Maria Predoi Trandafir Bălan

MATHEMATICAL ANALYSIS

VOL. I

DIFFERENTIAL CALCULUS

Craiova, 2005

PREFACE

Because of the numerous books that have already appeared about the classical Analysis, in principle it is very difficult to bring new facts in this field. However, the engineers, researchers in experimental sciences, and even the students actually need a quick and clear presentation of the basic theory, together with an extensive and efficient guidance to solve practical problems. Therefore, in this book we tried to combine the essential (but rigorous) theoretical results with a large scale of concrete applications of the Mathematical Analysis, and formulate them in nowadays language.

The content is based on a two-semester course that has been given in English to students in Computer Sciences at the University of Craiova, during a couple of years. As an independent work, it contains much more than the effective lessons can treat according to the imposed program.

Starting with the idea that nobody (even student) has enough time to read several books in order to rediscover the essence of a mathematical theory and its practical use, we have formulated the following objectives for the present book:

1. Accessible connection with mathematics in lyceum
2. Self-contained, but well referred to other works
3. Prominence of the specific structures
4. Emphasis on the essential topics
5. Relevance of the sphere of applications.

The first objective is assured by a large introductory chapter, and by the former paragraphs in the other chapters, where we recall the previous notions. To help intuition, we have inserted a lot of figures and schemes.

The second one is realized by a complete and rigorous argumentation of the discussed facts. The reader interested in enlarging and continuing the study is still advised to consult the attached bibliography. Besides classical books, we have mentioned the treatises most available in our zone, i.e. written by Romanian authors, in particular from Craiova.

Because Mathematical Analysis expresses in a more concrete form the philosophical point of view that assumes the continuous nature of the Universe, it is very significant to reveal its fundamental structures, i.e. the *topologies*. The emphasis on the structures is especially useful now, since the discrete techniques (e.g. digital) play an increasing role in solving practical problems. Besides the deeper understanding of the specific features, the higher level of generalization is necessary for a rigorous treatment of the fundamental topics like continuity, differentiability, etc.

To touch the fourth objective, we have organized the matter such that each chapter debates one of the basic aspects, more exactly continuity, convergence and differentiability in volume one, and different types of integrals in part two. We have explained the utility of each topic by plenty of historic arguments and carefully selected problems.

Finally, we tried to realize the last objective by lists of problems at the end of each paragraph. These problems are followed by answers, hints, and sometimes by complete solutions.

In order to help the non-native speakers of English in talking about the matter, we recommend books on English mathematical terms, including pronunciation and stress, e.g. the *Guide to Mathematical Terms* [BT_4]. Our experience has shown that most language difficulties concern speaking, rather than understanding a written text. Therefore we encourage the reader to insist on the phonetics of the mathematical terms, which is essential in a fluent dialog with foreign specialists.

In spite of the opinion that in old subjects like Mathematical Analysis everything is done, we still have tried to make our book distinguishable from other works. With this purpose we have pointed to those research topics where we have had some contributions, e.g. the quasi-uniform convergence in function spaces (§ II.3 in connection to [PM_2] and [PM_3]), the structures of discreteness (§ III.2 with reference to [BT_3]), the unified view on convergence and continuity via the intrinsic topology of a directed set, etc. We also hope that a note of originality there results from:

- The way of solving the most concrete problems by using modern techniques (e.g. local extrema, scalar and vector fields, etc.);
- A rigorous but moderately extended presentation of several facts (e.g. higher order differential, Jordan measure in $\mathbb{R}^n$, changing the variables in multiple integrals, etc.) which sometimes are either too much simplified in practice, or too detailed in theoretical treatises;
- The unitary treatment of the Real and Complex Analysis, centered on the analytic (computational) method of studying functions and their practical use (e.g. § II.4, § IV.5, Chapter X, etc.).

We express our gratitude to all our colleagues who have contributed to a better form of this work. The authors are waiting for further suggestions of improvements, which will be welcome any time.

