that moves in Y by the "law" f).

2. Any CONSTANT function is differentiable everywhere, with zero derivative.
$\lhd$ If $f \equiv c$, then $f(x + h) = f(x) + 0.h + 0$, and 0 is small (in any reasonable sense!) $\rhd$

3. Any continuous linear mapping l is differentiable, and its derivative at each point is equal to this mapping itself:

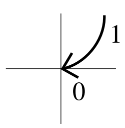
$$l' \equiv l \qquad (\text{that is, } \forall x \in X \,\vdots\, l'(x) = l).$$

$\lhd$ $l(x + h) = lx + lh + 0.$ $\rhd$

4. The function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|x\|_2^2 = x_1^2 + \cdots + x_n^2$ $\quad (x = (x_1, \ldots, x_n))$ is differentiable everywhere, and

$$f'(x) \cdot h = 2x \cdot h,$$

where the point to the right means scalar product. (Prove!)



5. Let the function $f : \mathbb{R}^2 \to \mathbb{R}$ is equal to 1 on the right branch of the parabola $\{(t, t^2)\}$ WITHOUT the origin ($t > 0$) and is equal to 0 at all rest points of the plane. Then f is G-differentiable at $0 (= (0, 0))$, with $f'(0) = 0$ $(\in \mathscr{L}(\mathbb{R}^2, \mathbb{R}))$, but is NOT F-differentiable. (Verify!)

Theorem 2.1.1. *F-differentiability implies G-differentiability, with the same derivative.*
◁ It is sufficient to show that if $r : X \to Y$ is F-small, then r is G-small. Let $r(0) = 0$ and

$$\frac{\|r(h)\|}{\|h\|} \xrightarrow[\|h\|\to 0]{} 0. \tag{5}$$

Then for each fixed $\hat{h} \in X \setminus 0$ (for short we write $X \setminus 0$ instead of $X \setminus \{0\}$)

$$\left\| \frac{r(t\hat{h})}{t} \right\| = \frac{\left\| r(t\hat{h}) \right\|}{\left\| t\hat{h} \right\|} \|\hat{h}\| \xrightarrow[t\to 0]{} 0,$$

since $\|t\hat{h}\| = |t| \left\|\hat{h}\right\| \xrightarrow[t\to 0]{} 0$ and hence, by (5), $\left\|r(t\hat{h})\right\| / \left\|t\hat{h}\right\| \xrightarrow[t\to 0]{} 0$. But this just means that

$$\frac{r(t\hat{h})}{t} \xrightarrow[t\to 0]{} 0.$$

Thus, r is G-small. ▷

Theorem 2.1.2. *G-derivative is defined uniquely.*
◁ We need to verify that if for a given $x \in X$

$$\forall h \in X \vdots \quad f(x) + l_1 h + r_1(h) = f(x) + l_2 h + r_2(h), \tag{6}$$

where $l_1, l_2 \in \mathscr{L}(X, Y)$ and r_1, r_2 are G-small, then $l_1 = l_2$, that is, for each $h \in X$ it holds $l_1 h = l_2 h$. But indeed (for $t \neq 0$)

$$l_1 h - l_2 h \overset{\text{trick}}{=} \frac{l_1(th) - l_2(th)}{t} \overset{(6)}{=} \frac{r_2(th) - r_1(th)}{t} = \frac{r_2(th)}{t} - \frac{r_1(th)}{t} \xrightarrow[t\to 0]{} 0,$$

whence it follows that $l_1 h - l_2 h = 0$. ▷

Corollary 2.1.3. *F-derivative is defined uniquely.*
◁ It follows from Theorems 2.1.1. and 2.1.2. ▷

Theorem 2.1.4. *If f is differentiable at x, then f is continuous at x.*
◁ By the condition, $f(x+h) = f(x) + lh + r(h)$, where $l \in \mathscr{L}(X, Y)$ and r is small. We need to verify that if $h \to 0$ then $lh + r(h) \to 0$. Since l is a continuous linear mapping, $lh \to 0$ as $h \to 0$. Now, for $h \neq 0$

$$\|r(h)\| = \underbrace{\frac{\|r(h)\|}{\|h\|}}_{\to 0} \|h\| \xrightarrow[\|h\|\to 0]{} 0$$

(for $h = 0$ we have $r(0) = 0$), which just means that $r(h) \to 0$ if $h \to 0$. ▷
NB G-differentiability does NOT imply continuity (see Example 5).

2.2 Computation Rule and directional differentiability.

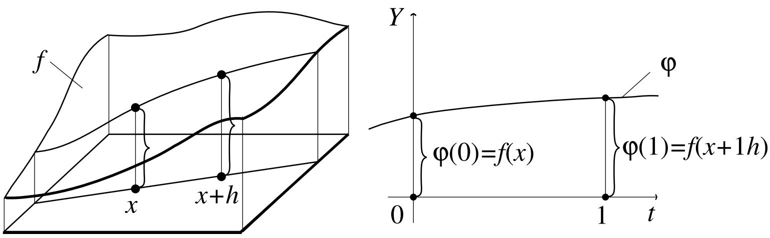
For practical computation of derivatives it is convenient to use the following
Computation Rule. Let $f : X \to Y$ be G-differentiable at a given point x, and let $h \in X$ be given. Put for $t \in \mathbb{R}$

$$\boxed{\varphi(t) := f(x + th)},$$

so that $\varphi : \mathbb{R} \to Y$. Then φ is differentiable at 0, and

$$\boxed{f'(x)h = \dot{\varphi}(0) = \left.\frac{\partial}{\partial t}\right|_{t=0} f(x+th)}.$$

(As to $\dot{\varphi}(0)$ see Example 1).)



$\lhd$ Let $f'(x) = l$. We have $\varphi(t) = f(x+th) = f(x) + l(th) + r(th)$, so (for $t \neq 0$)

$$\frac{\varphi(t)-\varphi(0)}{t} \underset{\varphi(0)=f(x)}{=} \frac{l(th)+r(th)}{t} \underset{l\in \mathrm{Lin}}{=} lh + \underbrace{\frac{r(th)}{t}}_{\to 0} \to lh$$

as $t \to 0$, which does mean that $\dot{\varphi}(0) = lh$. $\rhd$

The Computation Rule suggests the following definition. We say that a mapping $f : X \to Y$ is *differentiable* at a point x *in a direction* h if the function

$$\varphi : \mathbb{R} \to Y, \quad t \mapsto f(x+th)$$

is differentiable at 0. In such a case we call the vector $\dot{\varphi}(0) \in Y$ the *differential* of f at x *by the increment* h, and we denote this differential by $\mathrm{D}_h f(x)$. Thus

$$\boxed{\mathrm{D}_h f(x) := \dot{\varphi}(0) = \left.\frac{\partial}{\partial t}\right|_{0} f(x+th) = \lim_{t\to 0} \frac{f(x+th)-f(x)}{t}}. \tag{1}$$

Corollary 2.2.1. *If $f : X \to Y$ is G-differentiable at x then f is differentiable at x in each direction, and*

$$\forall h \in X\,\vdots \quad \boxed{\mathrm{D}_h f(x) = f'(x)h}.$$

Vice versa, if f is differentiable at x in each direction and the mapping

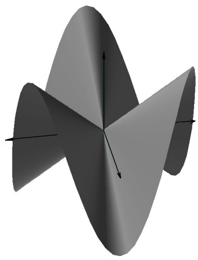
$$l : h \mapsto \mathrm{D}_h f(x), \quad X \to Y$$

is LINEAR *and* CONTINUOUS, *then f is G-differentiable at x, and $f'(x) = l$.*

Remark. The mapping l in Corollary 2.2.1. is ever HOMOGENUOUS. More precisely, if f is differentiable at x in a direction h, then for any real number c it is differentiable at x in the direction ch, and

$$\boxed{\mathrm{D}_{ch} f(x) = c\mathrm{D}_h f(x)}.$$

(This follows at once from the last expression for $\mathrm{D}_h f(x)$ in (1).) But this mapping l **can** be non-linear (that is, non-additive), as the following example shows.



Example. The function $f : \mathbb{R}^2 \to \mathbb{R}$ given (in the polar coordinates) by the formula

$$f(\varrho, \theta) = \varrho \sin 3\theta,$$

the graph of which is shown on the picture (from: M. Krupka, Matematická Analýza III, Opava 1999), is differentiable at EACH point in EACH direction, but is NOT G-differentiable at the origin. (Verify!)

The next lemma is an extension of (1).

Lemma 2.2.2. (on $f(x+th)$). *Let $f : X \to Y$. Put for given $x, h \in X$*

$$\varphi(t) := f(x+th) \qquad (t \in \mathbb{R}).$$

Then

$$\boxed{\dot{\varphi}(t) = \mathrm{D}_h f(x+th)}.$$

(If one side is defined then the other one is also defined, and they are equal.)

$$\begin{aligned} \lhd\ \mathrm{D}_h f(x+th) &\overset{(1)}{=} \lim_{\tau \to 0} \frac{f((x+th)+\tau h) - f(x+th)}{\tau} \\ &= \lim_{\tau \to 0} \frac{\varphi(t+\tau) - \varphi(t)}{\tau} = \dot{\varphi}(t). \ \rhd \end{aligned}$$

2.3 Rules of differentiation

First of all, differentiation is a linear operation:

Linearity of differentiation.*(a) If $f \in \mathrm{Dif}(x)$ then for each $c \in \mathbb{R}$ we have also $cf \in \mathrm{Dif}(x)$, and*

$$\boxed{(cf)'(x) = cf'(x)}.$$

(b) If $f_1, f_2 \in \mathrm{Dif}(x)$, then $f_1 + f_2 \in \mathrm{Dif}(x)$, and

$$\boxed{(f_1 + f_2)'(x) = f_1'(x) + f_2'(x)}.$$

$\lhd$ (a) We have $(cf)(x+h) = c(f(x+h)) = c(f(x) + f'(x)h + r(h)) = (cf)(x) + (cf'(x))h + (cr)(h)$, so we need to verify that cr is small. But indeed (for $h \neq 0$)

$$\frac{\|(cr)(h)\|}{\|h\|} = \frac{\|c(r(h))\|}{\|h\|} = |c| \underbrace{\frac{\|r(h)\|}{\|h\|}}_{\xrightarrow{r\text{ is small}} 0} \xrightarrow[\|h\| \to 0]{} 0.$$

(b) We have analogously (in obvious notations)

$$(f_1 + f_2)(x+h) = (f_1 + f_2)(x) + (f_1'(x) + f_2'(x))h + (r_1 + r_2)(h),$$

so we just need to verify that $r_1 + r_2$ is small. But indeed

$$\frac{\|(r_1+r_2)(h)\|}{\|h\|} = \frac{\|r_1(h)+r_2(h)\|}{\|h\|} \le \frac{\|r_1(h)\|+\|r_2(h)\|}{\|h\|}$$
$$= \frac{\|r_1(h)\|}{\|h\|} + \frac{\|r_2(h)\|}{\|h\|} \xrightarrow[\|h\|\to 0]{r_1, r_2 \text{ are small}} 0. \ \triangleright$$

Product Rule. *Let $X, Y_1, \dots, Y_m$ be normed spaces, and let $f_i : X \to Y_i,\ i = 1, \dots, m$. We denote by $(f_1, \dots, f_m)$ the product mapping $X \to Y_1 \times \dots \times Y_m$ defined by the formula*

$$\begin{array}{ccc} & & Y_1 \\ & f_1\nearrow & \\ X & \vdots & \\ & f_m\searrow & \\ & & Y_m \end{array}$$

$$(f_1, \dots, f_m)(x) := (f_1(x), \dots, f_m(x)).$$

The mapping $(f_1, \dots, f_m)$ is differentiable (resp., G-differentiable, differentiable in a direction h) at $x \in X$ iff each mapping f_i is differentiable (resp., G-differentiable, differentiable in h) at x, and

$$X \xrightarrow{(f_1,\dots,f_m)} Y_1 \times \ldots \times Y_m$$

$$\boxed{\begin{aligned} (f_1, \dots, f_m)'(x) &= (f_1'(x), \dots, f_m'(x)) \\ \mathrm{D}_h(f_1, \dots, f_m)(x) &= (\mathrm{D}_h f_1(x), \dots, \mathrm{D}_h f_m(x)) \end{aligned}}$$

(component-wise differentiation).

$\triangleleft$ Consider, e.g., the case of F-differentiation. We have

$$\begin{aligned} (f_1, \dots, f_m)(x+h) &= (f_1(x+h), \dots, f_m(x+h)) \\ &= (f_1(x) + f_1'(x)h + r_1(h), \dots, f_m(x) + f_m'(x)h + r_m(h)) \\ &= (f_1(x), \dots, f_m(x)) + (f_1'(x)h, \dots, f_m'(x)h) + (r_1(h), \dots, r_m(h)) \\ &= (f_1, \dots, f_m)(x) + (f_1'(x), \dots, f_m'(x))h + (r_1, \dots, r_m)(h). \end{aligned}$$

Now, $(f_1'(x), \dots, f_m'(x)) \in \mathscr{L}(X, Y_1 \times \dots \times Y_m)$ iff each $f_i \in \mathscr{L}(X, Y_i)$ (by the definitions of product vector space and product topology), and $(r_1, \dots, r_m)$ is small iff each r_i is small. Indeed

$$\frac{(r_1, \dots, r_m)(h)}{\|h\|} = \left(\frac{r_1(h)}{\|h\|}, \dots, \frac{r_m(h)}{\|h\|}\right) \xrightarrow[\|h\|\to 0]{} 0 \Leftrightarrow \forall i \vdots\ \frac{r_i(h)}{\|h\|} \xrightarrow[\|h\|\to 0]{} 0,$$

since convergence in a product space is just convergence of each component. $\triangleright$

Chain Rule. *Let $f : X \to Y$ be differentiable at a point $x \in X$, and let $g : Y \to Z$ be differentiable at the point $y := f(x)$. Then the composition $g \circ f$ is differentiable at x, and the derivative of $g \circ f$ is equal to the composition of derivatives of f and g:*

$$\begin{array}{ccccc} X & \overset{f}{\to} & Y & \overset{g}{\to} & Z \\ x & \mapsto & y & & \\ X & \xrightarrow{f'(x)} & Y & \xrightarrow{g'(x)} & Z \end{array}$$

$$\boxed{(g \circ f)'(x) = g'(y) \circ f'(x)}.$$

$\triangleleft$ 1° Put $f'(x) =: l,\ g'(y) =: m,\ g(y) =: z$. We have, by the conditions,

$$\forall\, \Delta x \in X \vdots\ \ f(x + \Delta x) = y + l\,\Delta x + r_f(\Delta x), \tag{1}$$

$$\forall\, \Delta y \in Y \vdots\ \ g(y + \Delta y) = z + m\,\Delta y + r_g(\Delta y), \tag{2}$$

where

$$\frac{\|r_f(\Delta x)\|}{\|\Delta x\|} \xrightarrow[\|\Delta x\|\to 0]{} 0, \tag{3}$$

$$\frac{\|r_g(\Delta y)\|}{\|\Delta y\|} \xrightarrow[\|\Delta y\|\to 0]{} 0. \tag{4}$$

We need to verify that

$$\forall\, \Delta x \in X \vdots \quad (g \circ f)(x + \Delta x) = z + (m \circ l)\Delta x + r(\Delta x), \tag{5}$$

where

$$\frac{\|r(\Delta x)\|}{\|\Delta x\|} \xrightarrow[\|\Delta x\|\to 0]{} 0. \tag{6}$$

2° But

$$r(\Delta x) = m r_f(\Delta x) + r_g(\Delta y), \tag{7}$$

where

$$\Delta y := l\,\Delta x + r_f(\Delta x). \tag{8}$$

Indeed

$$(g \circ f)(x + \Delta x) = g(f(x + \Delta x)) \overset{(1)}{=} g(y + \underbrace{l\,\Delta x + r_f(\Delta x)}_{\overset{(7)}{=}\Delta y})$$

$$\overset{(2)}{=} z + m(l\,\Delta x + r_f(\Delta x)) + r_g(\Delta y) \overset{(7)}{=} z + (m \circ l)\Delta x + r(\Delta x).$$

3° Now,

$$\frac{\|r(\Delta x)\|}{\|\Delta x\|} \overset{(7)}{=} \frac{\|m r_f(\Delta x) + r_g(\Delta y)\|}{\|\Delta x\|} \le \frac{\|m r_f(\Delta x)\| + \|r_g(\Delta y)\|}{\|\Delta x\|}$$

$$\overset{\text{BI}}{\le} \|m\| \underbrace{\frac{\|r_f(\Delta x)\|}{\|\Delta x\|}}_{\xrightarrow[\|x\|\to 0]{(3)} 0} + \frac{\|r_g(\Delta y)\|}{\underset{\text{trick}}{\|\Delta y\|}} \frac{\overset{\text{trick}}{\|\Delta y\|}}{\|\Delta x\|}.$$

So all will be proved if we show that
(a) $r_g(\Delta y)/\|\Delta y\| \to 0$ as $\|x\| \to 0$;
(b) $\|\Delta y\|/\|\Delta x\|$ is bounded for sufficiently small $\|\Delta x\|$.

4° Proof of (a): r_f is equal to 0 at 0 and is continuous (since f and l are). So $\|\Delta y\| \to 0$ if $\|\Delta x\| \to 0$, and (a) is true by (4).

5° Proof of (b): we have

$$\frac{\|\Delta y\|}{\|\Delta x\|} \overset{(8)}{=} \frac{\|l\,\Delta x + r_f(\Delta x)\|}{\|\Delta x\|} \le \frac{\|l\,\Delta x\| + \|r_f(\Delta x)\|}{\|\Delta x\|}$$

$$\overset{\text{BI}}{\le} \frac{\|l\|\,\|\Delta x\| + \|r_f(\Delta x)\|}{\|\Delta x\|} = \|l\| + \underbrace{\frac{\|r_f(\Delta x)\|}{\|\Delta x\|}}_{\xrightarrow[\|\Delta\|x\to 0]{(3)} 0} \le c$$

for some $c > 0$ and sufficiently small $\|\Delta x\|$. ▷

Important special cases.

1) If $f = l \in \mathscr{L}(X, Y)$ then $\boxed{(g \circ l)'(x) = g'(lx) \circ l}$.

2) If $g = l \in \mathscr{L}(Y, Z)$ then $\boxed{(l \circ f)'(x) = l \circ f'(x)}$ (we can “transfer” $l\circ$ through the brackets).

3) If $X = \mathbb{R}$, then $\boxed{(g \circ f)\dot{}(t) = g'(f(t)) \cdot \dot{f}(t)}$ (recall that $\dot{f}(t)$ is an element of Y).

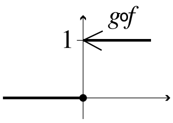
4) In particular, if $X = \mathbb{R}$ and $f = l$ then $\boxed{(g \circ l)'(t) = g'(lt) \cdot (l \cdot 1)}$, and

5) if $X = \mathbb{R}$ and $g = l$ then $\boxed{(l \circ f)\dot{}(t) = l \cdot \dot{f}(t)}$.

$\lhd$ All this follows from the facts that $l' \equiv l$ and that $\dot{f}(t) = f'(t) \cdot 1$. $\rhd$

Below we refer to Special Cases 1), 2), 4), 5) as to *l-Rule*.

NB For G-differentiability Chain Rule is NOT valid, as the following example shows.



Example. Let $f : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto (t, t^2)$, and let g be the function from 5) in (2.1) (where we used the letter f for it). Then f is G- (and even F-)differentiable at 0, g is G-differentiable at $f(0) = (0, 0)$, but $g \circ f$ is NOT G-differentiable at 0.

Lemma 2.3.1. (on evaluation). *Let X, Y, Z be normed spaces, and let a mapping $f : X \to \mathscr{L}(Y, Z)$ be differentiable at a point x. Let k be a fixed vector in Y, and let $g : X \to Z$ be defined by the formula*

$$g(x) := f(x) \cdot k \quad \textit{(the VALUE of } f(x) \textit{ at } k\textit{).}$$

Then g is differentiable at x, and

$$\forall h \in X \,\vdots \quad g'(x)h = (f'(x)h)k.$$

$\lhd$ Obviously, $g = \mathrm{ev}_k \circ f$ (recall that $\mathrm{ev}_k : l \mapsto l \cdot k$, see Chapter 1), hence, by l-Rule ($\mathrm{ev}_k \in \mathscr{L}(\mathscr{L}(Y, Z), Z)$),

$$g'(x)h = (\mathrm{ev}_k \circ f'(x))h = \mathrm{ev}_k(f'(x)h) = (f'(x)h)k. \rhd$$

2.4 Partial derivatives

Here we consider two related things: differentiation in a (vector) subspace and partial differentiation.

Differentiation in a subspace

Let $f : X \to Y$ be a mapping between normed spaces and let X_1 be a vector subspace in X (the notation $X_1 \Subset X$). We say that f is F- (resp., G-)*differentiable* at a point $x \in X$ *in the subspace* X_1 if f admits in $x + X_1$ the representation

$$\forall h_1 \in X_1 \,\vdots \quad f(x + h_1) = f(x) + l_1 h_1 + r_1(h_1),$$

where $l_1 \in \mathscr{L}(X_1, Y)$ and $r_1 : X_1 \to Y$ is F- (resp., G-) small. In such a case we write

$$f \in \mathrm{Dif}_{X_1}(x) \quad (\text{resp., } f \in G\text{-}\mathrm{Dif}_{X_1}(x))$$

and we call l_1 the *derivative* of f at x in the subspace X_1:

$$l_1 =: f'_{X_1}(x).$$

Example. A mapping $f : X \to Y$ is differentiable at a point x in a (non-zero) direction h iff f is differentiable at x in the one-dimensional subspace $\mathbb{R}h$ ($= \mathrm{lin}\{h\} \equiv \mathrm{span}\{h\}$), generated by h, and in such a case

$$f'_{\mathbb{R}h}(x) \cdot th = t\mathrm{D}_h f(x) \qquad (t \in \mathbb{R}).$$

(Verify!)

Theorem 2.4.1. *If $f : X \to Y$ is F- (resp., G-)differentiable at x, then f is F- (resp., G-)differentiable at x in each subspace $X_1 \Subset X$, and the derivative in X_1 is just the restriction of the "total" derivative:*

$$f'_{X_1}(x) = f'(x)|_{X_1}.$$

$\lhd$ This follows at once from the obvious facts, that the restriction of a continuous linear mapping onto a vector subspace is also a continuous linear mapping, and the restriction of a small mapping is also small. $\rhd$

Partial differentiation

Let $X_1, \ldots, X_n$ and Y be normed spaces. We say that a mapping $f : X_1 \times \ldots \times X_n \to Y$ is F- (resp., G-)*differentiable* at a point $x = (x_1, \ldots, x_n)$ *in the i-th coordinate* if the mapping

$$f(x_1, \ldots, x_{i-1}, \cdot, x_{i+1}, \ldots, x_n) : X_i \to Y, \quad \tilde{x}_i \mapsto f(x_1, \ldots, x_{i-1}, \tilde{x}_i, x_{i+1}, \ldots, x_n)$$

(that is, the mapping with all other coordinates FIXED) is F- (resp., G-)differentiable at the point x_i. We denote the corresponding derivative by

$$\frac{\partial f(x)}{\partial X_i} \quad (\in \mathscr{L}(X_i, Y))$$

and call it the *partial derivative* in X_i at the point x.

Theorem 2.4.2. *A mapping $f : X_1 \times \cdots \times X_n \to Y$ is differentiable (F- or G-) at x in the i-th coordinates iff f is differentiable (in the same sense) at x in the subspace*

$$0 \times \cdots \times 0 \times X_i \times 0 \times \cdots \times 0,$$

and

$$\forall h_i \in X_i \vdots \quad \frac{\partial f(x)}{\partial X_i} \cdot h_i = f'_{0\times\cdots\times X_i\times\cdots\times 0}(x) \cdot (0, \ldots, 0, h_i, 0, \ldots, 0).$$

$\lhd$ This follows immediately from the definitions. $\rhd$

Theorem 2.4.3. (on total and partial derivatives). *If a mapping $f : X_1 \times \cdots \times X_n \to Y$ is G-differentiable at a point $x = (x_1, \ldots, x_n)$ then f is G-differentiable at x in each coordinate, and*

$$\forall h = (h_1, \ldots, h_n) \in X_1 \times \cdots \times X_n \vdots \quad \boxed{f'(x) \cdot h = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial X_i} \cdot h_i},$$

or, more short,

$$\boxed{f'(x) = \frac{\partial f(x)}{\partial X_1} \oplus \cdots \oplus \frac{\partial f(x)}{\partial X_n} \equiv \bigoplus_{i=1}^{n} \frac{\partial f(x)}{\partial X_i}}.$$

Here $l_1 \oplus \cdots \oplus l_n$, for $l_i \in \mathscr{L}(X_i, Y)$, denotes the *direct sum* of the mappings l_i, defined by the formula

$$l_1 \oplus \cdots \oplus l_n : X_1 \times \cdots \times X_n \to Y, \quad (h_1, \ldots, h_n) \mapsto l_1 \cdot h_1 + \cdots + l_n \cdot h_n.$$

$$\begin{aligned} \lhd\ f'(x) \cdot h = f'(x) \cdot (h_1, \ldots, h_n) &= f'(x) \cdot \sum_{i=1}^{n} (0, \ldots, 0, h_i, 0, \ldots, 0) \\ &\overset{\text{Theorem 2.4.1.}}{=} \sum_{i=1}^{n} f'(x)(0, \ldots, 0, h_i, 0, \ldots, 0) \\ &\cdot \sum_{i=1}^{n} f'_{0 \times \ldots \times X_i \times \ldots \times 0}(x) \cdot (0, \ldots, 0, h_i, 0, \ldots, 0) \\ &\overset{\text{Theorem 2.4.2.}}{=} \sum_{i=1}^{n} \frac{\partial f(x)}{\partial X_i} \cdot h_i. \rhd \end{aligned}$$

2.5 Finite-dimensional case

We ever denote by $\mathrm{e}_1, \ldots, \mathrm{e}_n$ the standard basis in $\mathbb{R}^n$:

$$\mathrm{e}_i := (0 \ldots, 0, \underset{i\text{-th place}}{1}, 0, \ldots 0) \in \mathbb{R}^n.$$

For a mapping $f : \mathbb{R}^n = \mathbb{R} \times \ldots \times \mathbb{R} \to Y$ the partial derivative in the i-th coordinate applied to the "vector" $1 \in \mathbb{R}$ (that is, the "usual" partial derivative in the i-th coordinate) is traditionally denoted by

$$\frac{\partial f}{\partial x_i} \quad (\in \mathscr{L}(\mathbb{R}, Y) \approx Y).$$

By Theorem 2.4.2. (with $h_i = 1$), and by the Example in 2.4 (with $h = \mathrm{e}_i$ and $t = 1$),

$$\boxed{\frac{\partial f(x)}{\partial x_i} = \mathrm{D}_{\mathrm{e}_i} f(x)}. \tag{1}$$

(Emphasize that $\partial f(x)/\partial x_i$ is an element of Y.)

Jacobi matrix

Theorem 2.5.1. (on representation). *Let a mapping $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^n$ be G-differentiable at a point $x \in \mathbb{R}^n$. Then $f'(x)$ is represented as a linear mapping $\mathbb{R}^n \to \mathbb{R}^m$ by the matrix of partials derivatives*

$$\mathrm{J}_f(x) := \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \ldots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \ldots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix}.$$

This matrix is called *Jacobi matrix* of f at x.

◁ By linear algebra, we need to verify that the i-th column of the matrix represents the vector $f'(x) \cdot \mathrm{e}_i$. But indeed

$$f'(x) \cdot \mathrm{e}_i \overset{\text{Prod. Rule}}{=} (f_1'(x), \dots, f_m'(x)) \cdot \mathrm{e}_i = (f_1'(x) \cdot \mathrm{e}_i, \dots, f_m'(x) \cdot \mathrm{e}_i),$$

and

$$f_j'(x) \cdot \mathrm{e}_i = \mathrm{D}_{\mathrm{e}_i} f_j(x) \overset{(1)}{=} \frac{\partial f_j(x)}{\partial x_i}. \ \triangleright$$

Corollary 2.5.2. *In conditions of Chain Rule, for mappings between finite-dimensional spaces, Jacobi matrix of the composition is equal to the matrix product of Jacobi matrices of the composed mappings:*

$$\mathrm{J}_{g\circ f}(x) = \mathrm{J}_g(f(x))\mathrm{J}_f(x).$$

◁ This follows from Chain Rule and the fact that the matrix of a composition of two linear mappings is equal to the product of the matrices of these mappings. ▷

Example. How to compute the derivative of the function $f(x) = x^x$ $(x > 0)$? Represent f as a composition: $f = g \circ \triangle$, where $\triangle : \mathbb{R} \to \mathbb{R}^2, t \mapsto (t,t)$, $g : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto x^y$. By l-Rule,

$$\begin{aligned}\dot f(t) = g'(t,t) \cdot (1,1) &\overset{\text{Theorem 2.5.1.}}{=} \left(\frac{\partial g(t,t)}{\partial x}, \frac{\partial g(t,t)}{\partial y}\right)\begin{pmatrix}1\\1\end{pmatrix} = \frac{\partial g(t,t)}{\partial x} + \frac{\partial g(t,t)}{\partial y}\\ &= yx^{y-1}\Big|_{x=y=t} + (\ln x)x^y\Big|_{x=y=t} = (\ln t + 1)t^t.\end{aligned}$$

Gradient

In the special case of SCALAR functions $f : \mathbb{R}^n \to \mathbb{R}$ Jacobi matrix is the row

$$\begin{pmatrix}\dfrac{\partial f(x)}{\partial x_1} & \dots & \dfrac{\partial f(x)}{\partial x_n}\end{pmatrix}.$$

The corresponding vector in $\mathbb{R}^n$ is called the *gradient* of f at x and is denoted by grad $f(x)$:

$$\operatorname{grad} f(x) := \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n}\right) \qquad (\in \mathbb{R}^n).$$

In this situation Theorem 2.5.1. (Theorem on representation) says:

$$\boxed{f'(x) \cdot h = \operatorname{grad} f(x) \cdot h}, \tag{2}$$

where the point to the right means scalar product. Indeed

$$\begin{pmatrix}\dfrac{\partial f(x)}{\partial x_1} \cdots \dfrac{\partial f(x)}{\partial x_n}\end{pmatrix}\begin{pmatrix}h_1\\ \vdots \\ h_n\end{pmatrix} = \sum_{i=1}^n \frac{\partial f(x)}{\partial x_i} h_i.$$

For any UNIT vector $\vec{\nu} \in \mathbb{R}^n$ ($\|\vec{\nu}\|_2 = 1$), the differential of a mapping $f : \mathbb{R}^n \to \mathbb{R}$ at a point x by the increment $\vec{\nu}$ is called the *derivative* at x *in the direction* $\vec{\nu}$ and is denoted traditionally by $\partial f(x)/\partial \nu$:

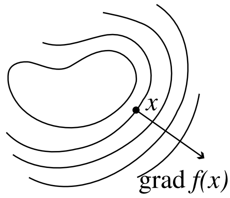
$$\frac{\partial f(x)}{\partial \nu} := \mathrm{D}_{\vec{\nu}} f(x).$$

Theorem 2.5.3. *If a function $f : \mathbb{R}^n \to \mathbb{R}$ is G-differentiable at a point x, then for any unit vector $\vec{\nu} \in \mathbb{R}^n$ it holds*

$$\frac{\partial f(x)}{\partial \nu} = \operatorname{grad} f(x) \cdot \vec{\nu}$$

(where the point means the scalar product).

$\lhd$ $\partial f(x)/\partial \nu = \mathrm{D}_{\vec{\nu}} f(x) = f'(x) \cdot \vec{\nu} \overset{(2)}{=} \operatorname{grad} f(x) \cdot \vec{\nu}$. $\rhd$



Corollary 2.5.4. *The gradient vector of f at x yields the direction of the greatest growth of f at x and is orthogonal to the level line of f passing through x.*

$\lhd$ The scalar product $\operatorname{grad} f(x) \cdot \vec{\nu}$ is maximal for the unit vector $\vec{\nu}$ that has the SAME direction as $\operatorname{grad} f(x)$ and is equal to 0 for $\vec{\nu}$ orthogonal to the gradient. $\rhd$

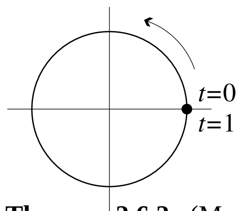
2.6 Mean Value Theorem

In classic differential calculus, the following result plays an important role:

Theorem 2.6.1. (Lagrange). *If a function $f : \mathbb{R} \to \mathbb{R}$ is continuous in the closed interval* $[0, 1]$ *and is differentiable in the open interval* $(0, 1)$, *then there exists $\theta \in (0, 1)$ such that*

$$f(1) - f(0) = \dot{f}(\theta).$$

This result is NOT true for functions with vector values, as the following example shows.



Example. Let $f : \mathbb{R} \to \mathbb{R}^2$, $t \mapsto (\cos 2\pi t, \sin 2\pi t)$. Then $f(1) - f(0) = (0, 0)$, but

$$\dot{f}(t) \overset{\text{Prod. Rule}}{=} (-2\pi \sin 2\pi t, 2\pi \cos 2\pi t),$$

hence $\left\| \dot{f}(t) \right\|_2 = 2\pi$, which is never zero.
But the following ESTIMATE of the increment is true:

Theorem 2.6.2. (Mean Value Theorem, (MVT)). *Let a function $\varphi : \mathbb{R} \to Y$ (where Y is a normed space) be continuous on* $[0, 1]$ *and differentiable in* $(0, 1)$. *Then*

$$\|\varphi(1) - \varphi(0)\| \leq \sup_{0<t<1} \|\dot{\varphi}(t)\| .$$

$\lhd$ 1° Put $y := \varphi(1) - \varphi(0)$. By Theorem 1.10.5. (Lemma from FA (see Chapter 1)) there exists $l \in \mathscr{L}(Y, \mathbb{R})$ such that

$$\|l\| = 1, \qquad ly = \|y\| . \tag{1}$$

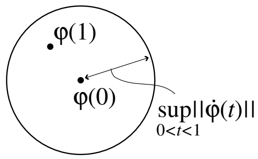
2° Consider the composition

$$\mathbb{R} \xrightarrow{\varphi} Y \xrightarrow{l} \mathbb{R}.$$

It is differentiable in $(0, 1)$ by Chain Rule.

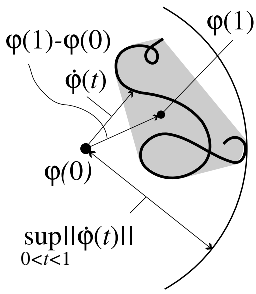
3° We have $\|y\| \overset{(1)}{=} ly = l(\varphi(1) - \varphi(0)) = (l \circ \varphi)(1) - (l \circ \varphi)(0)$. By Theorem 2.6.1. for some $\theta \in (0, 1)$ it holds

$$(l \circ \varphi)(1) - (l \circ \varphi)(0) = (l \circ \varphi)\dot{}(\theta) \overset{l\text{-Rule}}{=} l \cdot \dot{\varphi}(\theta) \overset{\text{BI}}{\leq} \underbrace{\|l\|}_{=1} \|\dot{\varphi}(\theta)\| \leq \sup_{0<t<1} \|\dot{\varphi}(t)\| . \ \rhd$$



CAR-INTERPRETATION. If $Y = \mathbb{R}^2$ and we consider $t \in \mathbb{R}$ as time, then φ describes a motion of a "car" in the plane, and $\dot{\varphi}(t)$ is the velocity of this car. The inequality in Theorem 2.6.2. (MVT) means that our car in one hour will be INSIDE the circle with the center at the original point, the radius of which is equal to the maximal value of the velocity of the car during this hour.

Remark. In fact the following much more strong result is true: By the conditions of MVT, the increment $\varphi(1) - \varphi(0)$ lies in the closed convex hull of the set $\{\dot{\varphi}(t) \mid 0 < t < 1\}$ (shadowed on the picture).



Corollary 2.6.3. *Let X, Y be normed spaces, let $x, h \in X$ and let a mapping $f : X \to Y$ has the following properties:*
a) *the restiction of f onto the closed interval $[x, x+h]$ $(:= \{x + th \mid t \in [0,1]\})$ is continuous and*
b) *f is differentiable in the direction h in the open interval $(x, x+h)$ $(:= \{x + th \mid t \in (0,1)\})$. Then*

$$\| f(x+h) - f(x) \| \leq \sup_{y \in (x, x+h)} \| \mathrm{D}_h f(y) \| .$$

$\lhd$ Put $\varphi(t) := f(x + th)$, $t \in \mathbb{R}$. By Lemma 2.2.2. (on $f(x+th)$), it holds $\dot{\varphi}(t) = \mathrm{D}_h f(x + th)$. So our assertion follows from MVT (Theorem 2.6.2.). $\rhd$

Corollary 2.6.4. *In the situation of Corollary 2.6.3., let f has the following properties:*
a) *the restiction of f onto the interval $[x, x+h]$ is comtinuous and*
b) *f is G-differentiable in $(x, x+h)$. Then*

$$\| f(x+h) - f(x) \| \leq \|h\| \sup_{y \in (x, x+h)} \left\| f'(y) \right\| .$$

$\lhd$ This follows from Corollary 2.6.3. and the fact that

$$\| \mathrm{D}_h f(y) \| = \left\| f'(y) \cdot h \right\| \overset{\mathrm{BI}}{\leq} \left\| f'(y) \right\| \|h\| . \ \rhd$$

2.7 Continuous differentiability

Let X, Y be normed spaces, let $x \in X$, and let $f : X \to Y$ be G-differentiable in an open neighbourhood U of x. We say that f is *continuously G-differentiable at x* and we write

$$f \in \mathrm{C}^1_G(x),$$

if the derivative mapping

$$f' : U \to \mathscr{L}(X, Y), \quad y \mapsto f'(y)$$

is continuous at x. Thus

$$f \in \mathrm{C}^1_G(x) :\Leftrightarrow \left\| f'(x+h) - f'(x) \right\| \xrightarrow[\|h\| \to 0]{} 0. \tag{1}$$

Theorem 2.7.1. (on Continuous Derivative). *If $f : X \to Y$ is continuously G-differentiable at a point $x \in X$, then f is F-differentiable at x.*

◁ 1° We have

$$f(x+h) = f(x) + f'(x)h + r(h), \qquad f'(x) \in \mathscr{L}(X,Y),\ r \in G\text{-small}.$$

Put for $t \in \mathbb{R}$

$$\psi(t) := r(th) = f(x+th) - f(x) - tf'(x)h. \tag{2}$$

Then

$$r(h) = \psi(1) - \psi(0). \tag{3}$$

2° We want to apply MVT, so compute the derivative of ψ. By Lemma 2.2.2. (on $f(x+th)$),

$$\dot\psi(t) = \underbrace{\mathrm{D}_h f(x+th)}_{=f'(x+th)h} - f'(x)h = (f'(x+th) - f'(x))h. \tag{4}$$

3° Now,

$$\begin{aligned}\frac{\|r(h)\|}{\|h\|} &\overset{(3)}{=} \frac{\psi(1)-\psi(0)}{\|h\|} \overset{\text{MVT}}{\le} \frac{1}{\|h\|}\sup_{0<t<1}\left\|\dot\psi(t)\right\| \\ &\overset{(4)}{=} \frac{1}{\|h\|}\sup_{0<t<1}\underbrace{\left\|(f'(x+th)-f'(x))h\right\|}_{\overset{\text{BI}}{\le}\|(f'(x+th)-f'(x))\|\|h\|} \\ &\le \sup_{0<t<1}\left\|(f'(x+th)-f'(x))\right\| \xrightarrow[\|h\|\to 0]{(1)} 0,\end{aligned}$$

since $\|th\| = |t|\,\|h\| \xrightarrow[\|h\|\to 0]{} 0$ uniformly in $0 < t < 1$. Thus r is F-small, so f is F-differentiable at x. ▷

Class C^1

Let X, Y be normed spaces and let U be an open set in X. We say that $f : X \to Y$ is *of class* C^1 *in* U an we write

$$f \in \mathrm{C}^1(U)$$

if f is differentiable at each point of U (that is, in U), and the derivative mapping

$$f' : U \to \mathscr{L}(X,Y), \quad x \mapsto f'(x)$$

is continuous.

Theorem 2.7.2. (C^1-Theorem on Continuous Derivative). *Let $f : X \to Y$ be continuously G-differentiable at each point of an open set U in X. Then f is of class C^1 in U.*
◁ This is an immediate corollary of Theorem 2.7.1. (on continuous derivative) ▷

2.8 Continuous partial derivatives

At first we prove one result on continuous differentiability in n fixed directions.

Continuous differentials

Lemma 2.8.1. *Let X, Y be normed spaces, and let $h_1, \ldots, h_n$ be fixed vectors in X. Let a mapping $f : X \to Y$ be differentiable in the directions $h_1, \ldots, h_n$ in some neighbourhood U of a point x, and let all the mappings*

$$\mathrm{D}_{h_i} f : \tilde x \mapsto \mathrm{D}_{h_i} f(\tilde x),\ U \to Y \qquad (i = 1, \ldots, n)$$

be continuous at x. Then f is G-differentiable at x in the vector subspace H in X, generated (spanned) by the vectors h_i ($H = \mathrm{lin}\{h_1, \ldots, h_n\} \equiv \mathrm{span}\{h_1, \ldots, h_n\}$).
$\lhd$ 1° If f IS G-differentiable at x in H, then

$$f'_H(x) \cdot h_i = \mathrm{D}_{h_i} f(x) \qquad (i = 1, \ldots, n),$$

so we must have $\forall c_1, \ldots, c_n \in \mathbb{R}$

$$f'(x) \cdot (c_1 h_1 + \ldots + c_n h_n) = c_1 f'(x) h_1 + \ldots + c_n f'(x) h_n = c_1 \mathrm{D}_{h_1} f(x) + \ldots + c_n \mathrm{D}_{h_n} f(x). \tag{1}$$

(Since H is finite-dimensional, the so defined linear mapping $f'_H(x)$ is CONTINUOUS.) We have to verify that for this $f'_H(x)$ it holds

$$\forall h \in H \vdots \quad f(x+h) = f(x) + f'(x)h + r(h), \qquad \frac{r(th)}{t} \xrightarrow[t \to 0]{} 0,$$

that is, that

$$\forall h \in H \vdots \quad \frac{f(x+th) - f(x) - t f'_H(x) h}{t} \xrightarrow[t \to 0]{} 0.$$

This means by (1) that we need to verify that $\forall c_1, \ldots, c_n \in \mathbb{R} \vdots$

$$\frac{f(x + t(c_1 h_1 + \ldots + c_n h_n)) - f(x) - t(c_1 \mathrm{D}_{h_1} f(x) + \ldots + c_n \mathrm{D}_{h_n} f(x))}{t} \xrightarrow[t \to 0]{} 0.$$

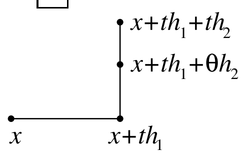
By homogeneity of $\mathrm{D}_h f(x)$ in h, it holds $c_i \mathrm{D}_{h_i} f(x) = \mathrm{D}_{c_i h_i} f(x)$ so without loss of generality we can assume that $c_1 = \ldots = c_n = 1$ (take $c_i h_i$ as NEW h_i). Further, by induction argument, it is sufficient to consider the case $n = 2$. Thus we need to verify that

$$\frac{f(x + t(h_1 + h_2)) - f(x)}{t} - (\mathrm{D}_{h_1} f(x) + \mathrm{D}_{h_2} f(x)) \xrightarrow[t \to 0]{} 0. \tag{2}$$

2° Adding and subtracting $f(x + th_1)$ in the numerator, we can write the left-hand side of (2) as the sum $\boxed{1} + \boxed{2}$, where

$$\boxed{1} = \frac{f(x + th_1) - f(x)}{t} - \mathrm{D}_{h_1} f(x),$$
$$\boxed{2} = \frac{f(x + th_1 + th_2) - f(x + th_1) - t\mathrm{D}_{h_2} f(x)}{t}.$$

So it is sufficient to verify that $\boxed{1} \to 0$ and $\boxed{2} \to 0$ as $t \to 0$.
3° $\boxed{1} \to 0$ as $t \to 0$ by the definition of D_h.



4° Putting (see the picture)

$$\varphi(\theta) := f(x + th_1 + \theta t h_2) - \theta t \mathrm{D}_{h_2} f(x) \qquad (\theta \in \mathbb{R}), \tag{3}$$

we can write $\boxed{2}$ in the form

$$\boxed{2} = \frac{\varphi(1) - \varphi(0)}{t}. \tag{4}$$

5° By Lemma 2.2.2. (on $f(x + th)$), we obtain from (3)

$$\dot\varphi(\theta) = \mathrm{D}_{th_2} f(x + th_1 + \theta t h_2) - t\mathrm{D}_{h_2} f(x)$$
$$\overset{\substack{\text{homog.}\\ \text{of } \mathrm{D_h}}}{=} t(\mathrm{D}_{h_2} f(x + th_1 + \theta t h_2) - \mathrm{D}_{h_2} f(x)). \tag{5}$$

6° At last,

$$\left\|\boxed{2}\right\| \overset{(4)}{=} \frac{1}{|t|}\,\|\varphi(1)-\varphi(0)\| \overset{\text{MVT}}{\leq} \frac{1}{|t|}\sup_{0<\theta<1}\|\dot{\varphi}(\theta)\|$$

$$\overset{(5)}{=} \sup_{0<\theta<1}\left\|\mathrm{D}_{h_2} f(x+th_1+\theta th_2)-\mathrm{D}_{h_2} f(x)\right\| \xrightarrow[t\to 0]{} 0,$$

since $\|th_1+\theta th_2\| \leq |t|(\|h_1\|+\theta\,\|h_2\|) \underset{0<\theta<1}{\leq} |t|(\|h_1\|+\|h_2\|) \xrightarrow[t\to 0]{} 0$ uniformly in θ and $\mathrm{D}_{h_2} f$ is continuous at x. Thus, $\boxed{2} \to 0$ as $t \to 0$. $\triangleright$

Continuous partial derivatives

Theorem 2.8.2. *Let $X_1, \ldots, X_n$ and Y be normed spaces and let a mapping $f : X_1 \times \ldots \times X_n \to Y$ have all the partial derivatives $\partial f/\partial X_i$ in an open set $U \subset X_1 \times \ldots \times X_n$ and these partial derivatives be continuous in U. Then f is of class C^1 in U.*
$\triangleleft$ For simplicity we consider only the case $X_1 = \cdots = X_n = Y = \mathbb{R}$ (that is, $f : \mathbb{R}^n \to \mathbb{R}$), which is the most important for us. In this case continuity of the partial derivatives $\partial f/\partial X_i$ means just continuity of the partial derivatives $\partial f/\partial x_i = \mathrm{D}_{\mathrm{e}_i} f$. By Lemma on Continuous Differentials we conclude that f is G-differentiable in U. By Theorem on Representation,

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

Since each component of this vector continuously depends on x, we conclude, that f' continuously depends on x. Hence by Theorem 2.7.2. (C^1-Theorem on Continuous Derivative), f is of class C^1 in U. $\triangleright$

Corollary 2.8.3. *Let $f = (f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$, and let all the partial derivatives $\partial f_j/\partial x_i$ ($i = 1, \ldots, n$; $j = 1, \ldots, m$) be continuous in $U \subset \mathbb{R}^n$. Then f is of class C^1 in U.*
$\triangleleft$ By the Product Rule, $\partial f/\partial x_i = (\partial f_1/\partial x_i, \ldots, \partial f_m/\partial x_i)$, so $\partial f/\partial x_i$ are continuous if all $\partial f_j/\partial x_i$ are. $\triangleright$

Chapter 3

Inverse Function Theorem

3.1 Lipschitz functions

Let X, Y be normed spaces, and let $A \subset X$. We say that a mapping $f : X \to Y$ is *Lipschitz on A with a constant $k > 0$*, and we write

$$f \in \mathrm{Lip}_A\, k$$

if

$$\forall x_1, x_2 \in A \colon \quad \| f(x_1) - f(x_2)\| \le k \|x_1 - x_2\|\,.$$

We say that f is *Lipschitz on A* and we write $f \in \mathrm{Lip}_A$ if f is Lipschitz with some constant k. If $A = X$ or if it is clear what A we mind, we omit A and write simply $f \in \mathrm{Lip}$.

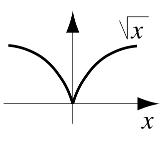
Examples.

1. $|\cdot| : \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$, is Lipschitz with the constant 1.

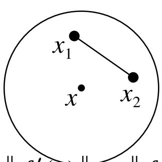
2. For any normed space X, the norm $\|\cdot\|$ is Lipschitz with the constant 1.

3. If $f \in \mathscr{L}(X, Y)$, then $f \in \mathrm{Lip}\, \|f\|$.
$\lhd$ $\|fx_1 - fx_2\| = \|f(x_1 - x_2)\| \le \|f\|\, \|x_1 - x_2\|$. $\rhd$

4. The function $x \mapsto \sqrt{|x|}$, $\mathbb{R} \to \mathbb{R}$, is NOT Lipschitz.



Theorem 3.1.1. *If f is Lipschitz on an open set U then f is continuous in U.*
$\lhd$ $\|f(x+h) - f(x)\| \le k\|h\| \xrightarrow[\|h\|\to 0]{} 0$. $\rhd$



Theorem 3.1.2. *If $f \in \mathrm{C}^1_G(x)$, then (for any $\varepsilon > 0$), f is Lipschitz in some neighbourhood of x with the constant $\|f'(x)\| + \varepsilon$.*
$\lhd$ By the definition of $\mathrm{C}^1_G(x)$, there exists $\delta > 0$ such that for any $y \in \mathrm{B}_\delta(x)$ it holds $\|f'(y) - f'(x)\| \le \varepsilon$ and hence

$$\|f'(y)\| = \|f'(y) - f'(x) + f'(x)\| \le \underbrace{\|f'(y) - f'(x)\|}_{\le\varepsilon} + \|f'(x)\| \le \varepsilon + \|f'(x)\| =: k. \tag{1}$$

Then $\forall x_1, x_2 \in \mathrm{B}_\delta(x)$ it holds

$$\|f(x_1) - f(x_2)\| \overset{\text{MVT}}{\le} \overbrace{\sup_{y\in[x_1,x_2]} \|f'(y)\|}^{\le k \text{ by (1)}} \|x_1 - x_2\| \le k\|x_1 - x_2\|\,. \rhd$$

3.2 Banach spaces

We say that a normed space X is a *Banach space* (in honour of a Polish mathematician Stefan Banach), and we write

$$X \in \mathrm{BS}$$

if X is complete as a metric space (with the metric $\varrho(x, y) := \|x - y\|$), that is, if any Cauchy sequence in X converges.

Note that in a normed space

$$\{x_n\} \in \text{Cauchy} \iff \|x_m - x_n\| \xrightarrow[m,n\to\infty]{} 0.$$

Examples.

1. $\mathbb{R}$, $\mathbb{R}^n$, $\mathrm{C}([0,1])$, ℓ_2 are Banach spaces.

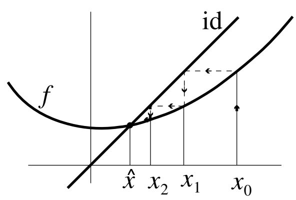
2. The vector subspace k in ℓ_2 which consists from all FINITE sequences, that is, sequences of the form $x = (x_1, \dots, x_n, 0, 0, \dots)$ (n depends on x), equipped with the norm form ℓ_2, is NOT a Banach space.

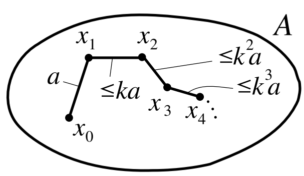
3.3 Contraction Lemma

It is the name of the following

Theorem 3.3.1. *Let $X \in \mathrm{BS}$, let A be a* CLOSED *subset in X, and let f be a mapping from A into itself, $f : A$. If $f \in \mathrm{Lip}_A\, k$ with $k < 1$ (strictly!), then the operator f has one and just one* FIXED POINT *$\hat{x}$, that is, a point such that*

$$f(\hat{x}) = \hat{x}. \tag{1}$$





(Note that we can rewrite (1) as $f(\hat{x}) = \mathrm{id}(\hat{x})$.) In this case we write

$$\boxed{\hat{x} \in \mathrm{Fix}\, f}.$$

$\triangleleft$ 0° The idea of the proof is clear from the picture: the broken line leads to $\hat{x}$.

1° Take an ARBITRARY point $x_0 \in A$ and put

$$x_1 = f(x_0),\ x_2 = f(x_1), \dots,\ x_{n+1} = f(x_n), \dots .$$

We have

$$\|x_1 - x_0\| =: a$$

$$\|x_2 - x_1\| = \|f(x_1) - f(x_0)\| \le k\,\|x_1 - x_0\| = ka, \tag{2}$$

$$\|x_3 - x_2\| = \|f(x_2) - f(x_1)\| \le k\,\|x_2 - x_1\| \overset{(2)}{\le} k^2 a,$$

$$\vdots$$

$$\|x_{n+1} - x_n\| \le k^n a. \tag{3}$$

2° $\{x_n\} \in$ Cauchy.

$$\lhd\!\lhd\ \|x_m - x_n\| \overset{\text{trick}}{\underset{\substack{\text{suppose}\\ m>n}}{=}} \|\underbrace{x_m - x_{m-1}} + \underbrace{x_{m-1} - x_{m-2}} + \cdots + \underbrace{x_{n+1} - x_n}\|$$

$$\leq \underbrace{\|x_m - x_{m-1}\|}_{\leq ak^{m-1} \text{ by (3)}} + \cdots + \underbrace{\|x_{n+1} - x_n\|}_{\leq ak^n \text{ by (3)}} \leq a(k^n + \cdots + k^{m-1}) \xrightarrow[m,n\to\infty]{\sum k^n < \infty} 0.\ \rhd\!\rhd$$

3° Put $\hat{x} := \lim x_n$ (the limit exists since $X \in$ BS). Then $\hat{x} \in A$, since A is closed. Further, since f is continuous (by Theorem 3.1.1.) it holds

$$f(\hat{x}) = \lim_{n\to\infty} \underbrace{f(x_n)}_{=x_{n+1}} = \lim_{n\to\infty} x_{n+1} \overset{\text{obv.}}{=} \lim_{n\to\infty} x_n = \hat{x},$$

hence $\hat{x}$ is a fixed point for f.
4° This fixed point is unique. $\lhd\!\lhd$ If x_1 and x_2 are both fixed points for f, then $\|\underbrace{f(x_1)}_{=x_1} - \underbrace{f(x_2)}_{=x_2}\| \leq \underbrace{k}_{<1} \|x_1 - x_2\| \Rightarrow \|x_1 - x_2\| = 0 \Rightarrow x_1 = x_2$. $\rhd\!\rhd$ $\rhd$

3.4 Isomorphisms

Let $X, Y \in$ NS, and let $l \in \mathscr{L}(X, Y)$. We say that l is an *isomorphism* and we write

$$l \in \mathrm{Iso}(X, Y) \qquad \text{(or simply } l \in \mathrm{Iso})$$

if l is a bijection, and if the inverse mapping l^{-1} (which is automatically linear, verify!) is also continuous.

We say that X and Y are *isomorphic* (as normed spaces) and we write

$$X \approx Y$$

if there exists an isomorphism from X onto Y.

Examples.

1. If $\|\cdot\|_1$ and $\|\cdot\|_2$ are two equivalent norms on a vector space X, then id : $(X, \|\cdot\|_1) \to (X, \|\cdot\|_2)$ is an isomorphism.
2. $\mathbb{R}^n \times \mathbb{R}^m \approx \mathbb{R}^{n+m}$.
3. $\mathscr{L}(\mathbb{R}, X) \approx X$.
4. $\mathscr{L}(\mathbb{R}^n, \mathbb{R}) \approx \mathbb{R}^n$.

NB $\exists X \in$ NS $\vdots\, \mathscr{L}(X, \mathbb{R}) \not\approx X$.

5. $\mathscr{L}(\ell_2, \mathbb{R}) \approx \ell_2$.
6. If $l \in \mathscr{L}(\mathbb{R}, \mathbb{R})$, then $l \in$ Iso $\Leftrightarrow l \neq 0$.
7. If $l \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$, then $l \in$ Iso $\Leftrightarrow \det l \neq 0$ (as is known from linear algebra).

3.5 Inverse Function Theorem

Here is a generalization of the classic Inverse Function Theorem:

Theorem 3.5.1. *Let* $X, Y \in \mathrm{BS}$, $f : X \to Y$, $\hat{x} \in X$, $\hat{y} := f(\hat{x})$. *If* $f \in \mathrm{C}^1_G(\hat{x})$ *and* $f'(\hat{x}) \in \mathrm{Iso}(X, Y)$, *then there exists a neighbourhood* U *of* $\hat{x}$ *such that* f *is a* HOMEOMORPHISM *of* U *onto* $f(U)$. *The inverse mapping* f^{-1} *is differentiable at* $\hat{y}$, *and*

$$(f^{-1})'(\hat{y}) = (f'(\hat{x}))^{-1}. \tag{1}$$

$\lhd$ 1° *Reduction to the case* $f(0) = 0$, $f'(0) = \mathrm{id}$. Without loss of generality we can assume that $X = Y$, $\hat{x} = \hat{y} = 0$, $f'(0) = \mathrm{id}$. $\lhd\!\lhd$ Put $\tilde{f}(h) := f(\hat{x} + h) - f(\hat{x})$. It is clear that $\tilde{f} : X \to Y$ satisfies the condition $\tilde{f}(0) = 0$ and has at 0 the same differentiability property as f has at $\hat{x}$. Thus, without loss of generality $\hat{x} = 0$, $\hat{y} = 0$.

Now put $l := f'(0)$ and $\tilde{f} := l^{-1} \circ f$ (recall that $l^{-1} \in \mathscr{L}(Y, X)$, since $l \in \mathrm{Iso}$). By Chain Rule, $\tilde{f}'(0) = l^{-1} \circ \underbrace{f'(0)}_{=l} = \mathrm{id}$. If the theorem is true for $\tilde{f}$, then it is true for $f = l \circ \tilde{f}$ (since $f^{-1} = (\tilde{f})^{-1} \circ l^{-1}$). Thus, without loss of generality $X = Y$, $f'(0) = \mathrm{id}$.

$$X \underset{g}{\overset{f}{\rightleftarrows}} Y \ \rhd\!\rhd$$

So, the decomposition $f(\hat{x} + h) = f(\hat{x}) + f'(\hat{x})h + r(h)$, $\frac{\|r(h)\|}{\|h\|} \xrightarrow[\|h\| \to 0]{} 0$, reduces to

$$f(h) = h + r(h), \qquad \text{or } f = \mathrm{id} + r, \tag{2}$$

where r satisfies the conditions

$$r(0) \overset{\text{as always}}{=} 0,\ r'(0) \overset{\text{as always}}{=} 0, r' \overset{f \in C^1_G(0)}{\in} \mathrm{Cont}(0), \frac{\|r(h)\|}{\|h\|} \xrightarrow[\|h\| \to 0]{} 0. \tag{3}$$

2° *Reduction to the fixed point problem.* Now note that to find the inverse function to f means to solve the equation

$$f(x) = y$$

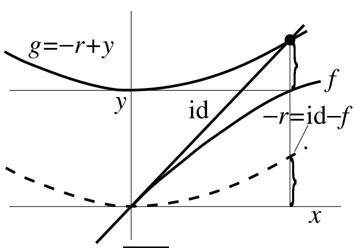
with respect to y. But for a FIXED y,

$$f(x) = y \Leftrightarrow x \in \mathrm{Fix}(\underbrace{\mathrm{id} - f}_{\overset{(2)}{=} -r} + y). \tag{4}$$

$\lhd\!\lhd$ $(\mathrm{id} - f + y)(x) = x \Leftrightarrow f(x) = y$. $\rhd\!\rhd$

So (in view of Contraction Lemma) our goal is to find a set, where the mapping

$$g := -r + y \tag{5}$$

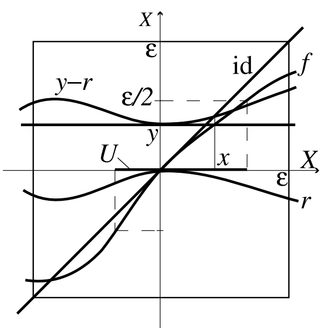


is Lipschitz.

3° By Theorem 3.1.2. applied to r, there exists $\varepsilon > 0$ such that $r \in \mathrm{Lip}_{\mathrm{B}_\varepsilon} \frac{1}{2}$. Put

$$U := (f|_{\mathrm{B}_\varepsilon})^{\boxed{-1}}(\mathrm{B}_{\varepsilon/2}) \qquad (\subset \mathrm{B}_\varepsilon).$$

(Here $\boxed{-1}$ denotes the pre-image, not the inverse mapping!)