The Authors

Craiova, September 2005

CONTENTS

VOL. I. DIFFERENTIAL CALCULUS

Chapter III. CONTINUITY

Chapter IV. DIFFERENTIABILITY

CHAPTER I. PRELIMINARIES

§ I.1. SETS, RELATIONS, FUNCTIONS

From the very beginning, we mention that a general knowledge of set theory is assumed. In order to avoid the contradictions, which can occur in such a "naive" theory, these sets will be considered *parts* of a total set T, i.e. elements of $\mathscr{P}(T)$. The sets are usually depicted by some specific properties of the component elements, but we shall take care that instead of *sets of sets* it is advisable to speak of *families of sets* (see [RM], [SO], etc).

When operate with sets we basically need one unary operation

$$A \mapsto \complement A = \{x \in T : x \notin A\} \text{ (complement)},$$

two binary operations

$$(A, B) \mapsto A \cup B = \{x \in T : x \in A \text{ or } x \in B\} \text{ (union)},$$

$$(A, B) \mapsto A \cap B = \{x \in T : x \in A \text{ and } x \in B\} \text{ (intersection)},$$

and a binary relation

$$A = B \Leftrightarrow x \in A \text{ iff } x \in B \text{ (equality)}.$$

1.1. Proposition. If $A, B, C \in \mathscr{P}(T)$, then:

(i) $A \cup (B \cup C) = (A \cup B) \cup C$; $A \cap (B \cap C) = (A \cap B) \cap C$ (*associativity*)

(ii) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$; $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (*distributivity*)

(iii) $A \cap (A \cup B) = A$; $A \cup (A \cap B) = A$ (*absorption*)

(iv) $(A \cap \complement A) \cup B = B$; $(A \cup \complement A) \cap B = B$ (*complementary*)

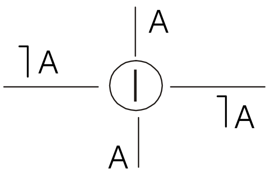
(v) $A \cup B = B \cup A$; $A \cap B = B \cap A$ (*commutativity*).

1.2. Remark. From the above properties (i)-(v) we can derive the whole set theory. In particular, the associativity is useful to define intersections and unions of a finite number of sets, while the extension of these operations to arbitrary families is defined by $\cap \{A_i : i \in I\} = \{x \in T : \forall i \in I \Rightarrow x \in A_i\}$ and $\cup \{A_i : i \in I\} = \{x \in T : \exists i \in I \text{ such that } x \in A_i\}$. Some additional notations are frequent, e.g. $\varnothing = A \cap \complement A$ for the (unique!) *void set*, $A \setminus B = A \cap \complement B$ for the *difference*, $A \Delta B = (A \setminus B) \cup (B \setminus A)$ for the *symmetric difference*, $A \subseteq B$ (defined by $A \cup B = B$) for the relation of *inclusion*, etc.

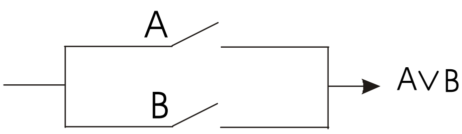
More generally, a non-void set $\mathscr{A}$ on which the equality = , and the operations $\rceil$, $\vee$ and $\wedge$ (instead of $\complement$, $\cup$, respectively $\cap$) are defined, such that conditions (i)-(v) hold as axioms, represents a *Boolean algebra.* Besides $\mathscr{P}(T)$, we mention the following important examples of Boolean algebras: the algebra of propositions in the formal logic, the algebra of switch nets, the algebra of logical circuits, and the field of events in a random process. The obvious analogy between these algebras is based on the correspondence of the following facts:

- a set may contain some given point or not;
- a proposition may be true or false;
- an event may happen in an experience or not;
- a switch may let the current flow through or break it;
- at any point of a logical circuit may be a signal or not.

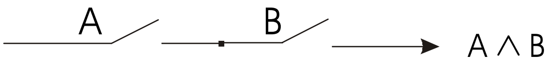
In addition, the specific operations of a Boolean algebra allow the following concrete representations in switch networks:



(double switch)

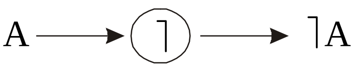


(parallel connection)

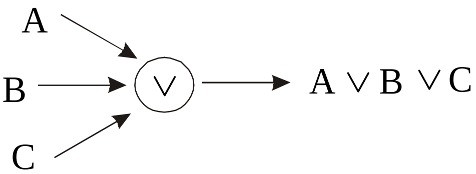


(serial connection)

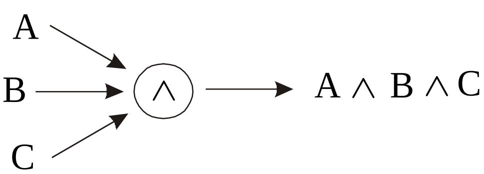
Similarly, in logical circuits we speak of "logical gates" like



(*non*-gate)



(*or*-gate)



(*and*-gate)

1.3. <u>The Fundamental Problems</u> concerning a practical realization of a switch network, logic circuits, etc., are the *analysis* and the *synthesis*. In the first case, we have some physical realization and we want to know how it works, while in the second case, we desire a specific functioning and we are looking for a concrete device that should work like this. Both problems involve the so-called *working functions*, which describe the functioning of the circuits in terms of values of a given formula, as in the table from bellow. It is advisable to start by putting the values *1, 0, 1, 0,…* for *A*, then *1, 1, 0, 0,…* for *B*, etc., under these variables, then continue by the resulting values under the involved connectors $\wedge$, $\vee$, $\rightarrow$, etc. by respecting the order of operations, which is specified by brackets. The last completed column, which also gives the name of the formula, contains the "truth values" of the considered formula.

As for example, let us consider the following disjunction, whose truth-values are in column *(9)*:

$(A$	$\wedge$	$B)$	$\vee$	$[(\rceil A$	$\rightarrow$	$C)$	$+$	$B]$
1	*1*	*1*	*1*	*0*	*1*	*1*	*0*	*1*
0	*0*	*1*	*0*	*1*	*1*	*1*	*0*	*1*
1	*0*	*0*	*1*	*0*	*1*	*1*	*1*	*0*
0	*0*	*0*	*1*	*1*	*1*	*1*	*1*	*0*
1	*1*	*1*	*1*	*0*	*1*	*0*	*0*	*1*
0	*0*	*1*	*1*	*1*	*0*	*0*	*1*	*1*
1	*0*	*0*	*1*	*0*	*1*	*0*	*1*	*0*
0	*0*	*0*	*0*	*1*	*0*	*0*	*0*	*0*
(1)	*(6)*	*(3)*	*(9)*	*(2)*	*(7)*	*(5)*	*(8)*	*(4)*

where *(1), (2),* etc. show the order of completing the columns.

The converse problem, namely that of writing a formula with previously given values, makes use of some *standard* expressions, which equal *1* only once (called *fundamental conjunctions*). For example, if a circuit should function according to the table from below,

A	*B*	*C*	*f(A,B,C)*	*fundamental conjunctions*
1	*1*	*1*	*1*	$A \wedge B \wedge C$
0	*1*	*1*	*0*	-
1	*0*	*1*	*0*	-
0	*0*	*1*	*1*	$\rceil A \wedge \rceil B \wedge C$
1	*1*	*0*	*0*	-
0	*1*	*0*	*1*	$\rceil A \wedge B \wedge \rceil C$
1	*0*	*0*	*1*	$A \wedge \rceil B \wedge \rceil C$
0	*0*	*0*	*0*	-

then one working function is

$$f(A,B,C) = (A\wedge B\wedge C)\vee(\rceil A\wedge \rceil B\wedge C)\vee(\rceil A\wedge B\wedge \rceil C)\vee(A\wedge \rceil B\wedge \rceil C).$$

This form of f is called *normal disjunctive* (see [ME], etc.).