4° $f \in \mathrm{Bij}(U, \mathrm{B}_{\varepsilon/2})$. $\lll$ Let $y \in \mathrm{B}_{\varepsilon/2}$. Consider the mapping

$$y - r: \ x \mapsto y - r(x), \ X \to X.$$

This mapping is Lipschitz on B_ε with the constant $\frac{1}{2}$, since r is Lipschitz. Moreover, $y - r$ maps B_ε into itself; indeed,

$$\|x\| \le \varepsilon \Rightarrow \|y - r(x)\| \le \underbrace{\|y\|}_{\le \varepsilon/2} + \underbrace{\|r(x)\|}_{\le \frac{1}{2}\underbrace{\|x\|}_{\le \varepsilon}} \le \varepsilon.$$

By Contraction Lemma, there exists one and just one fixed point of $y - r$. But by (4) this means that there exists one and just one point $x \in \mathrm{B}_\varepsilon$ such that $f(x) = y$, that is, by the definition of U, there exists one and just one point $x \in U$ such that $f(x) = 0$. $\rhd\!\!\rhd$
5° Let now f^{-1} denotes the (existing!) inverse mapping to $f : U \to \mathrm{B}_{\varepsilon/2}$. For convenience introduce the following notation: for $x \in U, \ y \in \mathrm{B}_{\varepsilon/2}$,

$$x \leftrightarrow y :\Leftrightarrow f(x) = y \Leftrightarrow x = f^{-1}(y).$$

6° $f^{-1} \in \mathrm{Lip}\, 2$. $\lll$ Let $x_1 \leftrightarrow y_1, \ x_2 \leftrightarrow y_2$. It holds

$$\|x_1 - x_2\| \overset{\mathrm{id}=f-r}{=} \|(\underbrace{f(x_1)}_{=y_1} - r(x_1)) - (\underbrace{f(x_2)}_{=y_2} - r(x_2))\| \le \|y_1 - y_2\| + \underbrace{\|r(x_2) - r(x_1)\|}_{\overset{(3)}{\le} \frac{1}{2}\|x_2 - x_1\|}$$

$$\le \|y_1 - y_2\| + \tfrac{1}{2}\|x_1 - x_2\|.$$

We conclude that $\|x_1 - x_2\| \le 2\|y_1 - y_2\|$, that is, $\left\| f^{-1}(y_1) - f^{-1}(y_2)\right\| \le 2\|y_1 - y_2\|$. $\rhd\!\!\rhd$

7° $f \in \mathrm{Homeo}(U, \mathrm{B}_{\varepsilon/2})$. $\lll$ $f = \mathrm{id} + r \overset{3^\circ}{\in} \mathrm{Cont}$; $f^{-1} \overset{6^\circ}{\in} \mathrm{Cont}$. $\rhd\!\!\rhd$
8° $(f^{-1})'(0) = \mathrm{id}\ (= (f'(0))^{-1})$. $\lll$ We need to verify that the mapping

$$s := f^{-1} - \mathrm{id}$$

is small (recall that $f^{-1}(0) = 0$, since $f(0) = 0$), that is, that

$$\frac{\left\| f^{-1}(k) - k\right\|}{\|k\|} \xrightarrow[\|k\| \to 0]{} 0,$$

or

$$\frac{\|h - k\|}{\|k\|} \xrightarrow[\|k\| \to 0]{} 0 \quad \text{if } h \leftrightarrow k.$$

But indeed

$$\frac{\|h - k\|}{\|k\|} = \underbrace{\frac{\|h - k\|}{\|h\|}}_{\overset{r=f-\mathrm{id}}{=} \|r(h)\|/\|h\| \xrightarrow[\|h\| \to 0]{} 0} \underbrace{\frac{\|h\|}{\|k\|}}_{\overset{6^\circ}{\le} 2} \xrightarrow[\|k\| \to 0]{} 0,$$

since $\|k\| \to 0 \Rightarrow \underbrace{\|h\|}_{\overset{6^\circ}{\le} 2\|k\|} \to 0$. $\rhd\!\!\rhd$ $\rhd$

Corollary 3.5.2. *Let $f = (f_1, \dots, f_n) : \mathbb{R}^n \to \mathbb{R}^n$ be continuously G-differentiable at x (this will be so, e.g., if all the partial derivatives $\partial f_j / \partial x_i$ are continuous in some neighbourhood of x), and let Jacobi matrix $\mathrm{J}_f(x)$ have non-zero determinant. Then there exists a neighbourhood U of x such that f is a homeomorphism of U onto $f(U)$, the inverse mapping f^{-1} is differentiable at $y := f(x)$, and*

$$\mathrm{J}_{f^{-1}}(y) = (\mathrm{J}_f(x))^{-1}.$$

$\lhd$ $\mathbb{R}^n$ is a Banach space, and $f'(x)$ is an isomorphism iff $\det \mathrm{J}_f(x) \neq 0$. $\rhd$

3.6 Implicite Function Theorem

An important corollary of the Inverse Function Theorem (3.5.1.) is:

Theorem 3.6.1. *Let $X, Y, Z \in \mathrm{BS}$, $F : X \times Y \to Z$. Put*

$$M := F^{-1}(0).$$

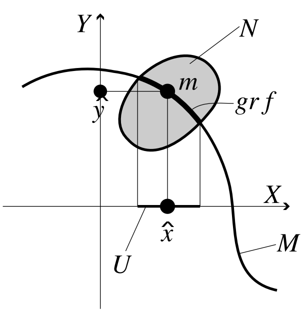
Let $m := (\hat{x}, \hat{y}) \in M$, that is, $F(\hat{x}, \hat{y}) = 0$, and let $F \in \mathrm{C}^1_G(m)$, $\partial F / \partial Y(m) \in \mathrm{Iso}(Y, Z)$ (so that Y and Z are isomorphic). Then there exists a neighbourhood U of $\hat{x}$ in X, a mapping $f : U \to Y$ and a neighbourhood N of m in $X \times Y$ such that

$$\mathrm{gr}\, f = M \cap N. \tag{1}$$

This mapping f is differentiable at $\hat{x}$, and

$$f'(\hat{x}) = -\left(\frac{\partial F(m)}{\partial Y}\right)^{-1} \circ \frac{\partial F(m)}{\partial X}. \tag{2}$$

In other words, since the condition $\mathrm{gr}\, f \subset M$ means that $F(x, f(x)) = 0$ (for $x \in U$), the theorem asserts that the equation



$$F(x, y) = 0$$

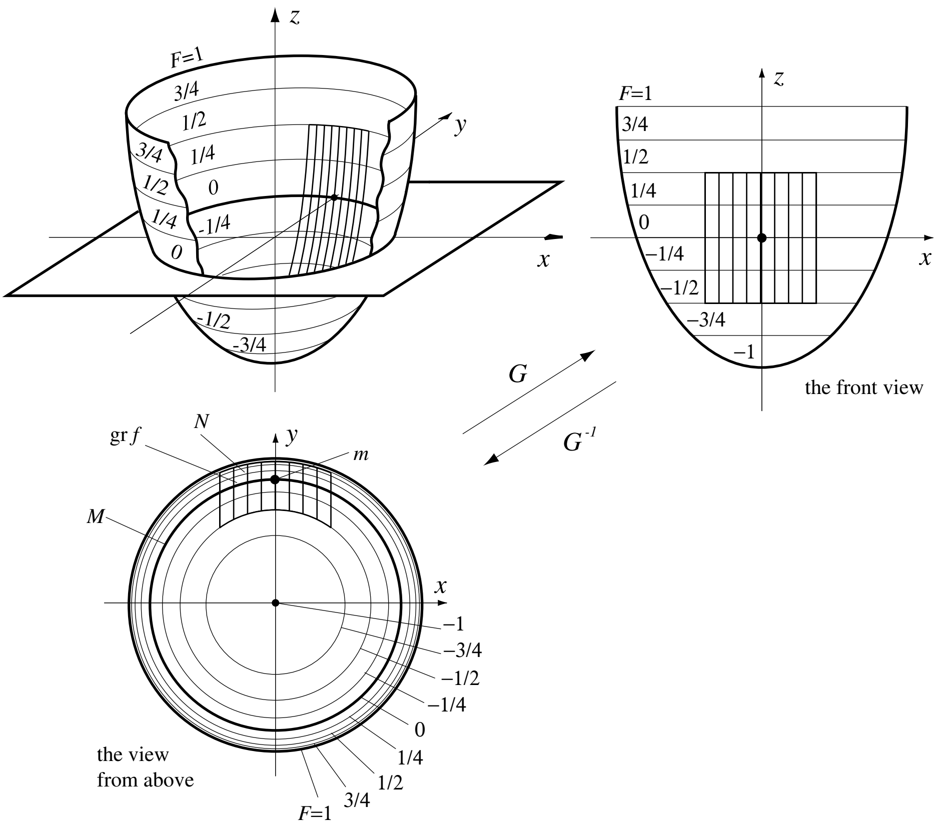
can be solved with respect to y:

$$y = f(x),$$

the resulting ("implicite") function f being differentiable at $\hat{x}$, and its derivative at $\hat{x}$ can be expressed in terms of partial derivatives of F at m.

Before the proof consider a model example.

Example. Let $X = Y = Z = \mathbb{R}$, $F(x, y) = x^2 + y^2 - 1$, $m = (0, 1)$. Here M is the unit circle with the center at 0. We have $\partial F / \partial x|_{(0,1)} = 0$, $\partial F / \partial y|_{(0,1)} = 2 \neq 0$. The set $M \cap N$ (see the picture below) is the graph of the function



$$f(x) = \sqrt{1 - x^2} \qquad (x \in U = (-\tfrac{1}{2}, \tfrac{1}{2})),$$

and

$$f'(0) = -\frac{\left.\dfrac{\partial F}{\partial x}\right|_{(0,1)}}{\left.\dfrac{\partial F}{\partial y}\right|_{(0,1)}} = 0.$$

For the proof we need some lemmas.

Lemma 3.6.2. *Let $X_1, \ldots, X_n$ be Banach spaces. Then their product $X_1 \times \ldots \times X_n$ is also a Banach space.*
$\lhd$ Completeness of all X_i implies completeness of the product, since convergence in the product is just convergence in each component. $\rhd$

Lemma 3.6.3. *Let $f = (f_1, \ldots, f_m) : X_1 \times \ldots \times X_n \to Y_1 \times \ldots \times Y_m$, and let $x \in (x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n$. Then*

$$\frac{\partial f_j(x)}{\partial X_i} = (\pi_j \circ f \circ \iota_i)'(x),$$

where ι_i and π_j are resp. the following imbeddings and projections:

$$\iota_i : X_i \to X_1 \times \ldots \times X_n, \quad \tilde{x} \mapsto (x_1, \ldots, x_{i-1}, \tilde{x}_i, x_{i+1}, \ldots, x_n),$$
$$\pi_j : Y_1 \times \ldots \times Y_m \to Y_j, \quad (y_1, \ldots, y_m) \mapsto y_j.$$

$\lhd$ This follows from the facts that

$$f_j = \pi_j \circ f$$

and $f_j \circ \iota_i$ is just the mapping f_j with all the arguments but i-th one FIXED to be equal the corresponding components of x. $\rhd$

Lemma 3.6.4. *Let X, Y, Z be normed spaces, Y and Z being isomorphic, and let*

$$A : X \times Y \to X \times Z$$

be the linear mapping represented by the matrix $\begin{pmatrix} \mathrm{id} & 0 \\ a & b \end{pmatrix}$:

$$A \sim \begin{pmatrix} \mathrm{id} & 0 \\ a & b \end{pmatrix},$$

where $\mathrm{id} = \mathrm{id}_X$, $0 \in \mathscr{L}(Y, Z)$, $a \in \mathscr{L}(X, Z)$, $b \in \mathrm{Iso}(Y, Z)$. *Then* $A \in \mathrm{Iso}(X \times Y, X \times Z)$*, and*

$$A^{-1} \sim \begin{pmatrix} \mathrm{id} & 0 \\ -b^{-1} \circ a & b^{-1} \end{pmatrix}.$$

Here

$$A \sim \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad (a_{ij} \in \mathrm{L}(X_j, Y_i)),$$

means of course that

$$Ah \in Y_1 \times Y_2 \quad \text{is represented by} \quad \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} := \begin{pmatrix} a_{11}h_1 + a_{12}h_2 \\ a_{21}h_1 + a_{22}h_2 \end{pmatrix},$$

if

$$h \in X_1 \times X_2 \quad \text{is represented by} \quad \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} \qquad (h_i \in X_i).$$

$\lhd$ The direct computation yields

$$\begin{aligned} \begin{pmatrix} \mathrm{id}_X & 0 \\ a & b \end{pmatrix} \circ \begin{pmatrix} \mathrm{id}_X & 0 \\ b^{-1} \circ a & b^{-1} \end{pmatrix} &= \begin{pmatrix} \mathrm{id} \circ \mathrm{id} - 0 \circ b^{-1} \circ a & \mathrm{id} \circ 0 + 0 \circ b^{-1} \\ a \circ \mathrm{id} - b \circ b^{-1} \circ a & a \circ 0 + b \circ b^{-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{id}_X & 0 \\ 0 & \mathrm{id}_Y \end{pmatrix} \sim \mathrm{id}_{X \times Y}, \end{aligned}$$

and analogously

$$\begin{pmatrix} \mathrm{id}_X & 0 \\ -b^{-1} \circ a & b^{-1} \end{pmatrix} \circ \begin{pmatrix} \mathrm{id}_X & 0 \\ a & b \end{pmatrix} = \begin{pmatrix} \mathrm{id}_X & 0 \\ 0 & \mathrm{id}_Y \end{pmatrix} \sim \mathrm{id}_{X \times Y}. \rhd$$

Proof of Theorem 3.6.1.

$\lhd$ 0° The idea is to extend F to a mapping G, to which we can apply the Inverse Function Theorem (see the picture in Example!).

1° Put

$$G := (\pi_1, F) : X \times Y \to X \times Z, \quad (x, y) \mapsto (x, F(x, y)).$$

Note that G does NOT change the first coordinate! (On the picture vertical lines remain vertical!)

2° We have

$$G' \overset{\text{Prod. Rule}}{=} (\pi_1', F') \overset{F \in C^1_G(m)}{\in} C^1_G(m)$$

(the projection $\pi_1 : X \times Y \to X$ is of course of class C^1 as a continuous linear mapping), and

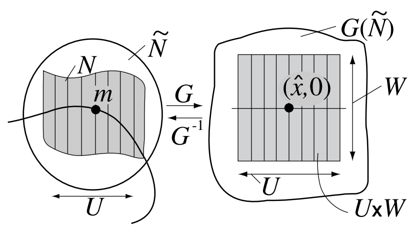
$$G'(m) \overset{\substack{\text{Theorem 2.5.1.}\\ \text{on represent.}}}{\sim} \begin{pmatrix} \frac{\partial \pi_1(m)}{\partial X} & \frac{\partial \pi_1(m)}{\partial Y} \\ \frac{\partial F(m)}{\partial X} & \frac{\partial F(m)}{\partial Y} \end{pmatrix} = \begin{pmatrix} \mathrm{id}_X & 0 \\ \frac{\partial F(m)}{\partial X} & \frac{\partial F(m)}{\partial Y} \end{pmatrix}.$$

3° By Lemma 3.6.4., $G' \in \mathrm{Iso}$, and

$$(G'(m))^{-1} \sim \begin{pmatrix} \mathrm{id}_X & 0 \\ -\left(\frac{\partial F(m)}{\partial Y}\right)^{-1} \circ \frac{\partial F(m)}{\partial X} & \left(\frac{\partial F(m)}{\partial Y}\right)^{-1} \end{pmatrix}.$$

4° By Lemma 3.6.2., both $X \times Y$ and $X \times Z$ are Banach spaces, and we can apply the Inverse Function Theorem. We conclude that there exists a neighbourhood $\tilde{N}$ of m in $X \times Y$ such that G is a homeomorphism of $\tilde{N}$ onto $G(\tilde{N})$, G^{-1} is differentiable at

$$F(m) = (\hat{x}, 0),$$



and

$$(G^{-1})'((\hat{x}, 0)) = (G'(m))^{-1}.$$

Note that G^{-1} does NOT change the first coordinate, since G does not.

5° By properties of product topology, there exist a neighbourhood U of $\hat{x}$ in X and a neighbourhood W of 0 in Z, such that $U \times W \subset G(\tilde{N})$. Put

$$N := G^{-1}(U \times W).$$

6° At last put

$$f := \pi_2 \circ G^{-1} \circ \iota_1,$$

where π_2 is the projection $X \times Y \to X$, $(x, y) \mapsto x$, and ι_1 is the imbedding $X \to X \times Z$, $x \mapsto (x, 0)$.

$$\begin{array}{ccc} X \times Y & \xleftarrow{G^{-1}} & X \times Z \\ \pi_2 \downarrow & & \uparrow \iota_1 \\ Y & \xleftarrow{f} & X \end{array}$$

7° By Change Rule, f is differentiable at $\hat{x}$, and

$$f'(\hat{x}) = (\pi_2 \circ G^{-1} \circ \iota_1)'(\hat{x}) \overset{\text{Lemma 3.6.3.}}{=} \frac{\partial (G^{-1})_2((\hat{x}, 0))}{\partial X}.$$

8° The latter partial derivative is just the (21)-element of the matrix representing

$$(G^{-1})'((\hat{x}, 0)) \overset{4°}{=} (G'(m))^{-1}.$$

9° By 3°, this element is equal to

$$-\left(\frac{\partial F(m)}{\partial Y}\right)^{-1} \circ \frac{\partial F(m)}{\partial X}.$$

Thus, (2) is proved.
10° And, by the very construction, $\mathrm{gr}\, f = M \cap N$. [Formal verification: Let $(x, y) \in M \cap N$ (so that $x \in U$). Then

$$\begin{aligned}(x, y) \in M &\Rightarrow F(x, y) = 0 \Rightarrow G(x, y) = (x, 0) \Rightarrow (x, y) = G^{-1}(x, 0) \\ &\Rightarrow y = (\underbrace{\pi_2 \circ G^{-1} \circ \imath_1}_{=f})(x) \Rightarrow (x, y) \in \mathrm{gr}\, f.\end{aligned}$$

Thus $N \cap M \subset \mathrm{gr}\, f$. Inverting the argument, we can analogously obtain the inverse inclusion.] $\triangleright$
NB Vice versa, Inverse Function Theorem can be deduced from Implicite Function Theorem. [HINT: the equation $y = f(x)$ can be written in the form $F(x, y) = 0$ with $F(x, y) := y - f(x)$.]

Chapter 4

Higher derivatives

4.1 Multilinear mappings

Let a mapping $f : X \to Y$ is differentiable (everywhere). Its derivative is a mapping from X into $\mathscr{L}(X, Y)$:

$$f' : X \to \mathscr{L}(X, Y),\ x \mapsto f'(x).$$

It is natural to define $f''(x)$ as $(f')'(x)$, so

$$f''(x) \in \mathscr{L}(X, \mathscr{L}(X, Y)).$$

It is also natural to consider the mapping

$$(h_1, h_2) \mapsto (\underbrace{f''(x) \cdot h_1}_{\in \mathscr{L}(X,Y)}) \cdot h_2,\ X \times X \to Y.$$

This mapping is BILINEAR, that is, linear in h_2 for fixed h_1 (evidently) and linear in h_1 for fixed h_2 (since $f''(x)$ is a linear mapping from X into $\mathscr{L}(X, Y)$).

Analogously higher derivatives lead to MULTILINEAR mappings. Let $X_1, \dots, X_n$ and Y be vector spaces. We say that a mapping $u : X_1 \times \dots \times X_n \to Y$ is *mutilinear* (or *n-linear*), and we write

$$u \in \mathrm{L}(X_1, \dots, X_n; Y),$$

if u is linear in each separate variable for fixed others, that is, if

$$\begin{aligned} \forall i \in \{1, \dots, n\}:\quad & u(x_1, \dots, x_{i-1}, \alpha x_i + \beta y_i, x_{i+1}, \dots, x_n) \\ &= \alpha u(x_1, \dots, x_i, \dots, x_n) + \beta u(x_1, \dots, y_i, \dots, x_n) \quad (\alpha, \beta \in \mathbb{R}). \end{aligned}$$

(For 2-linear mappings we say *bilinear*.)

For multilinear mappings one uses one of the following notations:

$$u(x_1, \dots, x_n) \equiv u \cdot x_1 \dots x_n \equiv u x_1 \dots x_n \equiv \langle u \mid x_1, \dots, x_n \rangle.$$

Examples.

1. The usual multiplication $\mathbb{R} \times \mathbb{R} \to \mathbb{R},\ (x, y) \mapsto xy$ is bilinear.

2. The multiplication $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R},\ (x, y, z) \mapsto xyz$ is 3-linear.

3. The scalar product $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},\ (x, y) \mapsto \sum_{i=1}^n x_i y_i$, where $x = (x_1, \dots, x_n),\ y = (y_1, \dots, y_n)$, is bilinear.

4. The vector product $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3,\ (\vec{x}, \vec{y}) \mapsto \vec{x} \times \vec{y}$ is bilinear.
5. The COMPOSITION

$$\mathrm{comp} : \mathrm{L}(X, Y) \times \mathrm{L}(Y, Z) \to \mathrm{L}(X, Z), (l, m) \mapsto m \circ l$$

and the EVALUATION

$$\mathrm{ev} : X \times \mathrm{L}(X, Y) \to Y, (x, l) \mapsto lx$$

are bilinear.
6. The DETERMINANT mapping

$$\det : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R},\ (\vec{x}, \vec{y}, \vec{z}) \mapsto \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}$$

($\vec{x} = (x_1, x_2, x_3), \ldots$) is 3-linear.

It is easy to verify(!) that $\mathrm{L}(X_1, \ldots, X_n; Y)$ is a VECTOR SPACE.

Operator norm

Now let X_i and Y be normed spaces. Then the vector subspace of $\mathrm{L}(X_1, \ldots, X_n; Y)$ consisting from all CONTINUOUS n-linear mappings we denote by

$$\mathscr{L}(X_1, \ldots, X_n; Y).$$

Put for each $u \in \mathrm{L}(X_1, \ldots, X_n; Y)$

$$\boxed{\|u\| := \sup_{\substack{\|x_1\| \le 1 \\ \cdots \\ \|x_n\| \le 1}} \|ux_1 \ldots x_n\|} \qquad (\textit{operator norm}).$$

Examples.
1. $\|\text{multiplication}\| = 1$;
2. $\|\text{scalar product}\| = 1$;
3. $\|\text{vector product}\| = 1$;
4. $\|\text{comp}\| \le 1$;
5. $\|\text{ev}\| \le 1$;
6. $\|\det\| = 1$.

Basic inequality. *Let $u \in \mathrm{L}(X_1, \ldots, X_n; Y)$. Then for any $(x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n$*

$$\boxed{\|ux_1 \ldots x_n\| \le \|x\|\, \|x_1\| \ldots \|x_n\|} \qquad (basic\ inequality\ (BI)).$$

$\lhd$ If $x_i = 0$ for some i then both sides are 0. Let none of x_i is 0. Then

$$\|ux_1 \ldots x_n\| \underset{\substack{u \text{ is} \\ \text{multilinear}}}{=} \underbrace{\left\| u \underbrace{\frac{x_1}{\|x_1\|}}_{\in \mathrm{B}_1^{X_1}} \ldots \underbrace{\frac{x_n}{\|x_n\|}}_{\in \mathrm{B}_1^{X_n}} \right\|}_{\underset{\substack{\text{def of} \\ \|u\|}}{\le} \|u\|} \|x_1\| \ldots \|x_n\| \le \|u\|\, \|x_1\| \ldots \|x_n\| . \rhd$$

Normed space $\mathscr{L}(X_1, \ldots, X_n; Y)$

Theorem 4.1.1. *Let $u \in \mathrm{L}(X_1, \ldots, X_n; Y)$. Then the following conditions are equivalent:*

a) $u \in \mathscr{L}(X_1,\dots,X_n;Y)$*, that is,* u *is continuous;*
b) u *is continuous at* 0*;*
c) $\|u\| < \infty$.

$\lhd$ 0° For short consider the case $n=2: u \in \mathrm{L}(X;Y;Z)$.
1° (a)$\Rightarrow$(b): obviously.
2° (b)$\Rightarrow$(a): Let (x,y) be an arbitrary point in $X \times Y$. We need to show that u is continuous at (x,y). We have

$$||\underbrace{u(x+h,y+k)}_{\overset{u \text{ is multilin.}}{=} \, u(x,y)+u(x,k)+u(h,y)+u(h,k)} - u(x,y)|| = \|uxk+uhy+uhk\| \le \underbrace{\|uxk\|}_{\boxed{1}} + \underbrace{\|uhy\|}_{\boxed{2}} + \underbrace{\|uhk\|}_{\boxed{3}}.$$

So it is sufficient to verify that $\boxed{1}, \boxed{2}, \boxed{3} \to 0$ as $\|h\|, \|k\| \to 0$. If $k=0$ then $\boxed{1}=0$, if $k \ne 0$ then

$$\boxed{1} = \left\| u\left(\sqrt{\|k\|}x\right)\left(\frac{k}{\sqrt{\|k\|}}\right)\right\|.$$

If $\|k\| \to 0$ then $\sqrt{\|k\|} \to 0$ and hence $\sqrt{\|k\|}x \to 0$; further $\left\|k/\sqrt{\|k\|}\right\| = \|k\|/\sqrt{\|k\|} = \sqrt{\|k\|} \to 0$. Thus $\boxed{1} \to 0$ as $\|k\| \to 0$, by (b).

Quite analogously $\boxed{2} \to 0$ as $\|h\| \to 0$. At last $\boxed{3} \to 0$ as $\|h\| \to 0$, $\|k\| \to 0$, by (b).
3° (b)$\Rightarrow$(c): By (b), there exists $\delta > 0$ such that

$$\|x\| \le \delta, \|y\| \le \delta \Rightarrow \|uxy\| \le 1. \tag{1}$$

Then

$$\|u\| = \sup_{\substack{\|x\|\le 1\\ \|y\|\le 1}} \underbrace{\|uxy\|}_{=\delta^{-2}\underbrace{\|u(\delta x)(\delta y)\|}_{\overset{(1)}{\le} 1 \text{ if } \|x\|\le 1, \|y\|\le 1}} \le \delta^{-2} < \infty.$$

4° (c)$\Rightarrow$(b): If $\|x\|, \|y\| \to 0$ then $\|uxy\| \overset{\text{BI}}{\le} \underbrace{\|u\|}_{\overset{(c)}{<}\infty} \|x\|\,\|y\| \to 0$. $\rhd$

Note that all the multilinear mappings from Examples 1)–6) are continuous (since all they have norms ≤ 1). As to mappings from 1)–4) and 6), their continuity follows also from

Theorem 4.1.2. *In finite-dimensional case all the multilinear mappings are continuous.*
$\lhd$ Analogously to the case of linear mappings. $\rhd$

Theorem 4.1.3. *The operator norm is really a norm in* $\mathscr{L}(X_1,\dots,X_n;Y)$.
We EVER consider $\mathscr{L}(X_1,\dots,X_n;Y)$ as a normed space with the operator norm!
$\lhd$ By Theorem 1, the operator norm is FINITE on the whole space $\mathscr{L}(X_1,\dots,X_n;Y)$, and it is easy to verify that all 3 axioms of a norm are fulfilled. $\rhd$

Canonical isomorphisms

Theorem 4.1.4. *For any natural* k *and* n*,* $k<n$*, it holds*

$$\mathscr{L}(X_1,\dots,X_n;Y) \approx \mathscr{L}(X_1,\dots,X_k;\mathscr{L}(X_{k+1},\dots,X_n;Y))$$

(the isomorphism of normed spaces), and the CANONICAL ISOMORPHISM

$$\mathscr{L}(X_1,\dots,X_n;Y)\to\mathscr{L}(X_1,\dots,X_k;\mathscr{L}(X_{k+1},\dots,X_n;Y)),\, u\mapsto\widetilde{u},$$

$$\widetilde{u}(x_1,\dots,x_k):=u(x_1,\dots,x_k,\underbrace{\cdot,\dots,\cdot}_{\substack{n-k\text{ “free”}\\ \text{arguments}}})$$

conserves the norm:

$$\|u\|=\|\widetilde{u}\|\,.$$

$\triangleleft$ 0° For short consider the case $n=2$, $k=1$. We need to verify that

$$\mathscr{L}(X_1,X_2;Y)\approx\mathscr{L}(X_1,\mathscr{L}(X_2,Y)).$$

1° ALGEBRAICAL ISOMORPHISM:

$$\mathrm{L}(X_1,X_2;Y)\overset{\text{alg.}}{\approx}\mathrm{L}(X_1,\mathrm{L}(X_2,Y)). \tag{2}$$

Put for $u\in\mathrm{L}(X_1,X_2;Y)$

$$\widetilde{u}(x_1):=u(x_1,\cdot)\qquad(\in\mathrm{L}(X_2,Y)),$$

and for $v\in\mathrm{L}(X_1.\,\mathrm{L}(X_2,Y))$

$$\widetilde{v}(x_1,x_2):=(v\cdot x_1)\cdot x_2\qquad(\in Y).$$

It is easy to see that the mapping

$$\widetilde{u}:X_1\to\mathrm{L}(X_2,Y)$$

is linear, that is, $\widetilde{u}\in\mathrm{L}(X_1,\mathrm{L}(X_2,Y))$, and the mapping

$$\widetilde{v}:X_1\times X_2\to Y$$

is bilinear, that is, $\widetilde{v}\in\mathrm{L}(X_1,X_2;Y)$, and that the mappings

$$u\mapsto\widetilde{u}\text{ and }v\mapsto\widetilde{v}$$

are linear and mutually inverse ($\approx{u}= u$, $\approx{v}= v$). Hence $u\mapsto\widetilde{u}$ is a linear bijection of $\mathrm{L}(X_1,X_2;Y)$ onto $\mathrm{L}(X_1,\mathrm{L}(X_2;Y))$, that is, (1) is true.

2° TOPOLOGICAL ISOMORPHISM: If $u\in\mathscr{L}(X_1,X_2,Y)$ then $\forall x_1\in X_1\vdots\ \widetilde{u}(x_1)=u(x_1,\cdot)\in\mathscr{L}(X_2,Y)$ (since u is continuous). Now the (linear) mapping

$$\widetilde{u}:X_1\to\mathscr{L}(X_2,Y)$$

is continuous since it has a finite norm equal to the norm of u:

$$\|\widetilde{u}\|=\sup_{\|x\|\le1}\|\widetilde{u}x_1\|=\underbrace{\sup_{\|x_1\|\le1}\sup_{\|x_2\|\le1}}_{\overset{\text{obv}}{=}\sup_{\|x_1\|\le1,\ \|x_2\|\le1}}||\underbrace{(\widetilde{u}x_1)\cdot x_2}_{=u(x_1,x_2)}||=\|u\|\,.$$

Thus, $\widetilde{u}\in\mathscr{L}(X_1,\mathscr{L}(X_2,Y))$.

Quite analogously it can be verified that if $v \in \mathscr{L}(X_1, \mathscr{L}(X_2, Y))$ then $\widetilde{v} \in \mathscr{L}(X_1, X_2; Y)$. We conclude that $u \mapsto \widetilde{u}$ is a linear bijection of $\mathscr{L}(X_1, X_2; Y)$ onto $\mathscr{L}(X_1, \mathscr{L}(X_2, Y))$. Since $\|u\| = \|\widetilde{u}\|$, both $u \mapsto \widetilde{u}$ and $v \mapsto \widetilde{v}$ have the norm 1 and hence are continuous. Thus they are isomorphisms of our normed spaces. $\triangleright$

Remark. For $X_1 = X_2 = \mathbb{R}^n$ and $Y = \mathbb{R}$, (1) is in fact the well-known (from linear algebra) isomorphism between bilinear forms and linear operators in $\mathbb{R}^n$.

Differentiation of multilinear mappings

Theorem 4.1.5. (Quasi-Leibniz Theorem (QL)). *Any mapping $u \in \mathscr{L}(X_1, \dots, X_n; Y)$ is differentiable, and its derivative is given by the formula*

$$\boxed{u'(x_1, \dots, x_n) \cdot (h_1, \dots h_n) = \sum_{i=1}^{n} u(x_1, \dots, x_{i-1}, h_i, x_{i+1}, \dots, x_n),}$$

or, more shortly

$$\boxed{u'(x_1, \dots, x_n) = \bigoplus_{i=1}^{n} u(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_n).} \tag{3}$$

(The definition of the DIRECT SUM $\oplus_{i=1}^{n} l_i \equiv l_1 \oplus \dots \oplus l_n$ see in Chapter 2.)

$\triangleleft$ This follows at once from Theorem on continuous partial derivatives. Indeed, u is linear and continuous in each its argument, hence $\forall i \in \{1, \dots, n\} \vdots$

$$\frac{\partial u}{\partial X_i}(x_1, \dots, x_n) := u(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_n) \qquad (\in \mathscr{L}(X_i, Y)).$$

Now, each partial derivative $\partial u / \partial X_i : X_1 \times \dots \times X_n \to \mathscr{L}(X_i, Y)$ is continuous as the composition of two continuous mappings:

$$\frac{\partial u}{\partial X_i} = u_i \circ \pi_i,$$

where

$$\pi_i : X_1 \times \dots \times X_n \to X_1 \times \dots \times X_i \times \dots \times X_n, (x_1, \dots, x_n) \mapsto (x_1, \dots, x_i, \dots, x_n),$$

$$u_i : X_1 \times \dots \times \overline{X_i} \times \dots \times X_n \to \mathscr{L}(X_i, Y),$$

$$(x_1, \dots, \overline{x_i}, \dots, x_n) \mapsto u(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_n);$$

π_i is continuous as any projection, and u_i is continuous by Theorem on canonical isomorphism (since u is). $\triangleright$

Remark. We can rewrite (3) so:

$$u' = \oplus \circ (u_1 \circ \pi_1, \dots, u_n \circ \pi_n),$$

where $\oplus$ denote the following mapping:

$$\oplus : \mathscr{L}(X, Y) \times \dots \times \mathscr{L}(X_n, Y) \to \mathscr{L}(X_1 \times \dots \times X_n, Y), (l_1, \dots, l_2) \mapsto \bigoplus_{i=1}^{n} l_i.$$

It is evident that this mapping $\oplus$ is linear, and it is easy to verify (!) that it is continuous.

Corollary 4.1.6. (Leibniz Theorem). *Let $f : X \to Y$ and $g : X \to Z$ be differentiable at a point x, and let $u \in \mathscr{L}(Y, Z; W)$. Then the composition $u \circ (f, g)$ is differentiable at x, and*

$$\boxed{(u \circ (f, g))'(x) = u\left(f'(x)h, g(x)\right) + u\left(f(x), g'(x)h\right).}$$

$$X \xrightarrow{(f,g)} Y \times Z \xrightarrow{u} W.$$

$\lhd$ This follows at once from Chain Rule and Quasi-Leibniz Theorem (QL):

$$(u \circ (f, g))'(x)h \overset{\substack{\text{Chain}\\\text{rule}}}{=} (u'(\underbrace{(f, g)(x)}_{=(f(x),g(x))}) \circ \underbrace{(f, g)'(x)}_{=(f'(x),g'(x))}) \cdot h$$

$$= u'(f(x), g(x)) \circ (f'(x)h, g'(x)h) \overset{\text{QL}}{=} u(f'(x)h, g(x)) + u(f(x), g'(x)h). \rhd$$

Examples.

1. If u is the usual multiplication $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, and if $X = \mathbb{R}$, we obtain the classic Leibniz rule: $(fg)' = f'g + fg'$.

2. For the mapping $q : \mathbb{R}^n \to \mathbb{R}, x \mapsto x_1^2 + \ldots + x_1^2 = x \cdot x \equiv x^2$, we have (here $f = g = \text{id}, u =$ scalar product)

$$g'(x) \cdot h = x \cdot h + h \cdot x = 2x \cdot h$$

(the first point denoting the application of a linear mapping, the other points denoting the scalar product!), so

$$g'(x) = 2x,$$

if we identify a vector x with the linear function $h \mapsto x \cdot h$. (Compare with the usual rule $(x^2)' = 2x$.)

4.2 Higher derivatives

Let a mapping $f : X \to Y$ be differentiable everywhere (or in an open set $U \subset X$). Then we can consider the derivative map

$$f' : X' \to \mathscr{L}(X, Y), x \mapsto f'(x).$$

We say that f is *two times differentiable* at a point x, and we write

$$f \in \text{Dif}^2(x),$$

if f' is differentiable at x; we define the *second derivative* $f''(x)$ of f at x as the derivative of f' at x:

$$f''(x) := (f')'(x) \qquad (\in \mathscr{L}(X, \mathscr{L}(X, Y))).$$

By induction, we put

$$f^{(n+1)}(x) := \left(f^{(n)}\right)'(x),$$

and we use in evident sense the notations

$$\text{Dif}^n(x), \qquad \text{Dif}^n(U), \qquad \text{Dif}^n.$$

Besides we put

$$\mathrm{Dif}^1 := \mathrm{Dif}, \qquad \mathrm{Dif}^\infty := \bigcap_{n=1}^{\infty} \mathrm{Dif}^n .$$

Thus we have

$$\begin{aligned} f &: X \to Y \\ f' &: X \to \mathscr{L}(X, Y) \\ f'' &: X \to \mathscr{L}(X, \mathscr{L}(X, Y)) \\ &\ \vdots \\ f^{(n)} &: X \to \mathscr{L}(X, \mathscr{L}(X, \dots, \mathscr{L}(X, Y) \dots)) \\ &\ \vdots \end{aligned}$$

The space $\mathscr{L}(X, \mathscr{L}(X, \dots, \mathscr{L}(X, Y) \dots))$ of values of the n-th derivative $f^{(n)}$ is isomorphic (by repeatedly applied Theorem on isomorphism, see 4.1) to the space of n-LINEAR mappings from $X \times \dots \times X$ (n times) into Y:

$$\mathscr{L}(X, \mathscr{L}(X, \dots, \mathscr{L}(X, Y) \dots)) \approx \mathscr{L}(\underbrace{X, \dots, X}_{n \text{ times}}; Y) =: \mathscr{L}(^n X; Y).$$

The multilinear mapping, corresponding to $f^{(n)}(x)$, is given by the rule

$$\widetilde{f^{(n)}(x)}(h_1, \dots, h_n) := (\dots((f^n(x) \cdot h_1) \cdot h_2) \dots) \cdot h_n.$$

Usually we IDENTIFY $f^{(n)}(x)$ and $\widetilde{f^{(n)}(x)}$, drop the wave and write

$$f^{(n)}(x)h_1, \dots, h_n.$$

Example. For $q : \mathbb{R}^n \to \mathbb{R}, x \mapsto x^2 := x \cdot x$ (see 4.1) we have $q'' \equiv$ (2 scalar product), that is, $\forall x \in \mathbb{R}^n \,\vdots\, q''(x)h_1h_2 = 2h_1 \cdot h_2$. (Prove!)

4.3 Rules of differentiation

They are in essence the same as for the first derivative.

Linearity. *If* $f, g \in \mathrm{Dif}^n(x)$ *then* $\forall \alpha, \beta \in \mathbb{R} \,\vdots\, \alpha f + \beta g \in \mathrm{Dif}^n(\alpha)$, *and*

$$(\alpha f + \beta g)^{(n)}(x) = \alpha f^{(n)}(x) + \beta g^{(n)}(x).$$

$\lhd$ By induction. $\rhd$

Product Rule. *Let* $f = (f_1, \dots, f_n) : X \to Y_1 \times \dots \times Y_m$. *Then* $f \in \mathrm{Dif}^n(x)$ *iff each* $f_i \in \mathrm{Dif}^n(x)$, *and*

$$f^{(n)}(x) = \left(f_1^{(n)}(x), \dots, f_m^{(n)}(x)\right).$$

$\lhd$ By induction. $\rhd$

Chain Rule. If $f \in \mathrm{Dif}^n(x)$ and $g \in \mathrm{Dif}^n(f(x))$, then $g \circ f \in \mathrm{Dif}^n(x)$. (The explicite formula for $(g \circ f)^{(n)}(x)$ is very cumbersome, and we drop it.)

$\lhd$ For simplicity we consider only the case $n = 2$. By Chain Rule for the first derivative,

$$(g \circ f)' = \mathrm{comp}(f', g' \circ f),$$

where

$$\mathrm{comp} : \mathscr{L}(X, Y) \circ \mathscr{L}(Y, Z) \to \mathscr{L}(X, Z), \quad (l, m) \mapsto m \circ l.$$

Since f and g are 2 times differentiable at x and at $f(x)$, resp., the mappings f' and g' are differentiable at x and $f(x)$, resp. The mapping comp is differentiable (everywhere) by Quasi-Leibniz Theorem. So $(g \circ f)'$ is differentiable at x by Chain and Product Rules for the first derivative. But this means that $g \circ f$ is 2 times differentiable at x. $\rhd$

Computation Rule. *Let $f : X \to Y$ be n times differentiable at x. Then for any $h_1, \ldots, h_n \in X$*

$$f^{(n)}(x)h_1, \ldots, h_n = \left.\frac{\partial^n}{\partial t_1 \ldots \partial t_n}\right|_{t_1 = \ldots = t_n = 0} f(x + t_1 h_1 + \ldots + t_n h_n)$$
$$:= \left.\frac{\partial}{\partial t_1}\right|_{t_1=0} \left(\left.\frac{\partial}{\partial t_2}\right|_{t_2=0} \left(\ldots \left(\left.\frac{\partial}{\partial t_n}\right|_{t_n=0} f(x + t_1 h_1 + \ldots + t_n h_n) \right) \ldots \right)\right).$$

For short we shall write the last expression as

$$\left.\frac{\partial}{\partial t_1}\right|_0 \ldots \left.\frac{\partial}{\partial t_n}\right|_0 f(x + t_1 h_1 + \ldots + t_n h_n).$$

$\lhd$ For simplicity consider the case $n = 2$. It holds

$$\left.\frac{\partial}{\partial t_1}\right|_0 \underbrace{\left.\frac{\partial}{\partial t_2}\right|_0 f(x + t_1 h_1 + t_2 h_2)}_{\underset{\substack{\text{C.R. for}\\ \text{the 1. der.}}}{=} f'(x+t_1h_1)h_2 \overset{\text{trick}}{=} \mathrm{ev}_{h_2} \cdot (f'(x+t_1h_1))}$$
$$= \left.\frac{\partial}{\partial t_1}\right|_0 \big(\mathrm{ev}_{h_2} \cdot f'(x + t_1 h_1)\big) \overset{\text{l−Rule}}{=} \mathrm{ev}_{h_2} \cdot \underbrace{\left.\frac{\partial}{\partial t_1}\right|_0 f'(x + t_1 h_1)}_{\underset{\substack{\text{C.R. for}\\ \text{the 1. der.}}}{=} (f')'(x) \cdot h_1}$$
$$= \mathrm{ev}_{h_2} \cdot (f''(x)h_1) = \big(f''(x)h_1\big) \cdot h_2 = f''(x)h_1 h_2. \rhd$$

(Recall that ev_h denotes the (continuous linear) mapping of evaluation at a given point h, see Chapter 1.)

l-Rules.

a) Let $X \xrightarrow{f} Y \xrightarrow{l} Z$, $f \in \mathrm{Dif}^p(x), l \in \mathscr{L}(Y, Z)$. Then

$$(l \circ f)^{(p)}(x)h_1 \ldots h_p = l \cdot \left(f^{(p)}(x)h_1 \ldots h_p \right),$$

or, shortly,

$$(l \circ f)^{(p)}(x) = l \circ (f^{(p)}(x)),$$

where we consider the p-th derivative $f^{(p)}(x)$ as a p-linear mapping.
In particular, if $X = \mathbb{R}$ then

$$(l \circ f)^{(p)}(x) = l \cdot f^{(p)}(x),$$

where we consider $f^{(p)}(x)$ as an element of Y.

b) Let $X \xrightarrow{l} Y \xrightarrow{f} Z, l \in \mathscr{L}(X,Y),\ f \in \mathrm{Dif}^{(p)}(lx)$. Then

$$(f \circ l)(x)h_1 \ldots h_p = f^{(p)}(lx)(lh_1)\ldots(lh_p).$$

$\lhd$ For short consider $p = 2$.

a)

$$\begin{aligned}(l \circ f)''(x)h_1h_2 &\overset{\text{Comp. Rule}}{=} \frac{\partial}{\partial t_1}\bigg|_0 \frac{\partial}{\partial t_2}\bigg|_0 (l \circ f)(x + t_1h_1 + t_2h_2)\\ &\overset{l-\text{rule for 1. der.}}{=} \frac{\partial}{\partial t_1}\bigg|_0 \left(l \cdot \frac{\partial}{\partial t_2}\bigg|_0 f(x + t_1h_1 + t_2h_2)\right)\\ &\overset{l-\text{rule for 1. der.}}{=} l \cdot \left(\frac{\partial}{\partial t_1}\bigg|_0 \frac{\partial}{\partial t_2}\bigg|_0 f(x + t_1h_1 + t_2h_2)\right)\\ &\overset{\text{Comp. Rule}}{=} l \cdot f''(x)h_1h_2.\end{aligned}$$

b)

$$\begin{aligned}(f \circ l)''(x)h_1h_2 &\overset{\text{Comp. Rule}}{=} \frac{\partial}{\partial t_1}\bigg|_0 \frac{\partial}{\partial t_2}\bigg|_0 \underbrace{(f \circ l)(x + t_1h_1 + t_2h_2)}_{\overset{l \text{ is linear}}{=} f(lx + t_1lh_1 + t_2lh_2)}\\ &\overset{\text{Comp. Rule}}{=} f''(lx)(lh_1)(lh_2). \rhd\end{aligned}$$

4.4 Higher partial derivatives

Let $f : X_1 \times \ldots \times X_n \to Y$. Of course we define partial derivatives of higher orders inductively:

$$\frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} := \frac{\partial}{\partial X_{i_1}}\left(\frac{\partial}{\partial X_{i_2}}\left(\ldots\left(\frac{\partial f}{\partial X_{i_p}}\right)\ldots\right)\right)(x),$$

or, shortly,

$$\frac{\partial^p}{\partial X_{i_1} \ldots \partial X_{i_p}} := \frac{\partial}{\partial X_{i_1}} \circ \cdots \circ \frac{\partial}{\partial X_{i_p}} \qquad (i_1, \ldots, i_p \in \{1, \ldots, n\}).$$

So

$$\begin{aligned}\frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} &\in \mathscr{L}(X_{i_1}, \mathscr{L}(X_{i_2}, \ldots, \mathscr{L}(X_{i_p}, Y)\ldots))\\ &\underset{\substack{\text{Th. on}\\ \text{isom.}}}{\approx} \mathscr{L}(X_{i_1}, \ldots, X_{i_p}; Y),\end{aligned}$$

and we have identity $\partial^p f(x)/\partial X_{i_1} \ldots \partial X_{i_p}$ with the corresponding p-linear mapping:

$$\frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} h_1 \ldots h_p \equiv \left(\ldots\left(\left(\frac{\partial^p f(x)}{\partial X_{i_1}} h_1\right) h_2\right)\ldots\right) h_p \qquad \left(h_k \in X_{i_k}\right).$$

As in the case of the first order, if each $X_i = \mathbb{R}$ (that is, $X_1 \times \ldots \times X_n = \mathbb{R}^n$), we put

$$\frac{\partial f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} := \left(\ldots\left(\left(\frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} \cdot 1\right) \cdot 1\right)\ldots\right) \cdot 1 \qquad (\in Y).$$

Lemma 4.4.1. *Let a mapping $f : X_1 \times \ldots \times X_n \to Y$ be p times differentiable at x. Then for any $i_1, \ldots, i_p \in \{1, \ldots, n\}$ and for any $h_k \in X_{i_k}$, $k = 1, \ldots, p$, it holds*

$$\frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} h_1 \ldots h_p = f^{(p)}(x) \hat{h_1} \ldots \hat{h_p},$$

where

$$\hat{h_k} := (0, \ldots, 0, \underset{i_k}{h_k}, 0, \ldots, 0),$$

that is, the partial derivative applied to the vectors $h_1, \ldots, h_p$ is just the "total" derivative applied to their images hat$h_1, \ldots, \hat{h}_p$ by the canonical imbeddings of X_{i_k} into the product space $X_1 \times \ldots \times X_n$.

$\lhd$ 1° We use below the following result: If $f : X \to \mathscr{L}(Y, Z)$ is differentiable at x then for any fixed $h \in Y$ the mapping $g : x \mapsto f(x)h$, $X \to Z$ is also differentiable at x, and

$$\forall k \in X \,\vdots\, g'(x)k = \big(f'(x)k\big)\, h.$$

(See Chapter 2.)

2° For short let $p = 2$. We have

$$\begin{aligned}
& f''(x)(0, \ldots, \underset{i_1}{h}, \ldots, 0)(0, \ldots, \underset{i_2}{h}, \ldots, 0) \\
& \overset{\substack{\text{Comp.}\\\text{Rule}}}{=} \frac{\partial}{\partial t_1}\Big|_0 \frac{\partial}{\partial t_1}\Big|_0 \underbrace{f(x + t_1(0, \ldots, h, \ldots, 0) + t_2(0, \ldots, k, \ldots, 0))}_{\overset{\text{Comp.Rule}}{=} (\partial f(x+t_1(0,\ldots,h,\ldots,0))/\partial X_{i_2})k} \\
& \overset{0^\circ}{=} \left(\frac{\partial}{\partial t_1}\Big|_0 \frac{\partial f(x + t_1(0, \ldots, h, \ldots, 0))}{\partial X_{i_2}}\right) k \\
& \overset{\substack{\text{Comp.}\\\text{Rule}}}{=} \left(\left(\frac{\partial}{\partial X_{i_1}} \frac{\partial}{\partial X_{i_2}}\right)(x)h\right) k \overset{\text{def}}{=} \frac{\partial^2 f(x)}{\partial X_{i_1} \cdot \partial X_{i_2}} hk. \rhd
\end{aligned}$$

Theorem 4.4.2. (on representation). *Let a mapping $f : X_1 \times \ldots \times X_n \to Y$ be p-times differentiable at x. Then its p-th derivative at x can be represented by the matrix of the partial derivatives:*

$$f^{(p)}(x) \sim \left(\frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}}\right)_{i_1,\ldots,i_p \in \{1,\ldots,n\}}$$

in sense that

$$\forall h^1, \ldots, h^p \in X_1 \times \ldots \times X_n, \quad h^k = (h^k_1, \ldots, h^k_n) \,\vdots$$

$$\boxed{f^{(p)}(x) h^1 \ldots h^p = \sum_{i_1,\ldots,i_p=1}^{n} \frac{\partial^p f(x)}{\partial X_{i_1} \ldots \partial X_{i_p}} h^1_{i_1} \ldots h^p_{i_p}.}$$

$\lhd$ In notations of the previous lemma,

$$f^{(p)}(x) h^1 \ldots h^p = (\hat{h^1_1} + \hat{h^1_n}) \ldots (\hat{h^p_1} + \hat{h^p_n})$$

$$\overset{f^{(p)}(x)\text{ is multilinear}}{=} \sum_{i_1,\dots,i_p=1}^{n} \underbrace{f^{(p)}(x)h^{1}_{i_1} \hat{+} h^{1}_{i_p}}_{\overset{4.4.1.}{=}(\partial^p(f(x))/\partial(X_{i_1}\dots X_{i_p}))h^1_{i_1}\dots h^p_{i_p}} . \triangleright$$

Corollary 4.4.3. *Let $f : \mathbb{R}^n \to Y$ be p-times differentiable at x. Then $f^{(p)}(x)$ can be represented by the following matrix with elements in Y:*

$$f^{(p)}(x) \sim \left(\frac{\partial^p f(x)}{\partial X_{i_1} \dots \partial X_{i_p}} \right)_{i_1,\dots,i_p \in \{1,\dots,n\}},$$

in the sense that for any $h^1, \dots h^p \in \mathbb{R}^n$ $\qquad (h^k = (h^k_1, \dots, h^k_n))$

$$\boxed{f^{(p)}(x)h^1 \dots h^p = \sum_{i_1,\dots,i_p=1}^{n} \underbrace{h^1_{i_1} \dots h^p_{i_p}}_{\in\mathbb{R}} \underbrace{\frac{\partial^p f(x)}{\partial X_{i_1} \dots \partial X_{i_p}}}_{\in Y}} \qquad (\in Y).$$

Here all $h^k_{i_k}$ are real numbers, and $h^1_{i_1} \dots h^p_{i_p}$ is just the usual product of real numbers.

Remark. For the case $p = 2$ and $Y = \mathbb{R}$ we obtain as the representative of $f''(x)$ a "usual" $n \times n$-matrix

$$\left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right)_{i,j \in \{1,\dots,n\}}.$$

This matrix is called the *Hesse matrix* of f at x (and its determinant is called the *Hessian* of f at x).

Example. For the mapping $q : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto x^2 \equiv x \cdot x$ we have

$$\forall x \in \mathbb{R}^n \colon\ f''(x) \sim \begin{pmatrix} 2 & & 0 \\ & \ddots & \\ 0 & & 2 \end{pmatrix} = 2 \cdot \mathbf{1}$$

where $\mathbf{1}$ denotes the unit matrix. It corresponds of course with the fact we know that $f'' \equiv 2\,\mathrm{id}$.

4.5 Class C^p

We say that a mapping $f : X \to Y$ is *p-times continuously differentiable* (resp., p times continuously differentiable *in an open set $U \subset X$* or *at a point $x \in X$*) or that f is *of class C^p* (resp., is of class C^p *in U* or *at x*), and we write

$$f \in C^p \left(\text{resp., } f \in C^p(U) \text{ or } f \in C^p(x) \right),$$

if f is p times differentiable everywhere (resp., in U or in a neighbourhood of x) and the derivative $f^{(p)}$ is continuous (resp., continuous in U or at x).

Thus,

$$\boxed{f \in C^p :\Leftrightarrow f^{(p)} \in \mathrm{Cont}.}$$

We put also

$$C^0 := \mathrm{Cont}, \qquad C^\infty := \bigcap_{p=0}^{\infty} C^p.$$

The mappings of class C^p we call also, for short, C^p*-mappings*.

Example. Any continuous linear mapping is of class C^∞. (Indeed, the first derivative is a constant mapping, and all the other derivatives are zeros.)

Lemma 4.5.1.

$$C^0 \supset C^1 \supset C^2 \supset \ldots \supset C^p \supset \ldots \supset C^\infty = \mathrm{Dif}^\infty .$$

$\lhd$ 1° $\mathrm{Dif}^{p+1} \subset C^p$. $\lhd\!\!\lhd$ $f \in \mathrm{Dif}^{p+1} \Rightarrow f^{(p)} \in \mathrm{Dif} \Rightarrow f^{(p)} \in \mathrm{Cont} \Rightarrow f \in C^p$. $\rhd\!\!\rhd$
2° $C^{p+1} \subset C^p$. $\lhd\!\!\lhd$ $f \in C^{p+1} \Rightarrow f \in \mathrm{Dif}^{p+1} \overset{1^\circ}{\Rightarrow} f \in C^p$. $\rhd\!\!\rhd$
3° $C^\infty \subset C^p$ $\lhd\!\!\lhd$ Obviously. $\rhd\!\!\rhd$
4° $C^\infty \subset \mathrm{Dif}^\infty$ $\lhd\!\!\lhd$ Obviously. $\rhd\!\!\rhd$
5° $\mathrm{Dif}^\infty \subset C^\infty$ $\lhd\!\!\lhd$ $f \in \mathrm{Dif}^\infty \Rightarrow \forall p \in \mathbb{N} \vdots\, f \in \mathrm{Dif}^{p+1} \overset{1^\circ}{\Rightarrow} \forall p \in \mathbb{N} \vdots\, f \in C^p \Rightarrow f \in C^\infty$. $\rhd\!\!\rhd$ $\rhd$

Lemma 4.5.2. *For any* $k \in \{0, 1, \ldots, p\}$

$$\boxed{f \in C^p \Leftrightarrow f^{(k)} \in C^{p-k}.}$$

$\lhd$ $\left(f^{(k)}\right)^{(p-k)} = f^{(p)}$. $\rhd$

Theorem 4.5.3. *Any continuous multilinear mapping is of class* C^∞.
$\lhd$ Use induction. For linear mappings the assertion is true, by Example above. Let our assertion is true for k-linear mappings with $k \le n-1$, and let

$$u \in \mathscr{L}(X_1, \ldots, X_n; Y).$$

By Remark after Quasi-Leibniz Theorem,

$$u' = \oplus \circ (u_1 \circ \pi_1, \ldots, u_n \circ \pi_n),$$

where u_i are continuous $(n-1)$-linear mappings, and $\oplus$ and π_i are continuous linear mappings. All these mappings are of class C^∞, by the inductive assumption and hence are infinitely differentiable. By Product and Chain Rules, u' is also infinitely differentiable. Hence u is infinitely differentiable and therefore (by Lemma 4.5.1.) is of class C^∞. $\rhd$
Remark. In fact the n-th derivative of a continuous n-linear mapping is a CONSTANT mapping, and hence all the subsequent derivatives are ZEROS:

$$u^{(n)} = \mathrm{const}, \qquad u^{(n+1)} = 0, \qquad u^{(u+2)} = 0, \ldots$$

Viz., if $u \in \mathscr{L}(X_1, \ldots, X_n; Y)$ then $\forall x \in X_1 \times \ldots \times X_n \vdots$

$$u^{(n)}(x)h^1 \ldots h^n =$$

$$= \sum_{\sigma \in \mathfrak{S}_n} u h_1^{\sigma(1)} \ldots h_n^{\sigma(n)} \quad \left(h^k = (h_1^k, \ldots, h_n^k) \in X_1 \times \ldots \times X_n\right) \tag{1}$$

where $\mathfrak{S}_n$ denotes the group of all permutations of the set $\{1, \ldots, n\}$. [This can be proved by using Computation Rule. (Prove!)] E.g., for bilinear u

$$u''(x)hk = uh_1k_2 + uk_1h_2. \tag{2}$$

Note that in the case of MULTIPLICATION ($u : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto xy$), Equation (1) follows at once from Representation Theorem:

$$u''(x) \sim \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \text{ hence } u''(x)hk = (h_1\ h_2) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \end{pmatrix} = h_1k_2 + k_1h_2.$$

Note also that it follows from (1) that $u^{(n)}(x)h^1 \ldots h^n$ does NOT change by any permutation of the vectors $h^1, \ldots, h^n$ (for bilinear case it is quite obvious, see(2)). This is no accident! See the next section.

Product

Theorem 4.5.4. $(f_1, \ldots, f_n) \in C^p \Leftrightarrow f_1, \ldots, f_m \in C^p$.
$\lhd$ This follows at once from Product rule for higher derivatives and from the topological fact that $\left(f_1^{(p)}, \ldots, f_m^{(p)}\right)$ is continuous iff each $f_i^{(p)}$ is. $\rhd$

Composition

Theorem 4.5.5. $f, g \in C^p \Rightarrow g \circ f \in C^p$.
$\lhd$ By induction. For $p = 0$ all is O.K. (the composition of continuous mappings is continuous). Let our assertion is true for $p - 1$. We have, by Chain Rule,

$$(g \circ f)' = \text{comp} \circ (f', g' \circ f).$$

The mapping comp : $(l, m) \mapsto m \circ l$ is a continuous bilinear mapping (see 4.1) and hence is of class C^∞ (by Theorem 4.5.3.). A fortiori it is of class C^{p-1}, by Lemma 4.5.1. The derivatives f' and g' are both of class C^{p-1}, by Lemma 4.5.2., and f is of class C^{p-1} by Lemma 4.5.1. Hence $g' \circ f$ is of class C^{p-1} by the induction assumption. Then, by Theorem 4.5.4. (on product), $(f', g' \circ f)$ is of class C^{p-1}. So, once again by the inductive assumption, the mapping comp $\circ(f', g' \circ f)$ is of class C^{p-1}. Thus, $(g \circ f)' \in C^{p-1}$, which means, by Lemma 4.5.2., that $g \circ f \in C^p$. $\rhd$

Case $\mathbb{R}^n \to \mathbb{R}^n$

Criterion. A mapping $(f_1, \ldots, f_m) : \mathbb{R}^n \to \mathbb{R}^m$ is of class C^p iff all the partial derivatives of the order $\leq p$ of each function f_j exist and are continuous.
$\lhd$ Analogously to the case $p = 1$. $\rhd$

4.6 Symmetry of higher derivatives

Here we prove that for C^p-mappings the derivative $f^{(p)}(x)$ is a *symmetrical* multilinear mapping.

A mapping $f : X^n \to Y$ (for arbitrary sets X and Y) is called *symmetrical* if its value does not change by any permutation of its arguments:

$$f \in \text{Sym} :\Leftrightarrow \forall x_1, \ldots, x_n \in X\ \forall \sigma \in \mathfrak{S}_n \vdots\ f(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) = f(x_1, \ldots, x_n).$$

(Recall that $\mathfrak{S}_n$ denotes the group of permutations of the set $\{1, \ldots, n\}$.) The set of all symmetrical n-linear mappings from $X^n \to Y$ (for vector spaces X and Y) we denote by

$$\mathrm{L}_{\mathrm{sym}}({}^n X; Y).$$

(Respectively, for continuous case we use the letter $\mathscr{L}$.)