The following type of subfamilies of $\mathscr{P}(T)$, where $T\neq\emptyset$, is frequently met in the Mathematical Analysis (see [BN_1], [DJ], [CI], [L-P], etc.):

1.4. <u>Definition</u>. A nonvoid family $\mathscr{F}\subset\mathscr{P}(T)$ is called (*proper*) *filter* if

[F_0] $\emptyset\notin\mathscr{F}$;

[F_1] $A, B\in\mathscr{F}\Rightarrow A\cap B\in\mathscr{F}$;

[F_2] ($A\in\mathscr{F}$ and $B\supseteq A$) $\Rightarrow B\in\mathscr{F}$.

Sometimes condition [F_0] is omitted, and we speak of filters in generalized (*improper*) sense. In this case, $\mathscr{F}=\mathscr{P}(T)$ is accepted as *improper* filter.

If family $\mathscr{F}$ is a filter, then any subfamily $\mathscr{B}\subseteq\mathscr{F}$ for which

[BF] $\forall A\in\mathscr{F}\ \exists B\in\mathscr{B}$ such that $B\subseteq A$,

(in particular $\mathscr{F}$ itself) is called *base of the filter* $\mathscr{F}$.

1.5. <u>Examples.</u> a) If at any fixed $x\in\mathbb{R}$ we define $\mathscr{F}\subset\mathscr{P}(T)$ by

$$\mathscr{F}=\{A\subseteq\mathbb{R}:\ \exists\varepsilon>0 \text{ such that } A\supseteq(x-\varepsilon, x+\varepsilon)\},$$

then $\mathscr{F}$ is a filter, and a base of $\mathscr{F}$ is $\mathscr{B}=\{(x-\varepsilon, x+\varepsilon):\varepsilon>0\}$. It is easy to see that $\cap\{A\subseteq\mathbb{R}: A\in\mathscr{F}\}=\{x\}$.

b) The family $\mathscr{F}\subset\mathscr{P}(\mathbb{N})$, defined by

$$\mathscr{F}=\{A\subseteq\mathbb{N}:\ \exists n\in\mathbb{N} \text{ such that } A\supseteq(n,\rightarrow)\},$$

is a filter in $\mathscr{P}(\mathbb{N})$ for which $\mathscr{B}=\{(n,\rightarrow): n\in\mathbb{N}\}$ is a base, and

$$\cap\{A\subseteq\mathbb{N}: A\in\mathscr{F}\}=\emptyset.$$

c) Let $\mathscr{B}\subset\mathscr{P}(\mathbb{R}^2)$ be the family of interior parts of arbitrary regular polygons centered at some fixed $(x, y)\in\mathbb{R}^2$. Then

$$\mathscr{F}=\{A\subseteq\mathbb{R}^2:\ \exists B\in\mathscr{B} \text{ such that } A\subseteq B\}$$

is a filter for which family $\mathscr{C}$, of all interior parts of the disks centered at (x, y), is a base (as well as $\mathscr{B}$ itself).

1.6. <u>Proposition.</u> In an arbitrary total set $T\neq\emptyset$ we have:

(i) Any base $\mathscr{B}$ of a filter $\mathscr{F}\subseteq\mathscr{P}(T)$ satisfies the condition

[FB] $\forall A, B\in\mathscr{B}\ \exists C\in\mathscr{B}$ *such that* $C\subseteq A\cap B$.

(ii) If $\mathscr{B}\subseteq\mathscr{P}(T)$ satisfies condition [FB] (i.e. together with [F_0] it is a *proper filter base*), then the family of oversets

$$\mathscr{G}=\{A\subseteq T:\ \exists B\in\mathscr{B} \textit{ such that } A\supseteq B\}$$

is a filter in $\mathscr{P}(T)$; we say that filter $\mathscr{G}$ is *generated by* $\mathscr{B}$.

(iii) If $\mathscr{B}$ is a base of $\mathscr{F}$, then $\mathscr{B}$ generates $\mathscr{F}$.

The proof is direct, and we recommend it as an exercise.

1.7. Definition. If A and B are nonvoid sets, their *Cartesian product* is defined by $A \times B = \{(a, b): a \in A, b \in B\}$.