Lemma 4.6.1. *Let for a mapping $\varphi : \mathbb{R}^2 \to \mathbb{R}$ the second partial derivative $\partial^2\varphi/\partial y \partial x$ is continuous at the origin. Then*

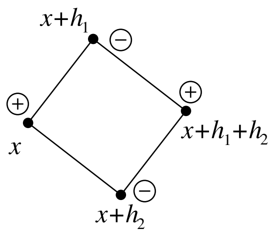
$$\frac{\partial^2\varphi(0,0)}{\partial y \partial x} = \lim_{t \downarrow 0} \frac{\Delta^2\varphi(0; (t,0), (0,t))}{t^2}.$$

Here $\Delta^2\varphi$ denotes the SECOND DIFFERENCE of φ. Recall that the *first difference* $\Delta\varphi(x; h)$ of φ at x by h is defined so:

$$\Delta\varphi(x; h) \equiv \Delta_h\varphi(x) := \varphi(x+h) - \varphi(x). \tag{1}$$

Higher differences $\Delta^n\varphi(x; h_1, \ldots, h_n)$ of φ at x by $h_1, \ldots, h_n$ are defined inductively. E.g.,

$$\begin{aligned} \Delta^2\varphi(x; h_1, h_2) &:= \Delta_{h_2}(\Delta_{h_1}\varphi)(x) = \Delta_{h_1}\varphi(x+h_2) - \Delta_{h_1}\varphi(x) \\ &\overset{1}{=} \varphi(x+h_1+h_2) - \varphi(x+h_2) - \varphi(x+h_1) + \varphi(x). \end{aligned} \tag{2}$$



Note that (as it is clear from (2)) the second difference is SYMMETRICAL in the increments:

$$\Delta^2\varphi(x; h_1, h_2) = \Delta^2\varphi(x; h_2, h_1). \tag{3}$$

$\lhd$ $t^{-2}\Delta^2\varphi(0; (t,0), (0,t)) = t^{-2}\Delta_{(t,0)}(\Delta_{(0,t)}\varphi)(0,0)$

$$\overset{\substack{\text{trick: } g(x):= \\ \Delta_{(0,t)}\varphi(x,0)= \\ \varphi(x,t)-\varphi(x,0)}}{=} t^{-2}(g(t) - g(0))$$

$$\overset{\substack{\text{Lagr. Th.;} \\ \text{for some} \\ \theta \in (0,t)}}{=} t^{-2}g'(\theta)t$$

$$\overset{\substack{g'(x)= \\ \frac{\partial\varphi(x,t)}{\partial x} - \\ \frac{\partial\varphi(x,0)}{\partial x}}}{=} t^{-1}\left(\frac{\partial\varphi(\theta,t)}{\partial x} - \frac{\partial\varphi(\theta,0)}{\partial x}\right)$$

$$\overset{\substack{\text{Lagr. Th.;} \\ \text{for some} \\ \tau \in (0,t)}}{=} t^{-1}\frac{\partial}{\partial y}\frac{\partial\varphi}{\partial x}(\theta,\tau)t$$

$$= \frac{\partial^2\varphi(\theta,\tau)}{\partial y \partial x} \xrightarrow[t \downarrow 0]{} \frac{\partial^2\varphi(0,0)}{\partial y \partial x},$$

since $\partial^2\varphi/\partial y \partial x$ is continuous at $(0,0)$ and $(\theta, \tau) \to (0,0)$ as $t \downarrow 0$. $\rhd$

Corollary 4.6.2. *Let for a function* $\varphi : \mathbb{R}^2 \to \mathbb{R}$ *the partial derivatives* $\partial^2\varphi/\partial y\partial x$ *and* $\partial^2\varphi/\partial x\partial y$ *are both continuous at* $(0,0)$. *Then they are equal there:*

$$\left.\frac{\partial^2}{\partial y\partial x}\right|_{(0,0)} = \left.\frac{\partial^2\varphi}{\partial x\partial y}\right|_{(0,0)}.$$

$\lhd$ This follows from Lemma 4.6.1., by symmetry of the second difference in the increments. $\rhd$

Lemma 4.6.3. *Let for a mapping* $\varphi : \mathbb{R}^2 \to Y$ *(where* Y *is a normed space) the partial derivatives* $\partial^2\varphi/\partial x_1\partial x_2$ *and* $\partial^2\varphi/\partial x_2\partial x_1$ *are continuous at a point* $\hat{x}$. *Then*

$$\left.\frac{\partial^2\varphi}{\partial x_1\partial x_2}\right|_{\hat{x}} = \left.\frac{\partial^2\varphi}{\partial x_2\partial x_1}\right|_{\hat{x}}.$$

$\lhd$ 1° Without loss of generality we can assume that $\hat{x} = 0$, since our function φ has the same differentiability properties at $\hat{x}$, as the function

$$\tilde{\varphi} : h \mapsto \varphi(\hat{x} + h), \mathbb{R}^2 \to Y$$

at 0. (This follows at once from Chain Rule, since the mapping $h \mapsto \hat{x} + h$ has at each point the derivative equal to id.)
2° Put

$$y := \left.\frac{\partial^2\varphi}{\partial x_1\partial x_2}\right|_0 - \frac{\partial^2\varphi}{\partial x_2\partial x_1} \qquad (\in Y).$$

Our aim is to show that $y = 0$.
3° By Lemma from Functional Analysis (see Chapter 1), there exist $l \in \mathscr{L}(Y, \mathbb{R})$ such that $\|l\| = 1$ and $ly = \|y\|$. Then

$$\|y\| = ly = l\left(\left.\frac{\partial^2\varphi}{\partial x_1\partial x_2}\right|_0 - \left.\frac{\partial^2\varphi}{\partial x_2\partial x_1}\right|_0\right) \overset{l-\text{Rule}}{=} \left.\frac{\partial^2(l\circ\varphi)}{\partial x_1\partial x_2}\right|_0 - \left.\frac{\partial^2(l\circ\varphi)}{\partial x_2\partial x_1}\right|_0 \overset{4.6.2.}{=} 0,$$

since the second partial derivative

$$\frac{\partial^2(l\circ\varphi)}{\partial x_i\partial x_j} \underset{\substack{l-\text{Rule for}\\ X=\mathbb{R}}}{=} l \circ \frac{\partial^2\varphi}{\partial x_i\partial x_j} \qquad (i, j = 1, 2)$$

is continuous at 0 together with $\partial^2\varphi/\partial x_i\partial x_j$. $\rhd$

Lemma 4.6.4. *Let* $\varphi : \mathbb{R}^n \to Y$ *be of class* C^p. *Then for any* $\sigma \in \mathfrak{S}_n$ *and any* $i_1, \ldots, i_p \in \{1, \ldots, n\}$ *it holds*

$$\frac{\partial^p\varphi}{\partial x_{i_{\sigma(1)}} \ldots x_{i_{\sigma(p)}}} = \frac{\partial^p\varphi}{\partial x_{i_1} \ldots x_{i_p}}.$$

In other words, partial derivatives do not depend on the order in which we differentiate.
$\lhd$ 1° It is sufficient to prove this for $p = 2$, since then we can TRANSPOSE any two NEIGHBOUR partial differentiations, and by such transpositions we can obtain any permutation.
2° for $p = 2$ our assertion follows from Lemma 4.6.3., since all partial derivatives at the second order of a C^2-mapping are continuous. $\rhd$

Theorem 4.6.5. *Let* $f : X \to Y$ *be of class* C^p. *Then for each* $x \in X$ *the p-th derivative* $f^{(p)}(x)$ *is symmetrical.*

$$\lhd\ f^{(p)}(x)h_{\sigma(1)}\dots h_{\sigma(p)} \overset{\substack{\text{Comp.}\\ \text{Rule}}}{=} \left.\frac{\partial^p}{\partial t_{\sigma(1)}\dots\partial t_{\sigma(n)}}\right|_0 f(x+\underbrace{t_{\sigma(1)}h_{\sigma(1)}+\dots+t_{\sigma(p)}h_{\sigma(p)}}_{\overset{\text{obv}}{=}t_1h_1+\dots+t_ph_p})$$

$$\overset{4.6.4.}{=} \left.\frac{\partial^p}{\partial t_1\dots\partial t_n}\right|_0 f(x+t_1h_1+\dots+t_ph_p) \overset{\substack{\text{Comp.}\\ \text{Rule}}}{=} f^{(p)}(x)h_1\dots h_p.\ \rhd$$

(We can apply Lemma 4.6.4., since the mapping $(t_1,\dots,t_n)\mapsto f(x+t_1h_1+\dots+t_nh_n)$, $\mathbb{R}^n\to Y$ is of class C^p as the composition of the C^∞-mapping $(t_1,\dots t_n)\mapsto x+t_1h_1+\dots+t_nh_n$ (a constant plus a (continuous) linear mapping) and $f\in C^p$.)

Corollary 4.6.6. *Let $f: X_1\times\dots\times X_n\to Y$ be of class C^p. Then partial derivatives*

$$\frac{\partial^p f}{\partial X_{i_1}\dots X_{i_p}} \qquad (i_1,\dots,i_n\in\{1,\dots,n\})$$

do NOT *depend on the order in which we differentiate.*

This means, e.g., that $\partial^2 f(x)/\partial X_1\partial X_2$ and $\partial^2 f(x)/\partial X_2\partial X_1$ define one and the same bilinear mapping $X_1\times X_2\to Y$:

$$\forall h_1\in X_1, h_2\in X_2: \frac{\partial^2 f(x)}{\partial X_1\partial X_2}h_1h_2 = \frac{\partial^2 f(x)}{\partial X_2\partial X_1}h_2h_1.$$

$\lhd$ This follows from Theorem in view of Lemma 4.4.1. $\rhd$

Remark. This corollary justifies notations of the type

$$\frac{\partial^3}{\partial X_1^2\partial X_2}.$$

4.7 Polynomials

Let X, Y be vector spaces. We say that a mapping $p: X\to Y$ is a *homogeneous polynomial* of *degree n*, and we write

$$p\in P_n(X,Y),$$

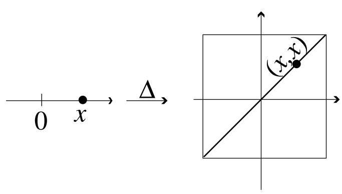
if there exists an n-linear mapping

$$u\in \mathrm{L}(^nX;Y) \qquad (=\mathrm{L}(\underbrace{X,\dots,X}_{n};Y))$$

such that

$$p = u\circ\triangle,$$

where $\triangle\equiv\triangle_n$ is the *diagonal mapping*, defined so:



$$\triangle: X\to X^n := \underbrace{X\times\dots\times X}_{n}, x\mapsto (x,\dots,x).$$

In other words,

$$p(x) = u(x,\dots,x).$$

In the case where X, Y are NORMED spaces, we say that p is a *continuous homogeneous polynomial* of degree n, and we write

$$p \in \mathcal{P}_n(X, Y),$$

if there exists $u \in \mathscr{L}(^n X; Y)$ with the same property.

If $p = u \circ \triangle$ we say that u *generates* p, or that p *is generated* by u.

We put also

$$\mathcal{P}_0(X, Y) := \{\text{all the CONSTANT mappings X} \to \text{Y}\}$$

(the polynomials of degree 0).

In what follows we consider only homogeneous polynomials, and we will drop "homogeneous".

Examples.

1. Each linear mapping is a polynomial of degree 1, that is, $P_1 = L$.

2. The power mapping $x \mapsto x^n, \mathbb{R} \to \mathbb{R}$ is a continuous polynomial of degree n.

3. The maping $(x, y) \mapsto x^3 + 4xy^2, \mathbb{R}^2 \to \mathbb{R}$ is a continuous polynomial of degree 3. $\lhd$ This polynomial is generated, e.g., by the following two 3-linear mappings $(\mathbb{R}^2)^3 \to \mathbb{R}$:

$$\begin{aligned}&((x_1, y_1), (x_2, y_2), (x_3, y_3)) \mapsto x_1x_2x_3 + \tfrac{4}{3}(x_1y_2y_3 + x_2y_3y_1 + x_3y_1y_2),\\ &((x_1, y_1), (x_2, y_2), (x_3, y_3)) \mapsto x_1x_2x_3 + 4x_1y_2y_3,\end{aligned}$$

the former being symmetrical, and the latter being not. $\rhd$

4. Each n-linear mapping is a continuous polynomial of degree n.
$\lhd$ For $n = 2$, e.g., a bilinear mapping $u : X_1 \times X_2 \to Y$ is generated by the following bilinear mapping $U : (X_1 \times X_2)^2 \to Y$:

$$U : ((x_1, y_1), (x_2, y_2)) \mapsto \tfrac{1}{2}(u(x_1, y_2) + u(x_2, y_1)). \rhd$$

5. The function $q : \mathbb{R}^n \to \mathbb{R}, x \mapsto x^2 \equiv x \cdot x = x_1^2 + \ldots + x_n^2$ is a continuous polynomial of degree 2 (generated by the scalar product).

6. More generally, for any linear operator $A : \mathbb{R}^n \to \mathbb{R}^n$, with the matrix (a_{ij}), the mapping (quadratic form)

$$\mathbb{R}^n \to \mathbb{R}, x \mapsto (Ax) \overset{\text{scal. prod.}}{\cdot} x = \sum_{i,j=1}^{n} a_{ij}x_ix_j \qquad (x = (x_1, \ldots, x_n))$$

is a polynomial of degree 2. (Prove!)

7. The function $C([0, 1]) \to \mathbb{R}, x \mapsto \int_0^1 x^2(t)\, dt$ is a polynomial of degree 2. (Prove!)

Symmetrization

For any mapping $f : X^n \to Y$, where Y is a VECTOR SPACE $(X^n := \overbrace{X \times \ldots \times X}^{n})$, we define its SYMMETRIZATION sym f by the formula

$$\forall x_1, \ldots, x_n \in X \vdots \ \boxed{(\text{sym } f)(x_1, \ldots, x_n) := \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})}.$$

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x - 2y$. Then

$$(\operatorname{sym} f)(x, y) = \tfrac{1}{2}(f(x, y) + f(y, x)) = \tfrac{1}{2}((x - 2y) + (y - 2x)) = -\frac{x + y}{2}.$$

Lemma 4.7.1. *Let* $u \in \mathrm{L}(^n; Y)$. *Then*

a) $\operatorname{sym} u \in \mathrm{L}_{\mathrm{sym}}(^n X; Y)$;

b) $u \in \mathrm{L}_{\mathrm{sym}}(^n X; Y) \Leftrightarrow \operatorname{sym} u = u$.

$\lhd$ a) For any $\tau \in \mathfrak{S}_n$ it holds

$$(\operatorname{sym} u)\left(x_{\tau(1)}, \dots, x_{\tau(n)}\right) \underset{\substack{\text{we replace } 1\\ \text{by } \sigma(1) \text{ etc}}}{\overset{\text{def}}{=}} \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} u(x_{\underbrace{\tau(\sigma(1))}_{=(\tau\circ\sigma)(1)}}, \dots, x_{\underbrace{\tau(\sigma(n))}_{=(\tau\circ\sigma)(n)}})$$

$$\underset{\substack{\text{if } \sigma \text{ runs over } \mathfrak{S}_n,\\ \tau\circ\sigma \text{ also runs over } \mathfrak{S}_n}}{\overset{\varrho:=\tau\circ\sigma}{=}} \frac{1}{n!} \sum_{\varrho \in \mathfrak{S}_n} u(x_{\varrho(1)}, \dots, x_{\varrho(n)})$$

$$\overset{\text{def}}{=} (\operatorname{sym} u)(x_1, \dots, x_n),$$

which means that $\operatorname{sym} u$ is symmetrical.

b) "$\Rightarrow$": if u is symmetrical then

$$(\operatorname{sym} u)h_1 \dots h_n \overset{\text{def}}{=} \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \underbrace{u h_{\sigma(1)} \dots h_{\sigma(n)}}_{= u h_1 \dots h_n} = \frac{1}{n!} u h_1 \dots h_n = u h_1 \dots h_n,$$

hence $\operatorname{sym} u = u$.

"$\Leftarrow$": if $\operatorname{sym} u = u$ then u is symmetrical by a). $\rhd$

Lemma 4.7.2. *If a polynomial is generated by u then it is also generated by* $\operatorname{sym} u$.

$\lhd$ Let $p = u \circ \triangle$. Then

$$((\operatorname{sym} u) \circ \triangle) = (\operatorname{sym} u)(x, \dots, x) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} u(x, \dots, x) = u(x, \dots, x) = p(x),$$

that is, $p = (\operatorname{sym} u) \circ \triangle$. $\rhd$

Lemma 4.7.3. *Each polynomial is generated by an* UNIQUE *symmetrical multilinear mapping.*

In other words, if $u_1 \circ \triangle = u_2 \circ \triangle$ and $u_1, u_2 \in \mathrm{Sym}$, then $u_1 = u_2$.

$\lhd$ For CONTINUOUS mappings this follows at once from Theorem on differention of polynomials it the next subsection, which says that if $p = u \circ \triangle$ and $u \in \mathscr{L}_{\mathrm{sym}}$, then $u = 1/n! \cdot p^{(n)}(0)$. For "algebraical case" we give below a scetch of the proof (you may omit it).

For any given $h_1, \dots, h_n$ put

$$\begin{aligned}\pi(t_1, \dots, t_n) :&= u(t_1 h_1 + \dots + t_n h_n, \dots, t_1 h_1 + \dots, t_n h_n)\\ &= t_1^n u h_1 \dots h_1 + \dots + t_n^n u h_n \dots h_n.\end{aligned}$$

Then $\partial^n \pi / \partial t_1 \dots \partial t_n|_0$ is equal to the coefficient by $t_1 t_2 \dots t_n$. This coefficient is equal, by symmetry of u, to $n! u h_1 \dots h_n$. Hence $u h_1 \dots h_n$ is uniquely defined by π. But π is uniquely defined by p, since $\pi(t_1, \dots, t_n) = p(t_1 h_1 + \dots + t_n h_n)$. $\rhd$

Corollary 4.7.4. *If a polynomial is generated by two multilinear mappings then they have one and the same symmetrization:*

$$u_1 \circ \triangle = u_2 \circ \triangle \Rightarrow \operatorname{sym} u_1 = \operatorname{sym} u_2.$$

$\lhd$ It follows from Lemmas 4.7.1.–4.7.3. $\rhd$

Differentiation of polynomials

The following theorem is a generalization of the fact that

$$(x^n)^{(k)} = n(n-1)\dots(n-k+1)x^{n-k}.$$

Theorem 4.7.5. *Let* $p \in \mathcal{P}_n(X, Y)$, $p = u \circ \triangle$, $u \in \mathscr{L}_{\text{sym}}({}^n X, Y)$. *Then*

a) *p is of class C^∞;*

b) *for any $k \in \{1, \dots, n\}$ it holds*

$$p^{(k)} \in \mathcal{P}_{n-k}(X, \mathscr{L}_{\text{sym}}({}^k X; Y)),$$

viz., for $k < n$

$$p^{(k)} = n(n-1)\dots(n-k+1)u_k \circ \triangle_{n-k}, \tag{1}$$

where

$$u_k(x_1, \dots, x_{n-k}) := u(x_1, \dots, x_{n-k}, \underbrace{\cdot, \dots, \cdot}_{k}), \tag{2}$$

that is,

$$p^{(k)}(x)h_1 \dots h_k = n(n-1)\dots(n-k+1)u(\underbrace{x, \dots, x}_{n-k}, h_1, \dots, h_k), \tag{3}$$

and for $k = n$

$$p^{(x)} \equiv n!u.$$

c) *For any natural $k > n$*

$$p^{(k)} = 0.$$

$\lhd$ 1° $p \in C^\infty$ by Theorem on composition of C^p-mappings, since $p = u \circ \triangle$, and both u (as a continuous multilinear mapping) and $\triangle$ (as a continuous linear mapping) are C^∞-mappings.

2° We have

$$\begin{aligned} p'(x)h = (u \circ \triangle)'(x)h &\overset{\substack{\text{Chain}\\\text{Rule}}}{=} u'(\triangle x) \cdot \triangle h = u'(x, \dots, x) \circ (h, \dots, h) \\ &\overset{\substack{\text{Quasi}-\\\text{Leibniz}}}{=} u(h, x, \dots, x) + \dots + u(x, \dots, x, h) \overset{u \in \text{Sym}}{=} nu(x, \dots, x, h), \end{aligned}$$

which means that for $k = 1$ the formula (1) is true. Let us fix this:

$$(u \circ \triangle_n)' = nu_1 \circ \triangle_{n-1}. \tag{4}$$

3° Let (1) is true for $1, \dots, k$. Then

$$p^{(k)} = n(n-1)\dots(n-k)u_k \circ \triangle_{n-k}.$$

Hence

$$\begin{aligned} p^{(k+1)} = (p^k)' &= n(n-1)\dots(n-k+1)(u_k \circ \triangle_{n-k})' \\ &\overset{\substack{(4),\text{applied to}\\ u_k \circ \triangle_{n-k}}}{=} n(n-1)\dots(n-k+1)(n-k)\ \underbrace{(u_k)_1}_{\overset{(2)}{=} u_{k+1}} \circ \underbrace{\triangle_{(n-k)-1}}_{=\triangle_{n-(k+1)}}, \end{aligned}$$

which means that (1) is true for $k+1$, and, by induction, b) is true.
4° Since $p^{(n)} = \text{const}$, all the subsequent derivatives are zero. $\rhd$
Remark. You can obtain (1) using Computation Rule. (Do it!)

Corollary 4.7.6. *Let p be a continuous polynomial of degree n, generated by a (continuous) symmetrical n-linear mapping u. Then $p(0) = 0$ and*

$$p^{(k)}(0) = \begin{cases} 0, & \text{if } k \neq n, \\ n!u, & \text{if } k = n. \end{cases}$$

4.8 Taylor Formula

At first a lemma:

Lemma 4.8.1. *Let a mapping $r : X \to Y$ be k times differentiable at 0, and let*

$$r(0) = 0, \qquad r'(0) = 0, \qquad r''(0) = 0, \ldots, r^{(k)}(0) = 0.$$

Then

$$r(h) = o\left(\|h\|^k\right) \left(:\Leftrightarrow \frac{r(h)}{\|h\|^k} \xrightarrow[\|h\|\to 0,\ \|h\|\neq 0]{} 0\right).$$

$\lhd$ By induction. 1° For $k = 1$ we have

$$r(\underbrace{0+h}_{=h}) = \underbrace{r(0)}_{=0} + \underbrace{r'(0)h}_{=0} + r(h)$$

(that is, r is equal to its rest term), so $r(h) = o(\|h\|)$ by the definition of differentiability.
2° Let for $k-1$ our assertion is true. Then

$$\frac{\|r(h)\|}{\|h\|^k} = \frac{\|r(h) - r(0)\|}{\|h\|^k} \overset{\text{MVT}}{=} \frac{1}{\|h\|^k} \sup_{0<t<1} \left\|r'(th)\right\| \|h\|$$
$$\overset{\text{trick}}{=} \sup_{0<t<1} t^{k-1} \frac{\left\|r'(th)\right\|}{\|th\|^{k-1}} \xrightarrow[\|h\|\to 0]{} 0,$$

by the induction assumption; indeed if $\|h\| \to 0$ then $\|th\| = |t|\,\|h\| \to 0$ uniformly in $t \in (0,1)$, and hence $\left\|r'(th)\right\| / \|th\|^{k-1} \to 0$, since r' satisfies the conditions of the lemma for $k-1$. $\rhd$

Taylor Formula

Theorem 4.8.2. *Let $f : X \to Y$ be of class C^k in $U \subset X$. Then for $x \in U$*

$$f(x+h) = f(x) + f'(x) + \tfrac{1}{2} f''(x)h^2 + \ldots + \frac{1}{k!} f^{(k)}(x)h^k + r(h), \tag{1}$$

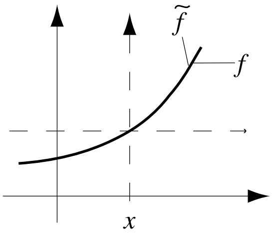
where

$$f^{(s)}(x)h^s := f^{(s)}(x)\underbrace{h\ldots h}_{s \text{ times}},$$

and

$$r(h) = o(\|h\|^k).$$

$\lhd$ 0° We can rewrite (1) in the form



$$r = \widetilde{f} - \sum_{i=1}^{k} \frac{1}{i!} p_i, \tag{2}$$

where

$$\widetilde{f}(h) := f(x+h) - f(x),$$

and p_i denotes the polynomial generated by $f^{(i)}(x)$ (which is a continuous symmetrical i-linear mapping, since f is of class C^p in U). Thus the graph of f (see the picture) is also the graph of $\tilde{f}$ when considered with respect to the translated (on the picture dotted) axes. By Lemma 4.8.1., all we need is to verify that $r(0) = 0, r'(0) = 0, \ldots, r^{(k)}(0) = 0$.
1° $r(0) = 0$, since $\widetilde{f}(0) = 0$ and each polynomial is equal to 0 at 0.
2° For any $j = 1, \ldots, k$

$$r^{(j)}(0) \overset{(2)}{=} \underbrace{\widetilde{f}^{(j)}(0)}_{\overset{\text{obv}}{=} f^{(j)}(x)} - \sum_{i=1}^{k} \frac{1}{i!} \underbrace{p_i^{(j)}(0)}_{\overset{4.7.6.}{=} \begin{cases} j!\,f^{(j)}(x) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}} = 0. \rhd$$

Case $\mathbb{R}^n \to \mathbb{R}$

Corollary 4.8.3. *Let a function $\mathbb{R}^n \to \mathbb{R}$ have continuous partial derivatives up to order k in an open set $U \subset \mathbb{R}^n$. Then for any $x \in U$*

$$\begin{aligned} f(x+h) = f(x) &+ \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} h_i + \frac{1}{2!} \sum_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j + \ldots \\ &+ \frac{1}{k!} \sum_{i_1,\ldots,i_k=1}^{n} \frac{\partial^k f(x)}{\partial x_{i_1} \ldots \partial x_{i_k}} h_{i_1} \ldots h_{i_k} + r(h), \\ &(h = (h_1, \ldots, h_n) \in \mathbb{R}^n) \end{aligned}$$

where $r(h) = o(\|h\|^k)$.

Here $\|\cdot\|$ is ANY norm in $\mathbb{R}^n$. (If $\|\cdot\|_1$ is another norm in $\mathbb{R}^n$, then $r(h) = o(\|h\|_1^k)$ also, since any two norms in $\mathbb{R}^n$ are equivalent.)

Chapter 5

Extreme Problems

I. Problems without constraints

5.1 Generalized theorem of Fermat

Definition. Let X be a topological space, and let f be a functional on X, $f : X \to \mathbb{R}$. We say that f has a *local minimum* at a point $\hat{x} \in X$, and we write

$$\hat{x} \in \mathrm{Locmin}\, f,$$

if f attains at $\hat{x}$ its minimal value in some neighbourhood of $\hat{x}$. Thus,

$$\hat{x} \in \mathrm{Locmin}\, f \;:\Leftrightarrow\; \exists U \in \mathrm{Nb}_{\hat{x}}\; \forall x \in U \,\vdots\; f(x) \geq f(\hat{x}).$$

Theorem 5.1.1. *Let X be a normed space, and let a functional $f : X \to \mathbb{R}$ has a local minimum at a point $\hat{x}$.*

a) *If for some $h \in X$ there exists $\mathrm{D}_h f(\hat{x})$, then $\mathrm{D}_h f(\hat{x}) = 0$.*
b) *If f is G-differentiable at $\hat{x}$, then $f'(\hat{x}) = 0$.*

◁ a) If $\hat{x} \in \mathrm{Locmin}\, f$ then $0 \in \mathrm{Locmin}\, \varphi$, where $\varphi : \mathbb{R} \to \mathbb{R}, t \mapsto f(\hat{x} + th)$. By the classic Fermat's theorem, $\dot{\varphi}(0) = 0$; but $\mathrm{D}_h f(\hat{x}) = \dot{\varphi}(0)$.
b) It follows from a), since $f'(\hat{x})h = \mathrm{D}_h f(x)$. ▷

Example. If $X = \mathbb{R}^n$ and f has the partial derivatives of the first order at the point $\hat{x} \in \mathrm{Locmin}\, f$, then all these partial derivatives are equal to 0:

$$\frac{\partial f(\hat{x})}{\partial x_1} = \ldots = \frac{\partial f(\hat{x})}{\partial x_n} = 0.$$

5.2 Necessary and sufficient conditions of locmin

Theorem 5.2.1. (on necessary conditions and sufficient conditions of the second order). *Let* $X \in \mathrm{NS}$, $f : X \to \mathbb{R}$, $f \in \mathbb{C}^2(\hat{x})$.

a) *(Necessary conditions) If* $\hat{x} \in \operatorname{Locmin} f$ *then* $f'(\hat{x}) = 0$*, and*

$$\forall h \in X \colon\ f''(\hat{x})h^2 \geq 0. \tag{1}$$

b) *(Sufficient conditions) If* $f'(\hat{x}) = 0$ *and* $\exists \alpha > 0$ *such that*

$$\forall h \in X \colon\ f''(\hat{x})h^2 \geq \alpha \left\| h \right\|^2, \tag{2}$$

then $\hat{x} \in \operatorname{Locmin} f$.

$\lhd$ By Taylor formula,

$$f(\hat{x}+h) = f(\hat{x}) + f'(\hat{x})h + \frac{1}{2} f''(\hat{x})h^2 + r(h), \tag{3}$$

where

$$r(h) = o(\|h\|^2). \tag{4}$$

a) Let $\hat{x} \in \operatorname{Locmin} f$. Then $f'(\hat{x}) \overset{\text{Theorem 5.1.1.}}{=} 0$. Further let $h \in X$. It holds $f(\hat{x}+th) \geq f(\hat{x})$ for all sufficiently small $t \in \mathbb{R}$. Hence,

$$\tfrac{1}{2}t^2 f''(\hat{x})h^2 + r(th) = \tfrac{1}{2} f''(\hat{x})(th)^2 + r(th) \overset{(3)}{=} \underbrace{f(\hat{x}+th) - f(\hat{x})}_{\geq 0} - \underbrace{f'(\hat{x})}_{0}(th) \geq 0 \tag{5}$$

for all sufficiently small t. But

$$r(th) = o(t^2), \tag{6}$$

since (without loss of generality $h \neq 0$)

$$\frac{|r(th)|}{t^2} = \|h\|^2 \underbrace{\frac{|r(th)|}{\|th\|^2}}_{\substack{\overset{(4)}{\to}\, 0 \\ t\to 0}} \xrightarrow[t\to 0]{} 0.$$

So (5) is possible, only if $f''(x)h^2 \geq 0$.

b) Let $f'(\hat{x}) = 0$, and let (2) be fulfilled. Then

$$f(\hat{x}+h) - f(\hat{x}) \overset{(3)}{=} \tfrac{1}{2} \underbrace{f''(\hat{x})h^2}_{\overset{(2)}{\geq} \alpha\|h\|^2} + r(h) \geq \tfrac{\alpha}{2} \|h\|^2 + o(\|h\|^2) \geq 0$$

for all sufficiently small $\|h\|$. Hence, $\hat{x} \in \operatorname{Locmin} f$. $\rhd$

Conditions (1) and (2) in (5.2.1.) are, respecively, the condition of non-negativity and the condition of strict positivity of the second derivative $f''(\hat{x})$ in the sense of the following definition:

Definition. Let $X \in \mathrm{NS}$, $u \in L(X, X; \mathbb{R})$ (bilinear functional).

a) u is said to be *non-negative* if the corresponding polynomial is non-negative, that is, if

$$\forall h \in X \colon\ u(h,h) \geq 0, \tag{7}$$

and *positive* if the corresponding polynomial is positive at any non-zero vector, that is, if

$$\forall h \in X\backslash 0 \colon\ u(h,h) > 0. \tag{8}$$

b) u is said to be *strictly positive* if

$$\exists \alpha > 0\ \forall h \in X \vdots\ u(h,h) \geq \alpha \|h\|^2 . \tag{9}$$

It is evident that

strict positivity $\Rightarrow$ positivity $\Rightarrow$ non-negativity

In general the inverse implications are not true:

(Counter-) examples.
1. The functional $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $((x_1, y_1), (x_2, y_2)) \mapsto x_1 x_2$ is non-negative, but is not positive.
2. The functional $\ell_2 \times \ell_2 \to \mathbb{R}$, $(x, y) \mapsto \sum_{i=1}^{\infty} (1/i!) x_i y_i$ is (evidently) positive, but is not strictly positive (verify!).

Finite-dimensional case

In FINITE-DIMENSIONAL *case positivity is equivalent to strict positivity: If* $u \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^n; \mathbb{R})$ *is positive, then u is strictly positive.*

$\lhd$ Denote by S the unit sphere in $\mathbb{R}^n$ (defined by the equation $\|x\| = 1$), put $p := u \circ \triangle$, and consider the restriction $p|_S$. It is clear that this restriction is continuous (since in finite dimensional case any bilinear functional is continuous). Further, S is compact, being closed ($S = \|\cdot\|^{-1}(1)$) and bounded. Hence, $p|_S$ attains its minimal value, say α. Since u is positive, we have $\alpha > 0$. Thus,

$$\|x\| = 1 \Rightarrow u(x, x) \geq \alpha > 0. \tag{10}$$

So for any $h \neq 0$

$$u(h,h) = u\left(\|h\| \frac{h}{\|h\|}, \|h\| \frac{h}{\|h\|}\right) = \|h\|^2 \underbrace{u\left(\frac{h}{\|h\|}, \frac{h}{\|h\|}\right)}_{\overset{(10)}{\geq}\ \alpha} \geq \alpha \|h\|^2 ,$$

which means that u is strictly positive. $\rhd$

Further, in FINITE-DIMENSIONAL case the positivity condition (2) takes the form

$$\forall h = (h_1, \ldots, h_n) \in \mathbb{R}^n \backslash 0 \vdots \sum_{i,j=1}^{n} \frac{\partial^2 f(\hat{x})}{\partial x_i \partial x_j} h_i h_j > 0. \tag{11}$$

This condition is none more then the condition of *positive definiteness* of Hesse matrix of the function f at the point $\hat{x}$. Thus, by 1.9, for $f : \mathbb{R}^n \to \mathbb{R}$, strict positivity of $f''(\hat{x})$ is equivalent to positive-definiteness of Hesse matrix of f at $\hat{x}$. The latter may be established with the aid of SILVESTER CRITERION from algebra:

SILVESTER CRITERION. *A squaire matrix A is positive definite iff all its principal minors* $\det A_k$ $(k = 1, \ldots, n)$ *are positive.*

$$\begin{array}{l|ccccc|}
\cline{2-6}
A_1 & a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\
A_2 & a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\
A_3 & a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
A = A_n & a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \\
\cline{2-6}
\end{array}$$

Remark. For the case of local MAXIMUM we have to replace all "$>$" in (1) and (2) by "$<$". The strict NEGATIVITY of $f''(x)$ is equivalent to NEGATIVE definiteness of the Hesse matrix A; the latter is equivalent to the conditions: $\det A_1 < 0$, $\det A_2 > 0$, $\det A_3 < 0$, $\det A_4 > 0, \dots$
$\lhd$ Apply Silvester criterion to $-A$. $\rhd$

II. PROBLEMS WITH CONSTRAINTS

5.3 Setting of the problem

Definition. Consider the following EXTREME PROBLEM WITH CONSTRAINTS (for definiteness, the case of minimum): for a given function $f : X \to \mathbb{R}$ ($X \in$ NS) and a given set $A \subset X$ (the *constraints*), to find all points in A, where the RESTRICTION $f|_A$ has its local minimum:

$$\operatorname{Locmin}(f|_A) = ? \tag{1}$$

Of course, we equippe A with the induced topology, so that

$$a \in \operatorname{Locmin} f|_A \Leftrightarrow \exists U \in \mathrm{Nb}_a(X) \forall x \in U \cap A \vdots\, f(x) \geq f(a) \qquad (a \in A).$$

If $A = X$, we obtain a problem without constraints.
Definition. By *smooth* (*extreme*) *problem* we shall mean a problem (1) with A given by an equation

$$A = g^{-1}(0), \tag{2}$$

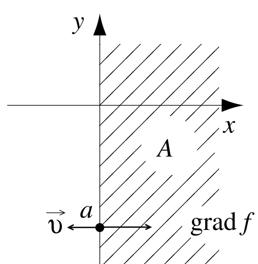
where $g : X \to Y$ is a (sufficiently) smooth (e.g., of class $\mathbb{C}^1$) mapping from our normed space X into some another normed space Y. In other words,

$$A = \{x \in X | \, g(x) = 0\}. \tag{3}$$

Example. The problem with the constraints $A \subset \mathbb{R}^2$ given as follows: $A = \{(x, y\} x = 1\}$ is a smooth problem with $Y = \mathbb{R}$ and $g(x) = x - 1$.

5.4 General (non-smooth) problems: necessary condition of locmin

At first consider a motivating example. Let $f : \mathbb{R}^2 \to \mathbb{R}$, $A = \{(x, y)| x \geq 0\}$, $f \in$ Dif, and let $a = (\hat{x}, \hat{y}) \in \operatorname{Locmin} f|_A$. Then



$$\begin{array}{ll} f'(a) = 0 & \text{if } a \in \operatorname{int} A \text{ (that is, if } \hat{x} > 0), \\ \dfrac{\partial f(a)}{\partial x} \geq 0, \ \dfrac{\partial f(a)}{\partial y} = 0 & \text{if } a \in \operatorname{fr} A \text{ (that is, if } \hat{x} = 0). \end{array}$$

This follows from a general theorem to be proved below, but it is clear by itself: in the first case we have in fact, locally (that is, in some neighbourhood of a), a problem without constraints, so Fermat theorem is applicable.
In the second case ($\hat{x} = 0$) our function f cannot have a strictly negative derivative in x at a, since it would mean that f STRICTLY decreases at a in x-direction, which contradicts

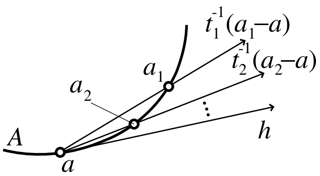
to local minimality at a. The conditions $\partial f(a)/\partial x \geq 0, \partial f(a)/\partial y = 0$ mean that the gradient $(\partial f(a)/\partial x, \partial f(a)/\partial y)$ of f and the unit outer NORMAL vector $\vec{v} = (-1, 0)$ to A at a have opposite directions. In general, as we shall see, the vector opposite to the gradient at a point of local minimum, must lie in the NORMAL CONE to A at this point.

Tangent vectors

Definition. Let X be a normed space, $A \subset X, a \in A$. We say that a vector $h \in X$ is *tangent* to A at a, and we write $h \in \mathrm{T}_a A$, if there exist a sequence $\{a_n\}$ of points in A and a sequence $\{\mathrm{T}_n\}$ of positive real number, such that a_n converges to a and $\mathrm{T}_n^{-1}(a_n - a)$ converges to h:

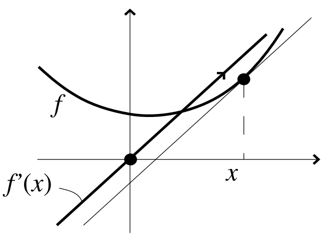
$$h \in \mathrm{T}_a A :\Leftrightarrow \exists\{\mathrm{T}_n\} \subset (0, +\infty)\exists\{a_n\} \subset A : a_n \to a, \frac{a_n - a}{\mathrm{T}_n} \to h.$$

It is obvious that always $0 \in \mathrm{T}_a A$ (take $a_n = a$) and that if $h \in \mathrm{T}_a A$ then $th \in \mathrm{T}_a A$ for any $t > 0$ (take $\mathrm{T}_n = t^{-1}\mathrm{T}_n$). This means that $\mathrm{T}_a A$ is a CONE with the vertex at 0 (the vertex belonging to the cone).



Examples.

1. For a motion $f : \underset{\text{“time”}}{\mathbb{R}} \to \mathbb{R}^3$ the velocity $f'(t)$ at a moment t is tangent to the trajectory at the point $f(t)$.



2. For a differentiable function $f : \mathbb{R} \to \mathbb{R}$ any vector of the graph of the derivative at a point x (considered as an element of $\mathscr{L}(\mathbb{R}, \mathbb{R})$) is tangent to the graph of f at the point $(x, f(x))$.

3. If A is open then ANY vector is tangent to A at each point:

$$\forall a \in A \vdots\ \mathrm{T}_a A = X \qquad \text{(verify!)}.$$

In particular

$$\forall x \in X \vdots\ \mathrm{T}_x X = X.$$

4. $\mathrm{T}_0\{0\} = \{0\}$.

5. If $Y \Subset X$ (this notation means that Y is a vector subspace in X), then

$$\forall y \in Y \vdots\ \mathrm{T}_y Y = Y.$$

Lemma 5.4.1. *Let $f : X \to Y$ be differentiable at a point a, and let*

$$a_n \to a, \frac{a_n - a}{\mathrm{T}_n} \xrightarrow[n\to\infty]{} h \qquad (a_n, h \in X, \mathrm{T}_n > 0). \tag{1}$$

Then

$$\frac{f(a_n) - f(a)}{\mathrm{T}_n} \xrightarrow[n\to\infty]{} f'(a)h. \tag{2}$$

$$\triangleleft \ \frac{f(a_n)-f(a)}{\mathrm{T}_n} \overset{f\in \mathrm{Dif}(a)}{=} \frac{(f(a)+f'(a)(a_n-a)+r(a_n-a))-f(a)}{\mathrm{T}_n}$$

$$\overset{\text{trick}}{=} \underbrace{f(a)\underbrace{\frac{a_n-a}{\mathrm{T}_n}}_{\xrightarrow{(1)} h}}_{\xrightarrow{f'(a)\in\mathrm{Cont}} f'(a)h} + \underbrace{\underbrace{\frac{r(a_n-a)}{\|a_n-a\|}}_{\substack{\|a_n-a\|\xrightarrow{(1)}0 \\ \xrightarrow{r\in\mathrm{Small}}0}} \underbrace{\frac{\|a_n-a\|}{\mathrm{T}_n}}_{\xrightarrow{(1)}\|h\|}}_{\to 0} \to f'(a)h. \ \triangleright$$

Normal vectors

NORMAL vectors to a set $A \subset X$ are not “in reality” vectors in X, they are COvectors, that is, elements of the space

$$X^* := \mathscr{L}(X, \mathbb{R})$$

(which is called the DUAL space to X; recall that $\mathbb{R}^* \approx \mathbb{R}$, $(\mathbb{R}^n)^* \approx \mathbb{R}^n$).

Definition. Let $X \in \mathrm{NS}$, $A \subset X$, $a \in A$. We say that an element $h^* \in X^*$ is *normal* to A at a, and we write

$$h^* \in \mathrm{N}_a\, A,$$

if h^* (as a linear function on X) is NON-POSITIVE on the tangent cone to A at a:

$$h^* \in \mathrm{N}_a\, A :\Leftrightarrow \forall h \in \mathrm{T}_a\, A \colon\ h^* \cdot h \le 0. \tag{3}$$

(Recall that we write $lh \equiv l \cdot x \equiv l(h)$ for linear l.)

In the case $X = \mathbb{R}^n$ you can IDENTIFY a linear function

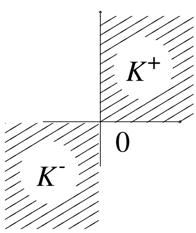
$$l(x_1, \ldots, x_n) = l_1 x_1 + \ldots + l_n x_n$$

with the vector $(l_1, \ldots, l_n)$ (in the SAME $\mathbb{R}^n$!) and think about $l \cdot h$ as about the SCALAR PRODUCT.

Once again, it is clear that $\mathrm{N}_a\, A$ is a cone, containing 0 as the vertex.

Examples.

1. For a (smooth) curve in $\mathbb{R}^3$, the normal cone at a point is the normal plane to the curve at this point.



2. Let K^+ and K^- be the positive and the negative quadrants in $\mathbb{R}^2$, resp. Then

$$\mathrm{N}_0\, K^+ = K^-, \qquad \mathrm{N}_0\, K^- = K^+.$$

3. For an OPEN set A the normal cone at any point is trivial:

$$\forall a \in A \colon\ \mathrm{N}_a\, A = \{0\} \qquad \text{(verify!)}.$$

In particular $\forall x \in X \colon\ \mathrm{N}_x\, X = \{0\}$.

4. $\mathrm{N}_0\{0\} = X$.

Necessary condition of locmin

Theorem 5.4.2. *Let* $X \in \mathrm{NS}$, $f : X \to \mathbb{R}$, $a \in A \subset X$, $f \in \mathrm{Dif}(a)$. *If* $a \in \mathrm{Locmin}\, f|_A$ *then*

$$-f'(a) \in \mathrm{N}_a\, A.$$

$\lhd$ Let us suppose that $-f'(a) \notin \mathrm{N}_a\, A$. Then, by definition, $\exists h \in \mathrm{T}_a\, A$:

$$-\ f'(a)h > 0. \tag{4}$$

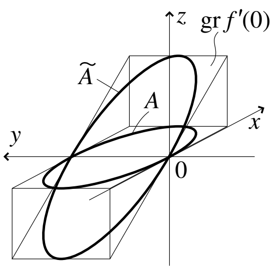
By definition of a tangent vector, $\exists a_n \to a$ $(a_n \in A)$ $\exists\, \mathrm{T}_n > 0$:

$$a_n \to a, \ \frac{a_n - a}{\mathrm{T}_n} \to h.$$

So

$$0 \overset{\substack{\text{for all sufficiently}\\ \text{great } n \text{ since}\\ a_n\to a \text{ and } a\in\mathrm{Locmin}\, f|_A\\ (\text{recall that } \mathrm{T}_n>0)}}{\leq} \frac{f(a_n)-f(a)}{\mathrm{T}_n} \xrightarrow{\text{Lm.5.4.1.}} f'(a)h \overset{(4)}{<} 0, \quad \text{a contradiction! } \rhd$$

Remark. That we require in the theorem $f \in \mathrm{Dif}(a)$, not merely $f \in \mathrm{Dif}_G(a)$, is essential, as the following counter-example shows:



Example. Let A be the circle in $\mathbb{R}^2$ shown on the picture, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by the rule

$$f(x, y) = \begin{cases} 0 \text{ if } (x, y) \in A \\ x \text{ if not.} \end{cases}$$

Then $f \in \mathrm{Dif}_G(0)$, with $f'(0) = (1, 0)$, and $0 \in \mathrm{Locmin}\, f|_A$, but

$$-f'(0) = (-1, 0) \notin \mathrm{N}_0\, A = \{(x, y) |\ x = 0\}, (= y\text{-axis}).$$

The point is that the set A is not “star-like”.

NB In this example $\mathrm{T}_0\, \mathrm{gr}\, f \neq \mathrm{gr}\, f'(0)$. $\lhd\!\lhd$ $\mathrm{gr}\, f = (A\times 0)\cup(\mathrm{gr}\, f'(0)\setminus\widetilde{A})$ (see the picture); $\mathrm{T}_0\, \mathrm{gr}\, f = \mathrm{gr}\, f'(0) \cup x$-axis $\rhd\!\rhd$

Remark. For $f \in \mathrm{Dif}$, the generalized Fermat theorem follows from Theorem 5.4.2. and Example 3 from prewious set of examples.

5.5 Smooth problems: sufficient conditions

IDEA OF LAGRANGE. The idea of Lagrange was to reduce the problem with constraints in question:

$$\mathrm{Locmin}\, f|_A = ? \qquad (A = g^{-1}(0))$$

to a problem WITHOUT constraints for some new function Φ (instead of f). The mosts simple way to construct $\Phi : X \to \mathbb{R}$, starting from f and g, is to consider some LINEAR (continuous) function $\lambda : Y \to \mathbb{R}$ and to put

$$\Phi := f + \lambda \circ g. \tag{1}$$

Such a function Φ is called LAGRANGE FUNCTION, and the functional $\lambda \in Y^*$ in (1) is called LAGRANGE MULTIPLIERS (plural!).

If $Y = \mathbb{R}$, then $\lambda \in \mathbb{R}^* \approx \mathbb{R}$ is just a number (Lagrange multiplier), if $Y = \mathbb{R}^n$, then $\lambda \in (\mathbb{R}^n)^* \approx \mathbb{R}^n$ is a vector $(\lambda_1, \dots, \lambda_n)$ (Lagrange multipliers).

Theorem 5.5.1. (sufficient conditions for a smooth problem). *Consider a smooth problem*

$$\operatorname{Locmin} f|_A = ?, \qquad A = g^{-1}(0).$$

If for some $\lambda \in Y^*$ Lagrange function $\Phi = f + \lambda \circ g$ has a local minimum at a point $a \in A$ then

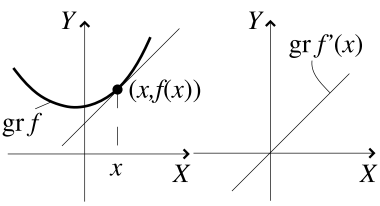
$$a \in \operatorname{Locmin} f|_A.$$

$$\lhd\ a \in \operatorname{Locmin}(f + \lambda \circ g) \overset{\text{obv.}}{\Rightarrow} a \in \operatorname{Locmin} \underbrace{(f + \lambda \circ g)|_A}_{\overset{\text{obv.}}{=} f|_A + \underbrace{(\lambda \circ g)|_A}_{\overset{g|_A = 0}{=} 0}} \Rightarrow a \in \operatorname{Locmin} f|_A.\ \rhd$$

Remark. The condition "$\exists \lambda \in Y^* : a \in \operatorname{Locmin}(f + \lambda \circ g)$" is NOT necessary for "$a \in \operatorname{Locmin} f|_A$", as the following counter-example shows:

Example. $X = \mathbb{R}^2,\ Y = \mathbb{R},\ f(x, y) = x,\ g(x, y) = x + x^2$; here $A = \{x = 0\} \cup \{x = -1\}$, $0 \in \operatorname{Locmin} f|_A$, but $\forall \lambda \in \mathbb{R} \vdots\ 0 \notin \operatorname{Locmin}(f + \lambda g)$. (Verify!)

5.6 Smooth problems: tangent cone to $A = g^{-1}(0)$



Theorem 5.6.1. (on the tangent cone to a graph). *Let $X, Y \in$ NS, $f : X \to Y$, and let $f \in \operatorname{Dif}(x)$. Then*

$$\mathrm{T}_{(x, f(x))} \operatorname{gr} f = \operatorname{gr} f'(x). \tag{1}$$

Here $\operatorname{gr} f$ dentotes the graph of f:

$$\operatorname{gr} f := \{(x,\ f(x)) |\ x \in X\} \subset X \times Y.$$

$\lhd$ 1° $\operatorname{gr} f'(x) \subset \mathrm{T}_{(x, f(x))} \operatorname{gr} f$. $\lhd\!\lhd$ Let $(h, k) \in \operatorname{gr} f'(x)$, that is, $k = f'(x)h$. Take ANY sequence $\mathrm{T}_n \downarrow 0$, and put $x_n := x + \mathrm{T}_n h$, $y_n := f(x_n)$. Then $(x_n, y_n) \in \operatorname{gr} f$, and

$$(x_n, y_n) = (x + \mathrm{T}_n h,\ f(x + \mathrm{T}_n h)) \xrightarrow[n \to \infty]{f \in \operatorname{Cont}(x), \mathrm{T}_n h \to 0} (x,\ f(x)),$$

$$\frac{(x_n, y_n) - (x,\ f(x))}{\mathrm{T}_n} = \frac{(x + \mathrm{T}_n h,\ f(x + \mathrm{T}_n h)) - (x,\ f(x))}{\mathrm{T}_n}$$
$$\overset{\text{obv}}{=} \left(h,\ \frac{f(x + \mathrm{T}_n) - f(x)}{\mathrm{T}_n h} \right) \xrightarrow[n \to \infty]{f \in \operatorname{Dif}(x)} (h,\ f'(x)h) = (h, k);$$

hence, $(h, k) \in \mathrm{T}_{(x, f(x))} \operatorname{gr} f$. $\rhd\!\rhd$
2° $\mathrm{T}_{(x, f(x))} \operatorname{gr} f \subset \operatorname{gr} f'(x)$.

$\lhd\!\lhd$ Let $(h, k) \in \mathrm{T}_{(x, f(x))} \operatorname{gr} f$, that is, $\exists \{x_n\} \subset X, \exists \{T_n\} \subset (0, +\infty)$:

$$(x_n,\ f(x_n)) \xrightarrow[n \to \infty]{} (x,\ f(x)), \tag{2}$$

$$\frac{(x_n,\ f(x_n)) - (x,\ f(x))}{T_n} \xrightarrow[n \to \infty]{} (h, k). \tag{3}$$

Relation (2) means that

$$x_n \to x, \quad f(x_n) \to f(x). \tag{4}$$

Relation (3) means that

$$\frac{x_n - x}{T_n} \to h, \quad \frac{f(x_n) - f(x)}{T_n} \to k. \tag{5}$$

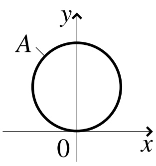
By Lemma 5.4.1., it follows from (4) and (5) that

$$\frac{f(x_n) - f(x)}{T_n} \to f'(x)h. \tag{6}$$

Comparing (6) with the second relation in (5), we conclude (by the uniqueness of limit in a Hausdorff space) that $k = f'(x)h$. But this means that $(h, k) \in \operatorname{gr} f'(x)$ ⊳⊳ ⊳
Remark. In Step 1° we used just G-differentiability of f at x, but in Step 2° we have used F-differentiability essentially, and this condition of F-differentiability is essential for validity of the theorem, as the following counter-example shows:

Example. Let A be a CIRCLE, shown on the picture, and let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by the rule

$$f(x, y) = \begin{cases} 0 \text{ if } (x, y) \notin A \\ x \text{ if } (x, y) \in A \end{cases}$$



Then $f \in \mathrm{Dif}_G(0)$, $f'(0) = 0$, $\operatorname{gr} f'(0) = \mathbb{R}^2 \times 0$, but $\mathrm{T}_0 \operatorname{gr} f = (\mathbb{R}^2 \times 0) \cup (\mathbb{R}(1, 0, 1))$. (Verify! Compare Ex. 2.9! That example also is suited!)

Theorem 5.6.2. (on the tangent cone to $g^{-1}(0)$). *Let $X, Y \in$ BS, $g : X \to Y$, $A = g^{-1}(0)$, $a \in A$ (that is, $g(a) = 0$), $g \in \mathbb{C}^1_G(a)$, $g'(a) \in$ Sur (that is, $g'(a)$ is* SURJECTIVE*: $g'(a)X = Y$), and let the kernel*

$$K := \ker g'(a) := \underbrace{(g'(a))^{-1}}_{\substack{\text{pre-image, rather} \\ \text{than the} \\ \text{inverse mapping!}}}(0) = \{k | \ g'(a)k = 0\} \tag{7}$$

SPLITS *the space X in the sense that there exists a vector subspace L in X such that:*

(i) *$K, L \in$ BS (when equipped by the induced norm);*
(ii) *$X = K \oplus L$ (that is, $X = K + L$ and $K \cap L = \{0\}$);*
(iii) *$X \approx K \times L$ (that is, more precisely, the mapping $(k, l) \mapsto k + l, K \times L \to X$) is a (linear) homeomorphism). Then*

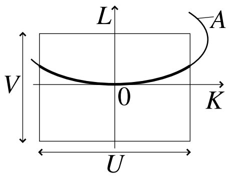
$$\boxed{\mathrm{T}_a A = \ker g'(a).} \tag{8}$$

Note, that K as the pre-image of a closed set is CLOSED, so the condition $K \in$ BS is fulfilled automatically (a closed set in a complete metric space is also complete, when equipped by the induced metric).

Note also, that in FINITE-DIMENSIONAL case ($X = \mathbb{R}^n$) ANY vector subspace splits the whole space [1], so you can forget about this condition if you wish deal just with finite-dimensional situation.

[1] ⊲ Choose an orthonormal base in K, and extend it to an orthonormal basis in $\mathbb{R}^n$; the subspace, generated by the “new” basis vectors, will be the desirable L. ⊳

$\lhd$ 1° Without loss of generality (wlog) we can assume that $a = 0$. Otherwise we consider a new mapping $\widetilde{g} : X \to Y$, defined by the rule $\widetilde{g}(h) = g(a + h)$. It is clear that $\widetilde{g}'(0) = g'(a)$, and that $\widetilde{A} := \widetilde{g}^{-1}(0) = g^{-1}(0) - a = A - a$, that is, $\widetilde{A}$ is the translation of A byf the vector $-a$, so that $\mathrm{T}_0\,\widetilde{A} = \mathrm{T}_a\,A$.



2° By (i)-(iii), we can assume that g is a mapping $K \times L \to Y$. Denote by g'_K and g'_L the corresponding partial derivatives:

$$g'_K(0)k \underset{(k\in K)}{=} g'(0)\cdot(k,0), \qquad g'_L(0)l \underset{(l\in L)}{=} g'(0)\cdot(0,l). \tag{9}$$

3° $g'_K(0) = 0$. $\lhd\!\lhd$ $K = \ker g'(0)$. $\rhd\!\rhd$

4° $g'_L(0) \in \mathrm{Sur}$. $\lhd\!\lhd$ Since $g'(0)$ is surjective

$$\forall y \in Y\ \exists (k,l) \in K \times L : y = g'(0)\cdot(k,l) \overset{(9)}{=} \underbrace{g'_K(0)k}_{\overset{3^\circ}{=}0} + g'_L(0)l = g'_L(0)l.$$

This means that $g'_L(0)$ is surjective. $\rhd\!\rhd$

5° $g'_L(0) \in \mathrm{Inj}$ (that is, is INJECTIVE). $\lhd\!\lhd$ Let $l \in L$ and $g'_L(0)l = 0$. Then

$$g'(0)\cdot(0,l) \overset{(9)}{=} \underbrace{g'_K(0)0}_{0} + \underbrace{g'_L(0)l}_{0} = 0,$$

which means that

$$(0,l) \in \ker g'(0) = K \times 0 \qquad \text{(we identify } K \text{ and } K \times 0!).$$

If follows (by (ii)) that $l = 0$. $\rhd\!\rhd$

6° By 4° and 5°, $g'_L(0) \in \mathrm{Bij}$ (is BIJECTIVE). Hence, $g'_L(0) \in \mathrm{Iso}(L, Y)$, in finite dimensional case automatically (any linear map is continuous!), and in general case by so-called Openness Principle from functional analysis.

7° By Implicit Function Theorem, $\exists U \in \mathrm{Nb}_0(K)\ \exists V \in \mathrm{Nb}_0(L)\ \exists \varphi : U \to V$:

1. $\mathrm{gr}\,\varphi = A \cap (U \times V)$,
2. $\varphi \in \mathrm{Dif}(0)$,
3. $\varphi'(0) = \underset{\text{inverse map!}}{-(g'_L(0))^{-1}} \circ \underbrace{g'_K(0)}_{\overset{3^\circ}{=}0} = 0.$

It follows from 3), that $\mathrm{gr}\,\varphi'(0) = K \times 0 = \ker g'(0)$.

8° By Theorem 5.6.1., $\mathrm{gr}\,\varphi'(0) = \mathrm{T}_0\,A$. $\rhd$

Theorem 5.6.2. says in particular that the tangent cone to A is a VECTOR SUBSPACE in X. In such a case any normal vector is ORTHOGONAL to each tangent vector:

Lemma 5.6.3. (on orthogonality). *Let $X \in \mathrm{NS}$, $A \subset X$, $a \in A$. If $\mathrm{T}_a\,A$ is a vector subspace in X, then*

$$\forall h \in \mathrm{T}_a\,A\ \forall h^* \in \mathrm{N}_a\,A \vdots\ h^* \cdot h = 0.$$

If $h^* \cdot h = 0$ then we say that h^* and h are *orthogonal* (in finite-dimensional case it is usual orthogonality).

$\lhd$ By the definition of a normal vector, $h^* \cdot h \le 0$. But we have also $-h \in \mathrm{T}_a\,A$ (since $\mathrm{T}_a\,A$ is a vector subspace), so it holds also $h^* \cdot (-h) \le 0$, that is, $h^* \cdot h \ge 0$. Hence, $h^* \cdot h = 0$. $\rhd$

Corollary 5.6.4. *In conditions of Lemma 5.6.3.,*

$$\forall h^* \in \mathrm{N}_a\,A \vdots\ \mathrm{T}_a\,A \subset \ker h^*.$$

5.7 Smooth problems: necessary condition

Theorem 5.7.1. (on Lagrange multipliers). *Let* $X, Y \in \mathrm{BS}$*, and let (see the diagram)* $A := g^{-1}(0), a \in A,\ f \in \mathrm{Dif}(a), g \in C^1_G(a), g'(a) \in \mathrm{Sur}$. *If*

$$a \in \operatorname{Locmin} f|_A, \tag{1}$$

then

$$\exists \lambda \in Y^*(= \mathscr{L}(Y, \mathbb{R})) : (\underbrace{f + \lambda \circ g}_{=:\Phi})'(a) = 0. \tag{2}$$

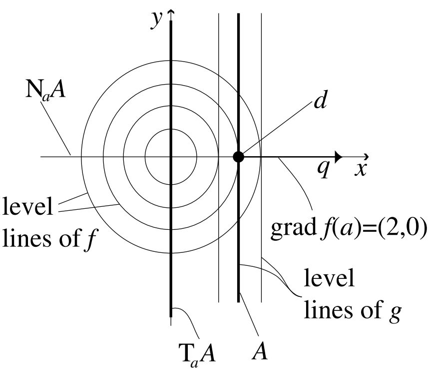
$$\begin{tikzcd} X \arrow[rr, "f"] \arrow[dr, "g"'] & & \mathbb{R} \\ & Y \arrow[ur, "\lambda"'] & \end{tikzcd}$$

Thus, the theorem says, that there exists Lagrange multipliers λ such that the corresponding Lagrange function satisfies at the point a Fermat condition.

Before the proof consider a model example:

Example. Let $X = \mathbb{R}^2,\ Y = \mathbb{R},\ f(x, y) = x^2 + y^2,\ g(x, y) = x - 1$. Here $A = \underbrace{\{(x, y)|x = 1\}}_{=:\{x=1\}}$, and $\forall b \in A\colon\ \mathrm{T}_b A = \{x = 0\},\ N_b A = \{y = 0\}$.

Now, $\lambda \in \mathbb{R}^* \approx \mathbb{R}$ may be here identified with a real numbers, so our Lagrange function has the form



$$\Phi(x, y) = x^2 + y^2 + \lambda(x - 1).$$

Condition (2) gives (for $a =: (\hat{x}, \hat{y})$)

$$\Phi'(a) = (2\hat{x} + \lambda, 2\hat{y}) = (0, 0). \tag{3}$$

Condition $a \in g^{-1}(0)$ gives

$$\hat{x} - 1 = 0. \tag{4}$$

It follows from (3) and (4) that

$$\hat{x} = 1,\ \hat{y} = 0 \text{ (that is, } a = (1, 0)),\ \lambda = -2.$$

Thus the unique candidate for a point of local minimum is $a = (1, 0)$, and it is easy to verify that really $a \in \operatorname{Locmin} f|_A$.

The necessary condition $-f'(a) \in \mathrm{N}_a A$ means here (since $\mathrm{T}_a A$ is a vector subspace of $\mathbb{R}^2$) that $\operatorname{grad} f|_a \perp \mathrm{T}_a A$. So $\operatorname{grad} f|_a$ is orthogonal at a both to the level line of f (as the gradient of f) and to the level line of g (which is just A). It follows (by the formula $\partial\varphi/\partial\nu = \operatorname{grad}\varphi \cdot \vec{\nu}$), that both f and g have zero derivative in the direction of the common tangent line to these level lines, that is, zero derivative in y: $\partial f/\partial y|_a = \partial g/\partial y|_a = 0$. Further, $\partial g/\partial x|_a = 1 \neq 0$, so far some $\widetilde{\lambda} \in \mathbb{R}$ it holds $\partial f/\partial x|_a = \widetilde{\lambda}\partial g/\partial x|_a$ (namely, for $\widetilde{\lambda} = 2$, for we have $\partial f/\partial x|_a = 2$). So, for this $\widetilde{\lambda}$, both f and $\widetilde{\lambda} g$ have ONE AND THE SAME partial derivatives at a and hence one and the same derivative at a. Hence their difference $f - \widetilde{\lambda} g$ has ZERO derivative at a.

We see that our desired Lagrange multiplier is

$$\lambda = -\widetilde{\lambda}.$$

Roughly speaking, by adding λg to f we "rotate" the graph of f around the HORIZONTAL line, passing through the point $(a, f(a))$ and parallel to the mentioned common tangent line, until we obtain the HORIZONTAL tangent plane to the graph.

THE PROOF. $\lhd$ 1° To avoid appealing functional analysis, we restrict ourselves by the FINITE-DIMENSIONAL case ($X = \mathbb{R}^n$, $Y = \mathbb{R}^m$).
2° By Theorem 5.2.1. $-f'(a) \in \mathrm{N}_a A$.
3° By Theorem 5.6.2., $\mathrm{T}_a A = \ker g'(a)$.
4° $\ker g'(a) \subset \ker f'(a)$. $\lhd\!\lhd$ $\ker g'(a) \overset{3^\circ}{=} \mathrm{T}_a A \overset{2^\circ;5.6.4.}{\subset} \ker(-f'(a)) \overset{\text{obv}}{=} \ker f'(a)$. $\rhd\!\rhd$
5° ALGEBRAICAL LEMMA (on passing through). Let X, Y, Z be vector spaces, and let $\varphi \in L(X, Z)$, $\gamma \in L(Y, Z)$. Let γ be SURJECTIVE. Then the following two conditions are equivalent:
(a) $\ker \gamma \subset \ker \varphi$;
(b) $\exists \widetilde{\lambda} \in L(Y, Z) : \varphi = \widetilde{\lambda} \circ \gamma$ (φ can be "passed through γ").

$$\begin{tikzcd} X \arrow[rr, "\varphi"] \arrow[dr, "\gamma"'] & & Z \\ & Y \arrow[ur, "\widetilde{\lambda}"'] & \end{tikzcd}$$

$\lhd\!\lhd$ (b)$\Rightarrow$(a): Let $x \in \ker \gamma$, that is, $\gamma x = 0$. Then $\varphi x \overset{\text{(b)}}{=} \widetilde{\lambda}(\underbrace{\gamma x}_{0}) = 0$, that is, $x \in \ker \varphi$.

(a)$\Rightarrow$(b): Take any element $y \in Y$. Since $\gamma \in \mathrm{Sur}$, $\exists x \in X : \gamma x = y$. Put

$$\widetilde{\gamma} y := \varphi x.$$

This definition is correct, that is, doesn't depend on the choice of x. Indeed, if we have another x' with the property $\gamma x' = y$ then

$$\gamma(x'-x) = \gamma x' - \gamma x = 0 \Rrightarrow x'-x \in \ker \gamma \overset{\text{(a)}}{\Rrightarrow} x'-x \in \ker \varphi \Rrightarrow \varphi(x'-x) = 0 \Rrightarrow \varphi x' = \varphi x.$$

By the very construction, $\varphi = \widetilde{\lambda} \circ \gamma$. $\rhd\!\rhd$
6° By 4°, we can apply 5° to the diagram

$$\begin{tikzcd} X \arrow[rr, "f'(a)"] \arrow[dr, "g'(a)"'] & & \mathbb{R} \\ & Y \arrow[ur, "\widetilde{\lambda}"'] & \end{tikzcd}$$

and conclude that $\exists \widetilde{\lambda} \in L(Y, \mathbb{R}) (= \mathscr{L}(Y, \mathbb{R})$, since $Y = \mathbb{R}^m) : f'(a) = \widetilde{\lambda} \circ g'(a)$.
7° Put $\lambda = -\widetilde{\lambda}$. Then

$$(f + \lambda \circ g)'(a) \overset{\substack{\text{Chaine}\\ \text{Rule}}}{=} f'(a) + \lambda \circ g'(a) = f'(a) - \widetilde{\lambda} \circ g'(a) \overset{6^\circ}{=} 0. \rhd$$

5.8 Problems with equations and inequalities

As an application consider a classic *extreme problem with equations and inequalities* to find local minimums of a given function $\mathbb{R}^n \to \mathbb{R}$ on the set

$$A = \{x \in \mathbb{R}^n \mid g_1(x) = 0, \ldots, g_k(x) = 0; g_{k+1}(x) \geq 0, \ldots, g_l(x) \geq 0\}.$$

(All the function are supposed to be sufficiently smooth.)
Description of a method. At a point $a \in \mathrm{Locmin}\, f|_A$ we have for $i = k+1, \dots, l$ either $g_i(a) = 0$ or $g_i(a) > 0$.

According to which of these two possibilities is realized, there 2^{l-k} possibilities. A method of solution the extreme problem is to consider one by another all the possibilities and apply to each of them Theorem on Lagrange multipliers (TLM) with an appropriate g.

We illustrate this method on the following simple example:

Example. Let
$$A = \{x |\ g_1(x) = 0,\ g_2(x) \geq 0\}.$$
Put
$$A_1 := g_1^{-1}(0),\ A_2 := g_2^{-1}(0),\ B_2 := g_2^{-1}((0, +\infty)).$$
The sets A_1 and A_2 are closed (as the pre-images of the closed set $\{0\}$), and the set B_2 is open (as the pre-image of the open set $(0, +\infty)$). It is clear that
$$A = (A_1 \cap A_2) \cup (A_1 \cap B_2),$$
the two intersections being disjoint. Let $a \in \mathrm{Locmin}\, f|_A$.

There are two possibilities: 1) $a \in A_1 \cap A_2$; 2) $a \in A_1 \cap B_2$.

In the first case
$$\left.\begin{matrix} a \in \mathrm{Locmin}\, f|_A \\ a \in A_1 \cap A_2 \end{matrix}\right\} \overset{A_1 \cap A_2 \subset A}{\Rightarrow} a \in \mathrm{Locmin}\, f|_{\underbrace{A_1 \cap A_2}_{\overset{\text{obv}}{=}(g_1, g_2)^{-1}(0)}}.$$
Thus we can apply TLM with $g = (g_1, g_2) : \mathbb{R}^n \to \mathbb{R}^2$.

In the second case
$$\left.\begin{matrix} a \in \mathrm{Locmin}\, f|_A \\ a \in A_1 \cap B_2 \end{matrix}\right\} \overset{A_1 \cap B_2 \subset A}{\Rightarrow} a \in \mathrm{Locmin}\, f|_{A_1 \cap B_2} \Rightarrow a \in \mathrm{Locmin}\, f|_{\underbrace{A_1}_{g_1^{-1}(0)}}.$$
(Proof of the last implication: since $B_2 \in \mathrm{Op}$, there exists $(U \in \mathrm{Nb}_a(\mathbb{R}^n) : U \subset B_2 \Rightarrow (A_1 \cap B_2) \cap U = A_1 \cap U$.)

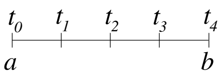
Thus we can apply TLM with $g = g_1 : \mathbb{R}^n \to \mathbb{R}$.

Chapter 6

Riemann integral in $\mathbb{R}^n$

6.1 Partitions and cubes

A *partition of a (bounded closed) interval* $I = [a, b]$ is a finite sequence $p = (t_0, t_1, \ldots, t_k)$, such that



$$a = t_0 \le t_1 \le \ldots \le t_k = b.$$

In such a case we write

$$p \in \operatorname{Part} I.$$

We say that the intervals $J_i = [t_{i-1}, t_i]$ are the *intervals of the partition* p, and we write

$$J_i \in \operatorname{Intv} p.$$

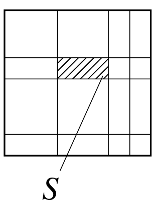
A *cube* Q in $\mathbb{R}^n$ is a product $I_1 \times \ldots \times I_n$ of n intervals $I_i = [a_i, b_i]$ (maybe $a_i = b_i$ for some i), we write

$$Q \in \operatorname{Cube} \mathbb{R}^n.$$

The *volume* of a cube is defined as the product of the lengths of its edges:

$$\operatorname{vol} Q := (b_1 - a_1) \ldots (b_n - a_n).$$

For example, any point $x \in \mathbb{R}^n$ considered as a one-point set $\{x\}$ is a cube of zero volume. A *partition* P *of a cube* $Q = I_1 \times \ldots \times I_n$ is a sequence $(p_1, \ldots, p_n)$, where p_i is a partition of the interval I_i:



$$P \in \operatorname{Part}(I_1 \times \ldots \times I_n) :\Leftrightarrow P = (p_1, \ldots p_n),\ p_i \in \operatorname{Part} I_i.$$

A *cube* S *of a partition* P is a product $J_1 \times \ldots \times J_n$, where each J_i is an interval of the partition p_i:

$$S \in \operatorname{Cube} P :\Leftrightarrow S = J_1 \times \ldots \times J_n,\ J_i \in \operatorname{Intv} p_i.$$

Let $P = (p_1, \ldots, p_n)$ and $P' = (p'_1, \ldots, p'_n)$ be two partitions of a cube Q. We say that P' is a *refinement* of P and we write

$$P' \succ P$$

if for such i the sequence p_i is a subsequence of p'_i.

6.2 Riemann integral

Let $f : M \to \mathbb{R}$, $M \subset \mathbb{R}^n$. In this chapter we ALWAYS suppose that f is *bounded*, that is, its image $f(M)$ is a bounded subset of $\mathbb{R}$:

$$f \in \mathrm{Bd}_M :\Leftrightarrow f(M) \in \mathrm{Bd}(\mathbb{R}).$$

If $f \in \mathrm{Bd}_M$ then we can EXTEND f to a bounded function on the WHOLE $\mathbb{R}^n$ by putting

$$f(x) = 0 \text{ for } x \in \mathbb{R}^n \setminus M.$$

So without loss of generality (wlog) we can (and we shall) assume that our functions are defined on the whole space.

For a given cube Q in $\mathbb{R}^n$ and a given partition P of Q we define the *lower sum* $\mathrm{L}_P\, f$ and the *upper sum* $\mathrm{U}_P\, f$ of f, corresponding to P, by the formulas

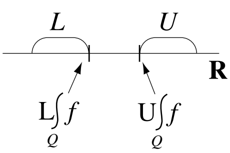
$$\mathrm{L}_P\, f := \sum_{S \in \mathrm{Cube}\, P} (\inf_S f) \operatorname{vol} S, \qquad \mathrm{U}_P\, f := \sum_{S \in \mathrm{Cube}\, P} (\sup_S f) \operatorname{vol} S.$$

(Here, e.g., $\inf_S f$ denotes the infimum of f on S, that is, $\inf(f(S))$.) By boundeness of f, both the $\inf_S f$ and $\sup_S f$ are ever FINITE.

The *lower integral* of f over Q is defined as the SUPREMUM of all lower sums, and the *upper integral* as the INFIMUM of all upper sums:

$$\mathrm{L}\int_Q f := \sup \underbrace{\{\mathrm{L}_P\, f \,|\, P \in \mathrm{Part}\, Q\}}_{=:L}, \qquad \mathrm{U}\int_Q f := \inf \underbrace{\{\mathrm{U}_P\, f \,|\, P \in \mathrm{Part}\, Q\}}_{=:U}.$$

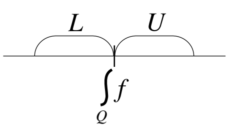
As we shall see in a minute (Lemma 6.3.3.), the set L lies TO THE LEFT of the set U, so both integrals are finite, and the lower one is less (by "less" we mean "$\leq$", for "$<$" we say "strictly less"):



$$\mathrm{L}\int_Q f \leq \mathrm{U}\int_Q f.$$

We say that f is *integrable over* Q *in the sense of Riemann* if the lower sum is EQUAL to the upper one:

$$f \in (R)\,\mathrm{Int}_Q :\Leftrightarrow \mathrm{L}\int_Q f = \mathrm{U}\int_Q f.$$



In such a case this common value is called the *Riemann integral* of f over Q and is denoted by

$$(R)\int_Q f \qquad \text{or} \qquad (R)\int_Q f(x_1, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

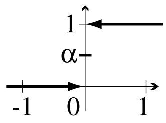
As a rule we shall drop (R) and "in the sense of Riemann."

Examples.

1. $f = \mathrm{const} = c$; $\int_Q c = c \operatorname{vol} Q$. $\lhd$ For any partition P of Q

$$\mathrm{L}_P\, c = \sum_{S \in \mathrm{Cube}\, P} c \operatorname{vol} S = c \sum_{S \in \mathrm{Cube}\, P} \overset{\text{obv.}}{=} c \operatorname{vol} Q,$$

and analogously for $\mathrm{U}_P\, c$. $\rhd$

2. $f =$ 

$\int_{[-1,1]} f = 1$ (does not depend on the value α at 0). ◁ Exercise. ▷

3. The *Dirichlet function* $f_{\mathrm{Dir}} : \mathbb{R} \to \mathbb{R}$, defined by the rule

$$f_{\mathrm{Dir}}(x) = \begin{cases} 1 \text{ if } x \notin \mathbb{Q} \\ 0 \text{ if } x \in \mathbb{Q} \end{cases}$$

is NOT integrable over, say, $[0, 1]$. ◁ $\forall P \in \mathrm{Part}[0, 1] \,\vdots\, \mathrm{L}_P\, f_{\mathrm{Dir}} = 0, \mathrm{U}_P\, f_{\mathrm{Dir}} = 1$, so $\mathrm{L}\int_{[0,1]} f_{\mathrm{Dir}} = 0, \mathrm{U}\int_{[0,1]} f_{\mathrm{Dir}} = 1$. ▷

4. Any function CONTINUOUS on a cube (that is, in each point of this cube) IS integrable over this cube. This follows from the LEBESGUE THEOREM below. For $n = 1$ we obtain the classic integral of one-dimensional analysis.

The Dirichlet function from example 3 is an example of so called indicator functions:

Definition. The *indicator* (or *characteristic*) *function* of a subset M of a set X is defined by the rule

$$\chi_M(x) := \begin{cases} 1 \text{ if } x \in M \\ 0 \text{ if } x \notin M \end{cases}$$

6.3 Criterion of existence of Riemann integral

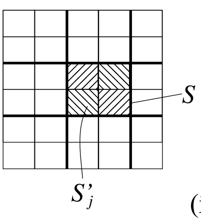
Let $f \in \mathrm{Bd}(\mathbb{R}^n)$, $Q \in \mathrm{Cube}\,\mathbb{R}^n$.

Lemma 6.3.1. $\forall P \in \mathrm{Part}\, Q \,\vdots\, \mathrm{L}_p\, f \le \mathrm{U}_P\, f$.

◁ $\forall S \in \mathrm{Cube}\, P \,\vdots\, \inf_S f \le \sup_S f$. ▷

Lemma 6.3.2. *If* $P, P' \in \mathrm{Part}\, Q$ *and* $P' \succ P$ *then*

$$\mathrm{L}_P\, f \le \mathrm{L}_{P'}\, f \le \mathrm{U}_{P'}\, f \le \mathrm{U}_P\, f.$$



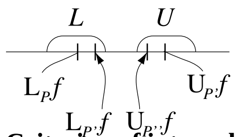
◁ The middle inequality is true by Lemma 6.3.1. Let us prove the left one. Any cube S of the partition P is built from some cubes $S'_1, \ldots, S'_k$ of the partition P' (k depends on S), and so

$$\mathrm{vol}\, S = \mathrm{vol}\, S'_1 + \ldots + \mathrm{vol}\, S'_k;$$

hence

$$(\inf_S f)\,\mathrm{vol}\, S' = \underbrace{(\inf_S f)}_{\le \inf_{S'_1} f}\ \mathrm{vol}\, S'_1 + \ldots + \underbrace{(\inf_S f)}_{\le \inf_{S'_k} f}\ \mathrm{vol}\, S'_k$$

If we sum these inequalities over all $S \in \mathrm{Cube}\, P$, we obtain $\mathrm{L}_P\, f \le \mathrm{L}_{P'}\, f$. The right inequality may be proved analogically. ▷

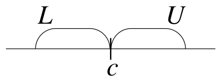


Lemma 6.3.3. $\forall P, P' \in \operatorname{Part} Q \vdots\ \mathrm{L}_P f \le \mathrm{U}_{P'} f$.
$\lhd$ Take a partition P'' of Q such that $P'' \succ P$ and $P'' \succ P'$. Then

$$\mathrm{L}_P f \overset{6.3.2.}{\le} \mathrm{L}_{P''} f \overset{6.3.1.}{\le} \mathrm{U}_{P''} f \overset{6.3.2.}{\le} \mathrm{U}_{P'} f. \rhd$$

Criterion of integrability. *A bounded function $f : \mathbb{R}^n \to \mathbb{R}$ is integrable over a cube Q in $\mathbb{R}^n$ if and only if*

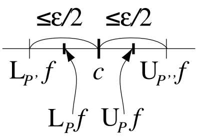
$$\forall \varepsilon > 0\ \exists P \in \operatorname{Part} Q : (0 \le) \underbrace{\mathrm{U}_P f - \mathrm{L}_P f}_{=:\triangle_P f} \le \varepsilon. \tag{1}$$



$\lhd$ 1° Let $f \in \operatorname{Int}_Q$, that is $\sup L = \inf U = c$, where

$$L := \{\mathrm{L}_P f | P \in \operatorname{Part} Q\}, \qquad U := \{\mathrm{U}_P f | P \in \operatorname{Part} Q\}.$$

Let $\varepsilon > 0$ be given. By the definitions of supremum and infimum



$$\exists P' \in \operatorname{Part} Q : c - \mathrm{L}_{P'} f \le \frac{\varepsilon}{2}, \tag{2}$$

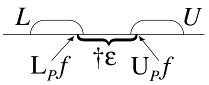
$$\exists P'' \in \operatorname{Part} Q : \mathrm{U}_{P''} f - c \le \frac{\varepsilon}{2}. \tag{3}$$

Let P be a refinement both of P' and P''. Then

$$\mathrm{U}_P f - \mathrm{L}_P f \overset{6.3.2.}{\le} \mathrm{U}_{P''} f - \mathrm{L}_{P'} f \overset{6.3.1.,6.3.2.}{\le} \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \qquad \text{O.K.}$$

2° Vice verse, let (1) be true. Then

$$\underbrace{\inf \mathrm{U}}_{\le \mathrm{U}_P f} - \underbrace{\sup \mathrm{L}}_{\ge \mathrm{L}_P f} \le \mathrm{U}_P f - \mathrm{L}_P f \le \varepsilon.$$



Since ε was arbitrary we conclude that $\inf \mathrm{U} - \sup \mathrm{L} \le 0$, that is,

$$\inf \mathrm{U} \le \sup \mathrm{L}\,.$$

But by Lemma 6.3.3.,

$$\inf \mathrm{U} \ge \sup \mathrm{L}\,.$$

Hence,

$$\inf \mathrm{U} = \sup \mathrm{L},$$

which means that $f \in \operatorname{Int}_Q$. $\rhd$

Remark that the difference $\mathrm{U}_P f - \mathrm{L}_P f$ which appears in Criterium, may be written in the form

$$\Delta_P f = \mathrm{U}_P f - \mathrm{L}_P f = \sum_{S \in \operatorname{Cube} P} (\sup_S f - \inf_S f) \operatorname{vol} S.$$

This justifies the following
Definition. Let $f \in \mathrm{Bd}_{\mathbb{R}^n}$, $M \subset \mathbb{R}^n$. We define the *oscillation* of the function f on the set M so:

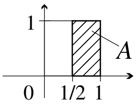
$$\Omega_M f := (\sup_M f) - (\inf_M f).$$

If $M = \mathbb{R}^n$ we omit M in the notation.

Example. $\Omega \sin = 2$.
Thus
$$\Delta_P f = \sum_{S \in \mathrm{Cube}\, P} (\Omega_S f) \operatorname{vol} S.$$

Lemma 6.3.4. (on monotony). *If* $P' \succ P$ *then* $\Delta_{P'} f \leq \Delta_P f$.
$\lhd$ It follows at once from Lemma 6.3.2. $\rhd$

Exercises.



1. $\int_{[0,1]^2} \chi_A = \frac{1}{2}$ (not depending on taking A WITH the boundary or WITHOUT).

2. $\int_{[a,b]} \mathrm{id} \left(= \int_a^b x\, dx\right) = \frac{1}{2}(b^2 - a^2)$ (do not use Newton-Leibniz formula!).
3. If $f, g \in \mathrm{Int}_Q$ then $f + g \in \mathrm{Int}_Q$, and $\int_Q (f+g) = \int_Q f + \int_Q g$.
[Hint: $\inf_S f + \inf_S g \leq \inf_S (f+g)$, $\sup_S f + \sup_S g \geq \sup_S (f+g)$.]

Below we omit for short "if... then...".

4. $\int_Q cf = c \int_Q f$.
5. $f \leq g \Rightarrow \int_Q f \leq \int_Q g$.
6. $\left|\int_Q f\right| \leq \int_Q |f|$. [Hint: $\Omega_S |f| \leq \Omega_S f$.]
7. $f = g$ on $Q \backslash F, \#F < \infty$ (F is FINITE) $\Rightarrow \int_Q f = \int_Q g$ (changing a function on a finite set does not change the integral).
8. $\forall P \in \mathrm{Part}\, Q \vdots \int_Q f = \sum_{S \in \mathrm{Cube}\, P} \int_S f$.

6.4 Null sets

We say that a set $N \subset \mathbb{R}^n$ is a set of *Lebesgue measure zero* or a *null* set if for any $\varepsilon > 0$ there exists (AT MOST) COUNTABLE family $\{Q_i\}$ of cubes in $\mathbb{R}^n$, which covers N and is such that the sum of the volumes of the cubes is less than ε:
$$N \in \mathrm{Null} :\Leftrightarrow \forall \varepsilon > 0\, \exists \{Q_i\}_{i \in \mathbb{N}} : Q_i \in \mathrm{Cube}\, \mathbb{R}^n, \quad \bigcup_{i \in \mathbb{N}} Q_i \supset N, \quad \sum_{i \in \mathbb{N}} \operatorname{vol} Q_i \leq \varepsilon.$$

(We can, without loss of generality, assume that the family is just countable, since adding to our family any countable number of one-point set does not change the sum of volumes.)
In the integration theory null sets are "negligible" in a sense, as we shall see.

Remarks.

1. Emphasize that Q_i may have zero volume.
2. A cube Q has POSITIVE volume iff it has the non-empty interior:
$$\operatorname{vol} Q > 0 \Leftrightarrow \overset{\circ}{Q} \neq \emptyset.$$
3. We obtain an EQUIVALENT definition if we replace the condition $\cup Q_i \supset N$ by
$$\bigcup_{i \in \mathbb{N}} \overset{\circ}{Q}_i \supset N.$$

(Prove as an EXERCISE. [Hint: for fixed $l_1, \dots, l_n$ (the lengths of the edges of a cube) the function $t \mapsto (l_1 + t) \dots (l_n + t), \mathbb{R} \to \mathbb{R}$ is continuous and strictly increasing at 0.])

Examples.

1. Any point (that is, a set $\{x\}$) is null. $\triangleleft$ $\{x\} \in \text{Cube}\,\mathbb{R}^n$, $\text{vol}\{x\} = 0$. $\triangleright$

2. Any finite set is null.

3. Any countable set is null. $\triangleleft$ Numerate the points of our set into a sequence $\{x_i\}$ and take $Q_i := \{x_i\}$. $\triangleright$

4. Any straight line in $\mathbb{R}^2$ is null. $\triangleleft$ EXERCISE. $\triangleright$

5. A set in $\mathbb{R}^n$ which has an INTERIOR point is NOT null: $\overset{\circ}{M} \neq \emptyset \Rightarrow M \notin \text{Null}$. In particular no cube with positive volume is null; in fact, vol $Q > 0 \Leftrightarrow Q \notin \text{Null}$ (prove!). (But there exist NOT-NULL sets (even in $\mathbb{R}$) with the EMPTY INTERIOR, cf. Exam. 6.7 2.)

Lemma 6.4.1. *Any subset of a null set is null.*
$\triangleleft$ Obviously. $\triangleright$

Lemma 6.4.2. *The union of a countable family of null sets is a null set.*
$\triangleleft$ Let $N_i \in \text{Null}$ for each $i \in \mathbb{N}$, and let $\varepsilon > 0$ be given. Let us write $\varepsilon = \varepsilon_1 + \varepsilon_2 + \dots$, where each $\varepsilon_i > 0$. For each i there exists a countable family $\{Q_{ij}\}_{j \in \mathbb{N}}$ of cubes, such that

$$\bigcup_j Q_{ij} \supset N_i, \qquad \sum_j \text{vol}\, Q_{ij} \leq \varepsilon_i.$$

Then the family $\{Q_{ij}\}_{i,j \in \mathbb{N}}$ (which is countable!) covers $\bigcup_i N_i$ and satisfies the inequality

$$\sum_{i,j} \text{vol}\, Q_{ij} = \sum_i \underbrace{(\sum_j \text{vol}\, Q_{ij})}_{\leq \varepsilon_i} \leq \sum_i \varepsilon_i = \varepsilon. \triangleright$$

Lemma 6.4.3. *If a null set N in $\mathbb{R}^n$ is* COMPACT *then for any $\varepsilon > 0$ there exists a* FINITE *family $Q_1, \dots, Q_k$ of cubes such that $\cup_{i=1}^k Q_i \supset N$, $\sum_{i=1}^k \text{vol}\, Q_i \leq \varepsilon$.*
$\triangleleft$ By Remark 3, there exists a countable family $\{Q_i\}$ of cubes such that

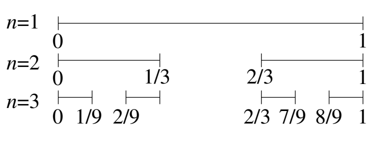
$$\bigcup_{i \in \mathbb{N}} \overset{\circ}{Q}_i \supset N, \qquad \sum_{i \in \mathbb{N}} \text{vol}\, Q_i \leq \varepsilon.$$

By compactness of N we can choose a finite subcovering, and this finite family is what we need. $\triangleright$

Remark 4. The compactness condition in Lemma 6.4.3. is essential (see Exercise 2 below).

Exercises

1. The Cantor set, the intersection of the sequence

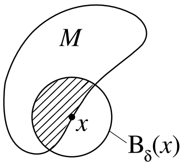


   is a (compact) null set.

2. Let M be the set of rational numbers between 0 and 1, $M := \mathbb{Q} \cap [0, 1]$. Then M is NULL as a countable set. Prove that there exists no FINITE family $I_1, \dots, I_k$ of intervals, such that $\cup I_i \supset M$ and $\sum$ length $I_i < 1$. [Hint: use induction in k.]

6.5 Oscillation

Let X be a normed space, let $M \subset X$, and let $f : X \to \mathbb{R}$ be a bounded function. The *oscillation of f on M at a point* $x \in X$ (usually $x \in M$) is defined by the formula



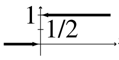
$$\omega_M f(x) := \lim_{\delta \downarrow 0} \Omega_{B_\delta(x) \cap M} f,$$

where Ω is the "global" oscillation, defined in Section 6.3:

$$\Omega_M f = \sup_M f - \inf_M f.$$

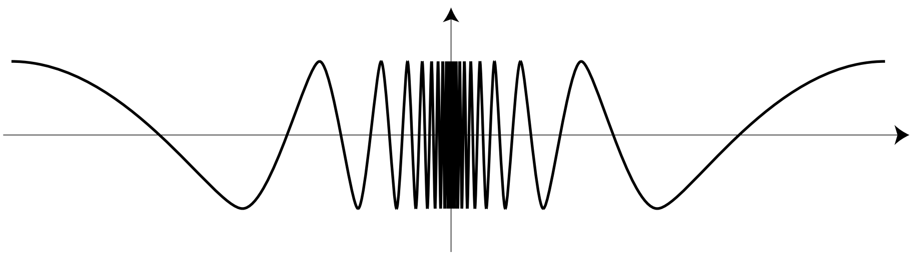
This limit exists since $\sup_{B_\delta(x)\cap M} f \downarrow$ and $\inf_{B_\delta(x)\cap M} f \uparrow$ as $\delta \downarrow 0$. If $M = X$ we omit the index M.

Examples.

1. $f =$ 

; $\omega f(0) = 1$, $\omega_{(-\infty,0)} f(0) = 0$, $\omega_{[-1,0]} f(0) = \frac{1}{2}$.

2.
$$f(x) = \begin{cases} \sin \frac{1}{|x|} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0; \end{cases} \qquad \omega f(0) = 2.$$



Remark. The value $\omega_M f(x)$ does not change if we replace the norm in X by any equivalent norm.

Lemma 6.5.1. *Let X be a normed space, $M \subset X, x \in M$, and let $f : X \to \mathbb{R}$ be a bounded function. Then f is continuous at x if and only if the oscillation of f at x is equal to zero:*

$$f|_M \in \mathrm{Cont}(x) \Leftrightarrow \omega_M f(x) = 0.$$

$\lhd$ "$\Rightarrow$": Let $f|_M$ is continuous at x. Consider arbitrary $\varepsilon > 0$. By supposed continuity there exists $\delta > 0$ such that

$$\forall y \in B_\delta(x) \cap M \,\vdots\, |f(y) - f(x)| \leq \tfrac{\varepsilon}{2}.$$

Then

$$\sup_{B_\delta(x)\cap M} f \leq f(x) + \tfrac{\varepsilon}{2}, \qquad \inf_{B_\delta(x)\cap M} f \geq f(x) - \tfrac{\varepsilon}{2},$$

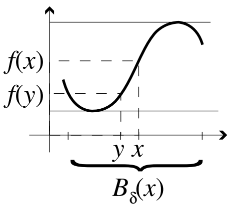
whence it follows that

$$\Omega_{B_\delta(x)\cap M} f \leq \varepsilon.$$

Hence

$$\omega_M f(x) \leq \varepsilon.$$

Since ε was arbitrary we conclude that $\omega_M f(x) = 0$.
"$\Leftarrow$": Let $\omega_M f(x) = 0$, and let $\varepsilon > 0$ be given. Then there exists $\delta > 0$ such that



$$\Omega_{\mathrm{B}_\delta(x)\cap M} f \le \varepsilon. \tag{1}$$

Hence for each $y \in \mathrm{B}_\delta(x) \cap M$ it holds

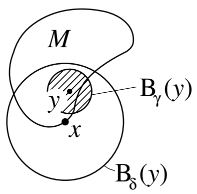
$$|f(x) - f(y)| \overset{\text{obv}}{\le} \sup_{\mathrm{B}_\delta(x)\cap M} f - \inf_{\mathrm{B}_\delta(x)\cap M} f \overset{(1)}{\le} \varepsilon,$$

which means that $f|_M$ is continuous at x. $\triangleright$

Lemma 6.5.2. *Let X be a normed space, $M \subset X$, and let $f : X \to \mathbb{R}$ be a bounded function. Then for any $\varepsilon > 0$ the set*

$$A_\varepsilon := \{x \in M | \omega_M f(x) < \varepsilon\} \qquad \textit{(strict inequality!)}$$

is OPEN IN M.
$\triangleleft$ Let $x \in A_\varepsilon$. We need to show that there exist $\delta > 0$ such that



$$\mathring{\mathrm{B}}_\delta(x) \cap M \subset A_\varepsilon. \tag{2}$$

But indeed (since $\omega_M f(x) < \varepsilon$) there exist $\delta > 0$ and $(0 <)\varepsilon' < \varepsilon$ such that

$$\Omega_{\mathrm{B}_\delta(x)\cap M} f \le \varepsilon'. \tag{3}$$

Let $y \in \mathring{\mathrm{B}}_\delta(x) \cap M$. Obviously there exists γ, $0 < \gamma \le \delta$, such that $\mathrm{B}_\gamma(y) \subset \mathrm{B}_\delta(x)$. Then $\mathrm{B}_\gamma(y) \cap M \subset \mathrm{B}_\delta(x) \cap M$, and hence

$$\Omega_{\mathrm{B}_\gamma(y)\cap M} f \le \Omega_{\mathrm{B}_\delta(x)\cap M} f \overset{(3)}{\le} \varepsilon',$$

which implies that

$$\omega_M f(y) \le \varepsilon' < \varepsilon.$$

This means that $y \in A_\varepsilon$, and (2) is true. $\triangleright$

Let us return in $\mathbb{R}^n$, equipped, say, by the Euclidean norm ($\|\cdot\|_2$).

Lemma 6.5.3. *Let Q be a cube in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded function such that*

$$\exists\varepsilon > 0\ \forall x \in Q \vdots\ \omega_Q f(x) < \varepsilon \quad (\textit{strict inequality}!). \tag{4}$$

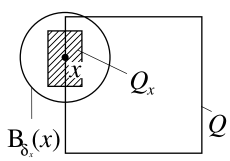
Then there exists a partition P of Q such that

$$\sum_{S\in\mathrm{Cube}\,P} (\Omega_S f)\,\mathrm{vol}\,S < \varepsilon\,\mathrm{vol}\,Q.$$

$\triangleleft$ 1° Consider arbitrary point $x \in Q$. By (4),

$$\exists\delta_x > 0 : \Omega_{\mathrm{B}_{\delta_x}(x)\cap Q} f < \varepsilon. \tag{5}$$

Let Q_x be a cube with the center at x such that $\mathring{Q}_x \ne \emptyset$ and $Q_x \subset \mathrm{B}_{\delta_x(x)}$. (Such a cube



exists, since $\|\cdot\|_\infty \sim \|\cdot\|_2$.) The cubes $\{\mathring{Q}_x\}$ form an open covering of Q. By compactness of Q, we can choose a finite subcovering, say

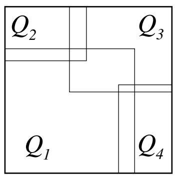
$$\{\mathring{Q}_{x_1}, \ldots, \mathring{Q}_{x_k}\} \quad (x_1, \ldots, x_k \in Q).$$

Thus, $\overset{\circ}{Q}_{x_1} \cup \ldots \cup \overset{\circ}{Q}_{x_k} \supset Q$. Therefore if we put $Q_i := Q_{x_i} \cap Q$, it holds

$$Q_1 \cup \ldots \cup Q_k = Q.$$

2° Obviously each set Q_i is a cube, and there exists a partition P of Q such that each cube S of P is cointained in some Q_i, that is,

$$S \subset Q_{x_i} \cap Q \subset \mathrm{B}_{\delta_{x_i}}(x_i) \cap Q.$$



Then

$$\Omega_S f \le \Omega_{\mathrm{B}_{\delta_{x_i}}(x_i)\cap Q} f \overset{(5)}{<} \varepsilon.$$

Hence,

$$\sum_{S\in\mathrm{Cube}\,P} \underbrace{(\Omega_S f)}_{<\varepsilon} \operatorname{vol} S < \varepsilon \sum_{S\in\mathrm{Cube}\,P} \operatorname{vol} S = \varepsilon \operatorname{vol} Q. \triangleright$$

6.6 Lebesgue theorem

The following result is fundamental.

Theorem 6.6.1. (Lebesgue). *Let Q be a cube in $\mathbb{R}^n$, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded function. Denote by* $\operatorname{discont}_Q f$ *the set of all points where $f|_Q$ is* **not** *continuous:*

$$\operatorname{discont}_Q f := \{x \in Q |\ f|_Q \notin \operatorname{Cont}(x)\}.$$

Then f is integrable over Q if and only if $\operatorname{discont}_Q f$ *is a null set:*

$$f \in \operatorname{Int}_Q \Leftrightarrow \operatorname{discont}_Q f \in \operatorname{Null}.$$

$\triangleleft$ For short put $A := \operatorname{discont}_Q f$, and put for each $\varepsilon > 0$

$$A_\varepsilon := \{x \in Q |\ \omega_Q f(x) \ge \varepsilon\}.$$

(This set is COMPLEMENTARY in Q to the set A_ε from Lemma 6.5.2. (with $M = Q$).) We have

$$A \overset{\text{Lm 6.5.2.}}{=} \{x \in Q |\ \omega_Q f(x) > 0\} \overset{\text{obv}}{=} \bigcup_{k=1}^{\infty} \{x \in Q |\ \omega_Q f(x) \ge \tfrac{1}{k}\},$$

that is,

$$A = \bigcup_{k\in\mathbb{N}} A_{1/k}. \tag{1}$$

"$\Rightarrow$" 1° Let $f \in \operatorname{Int}_Q$. We need verify that $A \in \operatorname{Nul}$. In view of (1) it is sufficient (by Lemma 6.4.2.) to show that for each $\delta > 0$

$$A_\delta \in \operatorname{Nul}. \tag{2}$$

Let ε be an arbitrary positive number. By (Corollary of) Criterium of integrability (Section 6.3) there exists a partition P of Q such that

$$\sum_{S\in\mathrm{Cube}\,P} (\Omega_S f) \operatorname{vol} S \le \varepsilon. \tag{3}$$

2° Denote by N the union of the boundaries of all cubes of P:

$$N := \bigcup_{S \in \mathrm{Cube}\, P} \mathrm{fr}\, S.$$

Obviously, $N \in$ Null; hence there exists a countable (end even finite, since N is compact; see Lemma 6.4.3.) family $\{Q_i\}$ of cubes in $\mathbb{R}^n$ such that

$$\bigcup Q_i \supset N, \qquad \sum \mathrm{vol}\, Q_i \le \varepsilon. \tag{4}$$

3° Now denote by $\mathcal{S}$ the set of all cubes S of P such that at least one INTERIOR point of S belongs to A_δ:

$$\mathcal{S} := \{S \in \mathrm{Cube}\, P | \overset{\circ}{S} \cap A_\delta \ne \emptyset\}. \tag{5}$$

It is clear that

$$\forall S \in \mathcal{S} \vdots \Omega_S f \ge \delta \tag{6}$$

(since for some $x \in \overset{\circ}{S}$ it holds $\omega f(x) \ge \delta$). Further,

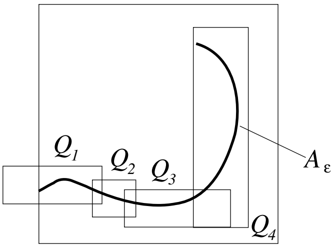
$$\begin{aligned} \textstyle\sum_{S\in\mathcal{S}} \mathrm{vol}\, S &\underset{\text{trick}}{\overset{(6)}{\le}} \textstyle\sum_{S\in\mathcal{S}} \delta^{-1}(\Omega_S f)\, \mathrm{vol}\, S = \delta^{-1} \sum_{S\in\mathcal{S}} (\Omega_S f)\, \mathrm{vol}\, S \\ &\underset{\text{obv}}{\le} \delta^{-1} \textstyle\sum_{S\in\mathrm{Cube}\, P} (\Omega_S f)\, \mathrm{vol}\, S \overset{(3)}{\le} \varepsilon\delta^{-1}. \end{aligned} \tag{7}$$

4° The cubes from $\{Q_i\}$ and from $\mathcal{S}$ altogether cover A_δ, since each point of A_δ either lies in N or is interior for some cube S, and

$$\sum \mathrm{vol}\, Q_i + \sum_{S\in\mathcal{S}} \mathrm{vol}\, S \overset{(4),(7)}{\le} \varepsilon(1+\delta^{-1}).$$

But here δ is fixed, and ε is arbitrary. We conclude that (2) is true.

"$\Leftarrow$" 5° Let $A \in$ Null. We prove that $f \in \mathrm{Int}_Q$, using the same Criterion. Let ε be an arbitrary positive number. The set $A_\varepsilon = \{x \in Q | \, \omega_Q f(x) \le \varepsilon\}$ is null (since $A_\varepsilon \subset A$) and is compact (A_ε is bounded since $A_\varepsilon \subset Q$, and A_ε is closed since $(A_\varepsilon)^c \cap Q = \{x \in Q | \, \omega_Q f(x) < \varepsilon\}$ is open in Q by Lemma 6.5.2., hence A_ε is closed in Q and therefore is closed (since Q is closed!)). By Lemma 6.4.3., there exists a finite number of cubes $Q_1, \ldots, Q_k$ such that



$$\bigcup_{i=1}^{k} \overset{\circ}{Q}_i \supset A_\varepsilon, \tag{8}$$

$$\sum_{i=1}^{k} \operatorname{vol} Q_i \le \varepsilon. \tag{9}$$

6° Put

$$\text{Black} := (\cup_{i=1}^{k} Q_i) \cap Q$$

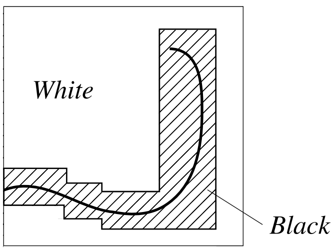
(shadowed on the second picture), and

$$\text{White} := Q \backslash (\cup_{i=1}^{k} \overset{\circ}{Q}_i) \qquad (= \operatorname{cl}(Q \backslash \text{Black})).$$

It is clear that

$$\forall x \in \text{White} \,\vdots\, \omega_Q f(x) < \varepsilon \tag{10}$$

(since White $\subset (A_\varepsilon)^c$).



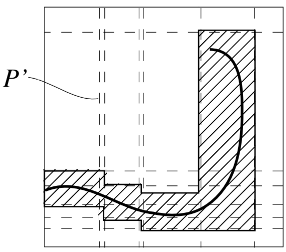
7° Obviously there exists a partition P' of Q such that each cube S' of P' lies either in White or in Black.

Put

$$\mathcal{W} := \{S' \in \operatorname{Cube} P' \,|\, S' \subset \text{White}\},$$
$$\mathcal{B} := \{S' \in \operatorname{Cube} P' \,|\, S' \subset \text{Black}\}.$$

It is clear from (9) that

$$\sum_{S' \in \mathcal{B}} \operatorname{vol} S' \le \varepsilon. \tag{11}$$



8° For each $S' \in \mathcal{W}$ there exists by Lemma 6.5.3. (in view of 10)) a partition $P_{S'}$ such that

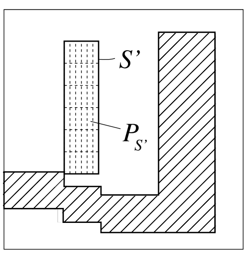
$$\Delta_{P_{S'}} f < \varepsilon \operatorname{vol} S'. \tag{12}$$

9° Finally, there exists a partition P of Q such that

$$P \succ P'$$

and

$$\forall S' \in \mathcal{W} \,\vdots\, P|_{S'} \succ P_{S'}.$$



(Here $P|_{S'}$ denotes, naturally, the "restriction" of the partition P to S'.) Let us show that P is what we need.

10° For this end put

$$(0 \le) M := \Omega_Q f$$

(M is finite, since $f \in \mathrm{Bd}$). It is clear that

$$\forall S \in \operatorname{Cube} P \,\vdots\, \Omega_S f \le M. \tag{13}$$

11° Now,

$$\Delta_P f = \sum_{S\in \mathrm{Cube}\, P} (\Omega_S f)\,\mathrm{vol}\, S = \underbrace{\sum_{S'\in\mathcal{W}} \sum_{\substack{S\in\mathrm{Cube}\,P\\ S\subset S'}} (\Omega_S f)\,\mathrm{vol}\, S}_{\boxed{1}} + \underbrace{\sum_{S'\in\mathcal{B}} \sum_{\substack{S\in\mathrm{Cube}\,P\\ S\subset S'}} (\Omega_S f)\,\mathrm{vol}\, S}_{\boxed{2}}.$$

We have

$$\boxed{1} = \Delta_{P|_{S'}} f \overset{6.3.4.}{\leq} \Delta_{P_{S'}} f \overset{(12)}{\leq} \varepsilon\,\mathrm{vol}\, S', \tag{14}$$

$$\boxed{2} \overset{(13)}{\leq} \sum_{S'\in\mathcal{B}} \sum_{\substack{S\in\mathrm{Cube}\,P\\ S\subset S'}} M\,\mathrm{vol}\, S = M \sum_{S'\in\mathcal{B}} \sum_{\substack{S\in\mathrm{Cube}\,P\\ S\subset S'}} \mathrm{vol}\, S \overset{\mathrm{obv}}{=} M \sum_{S'\in\mathcal{B}} \mathrm{vol}\, S' \overset{(11)}{\leq} M\varepsilon \tag{15}$$

Thus

$$\Delta_P f = \sum_{S'\in\mathcal{W}} \boxed{1} + \boxed{2} \overset{(14),(15)}{\leq} \sum_{S'\in\mathcal{W}} \varepsilon\,\mathrm{vol}\, S' + M\varepsilon \overset{\mathrm{obv}}{=} \varepsilon \underbrace{\sum_{S'\in\mathcal{W}} \mathrm{vol}\, S'}_{\overset{\mathrm{obv}}{\leq}\,\mathrm{vol}\, Q} + M\varepsilon \leq \varepsilon(\mathrm{vol}\, Q + M).$$

But ε was arbitrary small. So by Criterium, $f \in \mathrm{Int}_Q$. $\triangleright$

Exercises.

1. Let

$$f(x) = \begin{cases} \frac{1}{q} & \text{if } x\in\mathbb{Q} \text{ and } x = \frac{p}{q},\ p, q \text{ being} \\ & \text{mutually prime integers,} \\ 0 & \text{if} x \notin \mathbb{Q}. \end{cases}$$

Prove that discont $f = \mathbb{Q}$ (dense in $\mathbb{R}$!). So f is integrable over any (bounded) interval.

3. Let $f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1, \ldots, f_m)$, $Q \in \mathrm{Cube}(\mathbb{R}^n)$, and let each component function f_i is integrable over Q. Let further $g : \mathbb{R}^m \to \mathbb{R}$ be a CONTINUOUS function. Prove that the composition $g \circ f$

$$\mathbb{R}^n \xrightarrow{f} \mathbb{R}^m \xrightarrow{g} \mathbb{R}$$

is integrable over Q. (In particular the product $f_1 f_2$ of two integrable functions is integrable.) [Hint: discont$(g\circ f) \subset \cup_{i=1}^m$ Discont f_i.]

6.7 Jordan measurable sets

Now we define the integral over ARBITRARY (bounded) set. Let f be a bounded function on $\mathbb{R}^n$, and let M be a bounded set in $\mathbb{R}^n$. We say that f is *integrable over* M and we write $f \in \mathrm{Int}_M$, if the product $\chi_M f$ (recall that χ_M denotes the indicator function of M) is integrable over some cube $Q \supset M$, and in such a case we put

$$\int_M f := \int_Q \chi_M f.$$

(The result does not depend on the choice of Q, and sometimes we shall drop Q.)

Further we say, that M is *Jordan measurable* if the constant function 1 is integrable over M, and we define the *volume* of M as the corresponding integral:

$$M \in \text{JMeas} :\Leftrightarrow \exists \int_M 1 = \int \chi_M =: \text{vol}\, M.$$

This definition evidently agrees with our original definition of volume for cubes.

Theorem 6.7.1. *A bounded set M in $\mathbb{R}^n$ is Jordan measurable iff its boundary is a null set:*

$$M \in \text{JMeas} \Leftrightarrow \text{fr}\, M \in \text{Null}\,.$$

$\lhd$ This follows at once from Lebesgue Theorem, since Discont $\chi_M = \text{fr}\, M$. $\rhd$

Remark. A null set (and even countable!) may be **non**-Jordan-measurable (Example 1 below); an open set may be **non**-Jordan-measurable (Example 2). [All the null sets and all the open ones **are** LEBESGUE MEASURABLE.]

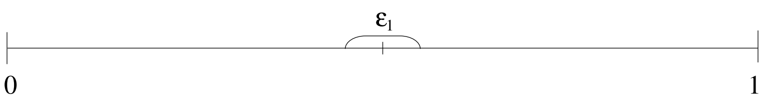
Example 1. The set $\mathbb{Q} \cap [0, 1]$ is not Jordan measurable. (Cf. Example 1.2 3.)

Example 2. We construct a bounded open set in $\mathbb{R}$ by the following procedure. Write

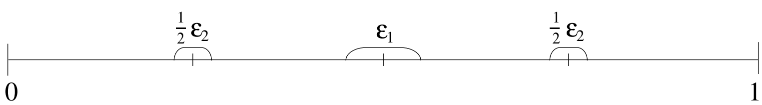
$$\tfrac{1}{2} = \varepsilon_1 + \varepsilon_2 + \ldots \quad (\varepsilon_i > 0)$$

(e.g. $\varepsilon_i = 2^{-i-1}$).

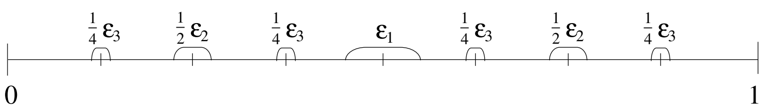
Step 1. Take the interval of the length ε_1 with the center common with the center of the interval [0, 1]:



Step 2. Take 2 open intervals, each of the length $\frac{1}{2}\varepsilon_2$ with the centers common, resp., with the centers of 2 intervals complementary in [0, 1] to the open interval constructed in Step 1:



Step 3. Take 4 open intervals, each of length $\frac{1}{4}\varepsilon_3$, with the centers common, resp., with the centers of 4 intervals complementary in [0, 1] to the open intervals constructed in Steps 1 and 2:



. . . .

The union M of all constructed by this procedure open intervals is a (bounded) open set, which is **not** Jordan measurable (but **is** LEBESGUE MEASURABLE, with LEBESGUE MEASURE $1/2$).

Exercises

1. Prove the assertion of Example 2. [Hint: prove at first that fr $M = [0,1]\backslash M$; then prove by induction in k, that there exists no finite covering of fr M by intervals with the sum of the lengths $< 1/2$ (*cf.* Lemma 6.4.3.).]
2. Any COMPACT null set **is** Jordan measurable (and its volume is equal to 0.)
3. Prove that
$$\operatorname{vol} M = 0 \Rightarrow M \in \text{Null}$$
and that if $M \in$ JMeas then
$$\operatorname{vol} M = 0 \Leftarrow M \in \text{Null}$$
4. $f \in \text{Bd}, \operatorname{vol} M = 0 \Rightarrow \int_M f = 0$.
5.
$$\begin{aligned}\operatorname{vol} M = 0 &\Leftrightarrow \forall \varepsilon > 0\, \exists k \in \mathbb{N}\, \exists Q_1, \dots, Q_k \in \text{Cube} : \textstyle\bigcup_{i=1}^k Q_i \supset M,\\ &\qquad \textstyle\sum_{i=1}^k \operatorname{vol} Q_i \le \varepsilon\\ &\Leftrightarrow \forall \varepsilon > 0\, \exists k \in \mathbb{N}\, \exists Q_1, \dots, Q_k \in \text{Cube} : \textstyle\bigcup \mathring{Q}_i \supset M,\\ &\qquad \textstyle\sum_{i=1}^k \operatorname{vol} Q_i \le \varepsilon\end{aligned}$$
6. $\operatorname{vol} M = 0 \Rightarrow \operatorname{vol} \bar{M} = 0$. (Remark that $M \in \text{Null} \not\Rightarrow \bar{M} \in \text{Null}$!)

6.8 Fubini Theorem

This theorem says about possibility to reduce calculation of the integral over a product to calculation of integrals over the factors.

Theorem 6.8.1. (Fubini). *Let A be a cube in $\mathbb{R}^n$, let B be a cube in R^m, and let $f : A \times B \to \mathbb{R}$ be a (bounded) integrable function. Put for each $x \in A$*

$$l(x) := \mathrm{L}\int_B f(x,\cdot), \qquad u(x) := \mathrm{U}\int_B f(x,\cdot).$$

Then both the functions l and u are integrable over A, and

$$\int_{A\times B} f = \int_A l = \int_A u.$$

(Recall that the "x-section" $f(x,\cdot)$ of f is the function $A \to \mathbb{R}$, $y \mapsto f(x,y)$.)

$\lhd$ 1° Obviously, any partition P of $A \times B$ may be written as a pair $P = (P_A, P_B)$, where $P_A \in \text{Part}\, A$, $P_B \in \text{Part}\, B$. We have

$$S \in \text{Cube}\, P \Leftrightarrow S = S_A \times S_B, \quad S_A \in \text{Cube}\, P_A, \quad S_B \in \text{Cube}\, P_B.$$

2° For any $P = (P_A, P_B) \in \text{Part}(A \times B)$

$$\begin{aligned}\mathrm{L}_P f &= \sum_{S \in \text{Cube}\, P} \left(\inf_S f\right) \operatorname{vol} S = \sum_{\substack{S_A \in \text{Cube}\, P_A\\ S_B \in \text{Cube}\, P_B}} \left(\inf_{S_A \times S_B} f\right) \underbrace{\operatorname{vol}(S_A \times S_B)}_{= \operatorname{vol} S_A \operatorname{vol} S_B}\\ &= \sum_{S_A \in \text{Cube}\, P_A} \underbrace{\left(\sum_{S_B \in \text{Cube}\, P_B} \left(\inf_{S_A \times S_B} f\right) \operatorname{vol} S_B\right)}_{\boxed{1}} \operatorname{vol} S_A.\end{aligned}$$

3° $\boxed{1} \le \inf_{S_A} l$. $\lhd\!\lhd$ $\forall x \in S_A \vdots \inf_{S_A\times S_B} f \le \inf_{\{x\}\times S_B} f = \inf_{S_B} f(x,\cdot)$. Hence

$$\forall x \in S_A \vdots \boxed{1} \le \sum_{S_B\in \mathrm{Cube}\, P_B} (\inf_{S_B} f(x,\cdot)) \operatorname{vol} S_B = \mathrm{L}_{P_B} f(x,\cdot) \le \mathrm{L}\int_B f(x,\cdot) = l(x).$$

We conclude that $\boxed{1} \le \inf_{S_A} l$. $\rhd\!\rhd$
4° $\mathrm{L}_P f \le \mathrm{L}_{P_A} l$.

$$\lhd\!\lhd\ \mathrm{L}_P f \overset{2^\circ}{=} \sum_{S_A\in \mathrm{Cube}\, P_A} \boxed{1} \operatorname{vol} S_A \overset{3^\circ}{\le} \sum_{S_A\in \mathrm{Cube}\, P_A} (\inf_{S_A} l) \operatorname{vol}_{S_A} = \mathrm{L}_{P_A} l.\ \rhd\!\rhd$$

5° $\mathrm{U}_P f \ge \mathrm{U}_{P_A} u$.$\lhd\!\lhd$ Analogously. $\rhd\!\rhd$
6° $\mathrm{L}_{P_A} l \le \mathrm{U}_{P_A} u$. $\lhd\!\lhd$ $L_{P_A} l \overset{l\le u}{\le} \mathrm{L}_{P_A} u \le \mathrm{U}_{P_A} u$. $\rhd\!\rhd$
7° By 3°–5°,

$$\mathrm{L}_P f \le \mathrm{L}_{P_A} l \le \mathrm{U}_{P_A} u \le \mathrm{U}_P f.$$

It follows that

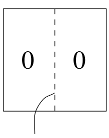
$$\underbrace{\sup_{P\in \mathrm{Part}(A\times B)} \mathrm{L}_P f}_{=\int_{A\times B} f} \le \underbrace{\sup_{P_A\in \mathrm{Part}(A)} \mathrm{L}_{P_A} l}_{=\mathrm{L}\int_A l} \le \underbrace{\inf_{P_A\in \mathrm{Part}(A)} \mathrm{U}_{P_A} l}_{=\mathrm{U}\int_A l} \le \underbrace{\inf_{P\in \mathrm{Part}(A\times B)} \mathrm{U}_P f}_{=\mathrm{U}\int_{A\times B} f}.$$

Therefore

$$\mathrm{L}\int_A l = \mathrm{U}\int_A l = \int_{A\times B} f,$$

which means that $\int_A l = \int_{A\times B} f$.
8° The other equation may be proved analogously. $\rhd$

Example. Let



Dirichlet function

$$f(x,y) = \begin{cases} 1 & \text{if } x = \frac{1}{2}, y \notin \mathbb{Q}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $l = 0$, $u = \chi_{\{1/2\}}$, and

$$\int_{[0,1]^2} f = \left\{ \begin{array}{l} \int_{[0,1]} l \ = \int_{[0,1]} 0 \\ \int_{[0,1]} u = \int_{[0,1]} \chi_{\{1/2\}} \end{array} \right\} = 0.$$

(Remark that f is integrable, though it has a NON-integrable section $f(1/2,\cdot)$.)
Notations. It is convenient to use the following “classic” notations:

$$\int f \equiv \int f(x,y)\,\mathrm{d}x\,\mathrm{d}y, \quad \int f(x,\cdot) \equiv \int f(x,y)\,\mathrm{d}y, \quad \int f(\cdot,y) \equiv \int f(x,y)\,\mathrm{d}x$$

(and analogously for $\mathrm{L}\int$, $\mathrm{U}\int$). E.g. we use these notations in the corollaries below.

Corollary 6.8.2. (change of the order of integrations).

$$\left(\int_{A\times B} f(x,y)\,\mathrm{d}x\,\mathrm{d}y =\right) \int_A (\mathrm{L}\int_B f(x,y)\,\mathrm{d}y)\,\mathrm{d}x = \int_B (\mathrm{L}\int_A f(x,y)\,\mathrm{d}x)\,\mathrm{d}y$$

(and any from two “L” or both of them may be replaced by “U”).

Corollary 6.8.3. (reduction of a double integral to a repeated one). *Let, in the conditions of Fubini Theorem, for each $x \in A$ the function $f(x,\cdot)$ is integrable. Then*

$$\int_{A\times B} f(x,y)\,\mathrm{d}x\mathrm{d}y = \int_A \left(\int_B f(x,y)\,\mathrm{d}y\right)\,\mathrm{d}x.$$

The condition of Corollary 6.8.3. is fulfilled, e.g., for continuous functions (since any section of a continuous function is also continuous).

In particular, for $f \in \mathrm{Cont}$ we have (by induction)

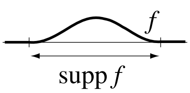
$$\begin{aligned}\int_{[a_1,b_1]\times\ldots\times[a_n,b_n]} f &\equiv \int_{a_1}^{b_1}\ldots\int_{a_n}^{b_n} f(x_1,\ldots,x_n)\,\mathrm{d}x_1\ldots\mathrm{d}x_n\\ &= \int_{a_n}^{b_n}\left(\ldots\left(\int_{a_1}^{b_1} f(x_1,\ldots,x_n)\,\mathrm{d}x_1\right)\ldots\right)\mathrm{d}x_n.\end{aligned}$$

Chapter 7

Partition of unity. Change of variables

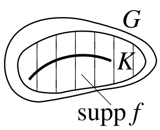
7.1 Smooth indicators

For a function $f : \mathbb{R}^n \to \mathbb{R}$ its *support* $\operatorname{supp} f$ is defined as the closure of the set, where f is not equal to 0:



$$\operatorname{supp} f := \operatorname{cl}\{x \in \mathbb{R}^n \,|\, f(x) \neq 0\}.$$

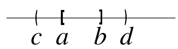
We say that a C^∞-function f is a *smooth indicator* of a set $A \subset \mathbb{R}^n$ if $f|_A = 1$.



Theorem 7.1.1. *For any open set G in $\mathbb{R}^n$ and any compact set $K \subset G$ there exists a smooth indicator f of K with the support in G:*

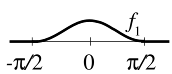
$$f|_K = 1, \qquad \operatorname{supp} f \subset G.$$

◁ 1° Theorem is true for $n = 1$, $K = [a, b]$, $g = (c, d)$.



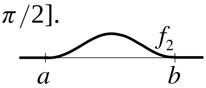
⧏◁ *Step 1.*

$$f_1(x) := \begin{cases} e^{-\operatorname{tg}^2 x} & \text{if } -\pi/2 < x < \pi/2, \\ 0 & \text{if not.} \end{cases}$$

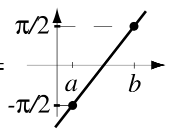


It is easy to verify that $f_1 \in C^\infty$ and $\operatorname{supp} f_1 = [-\pi/2, \pi/2]$.

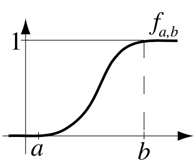
Step 2. $\forall a < b \; \exists f_2 \in C^\infty : \; f_2 \geq 0, \operatorname{supp} f_2 = [a, b]$.



⧏◁ Put $f_2 := f_1 \circ l$, where $l =$



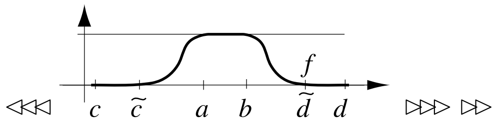
▷▷



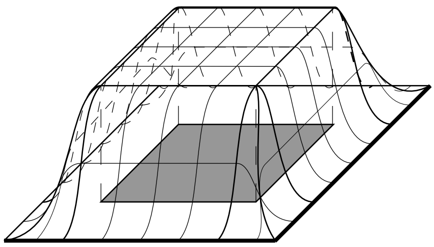
Step 3. $\forall a < b \; \exists f_{a,b} \in C^\infty : \; f_{a,b} \geq 0, \; f_{a,b}|_{(-\infty,a]} = 0$, $f_{a,b}|_{[b,+\infty)} = 1$.

$$⧏⧏◁ \quad f_{a,b}(x) := \left(\int_a^x f_2\right) \Big/ \left(\int_a^b f_2\right). \quad ▷▷▷$$

Step 4. Choose $\widetilde{c}$ and $\widetilde{d}$ such that $c < \widetilde{c} < a$, $b < \widetilde{d} < d$ and put $f := f_{\widetilde{c},d} - f_{b,\widetilde{d}}$.



2° Theorem is true for $K = Q_1$, $G = \overset{\circ}{Q}_2$, where $Q_1, Q_2 \in \mathrm{Cube}(\mathbb{R}^n)$.
⊲⊲ Let $Q_1 = [a_1, b_1] \times \ldots \times [a_n, b_n]$, $Q_2 = [c_1, d_1] \times \ldots \times [c_n, d_n]$. By 1°, for each $i = 1, \ldots, n$ there exists a smooth indicator f_i of $[a_i, b_i]$ with $\operatorname{supp} f_i \subset (c_i, d_i)$. Put

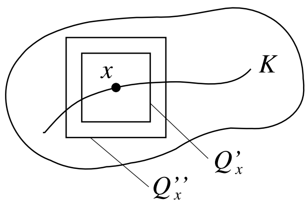


$$f := f_1 \otimes \ldots \otimes f_n,$$

that is,

$$f(x_1, \ldots, x_n) := f_1(x_1) \cdot \ldots \cdot f_n(x_n).$$

It is clear that f is what we need. ⊳⊳



3° *General case.* For any $x \in K$ there exist cubes Q'_x, Q''_x such that

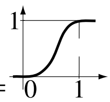
$$x \in \overset{\circ}{Q}{}'_x, \qquad Q'_x \subset \overset{\circ}{Q}{}''_x, \qquad Q''_x \subset G.$$

The cubes $\overset{\circ}{Q}{}'_x$ cover K. By compactness of K, we can choose a finite subcovering, say $\overset{\circ}{Q}{}'_1, \ldots, \overset{\circ}{Q}{}'_k$, with the corresponding "outer" cubes $Q''_1, \ldots, Q''_k$. By 2°, for each $i = 1, \ldots, k$ there exists a smooth indicator f_i of Q'_i with $\operatorname{supp} f_i \subset \overset{\circ}{Q}{}''_i$. Put

$$\widetilde{f} := \sum_{i=1}^{k} f_i.$$

It is clear that $\widetilde{f} \in C^\infty$, $\widetilde{f} \geq 0$, $\widetilde{f}|_K \geq 1$, $\operatorname{supp} \widetilde{f} \subset G$. At last, put

$$f := f_{0,1} \circ \widetilde{f},$$

where $f_{0,1} =$ 

(see Step 3 of 1°). It is obvious that f is what we need. ⊳

7.2 Partition of unity

Let $A \subset \mathbb{R}^n$, and let O be an open covering of A (the notation: $O \in \mathrm{OC}(A)$). A family Φ of C^∞-functions $\mathbb{R}^n \to \mathbb{R}$ is called a *partition of unity for A submitted to O* (the notation: $\Phi \in \mathrm{PU}(A, O)$), if

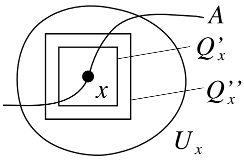
1) $\forall \varphi \in \Phi \,\vdots\, 0 \leq \varphi \leq 1$;
2) $\forall x \in A\ \exists U \in \mathrm{Nb}_x$ such that only FINITE number of functions from Φ are not identically zero on U (the condition of *local finiteness*);
3) $\forall x \in A \,\vdots\, \sum_{\varphi \in \Phi} \varphi(x) = 1$ (this sum is FINITE, by 2));
4) $\forall \varphi \in \Phi\ \exists U \in O : \operatorname{supp} \varphi \subset U$ (Φ is *submitted* to O).

Remark. For any COMPACT set $K \subset A$ there exists only FINITE number of functions $\varphi \in \Phi$ such that $\varphi|_K \neq 0$.

$\lhd$ This follows from 2) and from compactness of K. $\rhd$

Theorem 7.2.1. *For any $A \subset \mathbb{R}^n$ and any open covering O of A there exists a partition of unity for A submitted to O.*

$\lhd$ *Case 1. A is compact.* Without loss of generality we can assume that O is FINITE.



1° $\forall x \in A\ \exists U_x \in O\ \exists Q'_x, Q''_x \in \mathrm{Cube}(\mathbb{R}^n) : x \in \mathring{Q}'_x$, $Q_x \subset \mathring{Q}''_x$, $Q''_x \subset U_x$. The cubes $\mathring{Q}'_x$ cover A. By compactness of A we can choose a finite subcovering, say $\mathring{Q}'_1, \ldots, \mathring{Q}'_k$, with the corresponding outer cubes $Q''_1, \ldots Q''_k$.

By Theorem on smooth indicators applied to Q'_i and $\mathring{Q}''_i$, for each $i = 1, \ldots, k$ there exists a smooth indicator f_i for Q'_i with $\operatorname{supp} f_i \subset \mathring{Q}''_i$.

2° Put on $\mathring{Q}'_1 \cup \ldots \cup \mathring{Q}'_k =: G$

$$\Psi_i := \frac{f_i}{f_1 + \ldots + f_k} \qquad (i = 1, \ldots, k). \tag{1}$$

(Obviously, $f_1 + \ldots + f_k \geq 1$ on G, hence this definition is correct.)

3° Once again by Theorem on smooth indicators, applied this time to A and G, there exists a smooth indicator f_0 of A with $\operatorname{supp} f_0 \subset G$. Put

$$\varphi_i := \begin{cases} f_0 \psi_i & \text{on } G, \\ 0 & \text{on } G^c. \end{cases}$$

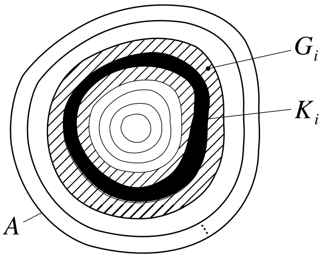
It is clear that $\varphi_1, \ldots, \varphi_k$ are just what we need. Indeed,

$$\operatorname{supp} \varphi_i \subset \operatorname{supp} f_i \subset \mathring{Q}''_i \overset{1^\circ}{\subset} U \text{ for some } U \in O,$$

and

$$\left.\left(\sum_{i=1}^k \varphi_i\right)\right|_A = \sum_{i=1}^k \Big(\underbrace{f_0|_A}_{=1}\Big)(\Psi_i|_A) = \left.\left(\sum_{i=1}^k \Psi_i\right)\right|_A \overset{(1),\ A \subseteq G}{=} 1.$$

Case 2. $A = \cup_{i=1}^\infty A_i$, $A_i \in \mathrm{Comp}$, $A_i \subset \mathring{A}_{i+1}$. Note that in this case A is open (since $A = \cup \mathring{A}_i$), each set $K_i := A_i \backslash \mathring{A}_{i-1}$ is compact (verify!), each set $G_i := \mathring{A}_{i+1}/A_{i-2}$ is open, and
$K_i \subset G_i$ (see the picture). Put



$$O_i := \{U \cap G_i | U \in O\}.$$

Then O_i is an open covering of K_i, and by Case 1, there exists a FINITE partition of unity Φ_i for K_i submitted to O_i. Now put

$$\psi := \sum_{i=1}^\infty \sum_{\varphi \in \Phi_i} \varphi.$$

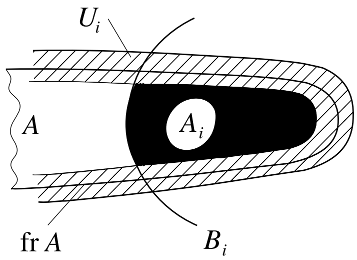
This definition is correct since each point of A lies in some G_i, and for each i all the functions from Φ_j with $j \geq i + 3$ have the supports OUTSIDE G_i, so on each G_i our ψ is the sum of a FINITE number of functions φ.

At last put for each $\varphi \in \cup_{i=1}^{\infty} \Phi_i$

$$\varphi' := \frac{\varphi}{\psi}.$$

The family of all such φ' is what we need.

Case 3. A is open. Put for $i = 1, 2, \ldots$



$$U_i := \{x \in \mathbb{R}^n \,|\, \mathrm{dist}(x, \mathrm{fr}A) < \tfrac{1}{i}\},$$

$$A_i = A \cap U_i^c \cap B_i.$$

Here B_i denotes the ball of radius i with the center at 0, and $\mathrm{dist}(x, Y)$ denotes the distance from a point x to a set Y, that is defined by the formula

$$\mathrm{dist}(x, Y) := \inf_{y \in Y} \|x - y\|.$$

For any fixed set Y the function

$$\varrho_Y := \mathrm{dist}(\cdot, Y)$$

is continuous (verify!).

We claim that

$$\forall i \,\vdots\, A_i \in \mathrm{Comp},\ A_i \subset \mathring{A}_{i+1},\ \text{ and } A = \bigcup_{i=1}^{\infty} A_i.$$

$\lhd\!\lhd$ We use the following simple fact from topology: *The difference of s set and an* OPEN *neighbourhood of its frontier is closed.* (Verify!)

By this fact $A \cap U_i^c \in \mathrm{Cl}$. Further, B_i is bounded and closed. Hence A_i is bounded and closed, that is, $A_i \in \mathrm{Comp}$. (Note that U_i is open since $U_i = \varrho_{\mathrm{fr}A}^{-1}((-\frac{1}{i}, +\frac{1}{i}))$ and $\varrho_{\mathrm{fr}A}$ is continuous function.) Other relations are obvious. $\rhd\!\rhd$

Hence we can apply Case 2.

General case. Put $G := \cup_{U \in O} U$. By Case 3, there exists a partition of unity for G submitted to O. It is of course also a partition of unity for A. $\rhd$

7.3 Partition of integral

Now we show that having a partition of unity Φ for A we can represent the integral $\int_A f$ as a sum of integrals $\int_A \varphi f$ over $\varphi \in \Phi$.

Lemma 7.3.1. *If A, B are Jordan measurable then*

$$A \cup B, \quad A \cap B, \quad A \backslash B, \quad B \backslash A \tag{1}$$

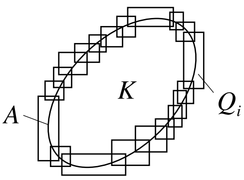
are also Jordan measurable.

$\lhd$ Recall that a set is Jordan measurable iff its frontier is a null set. Thus $\mathrm{fr}A$ and $\mathrm{fr}B$ are null sets and hence their union also is a null set. But the frontier of each from 4 sets in (1) lies in $\mathrm{fr}A \cup \mathrm{fr}B$ (verify!) and hence is null. $\rhd$

Lemma 7.3.2. *If A is a (bounded) Jordan measurable set then for any $\varepsilon > 0$ there exists a* COMPACT *Jordan measurable subset K of A such that*

$$\operatorname{vol}(A\setminus K) \leq \varepsilon.$$

(Note that $A\setminus K$ is Jordan measurable by Lemma 7.3.1.)
$\lhd$ fr$A$ is a compact null set, hence by Lemma 1.4.3 there exists a finite number of cubes $Q_1, \dots, Q_k$ such that



$$\bigcup_{i=1}^{k} \overset{\circ}{Q}_i \supset \operatorname{fr} A, \quad \sum_{i=1}^{k} \operatorname{vol} Q_i \leq \varepsilon.$$

Put

$$K := A \setminus \bigcup_{i=1}^{k} \overset{\circ}{Q}_i.$$

This set is bounded (obviously) and closed (as the difference of a set and an OPEN neigbourhood of its frontier, see the end of the previous section). Hence K is compact. By Lemma (7.3.1.), K is Jordan measurable (each $\overset{\circ}{Q}_i$ is obviously Jordan measurable). At last, $A\setminus A \subset \cup Q_i$, hence

$$\operatorname{vol}(A\setminus K) \leq \sum \operatorname{vol} Q_i \leq \varepsilon. \rhd$$

Theorem 7.3.3. *Let A be a (bounded) Jordan measurable set, and let f be a (bounded) function integrable over A. Let O be an open covering of A by Jordan measurable sets, and let Φ be a partition of unity for A submitted to O. Then*

$$\int_A f = \sum_{\varphi\in\Phi} \int_A \varphi f, \tag{2}$$

where the series converges ABSOLUTELY.

(Recall that a series $\sum_{\varphi\in\Phi} a_\varphi$ $(a_\varphi \in \mathbb{R})$ *converges absolutely* to s if for any $\varepsilon > 0$ there exists a FINITE set $\Phi_0 \subset \Phi$ such that for each FINITE set Φ', satisfying the condition $\Phi_0 \subset \Phi' \subset \Phi$, it holds

$$\left| s - \sum_{\phi\in\Phi'} a_\phi \right| \leq \varepsilon.$$

In such a case the series $\sum_{\phi\in\Phi} |a_\phi|$ also (absolutely) converges.)
$\lhd$ Consider an arbitrary $\varepsilon > 0$. By Lemma 7.3.2., there exists a compact Jordan measurable set $K \subset A$ such that

$$\operatorname{vol}(A\setminus K) \leq \varepsilon. \tag{3}$$

By Remark to the definition of a partition of unity, the set Φ_0 of all functions φ from Φ such that $\varphi|_K \neq 0$, is FINITE. For any finite Φ' such that $\Phi_0 \subset \Phi' \subset \Phi$ it holds

$$\left| \int_A - \sum_{\varphi\in\Phi'} \int_A \varphi f \right| \overset{\text{sum is finite}}{=} \left| \int_A \left(f - \sum_{\varphi\in\Phi'} \varphi f \right) \right| \leq \int_A \left| f \left(1 - \sum_{\varphi\in\Phi'} \varphi \right) \right|$$

$$\overset{\sum_{\varphi\in\Phi'}\varphi|_A=1}{\leq} \underbrace{\sup_A |f|}_{=:M} \int_A \left(\sum_{\varphi\in\Phi}\varphi - \sum_{\varphi\in\Phi'}\varphi\right) = M\int_A \sum_{\varphi\in\Phi\setminus\Phi'}\varphi$$

$$\underset{\substack{\varphi\in\Phi\setminus\Phi' \\ \Rightarrow\varphi|_K=0}}{=} M\int_{A\setminus K} \underbrace{\sum_{\varphi\in\Phi\setminus\Phi'}\varphi}_{\leq 1} \leq M\int_{A\setminus K} 1 = M\,\mathrm{vol}(A\setminus K) \overset{(3)}{\geq} M\varepsilon.$$

Since ε was arbitrary we conclude that (2) is true. $\triangleright$
Remark. 1. Since $\sum_{\varphi\in\Phi}\varphi|_A = 1$, we can rewrite (2) so:

$$\int_A \sum_{\varphi\in\Phi}\varphi f = \sum_{\varphi\in\Phi}\int_A \varphi f,$$

that is, we can change $\int$ and $\sum$ with places.
Remark. One can use (2) to EXTEND the definition of $\int_A f$ to **non**-Jordan-measurable or/and **non**-bounded sets A and **non**-bounded functions f. But we shall not need such an extension below.

7.4 Change of variables

The following result is a generalization of the known rule of classic analysis concerning a change of a variable in an integral.

By a *change of variables* in $\mathbb{R}^n$ we mean a (C^1-)*diffeomorphism* g of an open set $G \subset \mathbb{R}^n$ onto an open set $H \subset \mathbb{R}^n$, that is, a C^1-bijection $G \to H$ such that the inverse mapping $g^{-1} : H \to G$ is also of class C^1.

Since any C^1-mapping is continuous, both g and g^{-1} are continuous, thus g is a homeomorphism.

Remember: any (C^1-)diffeomorphism is a homeomorphism.

Since $g^1 \circ g = \mathrm{id}$ and $g \circ g^{-1} = \mathrm{id}$, it follows by Chain Rule that for any $x \in G$ and for $y := g(x)$

$$(g^{-1})'(y) \circ g'(x) = \mathrm{id}, \qquad g'(x) \circ (g^{-1})'(y) = \mathrm{id}\,.$$

Hence (remember!)

$$\forall x \in G \vdots\ g'(x) \in \mathrm{Iso}(\mathbb{R}^n)\ (\Leftrightarrow \det g'(x) \neq 0).$$

Theorem 7.4.1. *Let $g : G \to H$ ($G, H \subset \mathbb{R}^n$) be a ($C^1$-)diffeomorphism. Then for any integrable function $f : H \to \mathbb{R}$ it holds (below we prefer write $g(G)$ instead of H)*

$$G \overset{g}{\to} H \overset{f}{\to} \mathbb{R} \qquad \int_{g(G)} f = \int_G |f \circ g| \det g'|. \tag{1}$$

$\triangleleft$ I. PRELIMINARIES. 1° This theorem is true for integrals in the extended sense mentioned in last Remark. But we shall prove this theorem only for our "old" notion of the integral, and by this reason we shall suppose that our sets G and H are bounded and G is Jordan measurable. (It follows from (1) with $f = 1$ that H must then also be Jordan measurable.

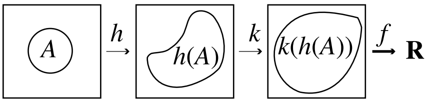
2° We say that a Jordan measurable set $A \subset G$ is *nice* for a diffeomorphism g and we write

$$A \in \mathrm{Nice}(g),$$

if for any (bounded) integrable function f (on H)

$$\int_{g(A)} = \int_A (f \circ g)|\det g'|. \tag{2}$$

It follows from (2) with $f = 1$ that $g(A)$ is then also to be Jordan measurable. Thus our aim is to prove that G is nice for g.



II. CONDITIONAL PART.
1° $A \in \mathrm{Nice}(h), h(A) \in \mathrm{Nice}(k) \Rightarrow A \in \mathrm{Nice}(k\circ h)$. ⊲⊲ For any f integrable on $k(h(A))$ it holds

$$\int_{k(h(A))} f = \int_{h(A)} (f \circ k)|\det k'| = \int_A (((f \circ k)|\det k'|) \circ h)|\det h'|$$

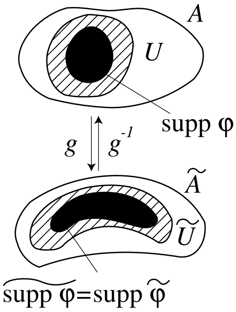
$$\overset{\substack{(\varphi\psi)\circ\gamma = \\ (\varphi\circ\gamma)(\psi\circ\gamma)}}{=} \int_A (f \circ k \circ h)|\det(k' \circ h)||\det h'|$$

$$\overset{\substack{\det(B\circ C) = \\ (\det B)(\det C)}}{=} \int_A (f \circ k \circ h)|\det \underbrace{((k' \circ h) \circ h')}_{\overset{\text{Ch.rule}}{=} (k\circ h)'}|.\ ⊳⊳$$

2° *If an open Jordan measurable set A admits an open covering O by sets each of which is a subset of A and is nice for a diffeomorphism g then A itself is nice for g.*
⊲⊲ For any set $S \subset G$ put for short

$$\widetilde{S} := g(S).$$

Since g is a homeomorphism, the sets $\widetilde{U}$ with $U \in O$ form an open covering of $\widetilde{A}$; note that each $\widetilde{U}$ is Jordan measurable, for U is nice (see I, 2°). Denote this covering by $\widetilde{O}$. By Theorem 2.2 there exists a partition of unity Φ for A submitted to O. For any $\varphi \in \Phi$ put $\widetilde{\varphi} := \varphi \circ g^{-1}$ (so that $\varphi = \widetilde{\varphi} \circ g$). It is clear that the functions $\widetilde{\varphi}$ with $\varphi \in \Phi$, form a partition of unity for $\widetilde{A}$ submitted to $\widetilde{O}$. Denote this partition by $\widetilde{\Phi}$. We have for any integrable function f



$$\int_{g(A)} f \overset{7.3.3.}{=} \sum_{\widetilde{\varphi}\in\widetilde{\Phi}} \int_{\widetilde{A}} f\widetilde{\varphi} \overset{\substack{\exists \widetilde{U}\in\widetilde{O} \\ \mathrm{supp}\,\widetilde{\varphi}\subset\widetilde{U}}}{=} \sum_{\widetilde{\varphi}\in\widetilde{\Phi}} \int_{\widetilde{U}} f\widetilde{\varphi}$$

$$\overset{U\in\mathrm{Nice}(g)}{=} \sum_{\widetilde{\varphi}\in\widetilde{\Phi}} \int_U \underbrace{((f\widetilde{\varphi}) \circ g)}_{=(f\circ g)\underbrace{(\widetilde{\varphi}\circ g)}_{=\varphi}} |\det g'|$$

$$\overset{\mathrm{supp}\subset U}{=} \sum_{\varphi\in\Phi} \int_A ((f \circ g)|\det g'|)\varphi \overset{7.3.3.}{=} \int_A (f \circ g)|\det g'|.\ ⊳⊳$$

3° *Let Q be a cube in $g(G)$. If for any cube $S \subset Q$*

$$\int_S 1 = \int_{g^{-1}(S)} |\det g'| \tag{3}$$

(that is, (2) is true for $g^{-1}(S)$ and $f = 1$), then both pre-images $g^{-1}(Q)$ and $g^{-1}(\overset{\circ}{Q})$ are nice for g.

◁◁ a) The pre-image for any cube $S \subset Q$ with $\operatorname{vol} S = 0$ also has zero volume. ◁◁◁

$$0 = \int_S 1 \overset{(3)}{=} \int_{g^{-1}(S)} \underbrace{|\det g'|}_{\geq m \overset{(*)}{>} 0} \geq m \int_{g^{-1}(S)} 1 \overset{m>0}{\Rightarrow} \int_{g^{-1}(S)} 1 = 0.$$

$(*)$: $|\det g'|$ is a continuous function on the compact $g^{-1}(S)$ which is nowhere 0. ▷▷▷

b) For any cube $S \subset Q$ and any bounded function f on G

$$\int_{\operatorname{fr} g^{-1}(S)} f = 0.$$

◁◁◁ $\operatorname{fr} g^{-1}(S) \overset{g\in\text{Homeo}}{=} g^{-1}(\operatorname{fr} S)$. Since $\operatorname{fr} S$ is a finite union of cubes with zero volume it follows by a) that $\operatorname{fr} g^{-1}(S)$ is a finite union of zero volume sets and hence has itself zero volume. But the integral of bounded function over a volume 0 set is equal to 0. ▷▷▷

c) For any integrable function f on Q it holds

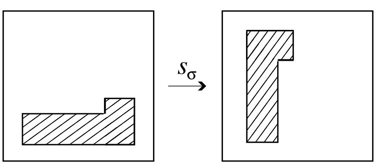
$$\forall P \in \operatorname{Part}(Q) \vdots L_P f = \sum_{S\in\operatorname{Cube} P} (\inf_S f) \underbrace{\operatorname{vol} S}_{=\int_S 1}$$

$$\overset{(3)}{=} \sum_{S\in\operatorname{Cube} P} \int_{g^{-1}(S)} \underbrace{(\inf_S f)}_{\leq (f\circ g)|_{g^{-1}(S)}} |\det g'| \overset{\text{b)}}{\leq} \int_{g^{-1}(Q)} (f \circ g)|\det g'|.$$

It follows that $\int_Q f \leq \int_{g^{-1}(Q)} (f \circ g)|\det g'|$. Analogously we conclude that the inverse inequality is true (consider $U_P f$). Hence $\int_Q f = \int_{g^{-1}(Q)} (f \circ g)|\det g'|$. By b) we conclude that also $\int_{\overset{\circ}{Q}} f = \int_{g^{-1}(\overset{\circ}{Q})} (f \circ g)|\det g'|$. ▷▷

III. ABSOLUTE PART. 1° *For any permutation $\sigma \in \mathfrak{S}_n$, each Jordan measurable set $A \subset \mathbb{R}^n$ is nice for s_σ, where*

$$s_\sigma(x_1, \ldots, x_n) := (x_{\sigma(1)}, \ldots, x_{\sigma(n)}).$$



(It is clear that s_σ is a linear bijection $\mathbb{R}^n \to \mathbb{R}^n$ and hence is a diffeomorphism.)

◁◁ The matrix of s_σ has evidently the determinant equal ± 1, so

$$\int_A (f \circ s_\sigma) \underbrace{|\det s'_\sigma|}_{=1} = \int_A f \circ s_\sigma = \int_A f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})\, \mathrm{d}x_1 \ldots \mathrm{d}x_n$$

$$\overset{\text{change the order of integrations by Fubini Theorem}}{=} \int_{s_\sigma(A)} f(y_1, \ldots, y_n)\, \mathrm{d}y_1 \ldots, \mathrm{d}y_n = \int_{s_\sigma(A)} f. \rhd\!\rhd$$

2° *Theorem is true for $n = 1$.*

$\lhd\!\lhd$ For any $[\alpha, \beta] \subset g(G)(\subset \mathbb{R})$ it holds

$$\int_{[\alpha,\beta]} 1 \overset{\text{obv}}{=} \beta - \alpha \overset{\text{Newton–Leibniz}}{=} \int_{g^{-1}(\alpha)}^{g^{-1}(\beta)} g' \overset{\text{consider two possible cases: } g'>0,\, g'<0}{=} \int_{g^{-1}([\alpha,\beta])} |g'|.$$

Hence by II, 3° the pre-image of any open interval in $g(G)$ is nice for g. Since these pre-images (which are open intervals) cover G we conclude by II, 2° that G is nice for g.

3° Now argue by induction. Let Theorem is true for $n - 1$.

4° For any point $\hat{x} \in G$ there exists an open neigbourhood $U_{\hat{x}}$ such that

$$g|_{U_{\hat{x}}} = k \circ h \circ s_\sigma, \tag{4}$$

where $\sigma \in \mathfrak{S}_n$, and k and h are diffeomorphism, each of which DOES NOT CHANGE AT LEAST ONE COORDINATE.

$\lhd\!\lhd$ 5° We have, putting $g =: (g_1, \ldots, g_n)$,

$$\det g'(\hat{x}) = \left.\begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{vmatrix}\right|_{\hat{x}} \overset{(*)}{=} \left.\left(\frac{\partial g_n}{\partial x_1} M_{n1} + \ldots + \frac{\partial g_n}{\partial x_n} M_{nn}\right)\right|_{\hat{x}}.$$

$(*)$: decomposition of the determinant corresponding to the last row.

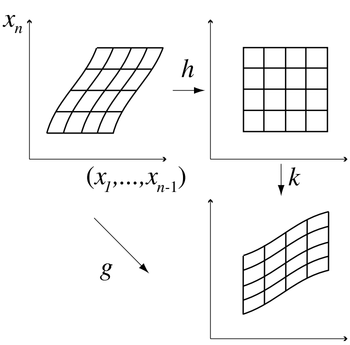
Since $\det g'(\hat{x}) \neq 0$, we have

$$\left.\frac{\partial g_n}{\partial x_i}\right|_{\hat{x}} \neq 0, \qquad M_{ni}|_{\hat{x}} \neq 0 \qquad \text{for some } i.$$

Take as σ the TRANSPOSITION of i and n. Then $g \circ s_\sigma =: \widetilde{g}$ satisfies the conditions

$$\left.\frac{\partial \widetilde{g}}{\partial x_n}\right|_{s_\sigma^{-1}(\hat{x})} \neq 0, \qquad \widetilde{M}_{nn}\big|_{s_\sigma^{-1}(\hat{x})} \neq 0.$$

If we decompose $\hat{g}$ into the composition of diffeomorphisms h and k as above, we obtain the diserable decomposition (4), since $\widetilde{g} \circ s_\sigma = g \circ \underbrace{s_\sigma \circ s_\sigma}_{=\mathrm{id}} = g$.



Thus without loss of generality we can assume that $i = n$, so that

$$\left.\frac{\partial g_n}{\partial x_n}\right|_{\hat{x}} \neq 0, \qquad M_{nn}|_{\hat{x}} \neq 0. \tag{5}$$

6° Put

$$h(x) := (g_1(x), \ldots, g_{n-1}(x), x_n), \tag{6}$$

so that h DOES NOT CHANGE THE LAST COORDINATE. We have

$$\det h'|_{\hat{x}} = \left|\begin{matrix} \frac{\partial g_1}{\partial x_1} & \dots & \frac{\partial g_1}{\partial x_{n-1}} & \frac{\partial g_1}{\partial x_n} \\ \vdots & \dots & \vdots & \vdots \\ \frac{\partial g_{n-1}}{\partial x_1} & \dots & \frac{\partial g_{n-1}}{\partial x_{n-1}} & \frac{\partial g_{n-1}}{\partial x_n} \\ 0 & \dots & 0 & 1 \end{matrix}\right|\Bigg|_{\hat{x}} = M_{nn}|_{\hat{x}} \overset{(5)}{\neq} 0.$$

By Inverse Function Theorem, there exists an open neighbourhood $U_{\hat{x}}$ of $\hat{x}$ such that

$$h|_{U_{\hat{x}}} \in \mathrm{Diffeo}\,.$$

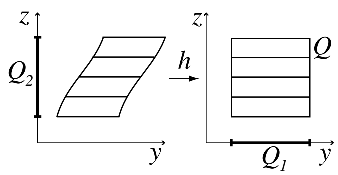
7° Now put

$$k := g \circ h^{-1} \tag{7}$$

(k is a diffeomorphism as a composition of two diffeomorphisms). This k DOES NOT CHANGE THE FIRST $n-1$ COORDINATES. Indeed, if $x = (x_1, \dots, x_n)$ and $h(x) = (y_1, \dots, y_n)$, then, by (6), $y_1 = g_1(x), \dots, y_{n-1} = g_{n-1}(x)$, $y_n = x_n$, so

$$k(y_1, \dots, y_n) \overset{(7)}{=} g(x) = (g_1(x), \dots, g_n(x)) = (y_1, \dots, y_{n-1}, g_n(x)).\ \rhd\!\rhd$$

8° $U_{\hat{x}} \in \mathrm{Nice}(h)$.



$\lhd\!\lhd$ For short put $x = (\underbrace{x_1, \dots, x_{n-1}}_{=:y}, \underbrace{x_n}_{=:z}) = (y, z)$, $h(y,z) =: (a(y,z), z)$. For any cube $Q = Q_1 \times Q_2$ in $h(U_{\hat{x}})$ it holds

$$\int_Q 1 \overset{\substack{\text{Fubini}\\ \text{Theorem}}}{=} \int_{Q_2} \mathrm{dz} \int_{Q_1} \mathrm{d}y \overset{3^\circ}{=} \int_{Q_2} \mathrm{dz} \int_{(a(\cdot,z)^{-1})(Q_1)} |\det \underbrace{(a(\cdot, z))'}_{=\partial h/\partial y}|$$

$$\overset{\det h' = \left|\begin{smallmatrix} \partial h/\partial y & \partial h/\partial z \\ 0 & 1 \end{smallmatrix}\right| = \det \partial h/\partial y}{=} \int_{Q_2} \mathrm{dz} \int_{(a(\cdot,z)^{-1})(Q_1)} |\det h'| \overset{\substack{\text{Fubini}\\ \text{Theorem}}}{=} \int_{h^{-1}(Q)} |\det h'|.$$

By II, 3° we conclude that for any cube $Q \subset h(U_{\hat{x}})$ we have $h^{-1}(\mathring{Q}) \in \mathrm{Nice}(h)$. But such the pre-images cover $U_{\hat{x}}$, so, by II, 2°, $U_{\hat{x}}$ is nice for h. $\rhd\!\rhd$
9° $h(U_{\hat{x}})$ Nice(k). $\lhd\!\lhd$ Quite analogously. $\rhd\!\rhd$
10° $U_{\hat{x}} \in \mathrm{Nice}(g)$. $\lhd\!\lhd$ This follows from (4), 1°, 5°, 6° and II, 1°. $\rhd\!\rhd$
11° $G \in \mathrm{Nice}(g)$ $\lhd\!\lhd$ This follows II, 2°, since the neighbourhoods $U_{\hat{x}}, \hat{x} \in G$, cover G and are nice for g, by 7°. $\rhd\!\rhd$ $\rhd$

Corollary 7.4.2. *Let g be a diffeomorphism of an open set $G \subset \mathbb{R}^n$ onto an open set $H \subset \mathbb{R}^n$, and let $A \subset G$. If* $\mathrm{vol}\, A = 0$ *then* $\mathrm{vol}\, g(A) = 0$.
$\lhd$ Exercise. $\rhd$
NB If g is merely a **homeo**morphism then $\mathrm{vol}\, A = 0 \not\Rightarrow \mathrm{vol}\, g(A) = 0$. (A counter-example can be constructed using two Cantor type sets, the usual one, with zero volume, and a modification, with a positive Lebesgue measure)

Chapter 8

Differential forms

8.1 Tensors

By *tensor* of rank k (or k-tensor), $k = 1, 2, \ldots$, on a vector space X we mean a k-LINEAR functional

$$u : \underbrace{X \times \ldots \times X}_{k-\text{times}} \to \mathbb{R}.$$

The set of all k-tensors on X we denote by $\mathrm{L}^k(X)$:

$$\mathrm{L}^k(X) := \mathrm{L}(\underbrace{X \times \ldots \times X}_{k}; \mathbb{R}).$$

It is convenient to put

$$\mathrm{L}^0(X) = \mathbb{R}.$$

Notations. Our main special case is $X = \mathbb{R}^n$, with points $x = (x_1, \ldots, x_n)$. We denote by $\mathrm{e}_1, \ldots, \mathrm{e}_n$ the *canonical basis* in $\mathbb{R}^n$:

$$\mathrm{e}_i := (0, \ldots, 0, \underset{[i]}{1}, 0, \ldots, 0),$$

and by π_i the canonical projections in $\mathbb{R}^n$:

$$\pi_i x := x_i.$$

It is clear that

$$\pi_i \mathrm{e}_j = \delta_{ij} := \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if not.} \end{cases} \tag{1}$$

Examples.

1. $\mathrm{L}^1(X) = \mathrm{L}(X, \mathbb{R}) =: X'$ (the *dual* vector space); 1-tensors are oft called *covectors*.
2. For a smooth function $f : \mathbb{R}^n \to \mathbb{R}$

$$f^{(k)}(x) \in \mathrm{L}^k_{\mathrm{sym}}(\mathbb{R}^n) \subset \mathrm{L}^k(\mathbb{R}^n),$$

where $\mathrm{L}^k_{\mathrm{sym}}(X)$ denotes the set of all SYMMETRIC k-linear functionals.

3. For any $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^n)$ we can define a 2-tensor u_A by the formula

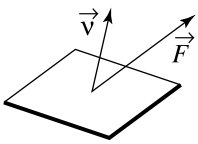
$$u_A(h, k) := \langle Ah, k \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^n$. (The correspondence $A \mapsto u_A$ is a bijection $\mathrm{L}(\mathbb{R}^n, \mathbb{R}^n) \to \mathrm{L}^2(\mathbb{R}^n)$).

4. (Elasticity theory.) Let $f(x)$ denotes the position of a point x of a (3-dimensional) body after a deformation. The 2-tensor generated (in the sense of previous Example) by the operator

$$\tfrac{1}{2}(f'(x) + f'(x)^T) - \mathrm{id}$$

(here $f'(x) \in \mathrm{L}(\mathbb{R}^3, \mathbb{R}^3)$; the symbol T denotes the transposed matrix; we identify linear operators in $\mathbb{R}^n$ with their matrices) is called *deformation tensor*.
The 2-tensor generated by the operator $\vec{\nu} \mapsto \vec{F}$, where $\vec{\nu}$ is the unit normal vector to some flat section of the body, and $\vec{F}$ is the force that acts "on 1 cm^2" of the surphace of that graf of our persected body for which $\vec{\nu}$ is OUTER normal vector, is called the *stress tensor*. The known *Hook law* says that the stress tensor at a point linearly depend on the deformation tensor at this point. The corresponding matrix describes the elasticity properties of our body at the point in question.



5. The determinat can be considered as a tensor:

$$\det(h_1, \dots, h_n) := \begin{vmatrix} h_{11} & \dots & h_{1n} \\ \dots & \dots & \dots \\ h_{n1} & \dots & h_{nn} \end{vmatrix}$$

(where $h_i = (h_{i1}, \dots, h_{in}) \in \mathbb{R}^n$).

The main operations over tensors are TENSOR PRODUCT and PULL-BACK.

Tensor product

Let $u \in \mathrm{L}^p(X)$, $v \in \mathrm{L}^q(X)$, $p, q \geq 1$. The *tensor product* $u \otimes v$ is defined by the formula

$$u \otimes v(h_1, \dots, h_{p+q}) := \underbrace{u(h_1, \dots, h_p)}_{\in\mathbb{R}} \cdot \underbrace{v(h_{p+1}, \dots, h_{p+q})}_{\in\mathbb{R}}.$$

It is clear that $u \otimes v \in \mathrm{L}^{p+q}(X)$.

For $t \in \mathrm{L}^0(X) = \mathbb{R}$ it is convenient to put

$$t \otimes u := tu.$$

NB In general $u \otimes v \neq v \otimes u$.

Example. In $\mathbb{R}^n$, $\pi_i \otimes \pi_j$ corresponds (in the sense of Example 3) above) to the operator A with the matrix with just one non-zero element which is equal to 1:

$$\begin{array}{cc} \begin{pmatrix} 0 & \vdots & 0 \\ \dots & 1 & \dots \\ 0 & \vdots & 0 \end{pmatrix} & [j] \\ [i] & \end{array}$$

$\lhd$ $(\pi_i \otimes \pi_j)(h, k) = (\pi_i h) \cdot (\pi_j k) = h_i k_j = \langle Ah, k \rangle$. $\rhd$

Theorem 8.1.1. *The operation $\otimes$ is distributive and associative:*

$$(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v, \qquad u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2,$$

$$t(u \otimes v) = (tu) \otimes v = u \otimes (tv), \qquad (u \otimes v) \otimes w = u \otimes (v \otimes w).$$

$\triangleleft$ Easy exercise. $\triangleright$

Theorem 8.1.2. (on basis of $\mathrm{L}^k(\mathbb{R}^n)$) *For any $k = 1, 2, \dots$ the products $\pi_{i_1} \otimes \dots \otimes \pi_{i_k}$ ($i_j \in \{1, \dots, n\}$) form a basis of the vector space $\mathrm{L}^k(\mathbb{R}^n)$. Hence,*

$$\dim \mathrm{L}^k(\mathbb{R}^n) = n^k.$$

$\triangleleft$ 1° let $u \in \mathrm{L}^k(\mathbb{R}^n)$. Then

$$\begin{aligned} u(h_1, \dots, h_k) &= u\left(\sum_{i_1=1}^{n} h_{1i_1} \mathrm{e}_{i_1}, \dots, \sum_{i_k=1}^{n} h_{ki_k} \mathrm{e}_{i_k}\right) \\ &\overset{u \in \mathrm{L}^k}{=} \sum_{i_1,\dots,i_k=1}^{n} \underbrace{h_{1i_1} \dots h_{ki_k}}_{=(\pi_{i_1} \otimes \dots \otimes \pi_{i_k})(h_1,\dots,h_k)} \underbrace{u(\mathrm{e}_{i_1}, \dots, \mathrm{e}_{i_k})}_{=:a_{i_1 \dots i_k}} \\ &= \left(\sum_{i_1,\dots,i_k=1}^{n} a_{i_1 \dots t_k} \pi_{i_1} \otimes \dots \otimes \pi_{i_k}\right)(h_1, \dots, h_k). \end{aligned}$$

Hence,

$$u = \sum_{i_1,\dots,i_k=1}^{n} \pi_{i_1} \otimes \dots \otimes \pi_{i_k},$$

that is, our products span $\mathrm{L}^k(\mathbb{R}^n)$.
2° They are linearly independent. Indeed, let

$$u = \sum_{i_1,\dots,i_k=1}^{n} a_{i_1 \dots i_k} \pi_{i_1} \otimes \dots \otimes \pi_{i_k} = 0$$

Applying this to $(\mathrm{e}_{i'_1}, \dots, \mathrm{e}_{i'_k})$, we obtain, by (1),

$$a_{i'_1 \dots i'_k} = 0. \triangleright$$

Pull-back

Let X, Y be a vector spaces, and let $l \in \mathrm{L}(X, Y)$. For any $v \in \mathrm{L}^k(Y)$ we define the *pull-back* $l^* v$ *of* v, putting

$$\begin{array}{ccc} X & \xrightarrow{\ l\ } & Y \\ \mathrm{L}^k(X) & \xleftarrow{\ l^*\ } & \mathrm{L}^k(Y) \end{array}$$

$$(l^* v)(h_1, \dots, h_k) := v(lh_1, \dots, lh_k).$$

It is clear that

$$l^* v \in \mathrm{L}^k(X).$$

So we “pull” the tensor v “back” to X.

$$\begin{array}{ccc} & \mathbb{R} & \\ l^*y' \nearrow & & \nwarrow y' \\ X & \xrightarrow{l} & Y \\ X' & \xleftarrow{l^*} & Y' \end{array}$$

Example. For $k = 1$ we obtain the operator $l^* : Y' \to X'$ which act so:

$$l^*y' = y' \circ l.$$

$\lhd$ $(l^*y')h = y'(lh) = (y' \circ l)h.$ $\rhd$

This operator between the DUAL spaces is called the *dual operator* to l. Note that l^* act in the OPPOSITE direction.

Theorem 8.1.3. *Pull-back* RESPECTS *tensor product:*

$$f^*(u \otimes v) = (f^*u) \otimes (f^*v).$$

$\lhd$ Easy exercise. $\rhd$

8.2 Asymmetric tensors

A tensor $u \in \mathrm{L}^k(X)$ $(k \geq 2)$ is called *antisymmetric* if it has value 0 at any point $(h_1, \ldots, h_k)$ which has two equal components. The set of all antisymmetric tensors we denote by $\Lambda^k(X)$. Thus,

$$u \in \Lambda^k(X) :\Leftrightarrow u(\ldots, h, \ldots, h, \ldots) = 0.$$

It is convenient to put

$$\Lambda^1(X) := \mathrm{L}^1(X)(= X'),$$
$$\Lambda^0(X) := \mathrm{L}^0(X)(= \mathbb{R}).$$

Remark. An equivalent description is such: a tensor is antisymmetric iff it changes the sign by any transposition of its arguments:

$$u \in \Lambda^k(X) \Leftrightarrow u(\ldots, h, \ldots, k, \ldots) = -u(\ldots, k, \ldots, h, \ldots)$$

(all others arguments remain unchanged).
$\lhd$ "$\Rightarrow$": $\underbrace{u(\ldots, h+k, \ldots, h+k, \ldots)}_{=0} = \underbrace{u(\ldots, h, \ldots, h, \ldots)}_{=0} + u(\ldots, h, \ldots, k, \ldots) +$ $u(\ldots, k, \ldots, h, \ldots) + \underbrace{u(\ldots, k, \ldots, k, \ldots)}_{=0},$
hence $u(\ldots, h, \ldots, k, \ldots) + u(\ldots, k, \ldots, h, \ldots) = 0.$
"$\Leftarrow$": $u(\ldots, h, \ldots, h, \ldots) = -u(\ldots, h, \ldots, h, \ldots)$, hence $u(\ldots, h, \ldots, h, \ldots) = 0.$ $\rhd$

Examples.

1. det.

2. Let $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^n)$, and let u_A be the corresponding 2-tensor. Then

$$u_A \in \Lambda^2(\mathbb{R}^n) \Leftrightarrow A^T = -A.$$

(We identify A with the corresponding matrix.) $\lhd$ Exercise. $\rhd$ Operators A satisfying the condition $A^T = -A$, are also called *antisymmetric*. A typical example is rotation by 90° in $\mathbb{R}^2$, e.g. counter-clock-wise, with the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. The *corresponding antisymmetric tensor is just* det.

$\lhd$ $u_A(h,k) = \langle Ah, k\rangle = (k_1, k_2)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = h_1k_2 - h_2k_1 = \det(h,k)$. $\rhd$

Thus,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \longleftrightarrow \det .$$

Operator alt

From any tensor we can make an antisymmetric one. Viz., put for $u \in \mathrm{L}^k(X)$

$$(\operatorname{alt} u)(h_1,\dots,h_k) := \frac{1}{k!}\sum_{\sigma\in\mathfrak{S}_k}(\operatorname{sgn}\sigma)u(h_{\sigma(1)},\dots,h_{\sigma(k)}).$$

(In the sum the signs alternate $(+,-,+,-,\dots)$, whence the notation.)

Recall that $\operatorname{sgn}\sigma$ denotes the sign of a permutation. It is clear that $\operatorname{alt} u$ is a k-tensor.

Examples.

1. $\operatorname{alt}\langle\cdot,\cdot\rangle = 0$. (Recall that $\langle\cdot,\cdot\rangle$ denotes the scalar product, which is a symmetric 2-tensor.) More generally, alt sends ANY symmetric tensor to 0:

$$u \in \mathrm{Sym} \Rightarrow \operatorname{alt} u = 0.$$

NB The inverse implication is not true! See Exercise 8.2.2. 3) below.

2. If $u \leftrightarrow A$ (that is $u = u_A$), then $\operatorname{alt} u \leftrightarrow \frac{1}{2}(A - A^T)$. (Verify!)

Theorem 8.2.1. *The operator* alt *has the following properties:*

a) $\operatorname{alt} \in \mathrm{L}(\mathrm{L}^k(X), \Lambda^k(X))$, *that is,* $\operatorname{alt} u$ *is an antisymmetric tensor, and the mapping* $u \mapsto \operatorname{alt} u$ *is linear;*
b) $u \in \Lambda^k \Rightarrow \operatorname{alt} u = u$, *that is,* Λ^k *is* INVARIANT *under* alt*;*
c) $\operatorname{alt}^2 = \operatorname{alt}$, *that is,* alt *is an* IDEMPOTENT *operator; (*$\operatorname{alt}^2 u := \operatorname{alt}(\operatorname{alt} u)$*);*
d) $\operatorname{alt} u = 0 \Rightarrow \forall v \vdots\ \operatorname{alt}(u\otimes v) = 0$ *("bad sheep principle": one bad sheep spoils all the crew).*

$\lhd$ a) The sum defining $(\operatorname{alt} u)(\dots,h,\dots,h\dots)$ can be splitted onto pairs of the form

$$+u(\dots,h,\dots,h\dots) - u(\dots,h,\dots,h\dots),$$

where our two h appear on one and the same pair of places (different for different pairs). Hence the sum is equal to 0.

b) $\forall u \in \Lambda^k \vdots$

$$\begin{aligned}\operatorname{alt} u(h_1,\dots,h_k) &= \frac{1}{k!}\sum_{\sigma\in\mathfrak{S}_k}(\operatorname{sgn}\sigma)\underbrace{u(h_{\sigma(1)},\dots,h_{\sigma(k)})}_{=(\operatorname{sgn}\sigma)u(h_1,\dots,h_k)} \\ &= \frac{1}{k!}(k!u(h_1,\dots,h_k)) = u(h_1,\dots,h_k).\end{aligned}$$

c) By a), $\operatorname{alt} u \in \Lambda^k$, hence

$$\operatorname{alt}(\operatorname{alt} u) \overset{\text{b)}}{=} \operatorname{alt} u.$$

d) $(\mathrm{alt}(u \otimes v))h_1 \dots h_{p+q}$

$$= \frac{1}{(p+q)!} \sum_{\sigma \in \mathfrak{S}_{p+q}} (\operatorname{sgn} \sigma)(uh_{\sigma(1)}, \dots, h_{\sigma(p)})(vh_{\sigma(p+1)}, \dots, h_{\sigma(p+q)}).$$

If, for a fixed σ_0 we consider all the σ such that

$$\{\sigma(1), \dots, \sigma(p)\} = \{\sigma_0(1), \dots, \sigma_0(p)\} \qquad \text{(as NON-ORDED set!)},$$

$$\sigma(p+1) = \sigma_0(p+1), \dots, \sigma(p+q) = \sigma_0(p+q),$$

then the sum over all such such σ is equal to 0, since $\mathrm{alt}\, u = 0$. But the whole sum splits onto such subsums. $\triangleright$

Exercise 8.2.2. 1) $\mathrm{alt} \circ \mathrm{sym} = 0$.
2) $\mathrm{sym} \circ \mathrm{alt} = 0$.
3) Give an example of $u \in \mathrm{L}^3(\mathbb{R}^3)$ such that $u \neq 0$, $\mathrm{sym}\, u = \mathrm{alt}\, u = 0$.
NB Such an u is **not** the sum of its symmetric part $\mathrm{sym}\, u$ and antisymmetric part $\mathrm{alt}\, u$. Only 2-tensors have this property.

Answer:

$$u\mathrm{e}_i\mathrm{e}_j\mathrm{e}_k = \begin{cases} \frac{2}{3} & \text{if } (i, j, k) = (1, 2, 3), \\ -\frac{1}{3} & \text{if } (i, j, k) = (3, 2, 1), \\ 0 & \text{otherwise.} \end{cases}$$

Using the operation alt, we can make from tensor product an operation over antisymmetric tensors.

Exterior product

For $u \in \Lambda^p(X)$ $(p \geq 1)$, $v \in \Lambda^q(X)$ $(q \geq 1)$, the *exerior product* $u \wedge v$ is defined by the formula

$$u \wedge v := \frac{(p+q)!}{p!\, q!} \mathrm{alt}(u \otimes v).$$

Remarks. 1) $u \otimes v$ itself is **not** in general antisymmetric (give an example!).
2) The coefficient in this formula is chosen to obtain the coefficient 1 in the formula (1) below.

Example. In $\mathbb{R}^2$, $\pi_1 \wedge \pi_2 = \det$. $\triangleleft$ $(\pi_1 \otimes \pi_2)(h, k) = \pi_1 h \cdot \pi_2 k = h_1 k_2$; hence $(\pi_1 \wedge \pi_2)(h, k) = (1+1)!/(1!\,1!)\,\mathrm{alt}(\pi_1 \otimes \pi_2)(h, k) = 2(\frac{1}{2}(\pi_1 \otimes \pi_2)(h, k) - \frac{1}{2}(\pi_1 \otimes \pi_2)(k, h)) = h_1 k_2 - k_1 h_2 = \det(h, k)$. $\triangleright$

Exercise 8.2.3. Prove that for $u \in \Lambda^p$, $u \in \Lambda^q$

$$(u \wedge v)h_1 \dots h_{p+q} = \sum_{\substack{\sigma \in \mathfrak{S}_{p+q} \\ \sigma(1) < \sigma(2) < \dots < \sigma(p) \\ \sigma(p+1) < \dots < \sigma(p+q)}} (\operatorname{sgn} \sigma)(uh_{\sigma(1)} \dots h_{\sigma(p)})(vh_{\sigma(p+1)} \dots h_{\sigma(p+q)}) \qquad (1)$$

Theorem 8.2.4. *The operation $\wedge$ has the following properties:*

a) $(u + v) \wedge v = u_1 \wedge v + u_2 \wedge v$, $\quad u \wedge (v_1 + v_2) = u \wedge v_1 + u \wedge v_2$,
$t(u \wedge v) = (tu) \wedge v = u \wedge (tv)$ $t \in \mathbb{R}$ *(distributivity);*

b) $u \wedge v = (-1)^{pq} v \wedge u$ *$(u \in \Lambda^p, v \in \Lambda^q)$* *("semi-commutativity");*

c) $(u \wedge v) \wedge w = u \wedge (v \wedge w) = (p+q+r)!/(p!q!r!)\,\mathrm{alt}(u \otimes v \otimes w) =: u \wedge v \wedge w$ *($u \in \Lambda^p, v \in \Lambda^q, w \in \Lambda^r$) (associativity);*

d) $f^*(v \wedge v) = (f^*u) \wedge (f^*v)$ *(pull-back* RESPECTS *exterior product).*

◁ a) Obvious.

b) Consider the permutation

$$\sigma_0 = \begin{pmatrix} 1 & \dots & p & p+1 & \dots & p+q \\ 1+q & \dots & p+q & 1 & \dots & q \end{pmatrix}$$

It is clear that $\operatorname{sgn}\sigma_0 = (-1)^{pq}$ (1 is transposed q times, then 2 is transposed q times, ... p is transposed q times).

Each permutation $\sigma \in \mathfrak{S}_{p+q}$ can be written an the form $\sigma = \sigma' \circ \sigma_0$. Then

$$\sigma(p+1) = \sigma'(\sigma_0(p+1)) = \sigma'(1), \dots, \sigma(1) = \sigma'(\sigma_0(q+1)) = \sigma'(q+1), \dots \quad (2)$$

and

$$\operatorname{sgn}\sigma = (\operatorname{sgn}\sigma')(\operatorname{sgn}\sigma_0) = (-1)^{pq}\operatorname{sgn}\sigma'. \quad (3)$$

Hence

$$\begin{aligned}
(u \wedge v)(h_1, \dots, h_{p+q}) &= \underbrace{\frac{(p+q)!}{p!\,q!}}_{=:c}\mathrm{alt}(u \otimes v)(h_1, \dots, h_{p+q}) \\
&= \frac{c}{(p+q)!}\sum_{\sigma \in \mathfrak{S}_{p+q}} (\operatorname{sgn}\sigma) u(h_{\sigma(1),\dots,h_{\sigma(p)}}) v(h_{\sigma(p+1)}, \dots, h_{\sigma(p+q)}) \\
&\overset{\substack{\sigma=\sigma'\circ\sigma_0 \\ (2),(3)}}{=} \frac{c}{(p+q)!}\sum_{\sigma' \in \mathfrak{S}_{p+q}} (-1)^{pq}(\operatorname{sgn}\sigma')(h_{\sigma'(q+1)}, \dots, h_{\sigma'(q+p)}) v(h_{\sigma'(1)}, \dots, h_{\sigma'(q)}) \\
&= (-1)^{pq} c\,\mathrm{alt}(v \otimes u)(h_1, \dots, h_{p+q}) = (-1)^{pq}(u \wedge v)(h_1, \dots, h_{p+q}).
\end{aligned}$$

c) To verify that

$$(u \wedge v) \wedge w = \frac{(p+q+r)!}{p!\,q!\,r!}\mathrm{alt}(u \otimes v \otimes w)$$

we need (after canceling constant factor) to verify that

$$\underbrace{\mathrm{alt}((\mathrm{alt}(u \otimes v)) \otimes v)}_{[1]} = \underbrace{\mathrm{alt}(u \otimes v \otimes w)}_{[2]}.$$

But $[1]-[2] \overset{\mathrm{alt}\in\mathrm{L}}{=} \mathrm{alt}((\mathrm{alt}(u \otimes v)) \otimes w - u \otimes v \otimes w) \overset{\mathrm{distr.}}{=} \mathrm{alt}(\underbrace{(\mathrm{alt}(u \otimes v) - u \otimes v)}_{[3]} \otimes w)$

$\overset{\text{bad Sheep Priciple}}{=} 0$, since $\mathrm{alt}[3] \overset{\mathrm{alt}\in\mathrm{L}}{=} \mathrm{alt}^2(u \otimes v) - \mathrm{alt}(u \otimes v) \overset{\mathrm{alt}^2=\mathrm{alt}}{=} 0.$

d) Easy exercise. ▷

Corollary 8.2.5. *For any antisymmetric tensor of* ODD *rank its exterior product by itself is equal to* 0.

◁ If $u \in \Lambda^k$, $k \in$ Odd, then $u \wedge u \overset{\mathrm{b)}}{=} \underbrace{(-1)^{k^2}}_{=-1}$, whence $2(u \wedge u) = 0$. ▷

Basis of $\Lambda^k(\mathbb{R}^n)$

Theorem 8.2.6. *For any* $1 \le k \le n$ *the set* $\{\pi_{i_1} \wedge \ldots \wedge \pi_{i_k} \mid i_1 < \ldots < i_k\}$ *is a basis of the vector space* $\Lambda^k(\mathbb{R}^n)$*. hence*

$$\dim \Lambda^k(\mathbb{R}^n) = \binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

$\lhd$ By Theorem on basis of $\mathrm{L}^k(\mathbb{R}^n)$, we can write any $u \in \Lambda^k(\mathbb{R}^n)$ as a linear combination of $\pi_{i_1} \otimes \ldots \otimes \pi_{i_k}$:

$$u := \sum_{i_1,\ldots,i_k=1}^{n} \ldots \pi_{i_1} \otimes \ldots \otimes \pi_{i_k} \tag{4}$$

(the first dots mean a number coefficient). It follows that

$$u \overset{\text{Th 8.2.1.}}{=} \operatorname{alt} u \overset{(4)}{=} \operatorname{alt} \sum \ldots \pi_{i_1} \otimes \ldots \otimes \pi_{i_k} \overset{\text{alt}\in\mathrm{L}}{=} \sum \ldots \operatorname{alt}(\pi_{i_1} \otimes \ldots \otimes \pi_{i_k})$$
$$\overset{\text{Th 8.2.4.}}{=} \sum \ldots \pi_{i_1} \wedge \ldots \wedge \ldots \pi_{i_k} \overset{\text{Th 8.2.4.,b),cor.}}{=} \sum_{i_1<\ldots<i_k} \ldots \pi_{i_1} \wedge \pi_{i_k}.$$

Thus, our set spans $\Lambda^k(\mathbb{R}^n)$. The linear independence can be proved just as in the case of $\mathrm{L}^k(\mathbb{R}^n)$. $\rhd$

Corollary 8.2.7. *The space* $\Lambda^n(\mathbb{R}^n)$ *is* 1*-dimensional. Hence (since* $\det \in \Lambda^n(\mathbb{R}^n)$*) any element of* $\Lambda^n(\mathbb{R}^n)$ *has the form*

$$c \det \qquad (c \in \mathbb{R}).$$

Corollary 8.2.8. *In* $\mathbb{R}^n$,

$$\boxed{\pi_1 \wedge \ldots \wedge \pi_n = \det.}$$

$\lhd$ By Corollary 8.2.7., $\pi_1 \wedge \ldots \wedge \pi_n = c \det$. Applying both sides to $(\mathrm{e}_1, \ldots, \mathrm{e}_n)$ and taking into account that $\pi_i \mathrm{e}_j = \delta_{ij}$, we conclude that $c = 1$. $\rhd$

Corollary 8.2.9. *For* $k > n$

$$\Lambda^k(\mathbb{R}^n) = \{0\}.$$

$\lhd$ This follows from the PROOF of Theorem 8.2.6. $\rhd$

Theorem on determinat

Theorem 8.2.10. *Let* $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^n)$*. Then*

$$\boxed{A^* \det = (\det A) \det}$$

Here $\det A$ denotes the determinant of the matrix of A in the canonical basis in $\mathbb{R}^n$.

NB Let X be an arbitrary n-dimensional vector space, and let A be a linear operator in X, $A \in \mathrm{L}(X, X)$. Then the determinant of the matrix of this operator in basis in X **does not depend** on the choice of the basis. $\lhd$ The matrix in a "new" basis has the form BMB^{-1}, where M is the matrix in the "old" basis, and B is the "transition matrix". But $\det(BMB^{-1}) = \det B \det M (\det B)^{-1} = \det M$. $\rhd$ So we can say about the *determinant of an operator* (in a finite-dimensional vector spaces).

$\lhd$ By Corollary 8.2.7., $A^* \det = c \det$. Hence

$$\underbrace{(A^* \det)\mathrm{e}_1 \dots \mathrm{e}_n}_{[1]} = c \underbrace{\det \mathrm{e}_1 \dots \mathrm{e}_n}_{=1} = c$$

$$\overset{[1],\text{ def of } A^*}{=} \det(A\mathrm{e}_1) \dots (A\mathrm{e}_n) \overset{\substack{\text{def. of the matrix}\\ \text{of an operator}}}{=} \det A$$

We conclude that $c = \det A$. $\triangleright$

Theorem 8.2.10. means that a linear operator A changes the volume by $\det A$ times. (It follows also from Theorem on Change!)

Corollary 8.2.11. *If* $\det A = 1$, *then* $A^* \det = \det$.

In other words, an operator with the unit determinant does not change the volume.

Examples.

1. For any $u_1, \dots, u_n \in \Lambda^1(\mathbb{R}^n)$ $(= \mathrm{L}(\mathbb{R}^n))$

$$u_1 \wedge \dots \wedge u_n = \begin{vmatrix} u_{11} & \dots & u_{1n} \\ \dots & \dots & \dots \\ u_{n1} & \dots & u_{nn} \end{vmatrix} \det,$$

where u_{ij} are the coefficients of the linear function $u_i : \mathbb{R}^n \to \mathbb{R}$,

$$u_i x = u_{i1}x_1 + \dots + u_{in}x_n.$$

2. For any $u_1, \dots, u_k \in \Lambda^1(\mathbb{R}^n)$, $\quad k \leq n$

$$(u_1 \wedge \dots \wedge u_k)h_1 \dots h_k = \begin{vmatrix} u_1 h_1 & \dots & u_1 h_k \\ \dots & \dots & \dots \\ u_n h_1 & \dots & u_n h_n \end{vmatrix}.$$

In particular

$$(\pi_{i_1} \wedge \dots \wedge \pi_{i_k})h_1 \dots h_k = \begin{vmatrix} h_{1i_1} & \dots & h_{1i_k} \\ \dots & \dots & \dots \\ h_{ki_1} & \dots & h_{ki_k} \end{vmatrix}.$$

(Note that the determinant of the transposed matrix is the same.)

8.3 Differential forms

Let X be a normed space, and let U be an open set in X. A *differential form* ω of *degree k* ($k = 0, 1, 2, \dots$) (or *k-form*) on U is a smooth (that is, of class C^p for some p) mapping

$$\omega : U \to \Lambda^k(X),$$

that is, a "tensor field" on U, all the tensor being antisymmetric. As to smoothness, we consider ω as a mapping into NORMED SPACE $\Lambda^k(X)$ (a vector subspace in $\mathrm{L}^k(X)$ equipped with the induced norm). We denote the set of all k-forms on U by

$$\Omega^k(U).$$

Examples.

1. Any smooth function $f : U \to \mathbb{R}$ is a 0-form.

2. For any smooth function $f : U \to \mathbb{R}$ its derivative f' is a 1-form.

3. As a special case of the previous example, $\pi_i' \in \Lambda^1(\mathbb{R}^n)$; for any point $x \in \mathbb{R}^n$ we have $\pi_i'(x) = \pi_i$. Thus π_i' is a CONSTANT 1-form on $\mathbb{R}^n$. It is denoted traditionally by

$$\mathrm{d}x_i .$$

So for any $x \in \mathbb{R}^n$ and any $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$

$$(\mathrm{d}x_i)(x) \cdot h = h_i .$$

4. On $\mathbb{R}^n$, a CONSTANT mapping $\omega(x) \equiv \det_k$ is k-form. We denote it also by $\det_k$, or simply det.

The operations $\wedge$ and $*$ for forms are "point-wise".

Exterior product

Let $\omega_1 \in \Omega^{p_1}(U)$, $\omega_2 \in \Omega^{p_2}(U)$. The *exterior product* $\omega_1 \wedge \omega_2$ is defined by the rule

$$\forall x \in U \vdots (\omega_1 \wedge \omega_2)(x) := (\omega_1(x)) \wedge (\omega_2(x)).$$

It is easy to verify that $\omega_1 \wedge \omega_2$ is a SMOOTH mapping $U \to \Lambda^{p_1+p_2}(X)$, so

$$\omega_1 \wedge \omega_1 \in \Omega^{p_1+p_2}(U).$$

Moreover we put for $f \in \Omega^0(U)$

$$f \wedge \omega := f\omega,$$

where

$$\forall x \in U (f\omega)(x) := \underbrace{(f(x))}_{\in\mathbb{R}}(\omega(x)).$$

Example. On $\mathbb{R}^n$,

$$\mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_n = \det .$$

$\lhd$ $\pi_i \wedge \ldots \wedge \pi_n = \det .$ $\rhd$

Theorem 8.3.1. *Any k-form ω on $\mathbb{R}^n$ can be written in the form*

$$\omega = \sum_{i_1<\ldots<i_k} f_{i_1\ldots i_k} \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_k},$$

where $f_{i_1\ldots i_k}$ are smooth (real-valued) functions.

$\lhd$ It follows at once from Theorem on basis on $\Lambda^k(\mathbb{R}^n)$. $\rhd$

Example. For $f \in \Omega^0$

$$f' = \frac{\partial f}{\partial x_1}\mathrm{d}x_1 + \ldots \frac{\partial f}{\partial x_n}\mathrm{d}x_n.$$

$\lhd$ Apply both sides to $h = (h_1, \ldots, h_n)$. $\rhd$

Pull-back

Let X, Y be normed spaces, let f be a smooth mapping from X into Y, and let $\omega \in \Omega^k(Y)$.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ \Omega^k(X) & \xleftarrow{f^*} & \Omega^k(Y) \end{array}$$

We define the *pull-back* $f^*\omega$ of ω by the rule

$$\forall x \in X \vdots (f^*\omega)(x) := (f'(x))^*\omega(f(x)),$$

where the star in the right-hand side means pull-back for tensors.

Thus, the value of the pull-back of ω by f at x is the tensor pull-back by the DERIVATIVE $f'(x)$ of the value of ω at $f(x)$.

More explicitly,

$$((f^*\omega)(x))h_1 \dots h_k := \omega(f(x))(f'(x)h_1)\dots(f'(x)h_n).$$

Moreover we put for $g \in \Omega^0(Y)$

$$f^*g = g \circ f.$$

Example. For a smooth mapping $f : \mathbb{R}^n \to \mathbb{R}^m$, $f = (f_1, \dots, f_m)$, it holds

$$f^*(\mathrm{d}y_i) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}\mathrm{d}x_j \qquad (i = 1, \dots, m)$$

$((x = x_1, \dots, x_n) \in \mathbb{R}^n, y = (y_1, \dots, y_m) \in \mathbb{R}^m)$.

$$\begin{aligned} \lhd\ f^*(\mathrm{d}y_i)(x)\cdot h &= (\mathrm{d}y_i)\cdot(f'(x)h) \\ &= \mathrm{d}y_i\left(\begin{pmatrix} \partial f_1/\partial x_1 & \dots & \partial f_1/\partial x_n \\ \dots & \dots & \dots \\ \partial f_m/\partial x_1 & \dots & \partial f_m/\partial x_n \end{pmatrix}\begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}\right) \\ &= \sum_{j=1}^{n}\frac{\partial f_i}{\partial x_j}\underbrace{h_j}_{=(\mathrm{d}x_j)h} = \left(\sum_{j=1}^{n}\frac{\partial f_i}{\partial x_j}\mathrm{d}x_j\right)h.\ \rhd \end{aligned}$$

Theorem 8.3.2. *The pull-back operation over forms has the following properties*

a) $f^*(\omega_1 + \omega_2) = f^*\omega_1 + f^*\omega_2$ *(linearity);*
b) $f^*(\omega_1 \wedge \omega_1) = (f^*\omega_1) \wedge (f^*\omega_2)$ *(* respects $\wedge$); in particular, for $g \in \Omega^0$*

$$f^*(g\omega) = (g \circ f)(f^*\omega).$$

$\lhd$ This follows at once from the definitions and the corresponding results for tensors. $\rhd$

Pull-back of determinant

Theorem 8.3.3. *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ (that is $f : \mathbb{R}^n \to \mathbb{R}^n$) be smooth, and let $g \in \Omega^0(\mathbb{R}^n)$. Then*

$\mathbb{R}^n \xrightarrow{f} \mathbb{R}^n \xrightarrow{g} \mathbb{R}^n$.

$$\boxed{f^*(g\det) = (g \circ f)(\det f')\det}$$

$\lhd$ 1° $f^*(\det) = (\det f')\det$.

$$\lhd\!\lhd\ \forall x \in \mathbb{R}^n \vdots (f^*\det)(x) = (f'(x))^*\underbrace{\det(f(x))}_{=\det} \overset{\text{Th. on det for tensors}}{=} (\det f'(x))\det\ \rhd\!\rhd.$$

2° $f^*(g\det) \overset{\text{Th 8.3.2., b)}}{=} (g \circ f)f^*(\det) \overset{1^\circ}{=} (g \circ f)(\det f')\det$. $\rhd$

NB In the special case where f is a diffeomorphism, this Theorem describes the change of a “weighted” volume by a change of variables—compare with Theorem on change of variables (where we write f instead of g and v.v.).

8.4 Exterion differentiation (operator d)

If we differentiate a form $\omega \in \Omega^k(X)$ (as a mapping $X \to \Lambda^k(X)$ between two normed spaces) then we obtain $\omega'(x) \in \mathrm{L}(X, \Lambda^k(X)) \subset \mathrm{L}(X, \mathrm{L}^k(X)) \sim \mathrm{L}^{k+1}(X)$. In general $\omega'(x)$ considered as an element of $\mathrm{L}^{k+1}(X)$ is **not** antisymmetric, that is, does **not** belong to $\Lambda^{k+1}(X)$. So we appeal the operator alt.

Operator d

Let X be a normed space, and let $\omega \in \Omega^k(X)$ we define the *exterior derivative* (or *exterior differential*) $\mathrm{d}\omega$ by the rule

$$\forall x \in X \vdots (\mathrm{d}\omega)(x) := (k+1)\,\mathrm{alt}\,\widetilde{\omega'(x)} \qquad (\in \Lambda^{k+1}(X)), \tag{1}$$

where $\widetilde{\omega'(x)}$ denotes the element of $\mathrm{L}^{k+1}(X)$ generated by $\omega'(x)$:

$$\widetilde{\omega'(x)} \cdot h_0 h_1 \dots h_k := \underbrace{(\omega'(x)h_0)}_{\in \Lambda^k(X)} \cdot h_1 \dots h_k. \tag{2}$$

Is is easy to verify that if $\omega \in \mathrm{C}^p$ then $\mathrm{d}\omega \in \mathrm{C}^{p-1}$, so is sufficiently smooth if ω is. Thus

$$\mathrm{d}\omega \in \Omega^{k+1}(X).$$

Remark. The factor $k+1$ is chosen to obtain the coefficient 1 in some formulas below.

Examples.

1. For $f \in \Omega^0$ we have $\mathrm{d}f = f'$.
2. For any $f \in \Omega^0$

$$\mathrm{d}^2 f = 0,$$

where

$$\mathrm{d}^2 f = \mathrm{d}(\mathrm{d}f).$$

$\lhd$ $(\mathrm{d}^2 f)(x) = (\mathrm{d}(\mathrm{d}f))(x) \overset{1)}{=} (\mathrm{d}(f'))(x) \overset{(1)}{=} (1+1)\,\mathrm{alt}\,\underbrace{\widetilde{(f')(x)}}_{=f''(x)} = 2\,\mathrm{alt}\,\underbrace{f''(x)}_{f''(x)\in\mathrm{Sym}} = 0.$ $\rhd$

3. In $\mathbb{R}^2$, $\mathrm{d}(\underbrace{x\mathrm{d}y}_{=:\omega}) = \det$. $\lhd$ We have $\omega : (x, y) \mapsto x\pi_2$, so that $\omega \in \mathrm{L}(\mathbb{R}^2, \mathrm{L}(\mathbb{R}^2, \mathbb{R}))$, hence $\omega' \equiv \omega$. So

$$(\mathrm{d}\omega)((x,y))\underbrace{h}_{=:(h_1,h_2)}\underbrace{k}_{=:(k_1,k_2)} = (1+1)\,\mathrm{alt}\,\underbrace{\widetilde{\omega'((x,y))}}_{=\omega} hk = 2\tfrac{1}{2}(\widetilde{\omega}hk - \widetilde{\omega}kh)$$

$$\overset{(2)}{=} \underbrace{(\omega h)}_{h_1\pi_2} k - \underbrace{(\omega k)}_{k_1\pi_2} h = h_1k_2 - kh_2 = \det(h,k). \ \rhd$$

Exercise 8.4.1. Let $A \in \mathrm{L}(\mathbb{R}^n, \mathbb{R}^n)$ and $\omega(x) = \langle An, \cdot\rangle$ $(\in \mathrm{L}(\mathbb{R}^n, \mathbb{R}) = \Lambda^1(\mathbb{R}^n))$. Prove that

$$\mathrm{d}\omega \equiv u_{A-A^T}.$$

Remark. It can be shown that

$$\boxed{\mathrm{d}\omega(x) \cdot h_0 \dots h_k = \sum_{i=0}^{k} (-1)^i (\omega'(x)h_i) \cdot h_0 \dots \overline{h_i} \dots h_k,}$$

where $\overline{h_i}$ means that this term is to be canceled.

Theorem 8.4.2. *The operator* d *has the following properties*

a) $\mathrm{d}(\omega_1+\omega_2)=\mathrm{d}\omega_1+\mathrm{d}\omega_2$ *(linearity);*
b) $\mathrm{d}(\omega_1\wedge\omega_2)=(\mathrm{d}\omega_1)\wedge\omega_2+(-1)^{\deg\omega_1}\omega_1\wedge\mathrm{d}\omega_2$ *("semi-leibniz" rule);*
c) $\mathrm{d}^2=0$ $(\mathrm{d}^2\omega=\mathrm{d})\mathrm{d}\omega)$ *("self-annihilation");*
d) $f^*(\mathrm{d}\omega)=\mathrm{d}(f^*\omega)$ *(pull-back respects* d*).*

Here $\deg\omega$ denotes the degree of ω.

◁ a) Obvious.
b) Let $\omega_1\in\Omega^{r_1}$, $\omega_2\in\Omega^{r_2}$. We have (for short we drop the argument x)

$$\begin{aligned}
&(\widetilde{\omega_1\otimes\omega_2})'h_0\dots h_{r_1+r_2}\overset{\text{Leibnitz rule}}{=}((\omega_1'h_0)\otimes\omega_2+\omega_1\otimes(\omega_2'h_0))h_1\dots h_{r_1+r_2}\\
&=((\omega_1'h_0)h_1\dots h_{r_1})(\omega_2h_{r_1+1}\dots h_{r_1+r_2})\\
&\quad+(\omega_1h_1\dots h_{r_1})((\omega_2'h_0)h_{r_1+1}\dots h_{r_1+r_2})\\
&=(\widetilde{\omega_1'}h_0\dots h_{r_1})(\omega_2h_{r_1+1}\dots h_{r_1-r_2})+(\omega_1h_1\dots h_{r_1})(\widetilde{\omega_2'}h_0h_{r_1+1}\dots h_{r_1+r_2})\\
&=(\widetilde{\omega_1'}\otimes\omega_2)h_0\dots h_{r_1+r_2}+(\omega_2\otimes\widetilde{\omega_1'})h_1\dots h_{r_1}h_0h_{r_1+1}\dots h_{r_1+r_2}.
\end{aligned}$$

To replace h_0 to the first place we have to make r_1 transpositions, which yields the factor $(-1)^{r_1}$. After alternating (applying of alt) we obtain what we need. We omitting the details.
c) The idea is the same as in Example 2) above. Let $\omega\in\Omega^r$. We have (omitting the details)

$$\begin{aligned}
(\mathrm{d}^2\omega)h_0\dots h_{r+1}&=\mathrm{const}(\mathrm{alt}\,(\widetilde{\mathrm{alt}\,\widetilde{\omega}'})')h_0\dots h_{r+1}\\
&=\mathrm{const}\sum_{\{h,k\}\subset\{h_0\dots h_{r+1}\}}\sum_{\cdots}((\omega''hk)\underbrace{\dots}_{*}-(\omega''kh)\underbrace{\dots}_{*})\\
&\quad *\text{ — other } r \text{ arguments in one and the same order}\\
&=\mathrm{const}\sum_{\cdots}(\underbrace{(\omega''hk-\omega''kh)}_{\omega''\in\mathrm{sym}\atop =0}\dots)=0
\end{aligned}$$

d) We consider the simplest case where $\omega\in\Omega^0$, that is, ω is a function g:

$$X\xrightarrow{f}Y\xrightarrow{g}\mathbb{R}.$$

We have

$$\begin{aligned}
((\mathrm{d}(f^*g))(x))h=(f^*g)(x)h=(g\circ f)'(x)h&\overset{\text{Chain rule}}{=}(\underbrace{g'}_{=\mathrm{d}g}(f(x)))\cdot(f'(x)h)\\
=((f^*(\mathrm{d}g))(x))h,&
\end{aligned}$$

whence $\mathrm{d}(f^*g)=f^*(\mathrm{d}g)$. In more general case the idea is the same. ▷

Remark. The semi-Leibniz rule is NON-symmetric w.r. to ω_1 and ω_2. Only $\deg\omega_1$ enters the rule. The matter is that in our definition of $\widetilde{\omega'(x)}$ we put the derivative to act onto the FIRST argument. But we could choose anyone. So in essence $\mathrm{d}\omega$ is defined just UP TO THE SIGN! "Physically", ω and $-\omega$ are same, that is why when integrating over manifolds (Chapter 6) we need to choose one of two possible ORIENTATIONS of the manifold in question.

Remark. It is instructive to see how semi-Leibniz rule interacts with semi-commutativity:

$$\begin{aligned}
\mathrm{d}(\omega_2 \wedge \omega_1) &= \mathrm{d}((-1)^{r_1 r_2}\omega_1 \wedge \omega_2) = (-1)^{r_1 r_2}(\mathrm{d}\omega_1\omega_2 + (-1)^{r_1}\omega_1\mathrm{d}\omega_2) \\
&= (-1)^{r_1 r_2}((-1)^{(r_1+1)r_2}\omega_2 \wedge \mathrm{d}\omega_1 + (-1)^{r_1}(-1)^{r_1(r_2+1)}\mathrm{d}\omega_2 \wedge \omega_1) \\
&= \mathrm{d}\omega_2 \wedge \omega_1 + (-1)^{r_2}\omega_2\mathrm{d}\omega_1,
\end{aligned}$$

r_1 “transforms” into r_2, as it need to be!
Exercise 8.4.3. Give the proof of d) with all details for the case $\omega \in \mathbf{\Omega}^1$.

d **in** $\mathbb{R}^n$

Theorem 8.4.4. *Let $\omega \in \Omega^r(\mathbb{R}^n)$. If the canonical representation of ω is*

$$\omega = \sum_{i_1<\ldots<i_r} f_{i_1\ldots i_r}\mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_r}$$

then $\mathrm{d}\omega$ *is given by the formula*

$$\mathrm{d}\omega = \sum_{i_1<\ldots<i_r} \mathrm{d}f_{i_1\ldots i_r} \wedge \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_r}$$

$\lhd$ This follows at once from semi-Leibnitz rule for forms and from the self-annihilation property (d^2); recall that $f\mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_r} = f \wedge \mathrm{d}x_{i_1} \wedge \ldots \wedge \mathrm{d}x_{i_r}$. $\rhd$

Example. In $\mathbb{R}^2$, $\quad \mathrm{d}(x\mathrm{d}y) = \mathrm{d}x \wedge \mathrm{d}y = \det$.

Closed and exact forms

If $\mathrm{d}\omega = 0$ (in an open set U) one can says that ω is *closed* (in U). If $\omega = \mathrm{d}\psi$ (in U) for some form ψ then one can says that ω is *exact* (in U).

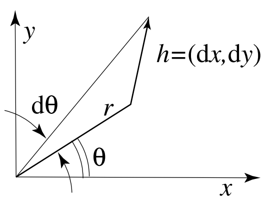
Each exact form is closed. $\lhd$ $\omega = \mathrm{d}\psi \Rightarrow \mathrm{d}\omega = \mathrm{d}^2\psi = 0$. $\rhd$

The inverse assertion is **not** in general true.

Example. The form

$$\omega = \frac{-y}{x^2+y^2}\mathrm{d}x + \frac{x}{x^2+y^2}\mathrm{d}y$$

on $\mathbb{R}^2 \setminus \{0\}$ is closed (verify!), but is not exact. (This form arises naturally if one consider



polar coordinates r, θ ($x = r\cos\theta$, $y = r\sin\theta$), and by this reason is usually denoted by $\mathrm{d}\theta$, and though ω is not exact! (But ω IS exact on somewhat less sets,as it follows, e.g., from Poicaré lemma:))

Poincaré lemma

Theorem 8.4.5. *Let U be an open ball in Banach space. If a fomr ω is closed in U, then it is exact in U.*

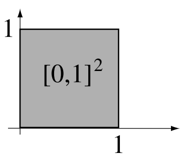
We omit the proof.

Chapter 9

Stokes Formula for Chains

9.1 Chains

By $[0,1]^k$ we denote the unit cube $[0,1] \times \ldots \times [0,1]$ (k-times) in $\mathbb{R}^k$. For $k = 0$ we put



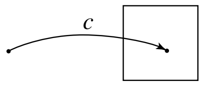
$$[0,1]^0 := \{0\}, \qquad \mathbb{R}^0 := \{0\}.$$

By a *curved k-cube c* in $\mathbb{R}^n$ we mean a continuous mapping

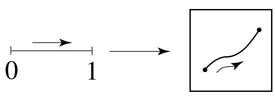
$$c : [0,1]^k \to \mathbb{R}^n.$$

For short we omit the word "curved" below. If $\operatorname{im} c \subset A \subset \mathbb{R}^n$ we say that c is a *cube in A*.

Examples.
1. A 0-cube is just a point:



2. A 1-cube is a curve:



3. The embeding $\mathrm{id} : [0,1]^k \to \mathbb{R}^k$ is called the *standard* k-cube and is denoted by I^k:

$$I^k := \mathrm{id}_{\mathbb{R}^k} |_{[0,1]^k}.$$

Chains

A *k-chain* (in A) is a formal (finite) sum of k-cubes (in A) with integer coefficients, that is, an expression of the form

$$2c_1 + 3c_2 - 5c_3 + 100c_4,$$

where c_i are k-cubes. We identify a k-cube c with multiplication with the chain $1 \cdot c$.

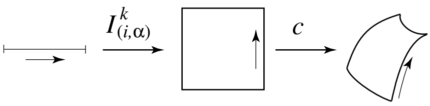
In natural way we define for k-chains multiplication by an integer number and addition.

Faces

For a standard cube I^n we define its *faces* $I^k_{(i,\alpha)}$, $\quad i = 1, \ldots, k, \quad \alpha = 0, 1$, by the rule

$$I^k_{(i,\alpha)} : [0,1]^{k-1} \to \mathbb{R}^k, (x_1, \ldots, x_{i-1}, \alpha, x_i, \ldots, x_{k-1}).$$

Thus, we insert α into i^{th} place and move all the "tail" to the right by one position.
For any k-cube c we define its face $c_{(i,\alpha)}$ by the rule



$$\boxed{c_{(i,\alpha)} := c \circ I^k_{(i,\alpha)}.}$$

Operator ∂

We define the *boundary* ∂c of a k-cube c as the following chain:

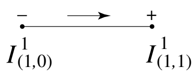
$$\boxed{\partial c := \sum_{i=1}^{k} \sum_{\alpha=0,1} (-1)^{i+\alpha} c_{(i,\alpha)}}$$

(an alternating sum of the faces). For chains we define the boundary "by linearity":

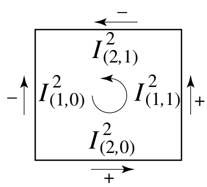
$$\boxed{\partial \sum a_i c_i := \sum a_i \partial c_i} \qquad (a_i \in \mathbb{Z}).$$

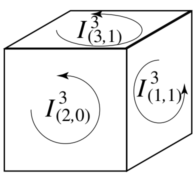
Examples.

1. $\partial I^1 = I^1_{(1,1)} - I^1_{(1,0)}$;



2. $\partial I^2 = I^2_{(2,0)} + i^2_{(1,1)} - I^2_{(2,1)} + I^2(1,0)$;





3. ∂I^3 three "visible" faces are positive (enter into sum with "$+$") three "non-visible" faces are negative.

Theorem 9.1.1. $\partial^2 = 0$ *(that is,* $\partial(\partial c) = 0$ *for any chain* c*).*

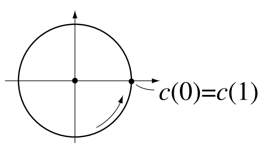
We do not need this result, so we omit the proof.

Exercise 9.1.2. Verify Theorem for I^3.

Closed and exact chains

A chain c is called *closed* if $\partial c = 0$, that is, if its boundary is the null chain; c is called *exact* (in A) if $c = \partial c'$ for some chain (in A), that is, if c is the boundary of some chain (in A).

By the Theorem above, each exact chain is closed. But not each closed is exact:



Example.
Consider the 1-cube $c : [0,1] \to \mathbb{R}^2$, $t \mapsto (\cos 2\pi t, \sin 2\pi t)$ (the unit circle in $\mathbb{R}^2 \setminus \{0\}$. We have $\partial c = 0$ (verify), so that c is closed, but c is not exact in $\mathbb{R}^2 \setminus \{0\}$, since c is the boundary of no chain in $\mathbb{R}^2 \setminus \{0\}$.

9.2 Integral over a chain

For any k-form ω in an open set U in $\mathbb{R}^k$, such that $U \supset [0,1]^k$, $k = 1, 2, \ldots$, we put

$$\boxed{\int_{I^k} \omega := \int_{[0,1]^k} f} \tag{1}$$

where f is the function $U \to \mathbb{R}$, uniquely defined by the relation $\omega = f\,\det$. (Recall that any k-form in $\mathbb{R}^k$ can be written (uniquelly) in such form; see Chapter 8.)

In more detailed record,

$$\int_{I^k} f\,\mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_n := \int_{[0,1]^k} f\,\mathrm{d}x_1 \ldots \mathrm{d}x_n. \tag{2}$$

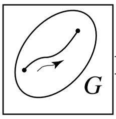
NB This definition DOES DEPEND on the ORDER of the basis vectors, since det does depend!

For $k = 0$ we put

$$\int_{i^0} \omega := \omega(0), \qquad \text{(here } \omega \in \Omega^0 \text{ is a FUNCTION).}$$

Now, let G be an open set in $\mathbb{R}^n$, let ω be a k-form in G, and let c be a k-cube in G. Then we put

$$\begin{array}{ccc} [0,1]^k & \xrightarrow{\;c\;} & G\ \mathbb{R}^n \\ c^*\omega \downarrow & & \downarrow \omega \\ \Lambda^k(\mathbb{R}^k) & & \Lambda^k(\mathbb{R}^n) \end{array}$$



$$\boxed{\int_c \omega := \int_{I^k} c^*\omega} \tag{3}$$

In particular, for a 0-cube c and a 0-form ω we have

$$\int_c \omega = \omega(c(0)) \qquad \text{(since } c^*\omega = \omega \circ c, \text{if } \omega \in \Omega^0\text{).}$$

At last for a chain $c = \sum a_i c_i$ we put

$$\boxed{\int_{\sum a_i c_i} \omega := \sum a_i \int_{c_i} \omega} \tag{4}$$

It is clear that so defined integral is LINEAR:

$$\int_c \alpha\omega = \alpha \int_c \omega, \qquad \int_c (\omega_1 + \omega_2) = \int_c \omega_1 + \int_c \omega_2 \qquad (\alpha \in \mathbb{R}).$$

Lemma 9.2.1. (on integral over the boundary). *Let c be a k-cube, and let ω be a k-form (both in G). Then*

$$\int_{\partial c} \omega = \int_{\partial [0,1]^k} c^*\omega.$$

◁ Exercise. ▷

9.3 Stokes formula

Similarly of properties of the operators d and ∂ ($\mathrm{d}^2 = 0$, $\partial^2 = 0$), and of the notions concerning them (closeness, exactness) is not an accident. There is a deep relation between d and ∂, which is expressed by the following

Theorem 9.3.1. *Let ω be a $k-1$-form in $G(\in \mathrm{Op}(\mathbb{R}^n))$, and let c be a k-chain in G. Then*

$$\boxed{\boxed{\int_c \mathrm{d}\omega = \int_{\partial c} \omega}} \qquad \textit{(Stokes formula)} \tag{1}$$

$\lhd$ 1° At first let us prove the folloving *fact about pull-back by the standard faces*:

$$\left(I^k_{(i,\alpha)}\right)^* \mathrm{d}x_j = \begin{cases} \mathrm{d}x_j & \text{if } j < i, \\ 0 & \text{if } j = i, \\ \mathrm{d}x_{j-1} & \text{if } j > i. \end{cases} \tag{2}$$

$\lhd\!\lhd$ $\left(I^k_{(i,\alpha)}\right)'(x)\underbrace{h}_{=(h_1,\ldots,h_{k-1})} \overset{\text{Exer.}}{=} (h_1,\ldots,h_{i-1},0,h_i,\ldots,h_{k-1})$ whence

$$\begin{aligned}((I^k_{(c,\alpha)})^*\mathrm{d}x_j) &= \mathrm{d}x_j(I^k_{(i,\alpha)})'(x)h = \mathrm{d}x_j \cdot (h_1,\ldots,h_{i-1},0,h_i,\ldots,h_{k-1}) \\ &= \begin{cases} h_j & \text{if } j < i, \\ 0 & \text{if } j = i, \\ h_{j-1} & \text{if } j > i. \end{cases} \quad \rhd\!\rhd\end{aligned}$$

2° *Case $c = I^k$, $\omega \in \Omega^{k-1}(\mathbb{R}^k)$*. Calculate the left-hand side in (1):

$$\begin{aligned}\int_{I^k} \mathrm{d}\omega &\overset{\text{Th 8.3.1.}}{=} \int_{I^k} \sum_{i=1}^k \mathrm{d} f_i \mathrm{d}x_1 \ldots \wedge \overline{\mathrm{d}x_i} \wedge \ldots \wedge \mathrm{d}x_k \\ &\overset{\text{Th 8.4.4.}}{=} \int_{I^k} \sum_{i=1}^k \underbrace{\mathrm{d} f_i}_{=\mathrm{D}_1 f_i \mathrm{d}x_1+\ldots+\mathrm{D}_k f_i \mathrm{d}x_k} \wedge \mathrm{d}x_1 \ldots \wedge \overline{\mathrm{d}x_i} \wedge \ldots \wedge \mathrm{d}x_k \\ &\overset{\text{Th 8.2.4. and its Cor.}}{=} \sum_{i=1}^k (-1)^{i-1} \int_{I^k} \mathrm{D}_i f_i \underbrace{\mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_k}_{=\det} \overset{(1)}{=} \sum_{i=1}^k (-1)^{i-1} \int_{[0,1]^k} \mathrm{D}_i f_i \\ &\overset{\text{Fubini Th}}{=} \sum_{i=1}^k (-1)^{i-1} \underbrace{\int_0^1 \ldots \int_0^1}_{k-1} \Bigg(\int_0^1 \underbrace{\mathrm{D}_i f_i(x_1,\ldots,x_k)\mathrm{d}x_i}_{\overset{\text{Newt.-Leib. Th}}{=} f_i(x_1,\ldots,\underset{i}{1},\ldots,x_k) - f_i(x_1,\ldots,\underset{i}{0},\ldots,x_k) =: \boxed{1}} \Bigg) \mathrm{d}x_1 \ldots \overline{\mathrm{d}x_i} \ldots \mathrm{d}x_k \\ &\overset{\text{Fubini+trick!}}{=} \sum_{i=1}^k (-1)^{i-1} \underbrace{\int_0^1 \ldots \int_0^1}_{k} \boxed{1} \mathrm{d}x_1 \ldots \mathrm{d}x_k.\end{aligned}$$

Calculate the right-hand side:

$$\begin{aligned}\int_{\partial I^k} \omega &\overset{\text{def of }\partial}{=} \int_{\sum_{i=1}^k \sum_{\alpha=0,1} (-1)^{i+\alpha} I^k_{(i,\alpha)}} \sum_{j=1}^k f_j \mathrm{d}x_1 \wedge \ldots \wedge \overline{\mathrm{d}x_j} \wedge \ldots \mathrm{d}x_k \\ &\overset{(4)}{=} \sum_{i,j=1}^k \sum_{\alpha=0,1} (-1)^{i+\alpha} \int_{I^k_{(i,\alpha)}} f_j \mathrm{d}x_1 \wedge \ldots \wedge \overline{\mathrm{d}x_j} \wedge \ldots \mathrm{d}x_k \\ &\overset{(3)}{=} \sum_{i,j,\alpha} (-1)^{i+\alpha} \int_{I^{k-1}} \left(I^k_{(i,\alpha)}\right)^* \left(f_j \mathrm{d}x_1 \wedge \ldots \wedge \overline{\mathrm{d}x_j} \wedge \ldots \mathrm{d}x_k\right) \\ &\overset{*\text{respects }\wedge}{=} \sum_{i,j,\alpha} (-1)^{i+\alpha} \int_{I^{k-1}} \Bigg(\underbrace{\left(I^k_{(i,\alpha)}\right)^* f_j}_{=f_j \circ I^k_{(i,\alpha)}} \Bigg) \\ &\wedge \underbrace{(((I^k_{(i,\alpha)})^*\mathrm{d}x_1) \wedge \ldots \wedge ((I^k_{(i,\alpha)})^*\mathrm{d}x_{j-1}) \wedge ((I^k_{(i,\alpha)})^*\mathrm{d}x_{j+1}) \wedge \ldots \wedge ((I^k_{(i,\alpha)})^*\mathrm{d}x_k))}_{\overset{(2)}{=} \begin{cases} 0 & \text{if } i \neq j \\ \mathrm{d}x_1 \wedge \ldots \wedge \mathrm{d}x_{k-1} & \text{if } i = j \end{cases}}\end{aligned}$$

$$= \sum_{i,\alpha}(-1)^{i+\alpha}\int_{I^{k-1}}\left(f_i\circ I^k_{(i,\alpha)}\right)\underbrace{\mathrm{d}x_1\wedge\ldots\wedge\mathrm{d}x_{k-1}}_{=\det}$$

$$\overset{(1)}{=}\sum_{i,\alpha}(-1)^{i+\alpha}\int_{[0,1]^{k-1}} f_i\circ I^k_{(i,\alpha)}$$

$$\overset{\text{def of } I^k_{(i,\alpha)}\text{+trick}}{=}\sum_{i,\alpha}(-1)^{i+\alpha}\left(\underbrace{\int_0^1\ldots\int_0^1}_{k-1} f_i(x_1,\ldots,\underset{i}{\alpha},\ldots,x_{k-1})\right)\underbrace{\left(\int_0^1\mathrm{d}x_k\right)}_{=1}$$

$$\overset{\text{Fubini Th}}{=}\sum_{i,\alpha}(-1)^{i+\alpha}\underbrace{\int_0^1\ldots\int_0^1}_{k} f_i(x_1,\ldots,\underset{i}{\alpha},\ldots,x_{k-1})\mathrm{d}x_1\ldots\mathrm{d}x_k$$

$$\boxed{y_1:=x_1,\ \ldots\ ,y_{i-1}:=x_{i-1},\ y_i:=x_k,\ y_{i+1}:=x_i,\ \ldots\ ,y_k:=x_{k-1}}$$

$$=\sum_{i,\alpha}(-1)^{i+\alpha}\int_0^1\ldots\int_0^1 f_i(y_1,\ldots,\underset{i}{\alpha},\ldots,y_{k-1})\mathrm{d}y_1\ldots\mathrm{d}y_k$$

$$\overset{\text{obv.}}{=}\sum_{i,\alpha}(-1)^{i+\alpha}\int_0^1\ldots\int_0^1(f_i(y_1,\ldots,1,\ldots,y_{k-1}) - f_i(y_1,\ldots,0,\ldots,y_{k-1}))\mathrm{d}y_1\ldots\mathrm{d}y_k.$$

The result is the same.

3° *Case $c\in k$-Cube(G),* $\omega\in\Omega^k(G)$.

$$\int_c\mathrm{d}\omega\overset{\text{def}}{=}\int_{I^k}c^*(\mathrm{d}\omega)\overset{*\text{ respects d}}{=}\int_{I^k}\mathrm{d}(c^*\omega)\overset{2^\circ}{=}\int_{\partial I^k}c^*\omega\overset{\text{Lm 9.2.1.}}{=}\int_{\partial c}\omega.\quad\text{O.K.}$$

4° General case $c=\sum\alpha_i c_i$, $c_i\in k$-Cube(G), $\omega\in\Omega^k(G)$.

$$\int_c\mathrm{d}\omega\overset{(4)}{=}\sum a_i\int_{c_i}\mathrm{d}\omega\overset{3^\circ}{=}\sum\alpha_i\int_{\partial c_i}\omega\overset{(4)}{=}\int_{\underbrace{\sum a_i\partial c_i}_{\text{def of}\partial\to=\partial c}}\omega=\int_{\partial c}\omega.\quad\text{O.K.}\ \triangleright$$

Chapter 10

Stokes Theorem for Manifolds

10.1 Manifolds in $\mathbb{R}^n$

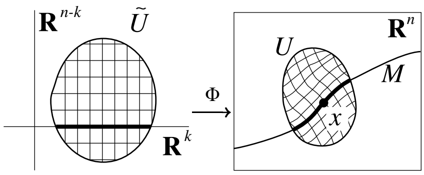
In this chapter by *smooth* mappings we mean $\mathbb{C}^\infty$-mappings.

We say that a subset M of $\mathbb{R}^n$ is a *k-dimensional manifold* (or simply *k-manifold*), and we write

$$M \in \mathrm{Mf}^k(\mathbb{R}^n),$$

if for each point $x \in M$ the following condition is fulfilled:

$$\begin{aligned} \exists U \in \mathrm{Op}\,\mathrm{Nb}_x(\mathbb{R}^n)\ \exists \tilde{U} \in \mathrm{Op}(\mathbb{R}^n)\ \ \exists \Phi \in \mathrm{Diffeo}(\tilde{U}, U) :& \\ U \cap M = \Phi(\tilde{U} \cap \mathbb{R}^n \times \underbrace{0}_{\in \mathbb{R}^{n-k}}).& \qquad (1) \end{aligned}$$

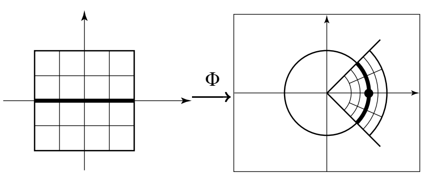


(In other words, M is locally, up to a diffeomorphism, a k-dimensional vector subspace in $\mathbb{R}^n$.) We call such a mapping Φ a *full chart* for M at x.

Examples.

1. Each single point set $\{x\}$ is a 0-dimensional manifold.

2. Each open set in $\mathbb{R}^n$ is an n-dimensional manifold.



3. The unit circle in $\mathbb{R}^2$ with the center at 0 is a 1-dimensional manifold. E.g. for the point $(1, 0)$ a full chart is shown on the picture. (Give an analytic expression for Φ!)

Exercise 10.1.1. Let $G \in \mathrm{Op}(\mathbb{R}^n)$, and let $g : G \to \mathbb{R}^p$ $(p \le n)$ be a smooth mapping. Put

$$M := g^{-1}(0).$$

If

$$\forall x \in M \vdots\ \ \mathrm{rank}\, g'(x) = p$$

then

$$M \in \mathrm{Mf}^{n-p}(\mathbb{R}^n).$$

Here rank denotes the rank of the corresponding matrix. Note that p is the maximal possible value for the rank of $n \times p$-matrix. [*Hint*: Let

$$g'(x) = \begin{pmatrix} \partial g_1/\partial x_1 & \dots & \partial g_1/\partial x_n \\ \dots & \dots & \dots \\ \partial g_p/\partial x_1 & \dots & \partial g_p/\partial x_n \end{pmatrix}$$

Wlog we can assume that it is the FIRST mirror that is not 0:

$$\begin{vmatrix} \partial g_1/\partial x_1 & \dots & \partial g_1/\partial x_p \\ \dots & \dots & \dots \\ \partial g_p/\partial x_1 & \dots & \partial g_p/\partial x_p \end{vmatrix} \neq 0$$

Put

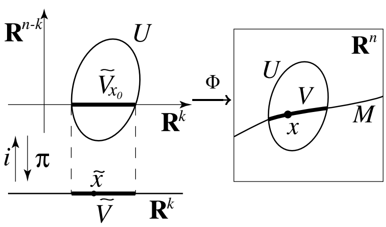
$$\Psi := (\mathrm{id}, g) : \mathbb{R}^n \to \mathbb{R}^{n-p} \times \mathbb{R}^p = \mathbb{R}^n,$$

that is,

$$\Psi(x_1, \dots, x_n) := (x_1, \dots, x_{n-p}, g_1(x), \dots, g_p(x)).$$

Verify that $\Psi'(x) \in \mathrm{Iso}$, and apply Inverse Function Theorem. The inverse mapping $\Phi := \Psi^{-1}$ is a full chart for M.]

Charts



Lemma 10.1.2. *Let $M \in \mathrm{Mf}^k(\mathbb{R}^n)$, and let $\Phi : \widetilde{U} \to U$ be a full chart for M at x. Put*

$$V := U \cap M, \quad \widetilde{V} := \pi(\widetilde{U} \cap (\mathbb{R}^k \times 0)), \quad (2)$$

$$\varphi := \Phi \circ i, \quad \psi := \pi \circ \Phi^{-1}, \quad (3)$$

where π and i are the canonical projection and inclusion, resp. ($\pi(x_1, \dots, x_n) = (x_1, \dots, x_k)$, $i(x_1, \dots, x_k) = (x_1, \dots, x_k, 0, \dots, 0)$). Then $\varphi : \widetilde{V} \to \mathbb{R}^n$ is a smooth map and is a bijection of $\widetilde{V}$ onto V. Moreover

$$\forall \widetilde{x} \in \widetilde{V} \vdots\ \varphi'(\widetilde{x}) \in \mathrm{Inj}\,\mathrm{L}(\mathbb{R}^k, \mathbb{R}^n) \quad (\Leftrightarrow \mathrm{rank}\,\varphi'(\widetilde{x}) = k). \quad (4)$$

We call φ a *chart* for M at x (generated by the full chart Φ), or a *coordinate system* on M at x. The element $\varphi^{-1}(x)$ (for any $x \in V$) is called the *representative* of x in the chart φ, and will be usually denoted by $\widetilde{x}$.
$\triangleleft$ All but the assertion on the rank is obvious. As to this assertion, we have

$$\psi \circ \varphi \overset{(3)}{=} \pi \circ \Phi^{-1} \circ \Phi \circ i = \pi \circ i = \mathrm{id}_{\mathbb{R}^k},$$

hence, by Chain Rule,

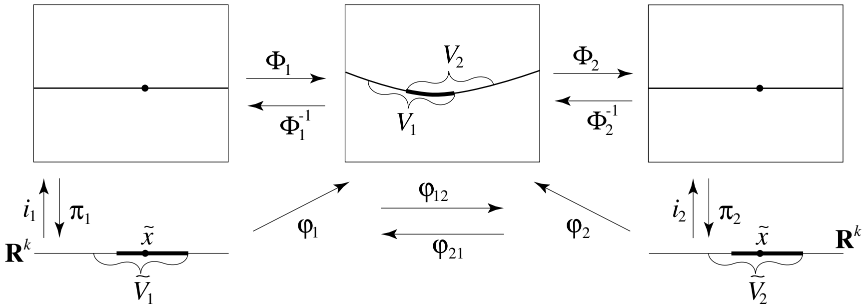
$$\psi'(x) \circ \varphi'(\widetilde{x}) = \mathrm{id}_{\mathbb{R}^k}.$$

This means that $\varphi'(\widetilde{x})$ is injective, that is, has the maximal possible rank, k. $\triangleright$

Example. If $M = \mathbb{R}^n$, then $\mathrm{id}_{\mathbb{R}^n}$ is a chart at all points at once.

Transition functions

Let we have two charts for M at x, φ_1 and φ_2 (see the diagram):



$$\varphi_1 : \widetilde{V}_1 \to V_1, \qquad \varphi_2 : \widetilde{V}_2 \to V_2.$$

Put $V := V_1 \cap V_2$ and

$$\varphi_{12} := \varphi_2^{-1} \circ \varphi_1 : \varphi_1^{-1}(V) \to \varphi_2^{-1}(V), \tag{5}$$

$$\varphi_{21} := \varphi_1^{-1} \circ \varphi_2 : \varphi_2^{-1}(V) \to \varphi_1^{-1}(V). \tag{6}$$

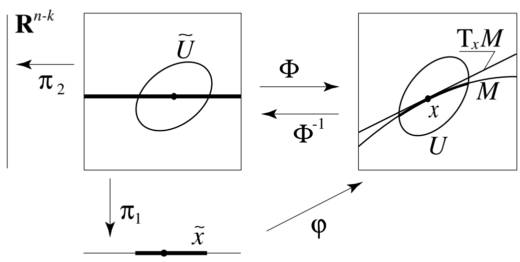
So defined φ_{12} and $\varphi 21$ are called the *transition functions* for these charts.

In other words, a transition function sends the representative of $x \in M$ in one chart into the representative of x in the other one.

Lemma 10.1.3. *The transition function φ_{12} and φ_{21} (see (5), (6)) are mutually inverse diffeomorphism.*
$\lhd$ It is clear from the diagram, that $\varphi_{12} = \pi_2 \circ \Phi_2^{-1} \circ \Phi_1 \circ i_1$, $\varphi_{21} = \pi_1 \circ \Phi_1^{-1} \circ \Phi_2 \circ i_2$. Hence both φ_{12} and φ_{21} are smooth as compositions of smooth mappings. Now, it is clear from (5), (6), that φ_{12} and φ_{21} are mutually inverse. Hence they are mutually inverse diffeomorphism. $\rhd$

10.2 Tangent space

Now we consider tangent vectors to a manifold in $\mathbb{R}^n$.



Theorem 10.2.1. *Let $M \in \mathrm{Mf}^k(\mathbb{R}^n)$. Then for each point $x \in M$ the tangent cone $\mathrm{T}_x\, M$ to M at x is a k-dimensional vector subspace in $\mathbb{R}^n$; viz., for any char φ for M at x*

$$\mathrm{T}_x\, M = \operatorname{im} \varphi'(\widetilde{x}),$$

where $\widetilde{x}$ is the representative of x in the chart φ ($\widetilde{x} = \varphi^{-1}(x)$).

$\lhd$ Put $g := \pi_2 \circ \Phi^{-1}$. Then obviously

$$M \cap U = g^{-1}(0).$$

By theorem on tangent cone to $g^{-1}(0)$,

$$\mathrm{T}_x\, M = \mathrm{T}_x(M \cap U) = \ker g'(x).$$

Thus we need to verify that

$$\ker g'(x) = \operatorname{im} \varphi'(\widetilde{x}). \tag{1}$$

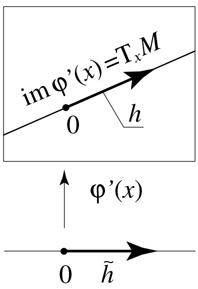
But, indeed, we have (since $\Phi^{-1}(U \cup M) \subset \mathbb{R}^k \times 0$)

$$g \circ \varphi = \pi_2 \circ \Phi^{-1} \circ \varphi = 0 \overset{\text{Chain Rule}}{\Rightarrow} g'(x) \circ \varphi'(\widetilde{x}) = 0 \Rightarrow \operatorname{im} \varphi'(\widetilde{x}) \subset \ker g'(x).$$

It remains to note that both $\operatorname{im} \varphi'(\widetilde{x})$ and $\ker g'(x)$ have the same dimension k (since $\ker \pi_2$ has dimension k, and since both $\varphi'(\widetilde{x})$ and $(\Phi^{-1})'(x)$ are injective). $\triangleright$
Remark. It is convenient to imagine a tangent vector h from $\mathrm{T}_x\, M$ as an arrow with the beginning at x and end at $x + h$ ($x \longrightarrow x+h$). Formally, starting from this point, we mean by a *tangent vector* to M at x a PAIR (x, h), where $h \in \mathrm{T}_x\, M$, and we consider the set $\mathrm{T}_x\, M$ and $\mathrm{T}_y\, M$ as DISJOINT (and if they may coincide as vector subspaces in $\mathbb{R}^n$.

Representatives of a tangent vectors

Let φ be a chart for M at x, and let $\widetilde{x}$ be the representative of x in φ. The pre-image by $\varphi'(x)$ of a tangent vector $h \in \mathrm{T}_x\, M$ (this pre-image is UNIQUE: by Lemma on charts, $\varphi'(\widetilde{x}) \in \mathrm{Inj}$) be called the *representative* of h in φ and will be denoted by $\widetilde{h}$).

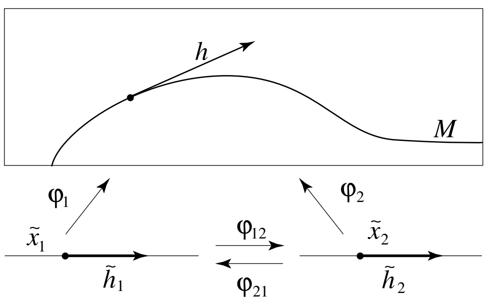


Example. For $\mathbb{R}^n$ as a manifold in $\mathbb{R}^n$ we have

$$\forall x \vdots \ \mathrm{T}_x\, \mathbb{R}^n = \mathbb{R}^n,$$

and in the chart id

$$\forall h \vdots \ \widetilde{h} = h.$$



Lemma 10.2.2. *Let $h \in \mathrm{T}_x\, M$, let $\varphi_1 \varphi_2$ be two chats for M at x let $\widetilde{x}_1, \widetilde{x}_2$ be the representatives of x in φ_1, φ_2, resp.; and let $\widetilde{h}_1, \widetilde{h}_2$ be the representatives of h. Then*

$$\varphi'_{12}(\widetilde{x})\widetilde{h}_1 = \widetilde{h}_2 \tag{2}$$

In other words, the derivative of transition function sends the representative of a tangent vector in one chart into the representative in the other one.

$\triangleleft\ \varphi_1 = \varphi_2 \circ \varphi_{12} \Rightarrow \varphi'_1(\widetilde{x}_1) = \varphi'_2(\widetilde{x}_2) \circ \varphi'_{12}(\widetilde{x}_1)$

$$\Rightarrow \underbrace{\varphi'_1(\widetilde{x}_1)\widetilde{h}_1}_{=h} = \varphi'_2(\widetilde{x}_2) \cdot \varphi'_{12}(\widetilde{x}_1)\widetilde{h}_1 \Rightarrow (2). \ \triangleright$$

10.3 Mapping between manifolds. Vector fields and forms

A *vector field* υ on a manifold M is a mapping from M into $\bigcup_{x \in M} \mathrm{T}_x\, M$ (all $\mathrm{T}_x\, M$ are mutually disjoint!), that sends a point $x \in M$ into a tangent vector to M at x:

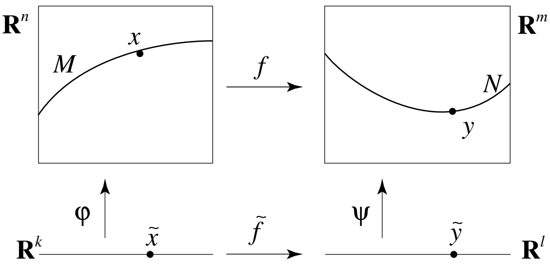
$$\upsilon : x \mapsto \upsilon(x) \in \mathrm{T}_x\, M.$$

A *k-form* ω on M is a mapping that sends $x \in M$ into an antisymmetric tensor on $\mathrm{T}_x\, M$:

$$\omega : x \mapsto \omega(x) \in \Lambda^k(\mathrm{T}_x\, M).$$

SMOOTHNESS of mappings between manifolds and of vector fields and forms on manifold we define as smoothness of their REPRESENTATIVES.

The *representative* $\widetilde{f}$ OF A MAPPING $f : M \to N$ ($M \in \mathrm{Mf}^k(\mathbb{R}^n)$, $N \in \mathrm{Mf}^l(\mathbb{R}^m)$) in charts φ for M at x and ψ for N at $y := f(x)$ is defined by the formula



$$\widetilde{f}(\widetilde{\xi}) := \widetilde{f(\xi)},$$

or, equivalently,

$$\widetilde{f} := \psi^{-1} \circ f \circ \varphi : \mathbb{R}^k \to \mathbb{R}^l$$

(for continuous f this last mapping is defined in an appropriate neighbourhood of x).

In particular, the representative of a mapping

$$f : M \to \mathbb{R}^m$$

in a chart φ for M and the identity chart for $\mathbb{R}^m$ is

$$\widetilde{f} = f \circ \varphi = \varphi^* f.$$

The *representative* $\widetilde{v}$ *of a vector field* v on M in a chart $\varphi : \widetilde{U} \to U$ we define as the mapping

$$\widetilde{v} : \widetilde{U} \to \mathbb{R}^k,\ \widetilde{x} \mapsto \widetilde{v(x)};$$

in other words,

$$\varphi'(\widetilde{x})\widetilde{v}(\widetilde{x}) = v(x).$$

The *representative* $\widetilde{\omega}$ *of a k-form* ω on M in a chart $\varphi : \widetilde{U} \to U$ is defined by the rule

$$\widetilde{\omega}(x)\widetilde{h}_1 \ldots \widetilde{h}_k = \omega(x)h_1 \ldots h_k,$$

or, equivalently,

$$\widetilde{\omega}(x)\widetilde{h}_1 \ldots \widetilde{h}_k = \omega(\varphi(\widetilde{x}))(\varphi'(\widetilde{x})\widetilde{h}_1) \ldots (\varphi'(\widetilde{x})\widetilde{h}_k),$$

that is,

$$\widetilde{\omega} = \varphi^* \omega.$$

Smoothness

We say that a *mapping* $f : M \to N$ ($M, N \in \mathrm{Mf}$) is *differentiable* at a point $x \in M$ (resp., is *smooth*), if for any charts φ at x and ψ at $y := f(x)$ the representative $\widetilde{f}$ of f in these charts is differentiable at $\widetilde{x}$ (resp., is smooth). We define the *derivative* $f_*(x)$ of f at x as the linear mapping from $\mathrm{T}_x M$ into $\mathrm{T}_y N$:

$$f_*(x) \in \mathrm{L}(\mathrm{T}_x M, \mathrm{T}_y N),$$

that acts by the rule

$$\widetilde{f_*(x)h} := \widetilde{f}'(\widetilde{x})\widetilde{h}.$$

In other words, $f_*(x)$ is the linear mapping, represented by the derivative of the representative of f at the representative of x.

(It is easy to verify (using Lemma 10.2.2., p.132), that if f has a differentiable representation in some two charts at x and y, then its representative in any other two charts at x and y will be also differentiable, and that our definition of the derivative does **not** depend on the choice of charts.)

Example. For any smooth mapping $\mathbb{R}^n \to \mathbb{R}^m$ its restriction to any manifold M in $\mathbb{R}^n$ is a smooth mapping $M \to \mathbb{R}^m$. (Verify!)

A *vector field* v on M is called *smooth* if the representative $\widetilde{v}$ of v in each chart for M is smooth. (It is easy to verify, that if v has smooth representatives in each chart from a family $\{\varphi_\alpha\})$, $\varphi_\alpha : \widetilde{U}_\alpha \to U_\alpha$ such that $\bigcup U_\alpha = M$, then the representative of v in any chart for M will be smooth, that is, v will be smooth.)

A *p-form* ω on M is smooth (the record: $\omega \in \Omega^p(M)$), if the representative $\widetilde{\omega}$ of ω in each chart for M is smooth. (Once again, it is easy to verify, that if ω has smooth representatives for some family of chart "covering" M, then ω is smooth.)

Example. For any smooth form on $\mathbb{R}^n$ its restriction to M

$$\omega|_M(x) := \omega(x)|_{\mathrm{T}_x M \times \ldots \times \mathrm{T}_x M}$$

is a smooth form on M.

Exterior derivative

We define the exterior derivative $\mathrm{d}\omega$ of a form ω on M as the form, the representative of which in any chart for M is the exterior derivative of the representative of ω:

$$\widetilde{\mathrm{d}\omega} := \mathrm{d}\widetilde{\omega}.$$

(it can be verified that this "chart-wise" definition is correct, that is, there exists just one smooth form on M with this property.)

Exterior product

Let ω_1, ω_2 be two forms on M. We define their *exterior product* point-wise:

$$\forall x \in M \vdots\ (\omega_1 \wedge \omega_2)(x) := \omega_1(x) \wedge \omega_2(x).$$

It is easy to verify that $\omega_1 \wedge \omega_2$ is also smooth, and in any chart for M

$$\widetilde{\omega_1 \wedge \omega_2} = \widetilde{\omega}_1 \wedge \widetilde{\omega}_2.$$

Pull-back

Let M, N be manifolds, let $f : M \to N$ be a smooth mapping, and let ω be a k-form on N. We define the *pull-back* $f^*\omega$ point-wise:

$$\forall x \in M \vdots\ (f^*\omega)(x) := (f_*(x))^*(\omega(f(x)))$$

(compare with the definition for forms on vector spaces), that is,

$$\forall h_1, \ldots, h_k \in \mathrm{T}_x M \vdots\ (f^*\omega)(x)h_1, \ldots, h_k = \omega(f(x))(f_*(x)h_1) \ldots (f_*(x)h_k).$$

Again, it can be verified that $f^*\omega$ is smooth, and that in any charts for M and N

$$\widetilde{f^*\omega} = (\widetilde{f})^*\widetilde{\omega}.$$

Remarks. 1. The representative $\widetilde{\omega}$ of a form ω on M in a chart φ for M is the pull-back:

$$\widetilde{\omega} = \varphi^*\omega.$$

2. Just as in the case of vector spaces, for manifolds also the operations $*$, d and $\wedge$ RESPECT each other.

10.4 Manifolds with a boundary

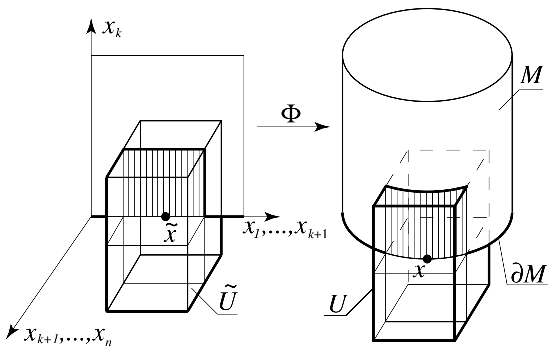
A subset M of $\mathbb{R}^n$ is a *k-dimensional manifold with a boundary*, or simply a k-manifold with a boundary (the notation: $M \in \mathrm{Mf}^k_\partial(\mathbb{R}^n)$) if for each $x \in M$ one of two condition (1) and (2), is fulfilled, where (1) is the condition from 10.1, p.129, and (2) is the following condition:

$$\exists U \in \mathrm{Op}\,\mathrm{Nb}_x(\mathbb{R}^n)\ \exists \widetilde{U} \in \mathrm{Op}(\mathbb{R}^n)\ \ \exists \Phi \in \mathrm{Diffeo}(\widetilde{U}, U):$$
$$\widetilde{x} := \Phi^{-1}(x) \in \mathbb{R}^{k-1} \times \underbrace{0}_{\in \mathbb{R}^{n-(k-1)}}, \tag{1}$$

and

$$U \cap M = \Phi(\widetilde{U} \cap (\mathbb{R}^{k-1} \times \mathbb{R}_+ \times \underbrace{0}_{\in \mathbb{R}^{n-k}})). \tag{2}$$

(Here $\mathbb{R}_+ := [0, +\infty)$.) In other words, M is up to a diffeomorphism, a k-dimensional **half**-space in $\mathbb{R}^n$, the point in question lying on the boundary of this half-space.



Note that (1) and (2) cannot be fulfilled simultaneously, since Φ is **homeo**morphism, and a half-space (closed!) and the whole space are not homeomorphic.

The set of all points $x \in M$, for which (2) is fulfilled, is called the *boundary* of M and is denoted by

$$\partial M.$$

Example. If $M \in \mathrm{Mf}^k(\mathbb{R}^n)$ then $\partial M = \emptyset$ (though $\mathrm{fr}\, M \neq \emptyset$ in general, e.g. for an open ball).

Exercise 10.4.1. Show that if $M \in \mathrm{Mf}^k_\partial(\mathbb{R}^n)$, then $\partial M \in \mathrm{Mf}^{k-1}(\mathbb{R}^n)$ and $M \setminus \partial M \in \mathrm{Mf}^k(\mathbb{R}^n)$.

All the notions introduced for manifold “without boundary” (full charts, charts, forms vector fields, representatives etc.) can be naturally extended to the case of manifolds with a boundary. Little complications arises with differentiation (”usual“ and exterior) at points of the boundary ∂M, but ∂M has zero volume (by Corollary from Theorem Change Variables), so when dealing with INTEGRALS over manifolds (to be defined below), the values at boundary points are not essential, and we can just ignore these points. [More accurately: we define a smooth form on a manifold with boundary as a form with smooth representatives, and we DEFINE a smooth (representing) form on a closed half-space as a RESTRICTION to this half space of a smooth form on an OPEN set that contains the half-space. This solves all mentioned problems.]

10.5 Orientation

Let X be an n-dimensional vector space, and let

$$A := \{a_1, \dots, a_n\} \qquad \text{and} \qquad B := \{b_1, \dots, b_n\}$$

be two bases for X. We define the *sign* of A with respect to B by the formula

$$\operatorname{sgn}_B A := \operatorname{sgn}\det_B a_1 \dots a_n,$$

where $\det_B$ denotes the *determinant with respect to basis B:*

$$a_j = \sum_{i=1}^{n} a_{ij} b_i \quad \Rightarrow \quad \det_B a_1 \dots a_n := \begin{vmatrix} a_{11} & \dots & a_{1n} \\ & \dots & \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

(Note that the last determinant cannot be equal to 0, since $a_1, \dots, a_n$ are linearly independent.)

The relation

$$A \sim B :\Leftrightarrow \operatorname{sgn}_B A = +1$$

is an equivalence relation (Verify!) An *orientation* of X is an equivalent class with respect to $\sim$. Obviously on X there are just 2 orientations. For a given basis, the orientation, containing B, we denote by

$$[B]$$

the other one by

$$-[B].$$

We say that X is *oriented* if there is chosen one of 2 possible orientation on X. We denoted this chosen class

$$\operatorname{or} X$$

For oriented X we say that a basis A is *positive* if $[A] = \operatorname{or} X$, and is *negative* $[A] = -\operatorname{or} X$.

For $\mathbb{R}^n$, the *canonical orientation* is $[\mathrm{e}_1, \dots, \mathrm{e}_n]$.

Examples.

1. $\forall \sigma \in \mathfrak{S}_n \vdots\ [\mathrm{e}_{\sigma(1)}, \dots, \mathrm{e}_{\sigma(n)}] = (\operatorname{sgn}\sigma)[\mathrm{e}_1, \dots, \mathrm{e}_n]$ (Prove!)
2. $[-\mathrm{e}_1, \mathrm{e}_2, \dots, \mathrm{e}_n] = -[\mathrm{e}_1, \dots, \mathrm{e}_n]$. (Prove!)

Positive linear bijections

Let X and Y be two n-dimensional ORIENTED vector spaces, $\operatorname{or} X = [A]$, $\operatorname{or} Y = [B]$, and let $l \in \mathrm{L}(X, Y)$ be a bijection. We put

$$\operatorname{sgn} l := \operatorname{sgn}\det l$$

where $\det l$ denotes the determinant of the matrix of l with respect to the bases A and B:

$$\det l := \det_B (la_1) \dots (la_n) \qquad (\{a_1, \dots, a_n\} = A).$$

(It is easy to verify, that this definition of $\operatorname{sgn} l$ does not depend on the choice of positive bases A and B.) We say that l is *positive* if $\operatorname{sgn} l = +1$ (resp., *negative*, if not).

Lemma 10.5.1. *A positive linear bijection l between oriented finite-dimensional vector spaces* RESPECTS *orientations, that is, sends each positive basis into a positive one.*

$\triangleleft$ Let $\operatorname{or} X = [A]$, $\operatorname{or} Y$. Let $\{c_1, \dots, c_n\}$ be a positive basis in X, that is,

$$\det_A c_1 \dots c_n > 0. \tag{1}$$

Then

$$\det_B(lc_1)\dots(lc_n) = (t^*\det_B)c_1\dots c_n \overset{\text{Th on det}}{=} \underbrace{(\det l)}_{\boxed{\operatorname{sgn} l=+1}>0}\ \underbrace{\det{}_A c_1\dots c_n}_{\boxed{(1)}>0} > 0,$$

which means that the basis $\{lc_1,\dots,lc_n\}$ is positive. ▷

Oriented manifolds

We say that a k-manifold M is *oriented*, and we write

$$M \in \operatorname{Or}\operatorname{Mf}^k,$$

if for each $x \in M$ the tangent space $\mathrm{T}_x M$ is oriented and if these orientations are *compatible* in the sense that for any chart $\varphi : \widetilde{V} \to V$ for M all all the mappings $\varphi'(x)$, $\widetilde{x} \in \widetilde{V}$ (which are linear bijections of $\mathbb{R}^k$ onto $\mathrm{T}_{\varphi(\widetilde{x})} M$) have one and the same sign (all are positive or all are negative), with respect to the canonical orientation of $\mathbb{R}^k$.

It is obvious that if M is *orientable* (can be oriented), there are just 2 orientation on M. To fix an orientation on an orientable M, it is sufficient to claim any one chart φ as *positive*, in the sense that for each point $\widetilde{x}$ from the domain of this chart $\varphi'(\widetilde{x})$ is positive.

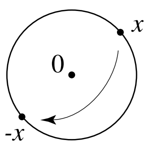
Examples.

1. For $n \geq 2$ all the $(n-1)$ dimensional spheres in $\mathbb{R}^n$ are orientable.

2. The famous Möbius band is not oriented. (Exercise: define the Möbius band using TWO full charts.)

Positive injections

Let $N, M \in \operatorname{Or}\operatorname{Mf}^k$, and let f be a (smooth) injection of M into N. We say that f is *positive* (resp., *negative*), if for any $x \in M$ the derivative $f_*(x) : \mathrm{T}_x M \to \mathrm{T}_{f(x)} N$ is positive (resp., negative). Thus, a positive mapping RESPECTS orientations.



Example. The mapping $f : \mathbb{S}^k \to \mathbb{S}^k$, $\quad x \mapsto -x$, where $\mathbb{S}^k$ denotes the unit sphere in $\mathbb{R}^{k+1}$, is positive if k is odd, and is negative if k is even.

Induced orientation

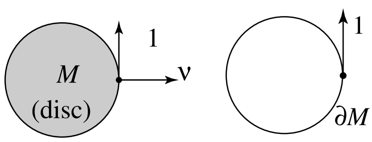
Let M be an oriented k-manifold with a boundary. The *induced orientation* on ∂M is given by the rule: for each $x \in \partial M$

$$[h_1,\dots,h_{k-1}] \in \operatorname{or}\mathrm{T}_x(\partial M) :\Leftrightarrow [\nu, h_1,\dots,h_{k-1}] = \operatorname{or}\mathrm{T}_x M,$$

where ν is the (uniquely defined) unit normal vector to M at x, such that $-\nu$ is a tangent vector to M at x (note that $T_x M$ is here a (k-dimensional) HALF-space).

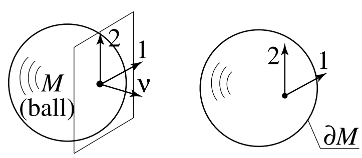
Examples.

1.



$[\nu, 1] = \operatorname{or} M$ $\qquad$ $[1] = \operatorname{or}\partial M$

2.



$[\nu, 1, 2] = \operatorname{or} M$ $\qquad$ $[1, 2] = \operatorname{or}\partial M$

Lemma 10.5.2. *Let in the cube* $[0,1]^k$ *its faces* $\operatorname{im} I^k_{(i,\alpha)}$ *are equipped with with the orientation induced by the canonical orientation in* $[0,1]^k$*. Then* $I^k_{(i,\alpha)}$ *is positive iff* $i+\alpha$ *is even:*

$$\operatorname{sgn} I^k_{(i,\alpha)} = (-1)^{1+\alpha}.$$

◁ EXERCISE. [*Hint*: see Example 2 (p. 136).] ▷

Corollary 10.5.3. *Let c be a k-cube in $\mathbb{R}^n$, and let* $\operatorname{im} c$ *be oriented by the condition that c is positive. Let the faces* $\operatorname{im} c_{(i,\alpha)}$ *be equipped by the induced orientation. Then $c_{(i,\alpha)}$ is positive iff $i + \alpha$ is even:*

$$\operatorname{sgn} c_{(i,\alpha)} = (-1)^{i+\alpha}.$$

10.6 Integral of a form on an oriented manifold

Throughout this section M denotes an ORIENTED k-manifold with a boundary.

Integral over a chain on a manifold

Let c be a k-cube in M (that is, a k-cube in $\mathbb{R}^n$, such that $\operatorname{im} c \subset M$), and let $\omega \in \Omega^k(M)$. We put just as for cubes in $\mathbb{R}^n$

$$\int_c \omega := \int_{I^k} c^*\omega.$$

Integrals over chains on M are defined once again ”by linearity“:

$$\int_{\sum a_i c_i} \omega := \sum a_i \int_{c_i} \omega \qquad \text{(the sums are FINITE)}.$$

Chart cubes

We say that a k-cube c in M is a *chart cube* if there exists a chart $\varphi : \widetilde{V} \to V$ for M such that

$$\widetilde{V} \supset [0,1]^k \qquad \text{and} \qquad c = \varphi|_{[0,1]^k}. \tag{1}$$

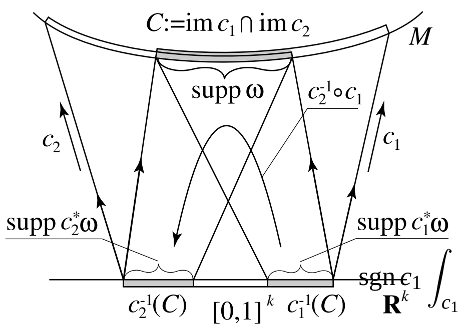
Thus each chart cube is an injection, so its SIGN is determined (with respect to the canonical orientation on $[0,1]^k$).

Let ω be a k-form on M. If there exists a chart k-cube c in M such that

$$\operatorname{supp} \omega \subset \operatorname{im} c \tag{2}$$

then we put

$$\boxed{\int_M \omega := \operatorname{sgn} c \int_c \omega} \tag{3}$$



Theorem 10.6.1. *This definition is correct, that is, does not depend on the choice of c: if c_1, c_2 are two chart k-cubes in M such that*

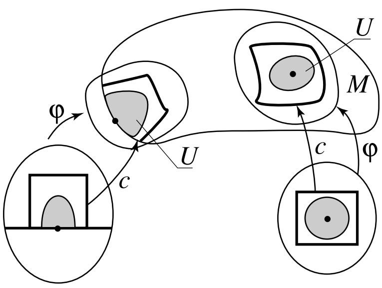
$$\operatorname{supp} \omega \subset \operatorname{im} c_1 \cap \operatorname{im} c_2,$$

then

$$\operatorname{sgn} c_1 \int_{c_1} \omega = \operatorname{sgn} c_2 \int_{c_2} \omega.$$

$$\begin{aligned}
\lhd \operatorname{sgn} c_1 \int_{c_1} \omega &= \operatorname{sgn} c_1 \int_{I^k} c_1^* \omega = \operatorname{sgn} c_1 \int_{I^k} (c_2 \circ \underbrace{(c_2^{-1} \circ c_1)}_{\substack{\text{defined only on } c_1^{-1}(C),\\ \text{but } \operatorname{supp}\omega \subset C}})^* \omega \\
&= \operatorname{sgn} c_1 \int_{I^k} (c_2^{-1} \circ c_1)^* \underbrace{c_2^* \omega}_{\substack{\text{as a } k\text{-form}\\ \text{on } \mathbb{R}^k\\ =\ g \det}} \\
&\overset{\text{Th on det}}{=} \operatorname{sgn} c_1 \int_{I^k} (g \circ c_2^{-1} c_1)\ \underbrace{\det(c_2^{-1} \circ c_1)'}_{=\underbrace{\operatorname{sgn}(c_2^{-1} \circ c_1)'}_{\substack{\text{Chain Rule and}\\ \text{def. of sgn}\\ = (\operatorname{sgn} c_2)(\operatorname{sgn} c_1)}} |\det(c_2^{-1} \circ c_1)|}\ \det \\
&= \operatorname{sgn} c_2 \int_{[0,1]^k} (g \circ (c_2^{-1} \circ c_1)) |\det(c_2^{-1} \circ c_1)'| \underset{\substack{\text{Th on change}\\ \text{of var's}}}{=} \operatorname{sgn} c_2 \int_{[0,1]^k} g \\
&\overset{c_2^*\omega = g \det}{=} \operatorname{sgn} c_2 \int_{I^k} c_2^* \omega = \operatorname{sgn} c_2 \int_{c_2} \omega. \rhd
\end{aligned}$$

Partitions of manifolds



Lemma 10.6.2. *Let M be compact. There exists a finite covering $\mathcal{O}$ of M by* RELATIVELY *open (that is, open in M equipped with the topology induced from $\mathbb{R}^n$) sets U, each of which is contained in the image of some chart cube.*

$\lhd$ It follows obviously from the fact that for each point of M it is fulfilled either (1) or (2) (see the picture). $\rhd$

Lemma 10.6.3. *For any compact M there exists a finite partition of unity Ψ on M, submitted to the covering $\mathcal{O}$ from Lemma 10.6.2.*

(The definition of a partition of unity for manifolds is the same as early, merely now functions $\varphi \in \Phi$ are functions on M; but we know what is a smooth function on a manifold.)

$\lhd$ This follows from the following lemma $\rhd$

Lemma 10.6.4. *For any M and any covering $\mathcal{O}$ of M by relatively open sets there exists a partition of unity on M, submitted to $\mathcal{O}$.*

$\lhd$ Each $U \in \mathcal{O}$ can be represented as $U' \cap M$, where U' is an open set in $\mathbb{R}^n$. The family $\mathcal{O}'$ of all such U' is an open covering of M in $\mathbb{R}^n$. Let Φ' be a partition of unity for M submitted to $\mathcal{O}'$ Then $\Phi := \{\varphi|_M : \varphi \in \Phi'\}$ is what we need. (Note that $\varphi|_M$ is a smooth function on M (see 10.3).) $\rhd$

Integral over a manifold

For a compact M and a k-form ω on M we put

$$\boxed{\int_M \omega := \sum_{\varphi \in \Phi} \int_M \varphi\omega}$$

where φ is a FINITE partition of unity for M submitted to a covering $\mathcal{O}$ of M by RELATIVELY OPEN sets, each of which is contained in the image of some CHART cube, and the integrals in the sum are defined by the formula (3) on p. 138.

Such a covering $\mathcal{O}$ and such a partition Φ do exists by Lemmas 10.6.2. and 10.6.3., and it can be shown that the so defined integral $\int_M \omega$ does not depend on the choice of $\mathcal{O}$ and Φ.

10.7 Stokes Theorem

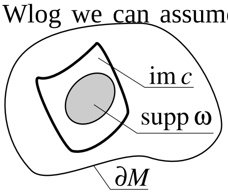
Theorem 10.7.1. (Stokes) *Let M be a compact oriented k-manifold with a boundary in $\mathbb{R}^n$, and let $\omega \in \Omega^{k-1}(M)$. Then*

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega$$

where ∂M is equipped by the induced orientation.

$\lhd$ *Case 1*: There exists a chart k-cube c in M such that $\operatorname{supp}\omega \subset \operatorname{int}(\operatorname{im} c)$.
(Of course $\operatorname{supp}\omega := \operatorname{cl}\{x \in M : \omega(x) \neq 0\}$, the closure IN M, but since M is compact, it is the same as the closure in $\mathbb{R}^n$.)

Wlog we can assume that c is positive (if there exists a negative chart cube with the mentioned property, then obviously there exists also a positive chart cube with the same property). We have



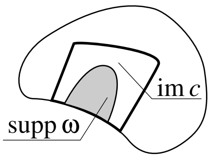
$$\int_M \mathrm{d}\omega = \int_c \mathrm{d}\omega = \int_{I^k} c^*\mathrm{d}\omega = \int_{I^k} \mathrm{d}(c^*\omega)$$
$$\overset{\substack{\text{Stokes Th}\\ \text{for chains}}}{=} \int_{\partial I^k} c^*\omega = \int_{\partial c} \omega = 0,$$

since $\omega = 0$ on

$$\operatorname{im}(\partial c) := \bigcup_{i,\alpha} \operatorname{im} c_{(i,\alpha)}.$$

(Note that Stokes Theorem for chains is applicable here, since $c^*\omega$ is, by the definition of a smooth form on manifold, smooth in some open neighbourhood of $[0,1]^k$.)

But also $\int_{\partial M} \omega = 0$, since $\omega = 0$ on ∂M O.K.



Case 2: There exists a chart k-cube c in M, such that

$$\partial M \cap \operatorname{im}(\partial c) = \operatorname{im} c_{(k,0)}, \tag{1}$$

$$\operatorname{supp}\omega \subset \operatorname{rel\,int}(\operatorname{im} c). \tag{2}$$

Again wlog c is positive, so that

$$\int_M \mathrm{d}\omega \overset{\text{as in Case 1}}{=} \int_{\partial c} \omega = \sum_{i,\alpha} (-1)^{1+\alpha} \int_{c_{(i,\alpha)}} \omega \overset{\substack{\omega = 0 \text{ on all the} \\ \text{faces but } c_{(k,0)}}}{=} (-1)^k \int_{c(k,0)} \omega$$

$$\overset{(3)}{=} (-1)^k \underbrace{\operatorname{sgn} c_{(0,k)}}_{\text{Lm } 10.5.2. \atop = (-1)^k} \int_{\partial M} \omega = \int_{\partial M} \omega. \quad \text{O.K.}$$

General case: We have

$$\int_M \mathrm{d}\omega \overset{\text{def}}{=} \overset{\text{finite!}}{\sum_{\varphi \in \Phi}} \int_M \varphi \mathrm{d}\omega \underset{\substack{\sum \mathrm{d}\varphi = 0, \\ \text{since } \sum \varphi = 1}}{\overset{\text{trick!}}{=}} \sum_{\varphi \in \Phi} \int_M (\mathrm{d}\varphi \wedge \omega + \varphi \mathrm{d}\omega) \overset{\text{Cases 1,2}}{=} \sum_{\varphi \in \Phi} \int_{\partial M} \varphi\omega$$

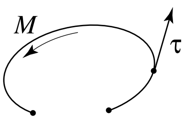
$$\overset{\text{def}}{=} \int_{\partial M} \omega. \ \rhd$$

10.8 Classical special cases

In this section we discuss classical notions of divergence and rotor. For this end we need some special 1- and 2-forms, named length element and area element, resp.

Length element

Let M be an oriented 1-manifold (maybe with a boundary) in $\mathbb{R}^3$, let τ denote the unit positive (that is, respecting the orientation) tangent vector to M.



The *length element* $\mathrm{d}s$ on M is the 1-form ON M, defined by the rule

$$\boxed{\mathrm{d}s(x)h := \langle \tau, h \rangle} \qquad (h \in \mathrm{T}_x\, M) \tag{1}$$

Theorem 10.8.1. *Let $M \in \operatorname{Or}\operatorname{Mf}^1_\partial(\mathbb{R}^3)$, let $\tau = (\tau_1, \tau_2, \tau_3)$ be the unit positive positive vector field on M, and let* $\mathrm{d}s$ *be the length element on M. Then*

a) $\mathrm{d}s = \tau_1 \mathrm{d}x + \tau_2 \mathrm{d}y + \tau_3 \mathrm{d}z$,
b) $\tau_1 \mathrm{d}s = \mathrm{d}x$, $\tau_2 \mathrm{d}s = \mathrm{d}y$, $\tau_3 \mathrm{d}s = \mathrm{d}z$.

Of course, here $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ denote the RESTRICTIONS on M of the 1-forms $\mathrm{d}x, \mathrm{d}y, \mathrm{d}z$ on $\mathbb{R}^3$.

$\lhd$ a) $\mathrm{d}s \cdot h \overset{(1)}{=} \langle \tau, h \rangle = \tau_1 h_1 + \tau_2 h_2 + \tau_3 h_3 = \tau_1 \mathrm{d}x \cdot h + \tau_2 \mathrm{d}y \cdot h + \tau_3 \mathrm{d}z \cdot h = (\tau_1 \mathrm{d}x + \tau_2 \mathrm{d}y + \tau_3 \mathrm{d}z)h$.

b) Let $h \in \mathrm{T}_x\, M$. Then $h = \alpha\tau$ for some $\alpha \in \mathbb{R}$. Hence

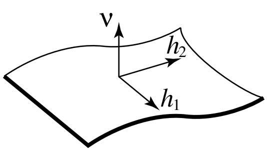
$$(\tau_1 \mathrm{d}s)h = \tau_1 (\mathrm{d}s \cdot h) \overset{(1)}{=} \underbrace{\tau_1}_{\mathrm{d}x \cdot \tau} \underbrace{\langle \tau, \alpha\tau \rangle}_{=\alpha} = \mathrm{d}x \cdot \underbrace{\alpha\tau}_{=h} = \mathrm{d}x \cdot h,$$

and analogously for $\mathrm{d}y$, $\mathrm{d}z$. $\rhd$

Definition. The *length* of a 1-dimensional compact oriented manifold M in $\mathbb{R}^3$ is $\int_M \mathrm{d}s$.

Area element

Let M be an oriented 2-manifold in $\mathbb{R}^3$ (maybe with a boundary), and let ν be the unit normal vector to M, *positive* in the sense that



$$[h_1, h_2] \in \operatorname{or} M \Rightarrow [\nu, h_1, h_2] = \operatorname{or} \mathbb{R}^3.$$

The *area element* $\mathrm{d}A$ on M is the 2-form on M, defined by the rule

$$\boxed{\mathrm{d}A(x)hk := \det(h, k, \nu)} \qquad (h, k \in \mathrm{T}_x\, M).$$

Theorem 10.8.2. *Let $M \in \operatorname{Or}\mathrm{Mf}^2_\partial(\mathbb{R}^3)$, let $\nu = (\nu_1, \nu_2, \nu_3)$ be the positive unit normal vector field on M, and let $\mathrm{d}A$ be the areal element on M. Then*

a) $\mathrm{d}A = \nu_1 \mathrm{d}y \wedge \mathrm{d}z + \nu_2 \mathrm{d}z \wedge \mathrm{d}x + \nu_3 \mathrm{d}x \wedge \mathrm{d}y$;
b) $\nu_1 \mathrm{d}A = \mathrm{d}y \wedge \mathrm{d}z, \quad \nu_2 \mathrm{d}A = \mathrm{d}z \wedge \mathrm{d}x, \quad \nu_3 \mathrm{d}A = \mathrm{d}x \wedge \mathrm{d}y.$

$\lhd$ 1° For $h, k \in \mathbb{R}^3$ define the *vector product* $h \times k$ by the rule

$$\forall t \in \mathbb{R}^3 : \ \langle h \times k, t\rangle := \det(h, k, t). \tag{2}$$

Applying (2) to $t = \mathrm{e}_i$, we conclude that

$$h \times k = \left(\begin{vmatrix} h_2 & h_3 \\ k_2 & k_3 \end{vmatrix}, \begin{vmatrix} h_3 & h_1 \\ k_3 & k_1 \end{vmatrix}, \begin{vmatrix} h_1 & h_2 \\ k_1 & k_2 \end{vmatrix} \right). \tag{3}$$

It follows at once from (2) that

$$h \times k \angle \operatorname{lin}\{h, k\}. \tag{4}$$

(Indeed if t is a linear combination of h and k, then $\det(h, k, t) = 0$.) Hence

$$\forall h, k \in \mathrm{T}_x\, M : \ h \times k = \alpha\nu \qquad \text{for some } \alpha \in \mathbb{R}. \tag{5}$$

2° Proof of a). $\lhd\!\lhd$ $\forall h.k \in \mathrm{T}_x\, M :$ $\mathrm{d}A(x)hk = \det(h, k, \nu) = \nu_1 \underbrace{\begin{vmatrix} h_2 & h_3 \\ k_2 & k_3 \end{vmatrix}}_{=(\mathrm{d}y\wedge\mathrm{d}z)hk} + \ldots =$ $(\nu_1 \mathrm{d}y \wedge \mathrm{d}z + \ldots)hk.$ $\rhd\!\rhd$

3° Proof of b). $\lhd\!\lhd$ $(\nu_1 \mathrm{d}A)hk = \nu_1(\mathrm{d}Ahk) = \nu_1 \det(h, k, \nu) \overset{(2)}{=} \nu_1 \langle h \times k, \nu\rangle \overset{(5)}{=} \alpha\nu_1 =$ $\alpha\langle \mathrm{e}_1, \nu\rangle = \langle \mathrm{e}_1, \alpha\nu\rangle \overset{(5)}{=} \langle \mathrm{e}_1, h \times k\rangle \overset{(3)}{=} \begin{vmatrix} h_2 & h_3 \\ k_2 & k_3 \end{vmatrix} = (\mathrm{d}y \wedge \mathrm{d}z)hk.$ $\rhd\!\rhd$ $\rhd$

Definition. The *area* of 2-dimenstional compact oriented manifold M in $\mathbb{R}^3$ is $\int_M \mathrm{d}A$.

Theorem on rotor

Let $M \in \operatorname{Or}\mathrm{Mf}^2_\partial(\mathbb{R}^3)$, and let $F = (F_1, F_2, F_3)$ be a vector field in $\mathbb{R}^3$. Put

$$\omega := F_1 \mathrm{d}x + F_2 \mathrm{d}y + F_3 \mathrm{d}z.$$

Then

$$\mathrm{d}\omega = (D_2F_3 - D_3F_2)\mathrm{d}y \wedge \mathrm{d}z + (D_3F_1 - D_1F_3)\mathrm{d}z \wedge \mathrm{d}x + (D_1F_2 - D_2F_1)\mathrm{d}x \wedge \mathrm{d}y.$$

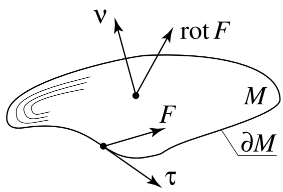
We define the *rotor* rot F of the vector field F as the vector vector formed by the coefficients of this form:

$$\operatorname{rot} F := \left(\begin{vmatrix} D_2 & D_3 \\ F_2 & F_3 \end{vmatrix}, \begin{vmatrix} D_3 & D_1 \\ F_3 & F_1 \end{vmatrix}, \begin{vmatrix} D_1 & D_2 \\ F_1 & F_2 \end{vmatrix} \right).$$

Stokes theorem yields

$$\int_M ((\underbrace{D_2F_3 - D_3F_2}_{(\operatorname{rot} F)_1})\underbrace{\mathrm{d}y \wedge \mathrm{d}z}_{\nu_1 \mathrm{d}A} + \ldots) = \int_{\partial M} (F_1 \underbrace{\mathrm{d}x}_{=\tau_1 \mathrm{d}s} + \ldots),$$

whence it follows that



$$\boxed{\int_M \langle \operatorname{rot} F, \nu\rangle \mathrm{d}A = \int_{\partial M} \langle F, \tau\rangle \mathrm{d}s.} \tag{6}$$

Physical meaning of (6) is: the FLOW of the rotor of a vector field through a surface is equal to the CIRCULATION of this vector field over the contour.

Remark. 1. Physically, rot F is **not** a vector, since the direction of so defined rot F depends on definition of d.

2. In English language literature one use more oft the termini “curl” instead of “rotor”.

3. The so-called Green formula $\int_{\partial M} \alpha \mathrm{d}x + \beta \mathrm{d}y = \int_M (\partial\beta/\partial x - \partial\alpha/\partial y)\mathrm{d}x\mathrm{d}y$ is a special case of (6), corresponding to $F_3 = 0$ and to a FLAT surface, parallel to x, y plane.

Theorem on divergence

Let now $M \in \operatorname{Or}\operatorname{Mf}^2_\partial(\mathbb{R}^3)$ (equipped with the orientation of $\mathbb{R}^3$), and let again F be a vector field in $\mathbb{R}^3$. Put

$$\omega := F_1 \mathrm{d}y \wedge \mathrm{d}z + F_2 \mathrm{d}z \wedge \mathrm{d}x + F_3 \mathrm{d}x \wedge \mathrm{d}y.$$

Then

$$\mathrm{d}\omega = (D_1F_1 + D_2F_2 + D_3F_3)\det.$$

The quantity in the brackets is called the *divergence* of the field F and is denoted by div F:

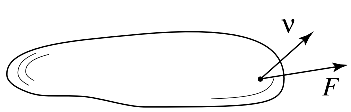
$$\operatorname{div} F := D_1F_1 + D_2F_2 + D_3F_3.$$

Stokes formula yields

$$\int_M (D_1F_1 + \ldots)\det = \int_\partial (F_1 \underbrace{\mathrm{d}y \wedge \mathrm{d}z}_{=\nu_1 \mathrm{d}A} + \ldots),$$

whence it follows that

$$\boxed{\int_M \operatorname{div} F = \int_{\partial M} \langle F, \nu\rangle \mathrm{d}A.} \tag{7}$$



The physical sense of (7) is: the flow of the vector field through a closed surface OUTSIDE is equal to the integral of the divergence of the field over the region inside this surface.

Chapter 11

Functions of complex variable

11.1 Analytic functions

We can identify the set $\mathbb{C}$ of complex numbers with the real plain $\mathbb{R}^2$. More precisely, if we put

$$\alpha : \mathbb{C} \to \mathbb{R}^2,\ z \mapsto (\mathcal{R}ez, \mathcal{I}mz) =: \widetilde{z} \quad \text{(we call } \widetilde{z} \text{ the } \textit{representative} \text{ of } z\text{)},$$

$$\beta : \mathbb{R}^2 \to \mathbb{C},\ (x, y) \mapsto x + \mathrm{i}y,$$

then

$$\alpha \circ \beta = \mathrm{id}, \qquad \beta \circ \alpha = \mathrm{id}\,.$$

By this identification, the norm in $\mathbb{C}$, defined by the formula

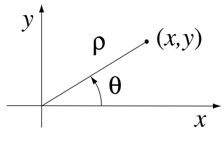
$$\|z\| := |z|,$$

coincides with Euclidean norm in $\mathbb{R}^2$. So we can identify $\mathbb{C}$ and $\mathbb{R}^2$ also as topological spaces.

If

$$x = \varrho\cos\theta, \qquad y = \varrho\sin\theta,$$

we say that ϱ, θ are *polar coordinates* of (x, y) and of $z = x + \mathrm{i}y$ and we write



$$(x, y) \sim \{\varrho, \theta\}, \quad \text{and} \quad z \sim \{\varrho, \theta\}.$$

E.g., $0 \sim \{0, 0\}$, $0 \sim \{0, 22\pi\}$, $\mathrm{i} \sim \{1, \pi/2\}$, $\mathrm{i} \sim \{1, -3\pi/2\}$.

Lemma 11.1.1. *If* $z_1 \sim \{\varrho_1, \theta_1\}$, $z_2 \sim \{\varrho_2, \theta_2\}$, *then* $z_1z_2 \sim \{\varrho_1\varrho_2, \theta_1 + \theta_2\}$.
$\lhd$ $(\varrho_1\cos\theta_1 + \mathrm{i}\varrho_1\sin\theta_1)(\varrho_2\cos\theta_2 + \mathrm{i}\varrho_2\sin\theta_2) = \varrho_1\varrho_2(\underbrace{\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2}_{=\cos(\theta_1+\theta_2)}) + \mathrm{i}\varrho_1\varrho_2(\underbrace{\sin\theta_1\cos\theta_2 - \cos\theta_1\sin\theta_2}_{=\sin(\theta_1+\theta_2)})$. $\rhd$

Lemma 11.1.2. *Let* $c = a + \mathrm{i}\varrho \sim \{\varrho, \theta\}$. *Consider the operator of multiplication by* c

$$A : \mathbb{C} \to \mathbb{C}, \qquad z \mapsto cz,$$

and denote by $\widetilde{A}$ the corresponding operator in $\mathbb{R}^2$:

$$\widetilde{A} := \alpha \circ A \circ \beta.$$

Then $\widetilde{A} \in \mathrm{L}(\mathbb{R}^2, \mathbb{R}^2)$, and the matrix of $\widetilde{A}$ is equal to

$$\begin{tikzcd} \mathbb{C} \arrow[r, "A"] & \mathbb{C} \arrow[d, "\alpha"] \\ \mathbb{R}^2 \arrow[u, "\beta"] \arrow[r, "\widetilde{A}"] & \mathbb{R}^2 \end{tikzcd}$$

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \varrho \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In other words, $\widetilde{A}$ is the composition of the TURN by the angle θ and blowing up with the coefficient ϱ (≥ 0).

$\lhd$ $(a+\mathrm{i}b)(x+\mathrm{i}y) = (ax-by) + \mathrm{i}(ay+bx)$, and $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax-by \\ ay+bx \end{pmatrix}$. $\rhd$

$\mathbb{C}$-differentiability

We say that a function $f : \mathbb{C} \to \mathbb{C}$ is $\mathbb{C}$-differentiable at a point $\hat{z} \in \mathbb{C}$, and we write

$$f \in \mathrm{Dif}_{\mathbb{C}}(\hat{z}),$$

if there exists the limit (in the norm)

$$\lim_{\substack{z \to \hat{z} \\ (z \neq \hat{z})}} \frac{f(z) - f(\hat{z})}{z - \hat{z}} =: f'(\hat{z}) \in \mathbb{C},$$

called the ($\mathbb{C}$-) *derivative* of f at $\hat{z}$. If G is an open subset of $\mathbb{C}$, and f is differentiable at each point of G, then we say that f is $\mathbb{C}$-*differentiable in G*, and we write

$$f \in \mathrm{Dif}_{\mathbb{C}}(G).$$

Examples.

1. $z \mapsto z^n$ is $\mathbb{C}$-differentiable everywhere if $n = 0, 1, 2 \ldots$, and is $\mathbb{C}$-differentiable in $\mathbb{C} \setminus 0$ if $n = -1, -2, \ldots$; $(z^n)' = nz^{n-1}$.

2. $z \mapsto \overline{z}$ is nowhere $\mathbb{C}$-differentiable. (Verify!)

$$\begin{tikzcd} \mathbb{C} \arrow[r, "f"] & \mathbb{C} \arrow[d, "\alpha"] \\ \mathbb{R}^2 \arrow[u, "\beta"] \arrow[r, "\widetilde{f}"] & \mathbb{R}^2 \end{tikzcd}$$

Notation: For any $f : \mathbb{C} \to \mathbb{C}$ we denote by $\widetilde{f}$ the corresponding "real" mapping:

$$\widetilde{f} := \alpha \circ f \circ \beta,$$

that is

$$\widetilde{f}(x, y) := (\underbrace{\mathcal{R}e(f(x+\mathrm{i}y))}_{=:u(x,y)}, \underbrace{\mathcal{I}m(f(x+\mathrm{i}y))}_{=:v(x,y)})$$

For short we write $f = u + \mathrm{i}v$. Thus

$$f = u + \mathrm{i}v :\Leftrightarrow \widetilde{f} = (u, v).$$

Theorem 11.1.3. *Let $f = u + \mathrm{i}v$. The following conditions are equivalent:*

a) $f \in \mathrm{Dif}_{\mathbb{C}}(z)$, $f'(z) = a + \mathrm{i}b \sim \{\varrho, \theta\}$;

b) $\widetilde{f} \in \mathrm{Dif}(\widetilde{z}),\ \widetilde{f}'(\widetilde{z}) = \begin{pmatrix} \partial u/\partial x & \partial u/\partial y \\ \partial v/\partial x & \partial v/\partial y \end{pmatrix}\Big|_{\widetilde{z}} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \varrho \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$

In other words, $\mathbb{C}$-differentiability is a SPECIAL case of differentiability where the derivative is the composition of turn an a blowing up with non-negative coefficient. $\lhd$ Just as in the classic real case, f is $\mathbb{C}$-differentiable at z iff it holds the decomposition

$$f(z+\zeta) = f(z) + f'(z)\zeta + r(\zeta), \qquad \frac{r(\zeta)}{|\zeta|} \xrightarrow[|\zeta|\to 0]{} 0.$$

In "$\mathbb{R}^2$-language" this means that

$$\widetilde{f}(\widetilde{z}+\widetilde{\zeta}) = \widetilde{f}(\widetilde{z}) + \widetilde{f'(z)}\widetilde{\zeta} + \widetilde{r}(\widetilde{\zeta}), \qquad \frac{\widetilde{r}(\widetilde{\zeta})}{\|\widetilde{\zeta}\|} \xrightarrow[\|\widetilde{\zeta}\|\to 0]{} 0;$$

here $\widetilde{f'(z)}$ denotes the linear operator $\mathbb{R}^2 \to \mathbb{R}^2$, corresponding to the operator of multiplication by $f'(z)$ (Lemma 11.1.2.). So our assertion follows from Lemma 11.1.2. $\rhd$

Cauchy-Riemann conditions

Let $G \in \mathrm{Op}(\mathbb{C})$. We say that a function $f : G \to \mathbb{C}$, $f = u + \mathrm{i}v$, satisfies *Cauchy-Riemann* (or *d'Alambert-Euler*) *conditions*, and we write

$$f \in \mathrm{CR}(G),$$

if $u, v \in \mathrm{C}^1(G)$ and

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}.$$

Analytic functions

Let $G \in \mathrm{Op}(\mathbb{C})$. We say that a function $f : G \to \mathbb{C}$, is *analytic*, and we write

$$f \in \mathrm{An}(G),$$

if $f \in \mathrm{Dif}_{\mathbb{C}}(G)$ and $f' : G \to \mathbb{C}$ is continuous.

NB It is possible to show that if f is $\mathbb{C}$-differentiable in G then f' IS AUTOMATICALLY continuous. But it is a hard theorem.

Example. $z \mapsto z^n$ is analytic in the whole $\mathbb{C}$ if $n \in \mathbb{N}$, and is analytic in $\mathbb{C} \setminus 0$ if $n \in \mathbb{Z} \setminus \mathbb{N}$.

Theorem 11.1.4. *A function $f : G \to \mathbb{C}$ is analytic iff it satisfies Cauchy-Riemann conditions:*

$$\mathrm{An}(G) = \mathrm{CR}(G).$$

$\lhd$ This follows at once from Theorem 11.1.3. $\rhd$

11.2 Complex forms

For any $M \in \mathrm{Mf}^k_\partial(\mathbb{R}^n)$ and any $p \in \{0, 1, \ldots, k\}$ we define

$$\Omega^p_{\mathbb{C}}(M)$$

as the set of pairs $(\omega_1, \omega_2) \in \Omega^p(M) \times \Omega^p(M)$ written as $\omega_1 + \mathrm{i}\omega_2$; for short we put

$$\omega + \mathrm{i}0 =: \omega.$$

We equipe this set by the following structure of vector space over $\mathbb{C}$ (below $a + \mathrm{i}b \in \mathbb{C}$):

$$(a+\mathrm{i}b)(\omega_1+\mathrm{i}\omega_2) := (a\omega_1+b\omega_2)+\mathrm{i}(b\omega_1+a\omega_2), \tag{1}$$
$$(\omega_1+\mathrm{i}\omega_2)+(\omega_3+\mathrm{i}\omega_4) := (\omega_1+\mathrm{i}\omega_3)+\mathrm{i}(\omega_2+\mathrm{i}\omega_4). \tag{2}$$

Define the *conjugate form* by the rule

$$\overline{\omega_1+\mathrm{i}\omega_2} := \omega_1-\mathrm{i}\omega_2, \tag{3}$$

and define the operations $\wedge$, d, *, $\int_M$ component-wise:

$$(\omega_1+\mathrm{i}\omega_2)\wedge(\omega_3+\mathrm{i}\omega_4) := (\omega_1\wedge\omega_3-\omega_2\wedge\omega_4)+\mathrm{i}(\omega_2\wedge\omega_3+\omega_1\omega_4), \tag{4}$$
$$\mathrm{d}(\omega_1+\mathrm{i}\omega_2) := \mathrm{d}\omega_1+\mathrm{i}\mathrm{d}\omega_2, \tag{5}$$
$$g^*(\omega_1+\mathrm{i}\omega_2) := g^*\omega_1+\mathrm{i}g^*\omega_2 \qquad (g: M\to N), \tag{6}$$
$$\int_M(\omega_1+\mathrm{i}\omega_2) := \int_M\omega_1+\mathrm{i}\int_M\omega_2. \tag{7}$$

Examples.

1. Let x and y denote, in accord with a classical tradition, the projections $(x, y) \mapsto x$ and $(x, y) \mapsto y$ in $\mathbb{R}^2$. Then

$$\begin{array}{ll} z := x+\mathrm{i}y \in \Omega^0_{\mathbb{C}}(\mathbb{R}^2), & \bar{z} = x-\mathrm{i}y \in \Omega^0_{\mathbb{C}}(\mathbb{R}^2), \\ \mathrm{d}z = \mathrm{d}x+\mathrm{i}\mathrm{d}y \in \Omega^1_{\mathbb{C}}(\mathbb{R}^2), & \mathrm{d}\bar{z} = \overline{\mathrm{d}z} = \mathrm{d}x-\mathrm{i}\mathrm{d}y \in \Omega^1_{\mathbb{C}}(\mathbb{R}^2). \end{array}$$

2. Each function $f : G \to \mathbb{C}$, $f : u + \mathrm{i}v$, $G \in \mathrm{Op}(\mathbb{C})$, with sufficiently smooth u, v may be considered as a complex 0-form:

$$f \in \Omega^0_{\mathbb{C}}(G).$$

(In the last relation we consider G as an open set in $\mathbb{R}^2$.) What means "sufficiently", depends on the context. Sometimes just continuity is sufficient.

NB All the result for "real" forms retain (as it is easy to verify) for complex ones, e.g. for $\mathbb{C}$-forms as for real ones, it holds

$$\mathrm{d}(\omega_1\wedge\omega_2) = \mathrm{d}\omega_1\wedge\omega_2+(-1)^{\deg\omega_1}\omega_1\wedge\mathrm{d}\omega_2, \tag{8}$$
$$\mathrm{d}^2 = 0, \tag{9}$$
$$\int_M \mathrm{d}\omega = \int_{\partial M}\omega \qquad \text{(for compact oriented } M\text{)}. \tag{10}$$

Theorem 11.2.1. *Let $G \in \mathrm{Op}(\mathbb{R}^2)$, and let $f, g \in \Omega^0_{\mathbb{C}}(G)$ (that is, f and g are functions $G \to \mathbb{C}$). Then the following assertions are true:*

a) $f \in \mathrm{An}(G) \Leftrightarrow \mathrm{d}(f\mathrm{d}z) = 0$;
b) $f \in \mathrm{An}(G) \Rightarrow \mathrm{d}f = f'\mathrm{d}z$;
c) $\mathrm{d}f = g\mathrm{d}z \Rightarrow f \in \mathrm{An}(G), g = f'$.

This can be summarized so:

$$\mathrm{d}(f\mathrm{d}z) = 0 \Leftrightarrow f \in \mathrm{An} \Leftrightarrow \mathrm{d}f = g\mathrm{d}z \Rightarrow g = f'. \tag{11}$$

NB The equation $\mathrm{d}(f\,\mathrm{dz}) = 0$ is equivalent to $\mathrm{d}f \wedge \mathrm{dz} = 0$ (since $\mathrm{d}^2 = 0$), and the equation $\mathrm{d}f = g\mathrm{dz}$ may be formally written as

$$\frac{\mathrm{d}f}{\mathrm{dz}} = g.$$

Both equations say that $\mathrm{d}f$ and dz are proportional ("parallel").
$\lhd$ Let $f = u + \mathrm{i}v$, $g = p + \mathrm{i}q$.

$$\begin{aligned}
\text{a) } \mathrm{d}(f\,\mathrm{dz}) &\overset{\mathrm{d}^2=0}{=} \mathrm{d}f \wedge \mathrm{dz} = (\mathrm{d}u + \mathrm{id}v) \wedge (\mathrm{d}x - \mathrm{id}y) \\
&\overset{(4)}{=} (\mathrm{d}u \wedge \mathrm{d}x - \mathrm{d}v \wedge \mathrm{d}y) + \mathrm{i}(\mathrm{d}v \wedge \mathrm{d}x + \mathrm{d}u \wedge \mathrm{d}y) \\
&\quad \boxed{\mathrm{d}u = (\partial u/\partial x)\mathrm{d}x + (\partial u/\partial y)\mathrm{d}y, \quad \mathrm{d}v = (\partial v/\partial x)\mathrm{d}x + (\partial v/\partial y)\mathrm{d}y} \\
&= \left(-\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right) + \mathrm{i}\left(\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right)\right)\mathrm{d}x \wedge \mathrm{d}y.
\end{aligned}$$

Hence

$$\mathrm{d}(f\,\mathrm{dz}) = 0 \Leftrightarrow \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = 0,\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0\right) \overset{\text{def.}}{\Leftrightarrow} f \in \mathrm{CR}(G) \overset{\text{Th.11.1.4.}}{\Leftrightarrow} f \in \mathrm{An}(G).$$

b) $f \in \mathrm{An}(G) \overset{\text{Th.11.1.4.}}{\Rightarrow} f \in \mathrm{CR}(G) \Rightarrow$

$$\begin{aligned}
\mathrm{d}f &= \mathrm{d}u + \mathrm{id}v = \left(\frac{\partial u}{\partial x}\mathrm{d}x + \frac{\partial u}{\partial y}\mathrm{d}y\right) + \mathrm{i}\left(\frac{\partial v}{\partial x}\mathrm{d}x + \frac{\partial v}{\partial y}\mathrm{d}y\right) \\
&\quad \boxed{\partial u/\partial x = \partial v/\partial y =: a, \quad \partial v/\partial x = -\partial u/\partial y =: b} \\
&= (a\mathrm{d}x - b\mathrm{d}y) + \mathrm{i}(b\mathrm{d}x + a\mathrm{d}y) \\
&\overset{(1)}{=} (a + \mathrm{i}b)(\mathrm{d}x + \mathrm{id}y) \overset{\text{Th.11.1.3.}}{=} f'(\mathrm{z})\mathrm{dz}.
\end{aligned}$$

c) $\mathrm{d}f = g\mathrm{dz} \Rightarrow$

$$((\partial u/\partial x)\mathrm{d}x + (\partial v/\partial y)\mathrm{d}y) + \mathrm{i}\,((\partial v/\partial x)\mathrm{d}x + (\partial v/\partial y)\mathrm{d}y) = (p + \mathrm{i}q)(\mathrm{d}x + \mathrm{id}y)$$

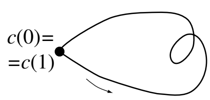
Compare $\mathcal{R}e$ and $\mathcal{I}m$

$$\begin{cases} (\partial u/\partial x)\mathrm{d}x + (\partial u/\partial y)\mathrm{d}y = p\mathrm{d}x - q\mathrm{d}y \\ (\partial v/\partial x)\mathrm{d}x + (\partial v/\partial y)\mathrm{d}y = q\mathrm{d}y + p\mathrm{d}y \end{cases}$$

Compare coeficients by $\mathrm{d}x$ and $\mathrm{d}y$

$$\begin{cases} \partial u/\partial x = \partial v/\partial y = p \\ -\partial u/\partial y = \partial v/\partial x = q \end{cases} \overset{\text{Th.11.1.3.}}{\Rightarrow} \begin{cases} f \in \mathrm{CR}(G) \\ f' = p + \mathrm{i}q = g. \end{cases} \quad \rhd$$

11.3 Integrals of complex 1-forms



Theorem 11.3.1. *Let $G \in \mathrm{Op}(\mathbb{C})$, let $c : [0,1] \to G$ be a closed curve (that is, a 1-cube with $c(0) = c(1)$), and let $f \in \Omega^0_{\mathbb{C}}(G)$ (that is, $f : G \to \mathbb{C}$). Then*

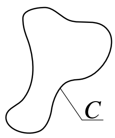
$$\int_c \mathrm{d}f = 0.$$

$\lhd$ $\int_c \mathrm{d}f = \int_{I^1} c^*\mathrm{d}f = \int_{I^1} \mathrm{d}\,\underbrace{(c^* f)}_{=f\circ c=:g} = \int_{I^1} g' = \int_{[0,1]} g' = \int_0^1 g'(t)\mathrm{d}t = g(1) - g(0) =$
$f(c(1)) - f(c(0)) = 0.$ $\rhd$

Example. For any $n \in \mathbb{Z} \setminus \{-1\}$, and for any closed curve in $\mathbb{C} \setminus \{0\}$

$$\int_c z^n \mathrm{d}z = 0.$$

$\lhd$ For $n \neq -1$ we have $z^n \mathrm{d}z = \mathrm{d} \underbrace{\left(\frac{1}{n+1} z^{n+1} \right)}_{\in \mathrm{An}(\mathbb{C} \setminus 0)}$. $\rhd$



Theorem 11.3.2. *Let* $C \in \mathrm{Comp}\,\mathrm{Or}\,\mathrm{Mf}^1(\mathbb{R}^2)$*, and let* $f \in \Omega^0_{\mathbb{C}}(C)$*. Then*

$$\int_C \mathrm{d}f = 0$$

$$\lhd \int_C \mathrm{d}f \overset{\text{Stokes Th}}{=} \int_{\partial C} f \overset{\partial M = \emptyset}{=} 0. \rhd$$

Let M be a compact 2-manifold in $\mathbb{R}^2$. Then M is orientable (verify!), and we always suppose that M is equipped with the orientation from $\mathbb{R}^2$. Further

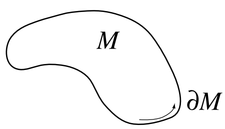
$$\mathring{M} = \mathrm{int}_{\mathbb{R}^2} M = M \setminus \partial M$$

(verify!). We say that a function $f : M \to \mathbb{C}$ is *analytic on* M, and we write

$$f \in \mathrm{An}(M),$$

if it is analytic in $\mathring{M}$ (we identify $\mathbb{R}^2$ and $\mathbb{C}$!) and is continuous on M. Thus,

$$\mathrm{An}(M) := \mathrm{An}(\mathring{M}) \cap \mathrm{C}(M).$$

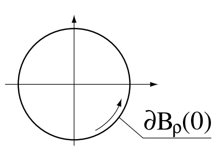


Theorem 11.3.3. (Cauchy) *Let* $M \in \mathrm{Comp}\,\mathrm{Mf}^2_{\partial}(\mathbb{R}^2)$*, and let* $f \in \mathrm{An}(M)$*. Then*

$$\int_{\partial M} f \mathrm{d}z = 0$$

$$\lhd \int_{\partial M} f \mathrm{d}z \overset{\text{Stokes Th.}}{=} \int_M \underbrace{\mathrm{d}(f \mathrm{d}z)}_{\boxed{\text{Th. 11.2.1.}} = 0} = 0. \rhd$$

If $f \notin \mathrm{An}(M)$, the integral can be non-zero, as the following example shows.



Example. Let $M = \mathrm{B}_\varrho(0)$ (the disc of radius ϱ with the center at 0). Then

$$\int_{\partial \mathrm{B}_\varrho(0)} \frac{\mathrm{d}z}{z} = 2\pi \mathrm{i}.$$

(Note that $(z \mapsto 1/z) \in \mathrm{An}(\mathbb{C} \setminus 0)$.)

$\lhd$ It is clear that

$$\int_{\partial \mathrm{B}_\varrho(0)} \frac{\mathrm{d}z}{z} = \int_c \frac{\mathrm{d}z}{z},$$

where $c : [0, 1] \to \mathbb{R}^2$, $t \mapsto (\underbrace{\varrho \cos 2\pi t}_{=:c_1}, \underbrace{\varrho \sin 2\pi t}_{=:c_2})$. But

$$\int_c \frac{\mathrm{d}z}{z} = \int_c \frac{\mathrm{d}x + \mathrm{i}\mathrm{d}y}{x + \mathrm{i}y} = \int_c \frac{(x - \mathrm{i}y)(\mathrm{d}x + \mathrm{i}\mathrm{d}y)}{x^2 + y^2} = \int_c c^* \left(\frac{(x - \mathrm{i}y)(\mathrm{d}x + \mathrm{i}\mathrm{d}y)}{x^2 + y^2} \right)$$
$$= \int_0^1 \frac{(c_1(t) - \mathrm{i}c_2(t))(c_1'(t)\mathrm{d}t + \mathrm{i}c_2'(t)\mathrm{d}t}{c_1^2(t) + c_2^2(t))}$$
$$= 2\pi\,\mathrm{i} \int_0^1 \underbrace{(\cos 2\pi t - \mathrm{i}\sin 2\pi t)(\cos 2\pi t + \mathrm{i}\sin 2\pi t)}_{=1} = 2\pi\,\mathrm{i}. \rhd$$

Lemma 11.3.4. *Let* $B := \mathrm{B}_\varrho(a)$, $f \in \mathrm{C}(\partial B)$. *Then*

$$\left| \int_{\partial B} f \mathrm{d}z \right| \le 2\pi \varrho \max_{\partial B} |f|.$$

$\lhd$ Let $a =: a_1 + \mathrm{i}a_2$. Put $c(t) := (\underbrace{a_1 + \varrho \cos 2\pi t}_{=:c_1}, \underbrace{a_2 + \varrho \sin \pi t}_{=:c_2})$. Then

$$\left| \int_{\partial B} f \mathrm{d}z \right| = \left| \int_c f \mathrm{d}z \right| = \left| \int_{I^1} (f \circ c) \underbrace{c^* \mathrm{d}z}_{=(c_1' + \mathrm{i}c_2')\mathrm{d}t} \right| = \left| \int_0^1 f(c(t))(c_1' + \mathrm{i}c_2')\mathrm{d}t \right|$$

$$\overset{\text{Exam. 11.3.5.}}{\le} \max_{0 \le t \le 1} \underbrace{|f(c(t))(c_1' + \mathrm{i}c_2'|}_{= |f(c(t))| \underbrace{|c_1' + \mathrm{i}c_2'|}_{=2\pi\varrho}} \le 2\pi\varrho \max_{\partial B} |f|. \rhd$$

Exercise 11.3.5. Prove, using Mean Value Theorem, that for any continuous function $f : [0,1] \to \mathbb{C}$

$$\left\| \int_0^1 f(t)\mathrm{d}t \right\| \le \max_{0 \le t \le 1} |f(t)|.$$

Exercise 11.3.6. More general, prove that if $C \in \mathrm{Comp\,Or\,Mf}^1(\mathbb{R}^2)$, $f \in \Omega^0_{\mathbb{C}}(C)$, then

$$\left\| \int_C f \mathrm{d}z \right\| \le \int_C |f| \mathrm{d}s.$$

11.4 Cauchy formula

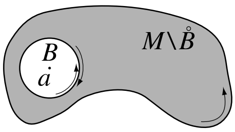
Example 11.3.6. is a special typical case of the following

Theorem 11.4.1. *Let* $M \in \mathrm{Comp\,Mf}^2_\partial(\mathbb{R}^2)$, *and let* $f \in \mathrm{An}(M)$. *Then for each point* $\overset{\circ}{M}(= M \setminus \partial M)$

$$\boxed{\int_{\partial M} \frac{f(z)\mathrm{d}z}{z - a} = 2\pi\,\mathrm{i}\, f(a)} \qquad \textit{(Cauchy Formula).}$$

In other words, the values of an analytic function inside a region are uniquely defined by the values on the boundary.

$\lhd$ Let $\varepsilon > 0$ be such that the disc $B := \mathrm{B}_\varepsilon(a)$ lies in $\overset{\circ}{M}$. Obviously $f(z)/(z - a) \in \mathrm{An}(M \setminus \overset{\circ}{M})$. Hence by Cauchy Theorem



$$0 = \int_{\partial(M\backslash\mathring{B})} \frac{f(z)\mathrm{d}z}{z-a} = \int_{\partial M} \frac{f(z)\mathrm{d}z}{z-a} - \int_{\partial B} \frac{f(z)\mathrm{d}z}{z-a}.$$

So it suffices to show that

$$\int_{\partial B} \frac{f(z)\mathrm{d}z}{z-a} \xrightarrow[\varepsilon\downarrow 0]{} 2\pi\mathrm{i}\, f(a). \tag{1}$$

Since $f \in \mathrm{Dif}_{\mathbb{C}}(a)$, we have

$$f(z) = f(a) + f'(a)(z-a) + r(z-a), \qquad \frac{r(\zeta)}{|\zeta|} \xrightarrow[|\zeta|\to 0]{} 0. \tag{2}$$

Therefore,

$$\int_{\partial B} \frac{f(z)\mathrm{d}z}{z-a} = f(a) \underbrace{\int_{\partial B} \frac{\mathrm{d}z}{z-a}}_{[1]} + \underbrace{\int_{\partial B} \underbrace{\left(f'(a) + \frac{r(z-a)}{z-a} \right)}_{=:g(z)} \mathrm{d}z.}_{[2]} \tag{3}$$

Just as in Example 11.3.6., $[1] = 2\pi\mathrm{i}$. As to [2], the function g is by (2) bounded (in the norm) in some neighbourhood of a, hence, by Theorem 11.3.1., $[2] \to 0$ as $\varepsilon \downarrow 0$, and (1) is proved. $\rhd$

Theorem 11.4.2. *Let* $M \in \mathrm{Comp\,Mf}^2_{\partial}(\mathbb{R}^2)$, $f \in \mathrm{An}(M)$. *Then* $f \in C^{\infty}_{\mathbb{C}}(\mathring{M})$, *and*

$$\forall n \in \mathbb{N}\, \forall a \in M \vdots\, f^{(n)}(a) = \frac{n!}{2\pi\mathrm{i}} \int_{\partial M} \frac{f(z)\mathrm{d}z}{(z-a)^{n+1}}. \tag{4}$$

Thus if f is just **one** time continuously $\mathbb{C}$-differentiable, it is **infinitely** $\mathbb{C}$-differentiable! $\lhd$ For $n = 1$ this this is Theorem 11.4.1.. Let it be true for $n-1$, that is,

$$\forall a \in M \vdots\, f^{(n-1)}(a) = \frac{n-1}{2\pi\mathrm{i}} \int_{\partial M} \frac{f(z)\mathrm{d}z}{(z-a)^{n}}. \tag{5}$$

Differentiation of (5) in a (which is possible as it can be shown) yields (4). $\rhd$
NB For ANY continuous function $\varphi : \partial M \to \mathbb{C}$, Cauchy formula

$$f(z) := \frac{1}{2\pi\mathrm{i}} \int_{\partial M} \frac{\varphi\,\mathrm{d}\zeta}{\zeta - z} \tag{6}$$

defines an analytic function $f : \mathring{M} \to \mathbb{C}$, but in general this f, extended to ∂M as φ, is NOT continuous! IF $\varphi = f|_{\partial M}$ for some $f \in \mathrm{An}(M)$, then Cauchy formula does reproduce the original function f.

11.5 Representation by series

We say that a series $\sum_{n=0}^{\infty} c_n$, $c \in \mathbb{C}$, *converges*, and we write

$$\sum_{n=0}^{\infty} c_n \rightsquigarrow,$$

if the sequence of partial sums $s_N := \sum_{n=0}^{N} c_n$ converges (in the norm $\|z\| = z$) to some $c \in C$, that is, if $|s_N - c| \xrightarrow[N\to\infty]{} 0$, and in such case we write $\sum_{n=0}^{\infty} c_n = c$.

Uniform convergence

Consider a FUNCTION SERIES $\sum_{n=0}^{\infty} f_n,\ f_n : \mathbb{C} \to \mathbb{C}$. If $\sum_{n=0}^{\infty} f_n \rightsquigarrow$ uniformly in $z \in A$ we write

$$\sum f_n \overset{\rightsquigarrow}{\rightrightarrows} A.$$

We say that a function series is *majorized* on $A \subset \mathbb{C}$ by a real series $\sum t_n,\ t_n \geq 0$, if $\forall z \in A\, \forall n \in \mathbb{N} \vdots\, |f_n(z)| \leq t_n$.

Lemma 11.5.1. *Let $\sum_{n=0}^{\infty} t_n$, be a* CONVERGING *real sequence, and let function series $\sum_{n=0}^{\infty} f_n,\ f_n : \mathbb{C} \to \mathbb{C}$, is majorized on $A \subset \mathbb{C}$ by $\sum t_n$. Then $\sum f_n \overset{\rightsquigarrow}{\rightrightarrows} A$.*

$\lhd\ \forall z \in A \vdots$

$$\left|\sum_{n=0}^{N} f(z) - \sum_{n=0}^{M}\right| \overset{\text{if e.g. } N > M}{=} \left|\sum_{n=M+1}^{N} f(z)\right| \leq \sum_{n=M+1}^{N} |f(z)| \leq \sum_{n=M+1}^{N} t_n \xrightarrow[N,M\to\infty]{} 0. \rhd$$

Member-wise integration and differentiation

Theorem 11.5.2. *Let $C \in \mathrm{Comp\,Or\,Mf}^1_\partial(\mathbb{R}^2)$, and let $f_n \in \Omega^0_{\mathbb{C}}(C)$ (that is, $f_n = u_n + \mathrm{i}v_n$, where u_n, v_n are (at least) continuous functions $C \to \mathbb{R}$). If $f_n \xrightarrow[n\to\infty]{} f$* UNIFORMLY *on C then*

$$\int_C f_n \mathrm{d}z \to \int_C f\,\mathrm{d}z.$$

$\lhd$ Taking into account the definition of $\int_M \omega$, this follows from the corresponding theorem for real-valued functions on $[0, 1]$. $\rhd$

Corollary 11.5.3. *Let $C \in \mathrm{Comp\,Or\,Mf}^1_\partial(\mathbb{R}^2),\ f \in \Omega^0_{\mathbb{C}}(C)$ and $\sum_{n=0}^{\infty} f_n \overset{\rightsquigarrow}{\rightrightarrows} C$. Then*

$$\int_C \left(\sum_{n=0}^{\infty} f_n\right) \mathrm{d}z = \sum_{n=0}^{\infty} \int_C f_n \mathrm{d}z.$$

Theorem 11.5.4. *Let $M \in \mathrm{Comp\,Mf}^1_\partial(\mathbb{R}^2)$, and let $f_n \in \mathrm{An}(M)$. If $f_n \xrightarrow[n\to\infty]{} f$* UNIFORMLY *on M, then also $f_n \in \mathrm{An}(M)$, and*

$$\forall a \in \overset{\circ}{M}\ \forall k \in \mathbb{N} \vdots\ f_n^{(k)}(a) \xrightarrow[n\to\infty]{} f^{(k)}(a).$$

$\lhd$ 1° $f \in \mathrm{C}(M)$ as an uniform limit of continuous function.

2° For each $a \in \overset{\circ}{M}$ we have obviously $f_n(z)/(z-a) \xrightarrow[n\to\infty]{} f(z)/(z-a)$ UNIFORMLY ON ∂M, hence

$$f(a) = \lim f_n(a) \overset{\text{Cauchy formula}}{=} \lim \frac{1}{2\pi\mathrm{i}} \int_{\partial M} \frac{f_n(z)\mathrm{d}z}{z-a} \overset{\text{Th 11.5.2.}}{=} \frac{1}{2\pi\mathrm{i}} \int_{\partial M} \frac{f(z)\mathrm{d}z}{z-a}.$$

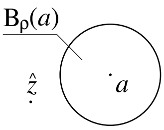
So by **NB** from 11.4, $f \in \mathrm{An}(\overset{\circ}{M})$. Thus, in view of 1°, $f \in \mathrm{An}(M)$. 3° $\forall a \in \overset{\circ}{M}\ \forall k \in \mathbb{N} \vdots$

$$f_n^{(k)}(a) \overset{\text{Th. 11.4.2.}}{=} \frac{k!}{2\pi\mathrm{i}} \int_{\partial M} \frac{f_n(z)\mathrm{d}z}{(z-a)^{k+1}} \xrightarrow{\text{Th 11.5.2.}} \frac{k!}{2\pi\mathrm{i}} \int_{\partial M} \frac{f(z)\mathrm{d}z}{(z-a)^{k+1}} \overset{\text{Th. 11.4.2.}}{=} f^{(k)}(a). \rhd$$

Corollary 11.5.5. *Let* $M \in \operatorname{Comp}\mathrm{Mf}^2_\partial(\mathbb{R}^2)$, $f_n \in \mathrm{An}(M)$, $\sum_{n=0}^{\infty} f_n \overset{\rightsquigarrow}{\rightrightarrows} M f$. *Then also* $f \in \mathrm{An}(M)$, *and*

$$\forall k \in \mathbb{N} \colon f^{(k)} = \sum_{n=0}^{\infty} f^k \quad \textit{in } \overset{\circ}{M}.$$

Convergence disc



Theorem 11.5.6. (Abel) *Let* $\hat{a}, z \in \mathbb{C}$, $\hat{a} \neq z$. *If a series* $\sum_{n=0}^{\infty} c_n(z-a)^n$ *converges for* $z = \hat{z}$ *then for each* $\varrho > 0$ *such that* $\varrho < |\hat{z} - a|$ *this series converges* UNIFORMLY *on* $\mathrm{B}_\varrho(a)$.

◁ 1° Since $\sum c_n(\hat{z} - a)^n \rightsquigarrow$, the sequence $|c_n(\hat{z} - a)^n|$ converges to zero and hence it is bounded, e.g. by $\alpha > 0$.

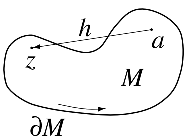
2° Now, let $z \in \mathrm{B}_\varrho(a)$. Then $|z - a|/|\hat{z} - a| \leq \varrho/|\hat{z} - a| =: k < 1$, hence

$$|c_n(z-a)^n| \leq \underbrace{|c_n(\hat{z} - a)^n|}_{\substack{1^\circ \\ \leq \alpha}} (\underbrace{|z - a|/|\hat{z} - a|}_{\leq k})^n \leq \alpha k^n.$$

Thus, our series is majorized on $\mathrm{B}_\varrho(r)$ by the converging real series $\sum \alpha k^n$, and our assertion follows from Lemma 11.5.1. ▷

It follows from this theorem that the set of all points z, where a power series $\sum c_n(z - a)^n$ converges, is either merely $\{a\}$, or it the whole $\mathbb{C}$, or lies between $\overset{\circ}{\mathrm{B}}_\varrho(a)$ and $\mathrm{B}_\varrho(a)$ for some $\varrho > 0$. In other words, the CONVERGENCE REGION is, up to the boundary, a DISC with the center at a. We call it the *convergence disc*.

Taylor formula



Lemma 11.5.7. *Let* $M \in \operatorname{Comp}\mathrm{Mf}^2_\partial(\mathbb{R}^2)$, *let* $z, a \in \overset{\circ}{M} (= M \setminus \partial M)$, $h := z - a$, *and let* $f \in \mathrm{An}(M)$. *Then for each* $n \in \mathbb{N}$

$$f(z) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a) h^k + \underbrace{\frac{h^{n+1}}{2\pi \mathrm{i}} \int_{\partial M} \frac{f(\zeta)\mathrm{d}\zeta}{(\zeta - z)(\zeta - a)^n + 1}}_{=: r_n(h)}.$$

$$\lhd\ f(z) \overset{\text{Cauchy formula}}{=} \frac{1}{2\pi \mathrm{i}} \int_{\partial M} \frac{f(\zeta)\mathrm{d}\zeta}{\zeta - z} \overset{\text{trick}}{=} \frac{1}{2\pi \mathrm{i}} \int_{\partial M} \frac{f(\zeta)\mathrm{d}\zeta}{(\zeta - a)\left(1 - \dfrac{z - a}{\zeta - a}\right)} =$$

$$\boxed{q := (z - a)/(\zeta - a) = h(\zeta - a), \quad (1 - q^{n+1})/(1 - q) = 1 + q + \ldots + q^n}$$

$$= \frac{1}{2\pi \mathrm{i}} \int_{\partial M} \frac{f(\zeta)}{\zeta - a} \left(\sum_{k=0}^{n} \left(\frac{h}{\zeta - a} \right)^n + \frac{1}{1 - \dfrac{h}{\zeta - a}} \left(\frac{h}{\zeta - a} \right)^{n+1} \right) \mathrm{d}\zeta$$

$$= \sum_{k=0}^{n} h^k \underbrace{\frac{1}{2\pi \mathrm{i}} \int_{\partial M} \frac{f(\zeta)\mathrm{d}\zeta}{(\zeta - a)^{k+1}}}_{\boxed{\text{Th.11.4.2.}} = f^{(k)}(a)/n} + \frac{h^{n+1}}{2\pi \mathrm{i}} \int_{\partial M} \frac{f(\zeta)\mathrm{d}\zeta}{(\zeta - z)(\zeta - a)^{n+1}}. \ \rhd$$

Definition. Let $G \in \mathrm{Op}(\mathbb{R}^2)$, $f \in \mathrm{An}(G)$, $a \in G$. The series $\sum_{n=0}^{\infty} 1/n! \cdot f^{(n)}(a)(z-a)^n$ is called *Taylor series* for f at a and is denoted by $\mathrm{ts}_a f$:

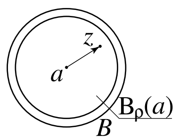
$$\mathrm{ts}_a f := \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(a)(z-a)^n.$$

Theorem 11.5.8. *Let B be a disc in $\mathbb{R}^2$ of a positive radius with center at a.*

a) *Let $f \in \mathrm{An}(\overset{\circ}{B})$. Then $\mathrm{ts}_a f \rightsquigarrow f$ in B.*

b) *Let $\sum_{n=0}^{\infty} c_n(z-a)^n \rightsquigarrow f$ in $\overset{\circ}{B}$. Then $f \in \mathrm{An}(\overset{\circ}{B})$, and $\mathrm{ts}_a f = \sum_{n=0}^{\infty} c_n(z-a)^n$.*

Thus an analytic function can be uniquely represented in each "disc of analyticity" by a power series w.r.t. to the center of the disc.



$\lhd$ a) Let $z \in \overset{\circ}{B}$, $h = z - a$. Choose $\varrho > 0$ such that $z \in \overset{\circ}{B}_\varrho(a) \subset \overset{\circ}{B}$. Then $f \in \mathrm{An}(B_\varrho(a))$ and $k := h/\varrho < 1$. Hence we have for the rest term r_n in Taylor formula

$$|r_n(h)| = \left| \frac{h^n + 1}{2\pi \mathrm{i}} \int_{\partial B_\varrho(a)} \frac{f(\zeta)\mathrm{d}\zeta}{(\zeta - z)(\zeta - a)^n} \right|$$

$$\overset{\text{Lm 11.3.4.}}{\leq} 2\pi \mathrm{i}\varrho \frac{|h|^{n+1}}{2\pi} \max_{\zeta \in \partial B_\varrho} \left| \frac{f(\zeta)}{\zeta - z} \right| \frac{1}{\varrho^{n+1}} = \mathrm{const} \cdot k^{n+1} \xrightarrow[n\to\infty]{} 0,$$

that is, $\mathrm{ts}_a f \rightsquigarrow f$ at z.

b) If $\sum c_n(z-a)^n \rightsquigarrow$ in $\overset{\circ}{B}$, then, by Abel Theorem, this series converges uniformly on each closed disc which is contained in $\overset{\circ}{B}$; hence, by Corollary 11.5.5., its sum f is analytic in $\overset{\circ}{B}$, and $\forall f \in \mathbb{N} \vdots\ f^{(k)}(a) = \sum_{n=0}^{\infty} \left.\frac{\mathrm{d}}{\mathrm{d}z^k}\right|_{z=a} (c_n(z-a)^n) = n! c_n$. It follows that $\mathrm{ts}_a f = \sum c_n(z-a)^n$. $\rhd$

NB This theorem allows to EXTEND analytic functions, if the convergence disc of the Taylor series is greater than the original disc of analyticity. These extensions can lead to DIFFERENT values at one and the same "outer" point!

11.6 Elementary functions

We DEFINE exp, sin, etc by the SAME series representation as in real analysis:

$$\mathrm{e}^z := 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \frac{z^5}{5!} + \dots, \tag{1}$$

$$\sin z := z - \frac{z^3}{3!} + \frac{z^5}{5!} - \dots, \tag{2}$$

$$\cos z := 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \dots, \tag{3}$$

$$\mathrm{sh}\, z := z + \frac{z^3}{3!} + \frac{z^5}{5!} + \dots, \tag{4}$$

$$\mathrm{ch}\, z := 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots. \tag{5}$$

Since all these series are majorized in each disc B_ϱ by the real series $\sum_{n=0}^{\infty} \varrho^n/n!$ all they converge in the whole $\mathbb{C}$ uniformly on each closed disc. So by Corollary 11.5.5., their sums are analytic functions in the whole $\mathbb{C}$.

It follows from (1)–(5) that

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} = -i\,\mathrm{sh}(iz), \qquad \mathrm{sh}\, z = \frac{e^z - e^{-z}}{2} = -i\sin(iz), \tag{6}$$

$$\cos z = \frac{e^{iz} + e^{-iz}}{2i} = \mathrm{ch}(iz), \qquad \mathrm{ch}\, z = \frac{e^z + e^{-z}}{2} = \cos iz, \tag{7}$$

and

$$e^z = \mathrm{ch}\, z + \mathrm{sh}\, z = \cos(iz) - i\sin(iz). \tag{8}$$

Putting $z := i\theta, \quad \theta \in \mathbb{R}$, we obtain

$$\boxed{e^{i\theta} = \cos\theta + i\sin\theta} \qquad (\textit{Euler formula}). \tag{9}$$

In particular

$$\boxed{\boxed{e^{i\pi} = -1.}} \tag{10}$$

This formula connect three the most fundamental numbers in mathematics, e, π and i, and is one of most beautiful mathematical formulas.

Relation (9) implies that $\varrho e^{i\theta} = \varrho\cos\theta + i\varrho\sin\theta$, which means that

$$\varrho e^{i\theta} \sim \{\varrho, \theta\} \qquad (\varrho \geq 0, \theta \in \mathbb{R}). \tag{11}$$

Further the following basic property of the exponent function remains true:

$$e^{z_1+z_2} = e^{z_1}e^{z_2}. \tag{12}$$

$\lhd$ Multiplying the absolutely converging series for e^{z_1} and e^{z_2}, we obtain

$$e^{z_1}e^{z_2} = \sum_{k=0}^{\infty}\frac{z_1^k}{k!}\sum_{l=0}^{\infty} f z_2^l l! = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\underbrace{\frac{1}{k!(n-k)!}}_{\frac{1}{n!}\binom{n}{k}} z_1^k z_2^{n-k}$$

$$= \sum_{n=0}^{\infty}\frac{(z_1+z_2)^n}{n!} = e^{z_1+z_2}. \rhd$$

Logarithm

At last we define $\ln z$ as a complex number c such that $e^c = z$. Such a number is defined non-uniquely. Indeed,

$$\forall n \in \mathbb{Z} \vdots e^{c+2\pi i n} \overset{(12)}{=} e^c \underbrace{e^{2\pi i}}_{=1} = e^c = z. \tag{13}$$

Since

$$e^{\ln\varrho + i\theta} \overset{(12)}{=} e^{\ln\varrho}e^{i\theta},$$

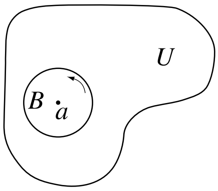
we see that if $z \sim \{\varrho, \theta\}$ then one of the values of $\ln z$ is $\ln\varrho + i\theta$. By (13), $\ln\varrho + i\theta + 2\pi i n$, $n \in \mathbb{Z}$,are also values of $\ln z$. It can be verified that there is no other ones. Thus

$$z \sim \{\varrho, \theta\} \Rightarrow \ln z = \ln\varrho + i(2\pi n + \theta). \tag{14}$$

The fact that ln is a MULTI-VALUED function is just another form of the fact that the representation $z \sim \{\varrho, \theta\}$ is non-unique.

11.7 Residues

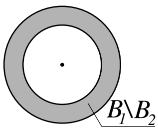
Let $a \in \mathbb{C}$, $U \in \operatorname{Op}\operatorname{Nb}_a(\mathbb{C})$, and $f \in \operatorname{An}(U \setminus \{a\})$. The *residue* of f at a is defined by the formula



$$\operatorname{res}_a f := \frac{1}{2\pi \mathrm{i}} \int_{\partial B} f(z)\mathrm{d}z, \tag{1}$$

where B is any disc with the center at a such that $B \subset U$ and $a \in \overset{\circ}{B}$.

This definition does not depend in the choice of B. Indeed if B_1 and B_2 are two different such discs, e.g. $B_1 \subset \overset{\circ}{B}_2$, then



$$0 \overset{\text{Cauchy Th}}{=} \int_{\partial(B_2 \setminus B_1)} f(z)\mathrm{d}z = \int_{\partial B_2} f(z)\mathrm{d}z - \int_{\partial B_1} f(z)\mathrm{d}z.$$

Examples.

1. If f is analytic in a neighbourbood of a, then

$$\operatorname{res}_a \frac{f(z)}{z-a} = f(a). \tag{2}$$

$\triangleleft$ This follows at once from Cauchy formula. $\triangleright$

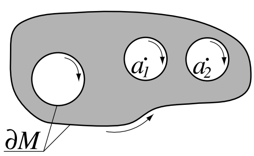
2. As a special case of previous example ($f = 1$) we obtain

$$\operatorname{res}_a \frac{1}{z-a} = 1. \tag{3}$$

3. For $k = 2, 3, \ldots$

$$\operatorname{res}_a \frac{1}{(z-a)^k} = 0. \tag{4}$$

(See Example on page 150.)



Theorem 11.7.1. (on residues) *Let $M \in \operatorname{Comp}\operatorname{Mf}^2_\partial(\mathbb{R}^2)$, let $a_1, \ldots, a_n \in \overset{\circ}{M} (= M \setminus \partial M)$, and let $f \in \operatorname{An}(M \setminus \{a_1, \ldots, a_n\})$. Then*

$$\int_{\partial M} f\,\mathrm{d}z = \sum_{k=1}^{n} \operatorname{res}_{a_k} f.$$

$\triangleleft$ Let B_k be mutually disjoint discs with the centers at a_k such that $B_k \subset \overset{\circ}{M}$. Then

$$0 \overset{\text{Cauchy Th}}{=} \int_{\partial(M \setminus \bigcup \overset{\circ}{B}_k)} f\,\mathrm{d}z = \int_{\partial M} f\,\mathrm{d}z - \sum \underbrace{\int_{\partial B_k} f\,\mathrm{d}z}_{2\pi \mathrm{i} \cdot \operatorname{res}_{a_k} f}$$

(note that ∂B_k as a part of $\partial(M \setminus \bigcup \overset{\circ}{B}_k)$ has the orientation opposite to orientation of ∂B_k itself). $\triangleright$

This theorem reduces calculation of integrals to calculation of residues.

Calculation of residues

A point $a \in \mathbb{C}$ is a *pole* of degree n (or n-*pole*), $n \in \mathbb{N}$, of a function $f : U \setminus a \to \mathbb{C}$, ($U \in \operatorname{Op}\operatorname{Nb}_a(\mathbb{C})$) if the function $(z-a)^n f$ admits an analytic extension to U, but the function $(z-a)^{n-1} f$ does not.

Examples.

1. 0 is a 1-pole for $1/z$, is a 2-pole for $1/z^2$, etc.
2. $1/\sin z$ has 1-poles at $z = 0, \pm\pi, \pm 2\pi, \ldots$.

Theorem 11.7.2.

a) *Let a be an n-pole for f. The there exist a closed disc B with the center at a such that $\mathring{B} \neq \emptyset$ and f can be (uniquely) represented in $\mathring{B} \setminus a$ by some series*

$$\sum_{k=-n}^{\infty} c_k (z-a)^k, \qquad (c_k \in \mathbb{C}),$$

called the Laurent series for f at a; that is,

$$f = \sum_{k=-n}^{\infty} c_k (z-a)^k \qquad \textit{in } \mathring{B} \setminus a.$$

b) *The residue of f at a is equal to the coefficient by $(z-a)^{-1}$:*

$$\boxed{\operatorname{res}_a f = c_{-1}.}$$

c) *This residue can be calculated by differentiation:*

$$\begin{aligned} \operatorname{res}_a f &= \frac{1}{(n-1)!} \left.\frac{\mathrm{d}^{n-1}}{\mathrm{d}z^{n-1}}\right|_a ((z-a)^n f) \\ &= \frac{1}{(n-1)!} \lim_{z \to a} \frac{\mathrm{d}^{n-1}}{\mathrm{d}z^{n-1}} \big((z-a)^n f(z)\big). \end{aligned}$$

◁ EXERCISE for you. [*Hint*: Apply Theorem 11.5.8. on Taylor series to $(z-a)^n f$ and then use Examples 2 and 3 on page 157.] ▷

Chapter 12

Ordinary differential equations

12.1 Analytic setting

A *differential equation* is an equation of the form

$$F\left(x, u(x), u'(x), u''(x), \ldots, u^{(p)}(x)\right) = 0, \tag{1}$$

where

$$u : X \to Y,$$

$$u' : X \to \mathscr{L}(X, Y)$$

$$\vdots$$

$$u^{(p)} : X \to \mathscr{L}(\underbrace{X, \ldots, X}_{p}; Y),$$

and

$$F : X \times Y \times \mathscr{L}(X, Y) \times \ldots \times \mathscr{L}(X, \ldots, X; Y) \to Z.$$

Here X, Y, Z are (say) normed spaces; U is the *unknown function,* and (1) is to be fulfilled at each point x of some open set $U \subset X$.

The order p of the highest derivative in (1) is called *order* of the equation.

Classification

If $X = \mathbb{R}$ we have an *ordinary* differential equation (ODE);
If $X = \mathbb{R}^n$ we have a *partial* differential equation (PDE);
If $Y = \mathbb{R}$ we have a *scalar* differential equation;
If $Y = \mathbb{R}^n$ we have a *vector* differential equation;
If $Z = \mathbb{R}$ we have ONE differential equation;
If $Z = \mathbb{R}^n$ we have a *system* of differential equations.

ODE's

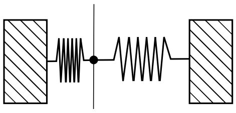
In ODE's the unknown function is a function of ONE independent variable, which physically can be interpreted as TIME (and is denoted usually by t), and (1) can be interpreted as a LAW of some process, of some evolution.

The derivatives in t are often denoted (after Newton) by dots above, e.g.,

$$\dot{x} := \frac{\mathrm{d}x}{\mathrm{d}t}, \qquad \ddot{x} := \frac{\mathrm{d}^2x}{\mathrm{d}t^2}.$$

Examples.

1. Inertial motion equation: $\ddot{x} = 0$, $x : \mathbb{R} \to \mathbb{R}^3$; this eqaution describes a free motion of a point in the space (no external forces).



2. Pendulum equation: $\ddot{x} + x = 0$, $x : \mathbb{R} \to \mathbb{R}$; this equation describes the motion of a point, on which the force act, that is proportional to the displacement of the point, from an equilibrium position and tends to return the point back.

3. Exponent equation: $\dot{x} = kx$, $x : \mathbb{R} \to \mathbb{R}$; it describes a process where the speed of grown of something is proportional to the present quantity of this something; e.g., for $k > 0$ it may be a chain reaction, for $k < 0$ it may be a nuclear decay.

Thus, a general ODE looks as follows (now we write t instead of x, and x instead of u, respectively X instead of Y):

$$F\left(t, x, \dot{x}, \ddot{x}, \dots, x^{(n)}\right) = 0. \tag{2}$$

Since all the derivatives of $x : \mathbb{R} \to X$ are again functions $\mathbb{R} \to X$, now we have

$$F : \mathbb{R} \times X^{n+1} \to Z.$$

A *solution* of (2) is an n times differentiable function $\varphi : (a, b) \to X$ (where (a, b) is an non-empty interval in $\mathbb{R}$, such that

$$\forall t \in (a, b) \colon \quad F\left(t, \dot{\varphi}(t), \ddot{\varphi}(t), \dots, \varphi^{(n)}(t)\right) = 0.$$

NB Not every equation, containig derivatives in t, is an ODE. E.g., the equation $\ddot{x}(\dot{x} = 0)$ is NOT.

Reduction to a first order equation

We shall consider ONLY ODE's SOLVED w.r. to the highest derivative:

$$\frac{\mathrm{d}^n x}{\mathrm{d}t^n} = F\left(t, x, \dot{x}, \ddot{x}, \dots, x^{(n-1)}\right) = 0. \tag{3}$$

Theorem 12.1.1. *Equation (3) is equivalent to the following system of n ODE's of the first order:*

$$\left.\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= F(t, x_1, x_2, \dots, x_n) \end{aligned}\right\} \tag{4}$$

$\lhd$ If $\varphi : (a, b) \to X$ is a solution of (3), then

$$\left(\varphi, \dot{\varphi}, \ddot{\varphi}, \dots, \varphi^{(n-1)}\right)$$

is a solution of (4); v.v., if

$$(\varphi_1, \varphi_2, \dots, \varphi_n) : (a, b) \to X^n$$

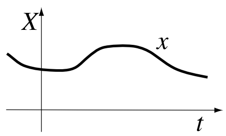
is a solution of (4), then φ_1 is a solution of (3). $\triangleright$

By this theorem we can restrict ourselves by ODE's of the first order

$$\boxed{\dot{x} = f(x, t)}, \qquad x : \mathbb{R} \to X, \qquad f : X \times \mathbb{R} \to X.$$

12.2 Geometric setting

Geometrically, the unknown function $x : \mathbb{R} \to X$ in an ODE $\dot{x} = f(x, t)$ is a CURVE in X. The space X is called the *phase space* of the equation. The graph of a solution φ is called an *integral curve* of the equation. The space $X \times \mathbb{R}$, where this graph lies, is called the *extend phase space*.



Examples.

1. Inertial motion equation $\ddot{x} = 0$ ($x : \mathbb{R} \to \mathbb{R}^3$) is equivalent to the system

$$\left.\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= 0 \end{aligned}\right\} \qquad (x_1, x_2 : \mathbb{R} \to \mathbb{R}^3),$$

which can be written as $\dot{x} = Ax$, where $x : \mathbb{R} \to \mathbb{R}^6$, and $A \in \mathscr{L}(\mathbb{R}^6, \mathbb{R}^6)$, viz.

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Here the phase space is $\mathbb{R}^6$ (position-velocity).

2. Pendulum equation $\ddot{x} + \dot{x} = 0$ is equivalent to the system

$$\left.\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -x_1 \end{aligned}\right\}$$

which can be written as $\dot{x} = Ax$, where $x : \mathbb{R} \to \mathbb{R}^2$,

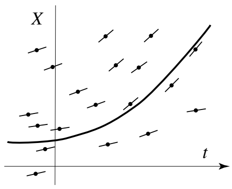
$$A = \begin{pmatrix} 0 & +1 \\ -1 & 0 \end{pmatrix} \in \mathscr{L}(\mathbb{R}^2, \mathbb{R}^2)$$

(the operator of rotation by 90° ↻) The phase space is $\mathbb{R}^2$.

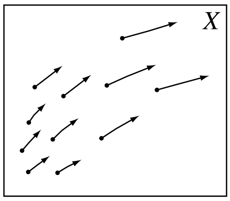
3. For the exponent equation $\dot{x} = kx$ the phase space is $\mathbb{R}$.

Fields corresponding to ODE's

A geometric interpretation of the equation $\dot{x} = f(x,t)$ is such. At each point of the extended phase space a DIRECTION is given (the derivative of a curve at a point "is" the tangent line to the graph). To find a solution of the equation means to find an integral curve of this *field of directions*, that is, to find a curve, such that at each point the tangent line coincides with the directions of the field at this point.

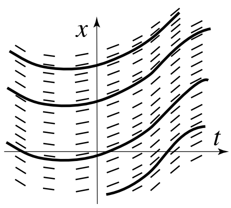


In the special case where f does NOT depend on the t (*autonomic* ODE's) another interpretation is possible, in the phase space itself. Viz., at each point x of the phase space a VECTOR $f(x)$ is given (the derivative can be considered as a vector in X (the velocity)). That is, we have a VECTOR FIELD. To find a solution of the equation $\dot{x} = f(x)$ means to find a motion of a point in the phase space such that the velocity at each moment of time coincides with the value of the vector field at the point where we come at this moment.



Examples.

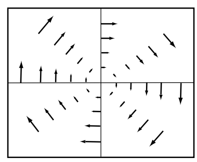
1. The simplest equation $\dot{x} = f(x)$, $x : \mathbb{R} \to \mathbb{R}$, $f : \mathbb{R} \to \mathbb{R}$; the field of directions does not depend on x. The solution is unique up to vertical translation of the graph. We know this already, of course:



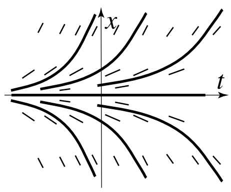
$$x(t) = \int_{x_0}^{t} f(\tau)\mathrm{d}\tau + \text{const} \qquad (\text{for any } t_0 \in I).$$

(we suppose that f is sufficiently "nice".)

2. $\ddot{x} + x = 0$; $(\dot{x}_1, \dot{x}_2) = (x_2, -x_1)$. The velocity field is drown on the picture. Obviously the solutions are motions along circles with center at 0, e.g., $(\cos t, -\sin t)$. Hence, the corresponding solutions of the original equation $\ddot{x} + x = 0$ are $\cos t$ and $\sin t$.



3. $\dot{x} = kx$. The field of directions for $k > 0$ looks as on the picture. Up to a horizontal translation of the graph there are just 3 solutions. Of course, we know them:



$$x = x_0 e^{kt}.$$

For $x_0 > 0$, $= 0$ or < 0 we obtain the 3 types of solutions.

Initial conditions

In general the function f in the right-hand side of our equation

$$\dot{x} = f(x,t) \tag{1}$$

is defined only on an open subset Ψ of $X \times \mathbb{R}$. We say that $\varphi : \mathbb{R} \to X$ is a *solution* of (1) in an interval (a,b) $(-\infty \le a < b \le +\infty)$, and we write

$$\varphi \in \mathrm{Sol}_{(a,b)}(1),$$

if $\varphi \in \mathrm{Dif}((a,b))$, $\{(\varphi(t),t) | tt \in (a,b)\} \subset \Psi$ and $\forall t \in (a,b) \vdots\ \dot{\varphi}(t) = f(\varphi, t)$.

We say that φ is a *solution* of (1), and we write

$$\varphi \in \mathrm{Sol}(1),$$

if $\varphi \in \mathrm{Sol}_{(a,b)}(1)$ for some a, b.

Let $(x_0, t_0) \in \Psi$. We say that φ is a *solution* of (1) (in (a, b)) *with the initial condition* (x_0, t_0), and we write

$$\varphi \in \mathrm{Sol}(1)_{t_0,x_0} \qquad (\text{resp., } \varphi \in \mathrm{Sol}_{(a,b)}(1)_{t_0,x_0})$$

if $\varphi \in \mathrm{Sol}(1)$ (resp., $\varphi \in \mathrm{Sol}_{(a,b)}(1)$) and

$$\varphi(t_0) = x_0.$$

This condition means physically that at given initial moment t_0 our process has value x_0, and means geometrically that the graph of φ, the integral curve, passes through the point (x_0, t_0).

12.3 Basic Theorem

Here we discuss the questions of existence and uniqueness of a solution of an equation

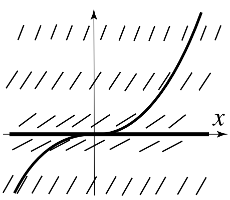
$$\dot{x} = f(x, t), \qquad x : \mathbb{R} \to X, \tag{1}$$

with a given initial condition, and dependence of the solution on the initial condition.

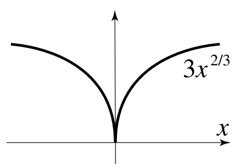
Theorem 12.3.1. (Peano) *Let* $X = \mathbb{R}^n$, $\Psi \in \mathrm{Op}(\mathbb{R}^n \times \mathbb{R})$. *If* $f \in \mathrm{C}(\Psi, X)$, *then*

$$\forall (x_0, t_0) \in \Psi\ \exists \varphi \in \mathrm{Sol}(1)_{t_0,x_0}.$$

In other words, if f is continuous then a solution of (1) always exists. (We omit the proof of this theorem). But in general a solution with a given initial condition is NOT unique, as the following example shows.



Example. The equation $\dot{x} = 3x^{2/3}$, $x : \mathbb{R} \to \mathbb{R}$, has 2 solution with the initial condition $(0, 0)$: $\varphi_1(t) = 0$ and $\varphi_2(t) = t^3$. The reason is that the right-hand side $x^{2/3}$ decrease too quickly as $x \to 0$:



If f is of CLASS C^1 in x, such a patology cannot occur, as we shall see.

Picard method

Consider the equation (1) for $X = \mathbb{R}$, $f \in \mathrm{C}(\mathbb{R} \times \mathbb{R}, \mathbb{R})$. The main idea of the method is: to find a solution of (1) means just to find a fixed point of some operator. More precisely:

$$\varphi \in \mathrm{Sol}(1)_{t_0,x_0} \Leftrightarrow \varphi \in \mathrm{Fix}\, A,$$

where A is defined by the formula

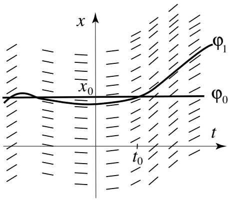
$$(A\varphi) := x_0 + \int_{t_0}^{t} f(\varphi(\tau), \tau)\mathrm{d}\tau. \tag{2}$$

$\lhd$ $(A\varphi)^\cdot(t) \overset{(2)}{=} f(\varphi(t),t)$; if $\varphi \in \mathrm{Sol}(1)_{t_0,x_0}$, that is, if $\dot\varphi = f(\varphi(t),t)$ and $\varphi(t_0) = x_0$, then $(A\varphi)^\cdot = \dot\varphi$ and $(A\varphi)(t_0) \overset{(2)}{=} x_0 = \varphi(t_0)$; hence $A\varphi = \varphi$, that is, $\varphi \in \mathrm{Fix}\,A$. V.v., if $\varphi \in \mathrm{Fix}\,A$, that is, $A\varphi = \varphi$, then $\dot\varphi = (A\varphi)^\cdot \overset{(2)}{=} f(\varphi(t),t)$ and $\varphi(t_0) = (A\varphi)(t_0) \overset{(2)}{=} x_0$. $\rhd$

Picard method is to construct a solution of (1) as a fixed point of A, that is, as a limit of sequence φ_0, $\varphi_1 := A\varphi_0$, $\varphi_2 := A\varphi_1, \ldots$, where φ_0 is an initial approximation to the solution.

Examples.

1. $\dot x = f(t)$, $x(t_0) = x_0$. Field of directions does not depend on x. Put



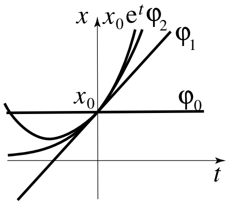
$$\varphi_0 :\equiv x_0.$$

Then already the FIRST approximation

$$\varphi_1(t) = (A\varphi)(t) = x_0 + \int_{t_0}^{t} f(\tau)\mathrm{d}\tau$$

yields the solution $\dot\varphi = f$, $\varphi_1(t_0) = x_0$.

2. $\dot x = x$, $x(0) = x_0$. Again put $\varphi_0 :\equiv x_0$. Then $\varphi_1(t) = x_0 + \int_{t_0}^{t} x_0\mathrm{d}\tau = x_0(1+t)$, $\varphi_2(t) = x_0 + \int_{t_0}^{t} x_0(1+t)\mathrm{d}\tau = x_0 + (1+t+t^2/2), \ldots, \varphi_n(t) = x_0(1+t+t^2/2+\ldots+t^n/n!)$, so that $\varphi_n(t) \to x_0\mathrm{e}^t$. And $x_0\mathrm{e}^t$ is indeed the solution.



To justify Picard method we shall show that in an appropriate space the Picard operator A is a contraction. We need for this end to define integral of a VECTOR function of real variable.

Integrals of vector functions

Let X be a Banach space (e.g., $\mathbb{R}^n$), and let $f \in \mathrm{C}(\mathbb{R}, X)$. The *integral*

$$\int_a^b f(t)\mathrm{d}t \qquad (\in X)$$

is defined just as usually (by means of partial sums).

Lemma 12.3.2.

$$\left\| \int_a^b f(t)\mathrm{d}t \right\| \le \left| \int_a^b \|f(t)\|\mathrm{d}t \right|.$$

(Here it may be $a > b$!)

$\lhd$ It follows from the corresponding inequality for partial sums:

$$\left\| \sum f(t_i)\Delta_i \right\| \le \sum \|f(t_i)\Delta_i\| = \left| \sum \|f(t_i)\|\Delta_i \right|. \rhd$$

Lemma 12.3.3.

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=\hat t} \int_a^b f(\tau)\mathrm{d}\tau = f(\hat t).$$

$\lhd$ Just as usually. $\rhd$

Lemma 12.3.4. (Newton-Leibniz Theorem). *Let* $\varphi \in \mathrm{C}^1(\mathbb{R}, X)$. *Then*

$$\int_a^b \dot\varphi(t)\mathrm{d}t = \varphi(b) - \varphi(a).$$

◁ As usually. ▷

Basic Theorem

Theorem 12.3.5. (on existence, uniqueness and continuous dependence on initial condition). *Let* $X \in \mathrm{BS}$, $W \in \mathrm{Op}(x \times \mathbb{R})$, $f \in \mathrm{C}^1(W, X)$, $(x_0, t_0) \in W$, *and we consider the equation*

$$\dot{x} = f(x, t). \tag{3}$$

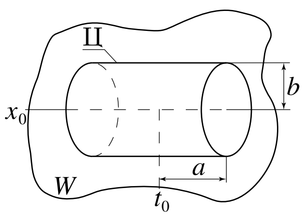
Then there exist an open interval I *with the center at* t_0 *and an open ball* B *in* X *with the center at* x_0*, such that*

$$\forall x \in B\ \exists! \varphi_x \in \mathrm{Sol}_I(3)_{t_0, x_0}$$

and for any CLOSED *interval* $J \subset I$ *the mapping*

$$x \mapsto \varphi_x|_J,\ B \to \mathrm{C}(J, X)$$

is continuous.



◁ 1° At first we construct a subspace M of of a Banach space, where a modification of Picard operator is a contraction Take $a > 0$, $b > 0$ such that the cylinder

$$\text{П} := \mathrm{B}_b(x_0) \times \mathrm{I}_a(t_0)$$

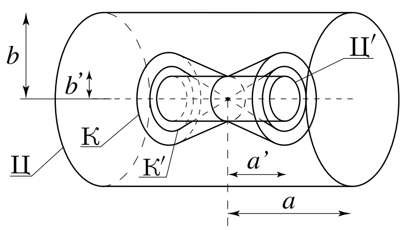
lies in W. (We denote by $\mathrm{I}_a(t)$ the closed ball $\mathrm{B}_a(t)$ in $\mathbb{R}$.) Put

$$S := \sup_{(x,t) \in \text{П}} \| f(x, t) \|, \tag{4}$$

$$L := \sup_{(x,t) \in \text{П}} \| \mathrm{D}_1 f(x, t) \| \tag{5}$$

(where $\mathrm{D}_1 f \equiv \partial f / \partial x : \Psi \to \mathscr{L}(X, X)$). These supremums are finite and attained since П is compact.

Now choose $a' > 0$ and $b' > 0$ so that the cone $\text{К}' := \{(x, t) : |t - t_0| \le a', \|x - x_0\| \le S|t - t_0|\}$ and all its translations by the vectors $(b, 0)$, $b \in \mathrm{B}_{b'}(x_0)$, lies in the cylinder П:



$$\text{К} := \text{К}' + (\mathrm{B}_{b'}(x_0) \times \{0\}) \subset \text{П}. \tag{6}$$

Consider the "small" cylinder

$$\text{П}' := \mathrm{B}_{b'}(x_0) \times \mathrm{I}_{a'}(t_0) \subset \text{П}.$$

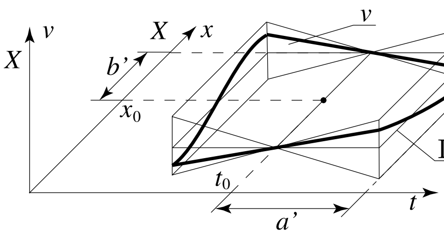
We put

$$\mathrm{M} := \{ v \in \mathrm{C}(\text{П}, X) | \forall (x, t) \in \text{П}' \colon \|v(x, t)\| \le S|t - t_0| \}.$$

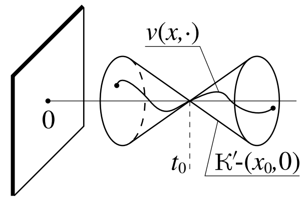
In particular

$$\forall v \in \mathrm{M} \colon v(\cdot, t_0) = 0. \tag{7}$$

An element of M is shown on the following two pictures (note that on the left picture the phase space X is represented by a LINE, and on the right one by a PLAIN):



The graph of v.



The graph of $v(x,\cdot)$ for fixed x.

Emphasize that M depends on a', b', S.

2° Now define a modified Picard operator A on M by the formula

$$(Av)(x,t) := \int_{t_0}^{t} f(x+v(x,\tau),\tau)\mathrm{d}\tau \qquad ((x,t)\in \Pi'). \tag{8}$$

This definition is correct since, by (6), the argument of f lies in Ψ.

3° *A maps* M *into itself.* $\lhd\!\lhd\ \forall (x,t)\in\Pi'\ \vdots$

$$\begin{aligned}\|Av(x,t)\| &= \left\|\int_{t_0}^{t} f(x+v(x,\tau),\tau)\mathrm{d}\tau\right\| \\ &\overset{\text{Lm12.3.2.}}{\le} \left\|\int_{t_0}^{t} \underbrace{\|f(x+v(x,\tau),\tau)\|}_{\le S}\mathrm{d}\tau\right\| \le S|t-t_0|.\ \rhd\!\rhd\end{aligned}$$

4° *A is a contraction for sufficiently small* a'. Indeed $A\in \mathrm{Lip}_{La'}$, where L is from (5). $\lhd\!\lhd\ \forall v_1, v_2\in\mathrm{M}\ \vdots$

$$\|Av_1-Av_2\| = \sup_{(x,t)\in\Pi'} \|\underbrace{\underbrace{(Av_1-Av_2)(x,t)}_{\overset{(8)}{=}\int_{t_0}^{t}(f(x+v_1(x,\tau),\tau)-f(x+v_2(x,\tau),\tau))\mathrm{d}\tau}}_{\boxed{1}}\| \le \sup_{(x,t)\in\Pi'}\left|\int_{t_0}^{t} \underbrace{\left\|\boxed{1}\right\|}_{\overset{\text{MVT}}{\le} L\underbrace{\|v_1(x,\tau)-v_2(x,\tau)\|}_{\overset{\text{obv.}}{\le}\|v_1-v_2\|}} \mathrm{d}\tau\right|$$

$$\overset{\text{Lm 12.3.2.}}{\le} \sup_{(x,t)\in\Pi'} L\underbrace{|t-t_0|}_{\le a'}\|v_1-v_2\| \le La'\|v_1-v_2\|.\ \rhd\!\rhd$$

5° By Fixed point Theorem (= Contraction Lemma) there exists $v\in\mathrm{M}$ such that $Av=v$. Put

$$u(x,t) := x+v(x,t) \qquad ((x,t)\in\Pi').$$

For any given $x\in \mathrm{B}_{b'}(x_0)$ *(an initial value) we have*

$$u(x,\cdot)\in \mathrm{Sol}(3)_{t_0,x_0}.$$

$$\begin{aligned}\lhd\!\lhd\ \frac{\mathrm{d}}{\mathrm{d}t}u(x,t) &= \frac{\mathrm{d}}{\mathrm{d}t}(x+\underbrace{v(x,t)}_{=Av(x,t)}) \\ &\overset{(8)}{=} \frac{\mathrm{d}}{\mathrm{d}t}\left(x+\int_{t_0}^{t} f(x+v(x,\tau),\tau)\mathrm{d}\tau\right) \overset{\text{Lm 12.3.3.}}{=} f(x+v(x,\tau),\tau) \\ &= f(u(x,\tau),\tau),\end{aligned}$$

and $u(x,t_0) = x+\underbrace{v(x,t_0)}_{(7)=0} = x$. $\rhd\!\rhd$

6° Our solution depends continuously on the initial value x since υ is a continuous mapping.
7° Uniqueness: Take $b' := 0$, and consider the corresponding set M and operator A. (Now M consists from functions defined just on $\mathrm{I}_{a'}(t_0)$.) Obviously,

$$\varphi \in \mathrm{Sol}(3)_{t_0,x_0} \Leftrightarrow \varphi - x_0 \in \mathrm{Fix}\, A,$$

but the fixed point of A is unique (by Contraction Lemma). Hence (on the interval $\overset{\circ}{\mathrm{I}}_{a'}(t_0)$) the solution is unique. $\triangleright$

12.4 Methods of solutions

As Liouville showed, in general it is impossible to solve a given ODE in explicite form (in "quadratures"), that is, in form of finite combination of elementary and algebraic functions and of integrals of them. E.g., such a simple equations as $\mathrm{d}y/\mathrm{d}x = y^2 - x$ cannot be solved in quadratures.

There are general methods of APPROXIMATIVE solution of ODE's, in particular methods based on Picard approximations.

Rather full theory of explicite solution is only for LINEAR ODE's, which we consider in the next two sections.

Here we discuss special but important case where solutions can be calculated rather explicitly.

No dependence on x

$$\dot{x} = f(t).$$

This equation describes a process, the speed of which does not depend on its state, but is fully determined "from outside". The solution satisfying an initial condition $x(t_0) = x_0$ is given by the "classic" formula of analysis

$$x(t) = x_0 + \int_{t_0}^{t} f(\tau)\mathrm{d}\tau.$$

No dependence on t

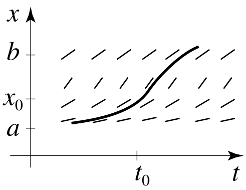
$$\dot{x} = f(x) \qquad (X = \mathbb{R}). \tag{1}$$

This equation describes an "automatic" process where the behavior of the process is defined entirely by its present state.

Theorem 12.4.1. *Let* $f \in \mathrm{C}^1((a,b))$, $x_0 \in (a,b)$, $f(x_0) \neq 0$ $(-\infty \leq a < b \leq +\infty)$. *Then for any* $t_0 \in \mathbb{R}$ *the solution* φ *of equation (1) with initial condition* (x_0, t_0) *(which does exist by Basic Theorem) satisfies the relation*

$$t - t_0 = \int_{x_0}^{\varphi(t)} \frac{\mathrm{d}\xi}{f(\xi)}. \tag{2}$$

In other words, our solution $x = \varphi(t)$ can be found by solving of the equation



$$t - t_0 = \int_{x_0}^{x} \frac{\mathrm{d}\xi}{f(\xi)}$$

w.r.t. to x.
$\lhd$ Let $\varphi \in \mathrm{Sol}(1)_{t_0,x_0}$. Then $\dot\varphi(t_0) = f(x_0) \neq 0$. By inverse Function Theorem, the inverse function $\varphi^{-1} =: \psi$ is defined locally (in a neighbourhoodof x_0).
We have $\psi(x_0) = t_0$, and

$$\left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{\xi} = \frac{1}{f(\xi)}.$$

Since $f(x_0) \neq 0$, the function $1/f(\xi)$ is continuous in a neighbourhood of x_0. By Newton-Leibniz Theorem,

$$\psi(x) - \psi(x_0) = \int_{x_0}^{x} \frac{(\mathrm{d}\xi)}{f(\xi)}.$$

Putting here $x = \varphi(t)$, we obtain (2). $\rhd$

Separable variables

$$\dot x = \frac{g(x)}{f(t)} \qquad (X = \mathbb{R}).$$

Here x and t enter "separately". For better symmetry let us write y instead of x and x instead of t:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{g(y)}{f(x)}. \tag{3}$$

Theorem 12.4.2. *Let f and g are of class C^1 in some neighbourhoods of points x_0 and y_0, resp.; let $f(x_0) \neq 0$, $g(y_0) \neq 0$, and let $y = F(x)$ be a solution of (3) with the initial condition $F(x_0) = y_0$ (such solution does exist by Basic Theorem). Then F is given implicitly by the equation*

$$\int_{x_0}^{x} \frac{\mathrm{d}\xi}{f(\xi)} = \int_{y_0}^{y} \frac{\mathrm{d}\eta}{g(\eta)}.$$

Thus the solution method is: to rewrite (3) formally as $\mathrm{d}x/f(x) = \mathrm{d}y/g(y)$ and to integrate in the corresponding limits.
$\lhd$ Consider two new ODE's

$$\dot x = f(x), \tag{4}$$

$$\dot y = g(y). \tag{5}$$

By Basic Theorem, there exist $\varphi \in \mathrm{Sol}(4)_{0,x_0}$, $\psi \in \mathrm{Sol}(5)_{0,y_0}$, defined on a (w.l.o.g.) COMMON open interval I:

$$\dot\varphi(t) = f(\varphi(t)), \qquad \varphi(0) = x_0, \tag{6}$$

$$\dot\psi(t) = f(\psi(t)), \qquad \psi(0) = x_0. \tag{7}$$

We have $\dot\varphi(0) = f(x_0) \neq 0$, $\dot\psi(0) = g(y_0) \neq 0$. By Inverse Function Theorem, there exist (locally) the inverse functions, φ^{-1}. We claim that

$$\begin{tikzcd} & t \arrow[dl, shift left, "\varphi^{-1}"] \arrow[dr, "\psi"] & \\ x \arrow[ur, shift left, "\varphi"] \arrow[rr, mapsto, "u"'] & & y \end{tikzcd}$$

$$u := \psi \circ \varphi^{-1} \in \mathrm{Sol}(3)_{x_0,y_0}.$$

Indeed

$$\dot u(x) \underset{t:=\varphi^{-1}(x)}{=} \dot\psi(t)(\varphi^{-1})^{\cdot}(x) = \frac{\dot\psi(t)}{\dot\varphi(t)} \underset{y:=u(x)=\psi(t)}{\overset{(6),(7)}{=}} \frac{g(y)}{f(x)},$$

$$u(x_0) = \psi(\varphi^{-1}(x_0)) \overset{(6)}{=} \psi(0) \overset{(7)}{=} y_0.$$

But by Theorem 12.4.1.,

$$\int_{x_0}^{\varphi(t)} \frac{d\xi}{f(\xi)} = t - t_0 = \int_{y_0}^{\psi(t)} \frac{d\eta}{g(\eta)}.$$

Putting $\varphi(t) = x$, $\psi(t) = y$, we see that $u = \psi \circ \varphi^{-1} : x \mapsto y$, and

$$\int_{x_0}^{x} \frac{d\xi}{f(\xi} = \int_{y_0}^{y} \frac{d\eta}{g(\eta)},$$

which is what we need. $\rhd$

12.5 Linear equations

By a *linear* (*homogenious*) *ODE* we mean an equation

$$\dot x = A(t)x \qquad (x : I \to X,\ X \in \mathrm{NS}), \tag{1}$$

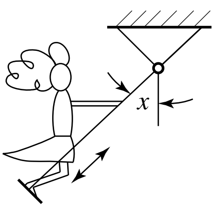
where $A(t)$ for each $t \in I$ is a continuous LINEAR operator in X, and the mapping

$$A : I \to \mathscr{L}(X, X)$$

is sufficiently smooth. Thus the right-hand side of (1) is linear (and continuous) in x. In the case $X = \mathbb{R}^n$ the equation (1) takes the form

$$\left.\begin{aligned} \dot x_1 &= a_{11}(t)x_1 + \ldots + a_{1n}(t)x_n \\ &\vdots \\ \dot x_n &= a_{n1}(t)x_1 + \ldots + a_{nn}(t)x_n \end{aligned}\right\} \tag{2}$$

so usually one calls (1) a linear ODE *with variable coefficients*.



Example. Pendulum of variable length: $\ddot x = -\omega^2(t)x$, $(x : \mathbb{R} \to \mathbb{R})$. This equation when written in the form (2) is

$$\left.\begin{aligned} \dot x_1 &= x_2 \\ \dot x_2 &= -\omega^2(t)x_1 \end{aligned}\right\} \tag{3}$$

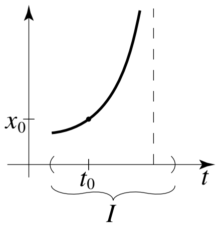
In the form (1) it looks as

$$\dot x = A(t)x, \qquad \text{where } A(t) = \begin{pmatrix} 0 & 1 \\ -\omega^2(t) & 0 \end{pmatrix},\ x : \mathbb{R} \to \mathbb{R}^2.$$

A very pleasant feature of linear equations is that they have solutions defined in the whole interval I:

Theorem 12.5.1. *Any solution of (1) can be extended to I.*
$\lhd$ The idea of the proof is such. Since on any COMPACT subinterval J of I the norm $\|A\|$ is bounded (as a continuous function), we have $\|\dot x\| = \underbrace{\|A(t)x\|}_{\le C} \|x\| \le C\|x\|$. on J. It

follows that any solution grows not faster than e^{Ct} (in the norm) on J, and hence cannot go away to infinity on J. An accurate proof see e.g., in [9, p. 196]. $\rhd$

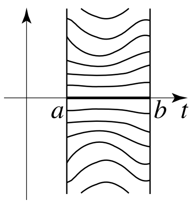


NB For non-linear equations it can be that a solution does not admit an extension to whole I. E.g., for the equation $\dot{x} = x^2$ one such solution $x(t) = -1/(t-1)$ is shown on the picture.

Vector space of solutions

Theorem 12.5.2. *The set S of all the solutions of (1) (defined in the whole I) is a vector space. This space is isomorphic to the phase space X.*

$\lhd$ 1° If $\varphi_1, \varphi_2 \in \mathrm{Sol}(1)$ then $\forall \alpha_1, \alpha_2 \in \mathbb{R}$: $(\alpha_1\varphi_1 + \alpha_1\varphi_1)^\cdot = \alpha_1\dot{\varphi}_1 + \alpha_1\dot{\varphi}_2 \overset{(1)}{=} \alpha_1 A\varphi_1 + \alpha_2 A\varphi_2 = \boxed{A \text{ is linear}} = A(\alpha_1\varphi_1 + \alpha_2\varphi_2)$, that is, $\alpha_1\varphi_1 + \alpha_2\varphi_2 \in \mathrm{Sol}(1)$. Thus, S is a vector space.



In particular

$$0 \in \mathrm{Sol}(1),$$

and

$$\varphi \in \mathrm{Sol}(1) \Rightarrow -\varphi \in \mathrm{Sol}(1);$$

the picture of integral curves is SYMMETRIC (see the picture).

2° Fix any $t \in I$ and consider the mapping

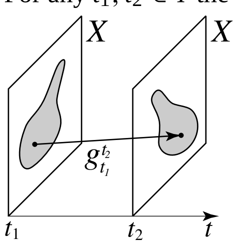
$$\delta_t S :\to X \qquad \varphi :\mapsto \varphi(t),$$

which sends each solution φ into its value at the moment t. Obviously, δ_t is linear. The image of δ_t is the whole X, since by Basic Theorem for any $x \in X$ there exists a solution φ with $\varphi(t) = x$. The kernel of δ_t is $\{0\}$, since again by Basic Theorem, there exists just one solution φ with $\varphi(t) = 0$, and this solution is evidently $\varphi = 0$. Thus δ_t is both surjective and injective. $\rhd$

Fundamental system of solutions

Let $X = \mathbb{R}^n$. Then, by Theorem 12.5.2., $S \approx \mathbb{R}^n$. Any basis $\varphi_1, \ldots, \varphi_n$ of S is called a *fundamental system of solutions* for (1). Thus:

a) Each Equation (1) (in $\mathbb{R}^n$) has a fundamental system of solutions.
b) If $\varphi_1, \ldots, \varphi_n$ is a fundamental system of solutions then any solution φ is a linear combination of $\varphi_1, \ldots, \varphi_n$.
c) Any $n+1$ solutions are linearly depend.
d) For any $t_1, t_2 \in I$ the mapping



$$g_{t_1}^{t_2} := \delta_{t_2} \circ (\delta_{t_1})^{-1} : X \to X$$

(the transformation of the phase space in the time from t_1 up to t_2 is a linear isomorphism.

Example. For the pendulum equation $\dot{x} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} x$ $(x : \mathbb{R} \to \mathbb{R}^2)$ the system

$$\{(\cos t, -\sin t), (\sin t, \cos t)\}$$

is a fundamental system of solutions. (Verify! [Hint: $\begin{vmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{vmatrix} = 1$.])

Scalar linear equation of the n-th order

Consider a *linear* (in x) *homogeneous equation* of the n-th order with variable coefficients

$$x^{(n)} = a_1(t)x^{(n-1)} + \ldots + a_n(t)x, \qquad (x : I \to \mathbb{R}, a_i \in \mathrm{C}(I, \mathbb{R})). \tag{4}$$

We know from Theorem 12.1.1. that (4) is equivalent to an equation

$$\dot{\vec{x}} = A(t)\vec{x}, \qquad (\vec{x} : I \to \mathbb{R}^n, A \in \mathrm{C}(I, \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)).$$

In view of this equivalence, it follows from Theorem 12.5.2. that the following result is true:

Theorem 12.5.3. *The set S of all solutions of (4) (defined on the whole I) is a vector space, which is isomorphic to $\mathbb{R}^n$. This isomorphism is realized by the mapping*

$$S \to \mathbb{R}^n, \ \varphi \mapsto (\varphi(t), \dot{\varphi}(t), \ldots, \varphi^{(n-1)}(t)),$$

where t is an arbitrary fixed point in I.

Any basis of this n-dimensional vector space is called *fundamental system of solutions* for (4).

Example. For the pendulum equation $\ddot{x} + x = 0$ the functions $\cos t$, $\sin t$ form a fundamental system of solutions (see example on 171).

Finding solutions with given initial conditions

Let φ be a solution of (4). We say that φ satisfies an initial condition

$$(\vec{x}_0, t_0) \in \mathbb{R}^n \times I$$

if

$$(\varphi(t_0), \dot{\varphi}(t_0), \ldots, \varphi^{(n-1)}(t_0) = \vec{x}_0,$$

that is,

$$\left.\begin{aligned} \varphi(t_0) &= x_{01} \\ \dot{\varphi}(t_0) &= x_{02} \\ &\vdots \\ \varphi^{(n-1)}(t_0) &= x_{0n} \end{aligned}\right\} \tag{5}$$

Let we know a fundamental system of solutions $\varphi_1, \ldots, \varphi_n$ for (4). If we need to find the solution with initial conditions (5), we look for the solution in the form

$$\varphi = c_1\varphi_1 + \ldots + c_n\varphi_n \qquad (c_i \in \mathbb{R}).$$

Then (5) yields

$$\left.\begin{aligned} c_1\varphi_1(t_0) + \ldots + c_n\varphi_n(t_0) &= x_{01} \\ &\vdots \\ c_1\varphi_1^{(n-1)}(t_0) + \ldots + c_n\varphi_n^{(n-1)}(t_0) &= x_{01} \end{aligned}\right\} \tag{6}$$

Solving this linear algebraic system, we obtain the desired values $c_1, \dots, c_n$.
Remark. The determinant of the system (6)

$$\begin{vmatrix} \varphi_1(t_0) & \dots & \varphi_n(t_0) \\ \dot\varphi_1(t_0) & \dots & \dot\varphi_n(t_0) \\ \vdots & & \vdots \\ \varphi_1^{(n-1)}(t_0) & \dots & \varphi_n^{(n-1)}(t_0) \end{vmatrix}$$

is called *Wronskian* of the system $\{\varphi_1, \dots, \varphi_n\}$. Since the solution of (6) MUST exist for any $\vec{x}_0$ we conclude that Wronskian of any fundamental system of solutions of (4) is NO-ZERO for each $t \in T$.

Variation of constants

For solving of NON-homogenious linear equations the following METHOD OF VARIATION OF CONSTANTS is available:

In order to solve an equation

$$\dot x = A(t)x + h(t), \qquad x : I \to \mathbb{R}^n, A \in \mathrm{C}(I, \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n), h \in \mathrm{C}(I, \mathbb{R}^n),$$

supposing we know a fundamental system $\varphi_1, \dots, \varphi_n$ of solutions of the corresponding homogenious equation $\dot x = A(t)x$, we look for the solutions in the form

$$\varphi(t) = c_1(t)\varphi_1(t) + \ldots + c_n(t)\varphi(t) \qquad (\varphi, \varphi_i : I \to \mathbb{R}^n)$$

(with VARIABLE “constants” c_i!). Then we obtain these unknown functions

$$(c_1, \dots, c_n) := c : I \to \mathbb{R}^n$$

a “simplest” linear equation of the form

$$\dot c = f(t) \qquad (f : I \to \mathbb{R}^n),$$

which we know to solve.

Example. Consider the equation

$$\ddot x + x = f(t) \qquad x : I \to \mathbb{R},\ f \in \mathrm{C}(I, \mathbb{R}), 0 \in I, \tag{7}$$

with initial condition $(0, a)$, $a = (a_1, a_2) \in \mathbb{R}^2$, that is

$$x(0) = a_1, \qquad \dot x(0) = a_2. \tag{8}$$

Reduction to a 1. order system yields

$$\left.\begin{aligned} \dot x_1 &= x_2 \\ \dot x_2 &= -x_1 + f \end{aligned}\right\},\ \left.\begin{aligned} x_1(0) &= a_1 \\ x_2(0) &= a_2 \end{aligned}\right\}. \tag{9}$$

The corresponding homogenious system

$$\begin{cases} \dot x_1 = x_2 \\ \dot x_2 = -x_1 \end{cases}$$

has a well known (Example on page 171) a fundamental system of solutions

$$\{(\cos t, -\sin t), (\sin t, \cos t)\}.$$

So look for the solution in the form

$$(x_1, x_2) = c_1(t)(\cos t, -\sin t) + c_2(t)(\sin t, \cos t).$$

The substitution into (9) gives after simplifications

$$\left.\begin{aligned} \dot c_1 \cos t + \dot c_2 \sin t &= 0 \\ -\dot c_1 \sin t + \dot c_2 \cos t &= f \end{aligned}\right\}, \quad \left.\begin{aligned} c_1(0) &= a_1 \\ c_2(0) &= a_2 \end{aligned}\right\},$$

whence we obtain

$$\begin{aligned} &\dot c_1 = -f \sin t,\ c_1(0) = a_1 \Rightarrow c_1 = a_1 - \textstyle\int_0^t f(\tau)\sin\tau\,\mathrm{d}\tau, \\ &\dot c_2 = f \cos t, \quad c_2(0) = a_2 \Rightarrow c_2 = a_2 + \textstyle\int_0^t f(\tau)\cos\tau\,\mathrm{d}\tau. \end{aligned}$$

If follows that the answer is ($x(t) = x_1(t)$!):

$$x(t) \overset{(8)}{=} \left(x(0) - \int_0^t f(\tau)\sin\tau\,\mathrm{d}\tau\right)\cos t + \left(\dot x(0) + \int_0^t f(\tau)\cos\tau\,\mathrm{d}\tau\right)\sin t.$$

12.6 Linear equations with constant coefficients

Here we study an equation

$$\dot x = Ax, \qquad x : \mathbb{R} \to X, X \in \mathrm{NS}, A \in \mathscr{L}(X, X). \tag{1}$$

We suppose that the space $\mathscr{L}(X, X)$ (with the operator norm) is COMPLETE; for example, it is ever true for $X = \mathbb{R}^n$.

In the simplest case $X = \mathbb{R}$ we have an equation

$$\dot x = ax, \qquad x : \mathbb{R} \to \mathbb{R}, a \in \mathbb{R}.$$

The solution is well-known

$$x = x_0 \mathrm{e}^{at}, \qquad x_0 = x(0).$$

In the general case the result is just the same:

Theorem 12.6.1. *Any solution x of (1) can be extended to the whole $\mathbb{R}$ and is given by the formula*

$$x = \mathrm{e}^{At}x_0, \qquad x_0 = x(0).$$

Here for any operator $A \in \mathscr{L}(X, X)$ we put

$$\mathrm{e}^A := \mathrm{id} + A + \tfrac{1}{2}A^2 + \tfrac{1}{3!}A^3 + \ldots \in \mathscr{L}(X, X), \qquad A^k := \underbrace{A \circ \ldots \circ A}_{k\text{-times}}$$

This series converges in $\mathscr{L}(X, X)$, since it is majorized by the converging non-negative real series $\sum_{k=0}^{\infty} \|A\|^k/k!$ (indeed, $\|A^k\| = \|\underbrace{A \circ \ldots \circ A}_{k\text{-times}}\| \le \|A\|^k$).

$\lhd$ The theorem can be proved essentially in the same way as in classic 1-dimensional case (using member wise differentiation of series). $\rhd$

Thus principally we know the solution of (1), but the problem is how to CALCULATE e^{At} for concrete A. Even for $X = \mathbb{R}$ it is non-trivial problem.

Case of diagonal operators

Let $X = \mathbb{R}^n$. We shall identify an operator $A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$ with its matrix. If A is diagonal, that is,

$$A = \begin{pmatrix} \lambda_1 & & 0 \\ & \dots & \\ 0 & & \lambda_n \end{pmatrix},$$

then it is easy to calculate e^{At}:

$$e^{At} = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \dots & \\ 0 & & e^{\lambda_n t} \end{pmatrix}.$$

$\lhd$ $A^k = \begin{pmatrix} \lambda_1^k & & 0 \\ & \dots & \\ 0 & & \lambda_n^k \end{pmatrix}$, hence

$$\sum \frac{A^k t^k}{k!} = \sum \begin{pmatrix} \frac{\lambda_1^k t^k}{k!} & & 0 \\ & \dots & \\ 0 & & \frac{\lambda_n^k t^k}{k!} \end{pmatrix} = \begin{pmatrix} e^{\lambda_1 t} & & 0 \\ & \dots & \\ 0 & & e^{\lambda_n t} \end{pmatrix}. \rhd$$

Hence the solution of equation

$$\dot{x} = Ax, \qquad x(0) = x_0 = (x_{01}, \dots, x_{0n})$$

is

$$x = e^{At} x_0 = (x_{01} e^{\lambda_1 t}, \dots, x_{0n} e^{\lambda_n t}).$$

Note that λ_k are just the EIGENVALUES of our diagonal operator. Thus, each component of the solution has the form

$$c\, e^{\lambda t}$$

where λ is an eigenvalue of A.

NB In OTHER bases the components of the solutions will be LINEAR COMBINATIONS of the exponents $e^{\lambda_k t}$.

General case

In occurs that in general case each component of the solution of an equation

$$\dot{x} = Ax, \qquad x : \mathbb{R} \to \mathbb{R}^n,\ A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n), \tag{2}$$

is a linear combination of n members of the form

$$t^m \mathcal{R}e\, e^{\lambda t} \qquad \text{or} \qquad t^m \mathcal{I}m\, e^{\lambda t}$$

where λ is an eigenvalue of A ($\lambda \in \mathbb{C}$) and m is a natural number less than multiplicity of λ.

Recall that the eigenvalues of a are roots of the *characteristic equation*

$$\det(A - \lambda E) = 0 \qquad E \text{ denotes the unit matrix}$$

and that the multiplicity of an eigenvalue is just the multiplicity of the root.

Case of one scalar equation of n-th order

$$x^{(n)} = a_1 x^{(n-1)} + \ldots + a_n x, \qquad x : \mathbb{R} \to \mathbb{C},\ a_j \in \mathbb{C}. \tag{3}$$

As we know (3) can be considered as a special case of (2). It follows that:

Any solution of (3) has the form

$$x(t) = \sum_{l=1}^{k} \mathrm{e}^{\lambda_l t} p_l(t), \tag{4}$$

where $\lambda_1, \ldots, \lambda_k$ are the different roots of the *characteristic equation*

$$\lambda^n = a_1 \lambda^{n-1} + \ldots + a_n, \tag{5}$$

and p_l is a polynomial of degree less than the multiplicity of the root λ_l.
NB This result remains true for **non**-homogenious equations $x^{(n)} = a_1 x^{(n-1)} + \ldots + a_n x + f(t)$, if $f(t)$ has the form (4).

Examples.

1. $\ddot{x} + x = 0$. The characteristic equation $\lambda^2 + 1$ has the roots $\pm\mathrm{i}$; we have $\mathcal{R}e\,\mathrm{e}^{\pm \mathrm{i}t} = \cos t$, $\mathcal{I}m\,\mathrm{e}^{\pm \mathrm{i}t} = \pm \sin t$. The functions $\cos t, \sin t$ form a fundamental system solutions. The general solution is $c_1 \cos t + c_2 \sin t$.

2. $\ddot{x} - x = 0$. The characteristic equation $\lambda^2 - 1$ has the roots ± 1. The corresponding functions e^t and e^{-t} form a fundamental system of solutions. The general solution is $c_1 \mathrm{e}^t + c_2 \mathrm{e}^{-t}$. (In particular $\mathrm{sh}\, t$ and $\mathrm{ch}\, t$ are solutions.)

3. $\ddot{x} = 0$. The characteristic equation $\lambda^2 = 0$ has one 2-multiple root 0. The corresponding functions from (4) are 1 and t. They form a fundamental system of solutions. The general solution is $c_1 + c_2 t$ (as well known, of course).