Any part $\mathcal{R} \subseteq A \times B$ is called *binary relation between A and B.* In particular, if $\mathcal{R} \subseteq T \times T$, it is named *binary relation on T.* For example, the *equality* on T is represented by the *diagonal* $\delta = \{(x, x): x \in T\}$.

If $\mathcal{R}$ is a relation on T, its *inverse* is defined by

$$\mathcal{R}^{-1} = \{(x, y): (y, x) \in \mathcal{R}\}.$$

The composition of two relations $\mathcal{R}$ and $\mathcal{S}$ on T is noted

$$\mathcal{R} \circ \mathcal{S} = \{(x, y): \exists z \in T \textit{ such that } (x, z) \in \mathcal{S} \text{ and } (z, y) \in \mathcal{R}\}.$$

The section (*cut*) of $\mathcal{R}$ at x is defined by

$$\mathcal{R}[x] = \{y \in T : (x, y) \in \mathcal{R}\}.$$

Most frequently, a binary relation $\mathcal{R}$ on T may be:

Reflexive: $\delta \subseteq \mathcal{R}$;

Symmetric: $\mathcal{R} = \mathcal{R}^{-1}$;

Antisymmetric: $\mathcal{R} \cap \mathcal{R}^{-1} = \delta$;

Transitive: $\mathcal{R} \circ \mathcal{R} \subseteq \mathcal{R}$;

Directed: $\mathcal{R}[x] \cap \mathcal{R}[y] \neq \emptyset$ for any $x, y \in T$.

The reflexive, symmetric and transitive relations are called *equivalences*, and usually they are denoted by $\sim$. If $\sim$ is an equivalence on T , then each $x \in T$ generates a *class of equivalence*, noted $x^\wedge = \{y \in T : x \sim y\}$.

The set of all equivalence classes is called *quotient set,* and it is noted $T/\sim$.

The reflexive and transitive relations are named *preorders*.

Any antisymmetric preorder is said to be a *partial order*, and usually it is denoted by $\leq$. We say that an order $\leq$ on T is *total* (or, equivalently, $(T, \leq)$ is *totally, linearly ordered*) iff for any two $x, y \in T$ we have either $x \leq y$ or $y \leq x$. Finally, $(T, \leq)$ is said to be *well ordered* (or $\leq$ is a *well ordering* on T) iff $\leq$ is total and any nonvoid part of T has a smallest element.

1.8. Examples. *(i)* Equivalences:

1. The equality (of sets, numbers, figures, etc.);
2. $\{((a, b), (c, d)) \in \mathbb{N}^2 \times \mathbb{N}^2: a + d = b + c\}$;
3. $\{((a, b), (c, d)) \in \mathbb{Z}^2 \times \mathbb{Z}^2 : ad = bc\}$;

4. $\{(A, B) \in \mathcal{M}_n(\mathbb{R}) \times \mathcal{M}_n(\mathbb{R}) : \exists T \in \mathcal{M}_n(\mathbb{R})$ such that $B = T^{-1} A T\}$.

The similarity of the figures (triangles, rectangles, etc.) in $\mathbb{R}^2$, $\mathbb{R}^3$, etc., is an equivalence especially studied in Geometry.

***(ii)* Orders and preorders:**

1. The inclusion in $\mathscr{P}(T)$ is a partial order;
2. $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are totally ordered by their natural orders $\leq$;
3. $\mathbb{N}$ is well ordered by its natural order;
4. If T is (totally) ordered by $\leq$, and S is an arbitrary nonvoid set, then the set $\mathcal{F}_T(S)$ of all functions $f : S \to T$, is partially ordered by

$$\mathcal{R} = \{(f, g) \in \mathcal{F}_T(S) \times \mathcal{F}_T(S) : f(x) \leq g(x) \text{ at any } x \in S\}.$$

This relation is frequently called *product order* (compare to the examples in problem 9, at the end of the paragraph).

***(iii)* Directed sets** (i. e. preordered sets $(D, \leq)$ with directed $\leq$):

1. $(\mathbb{N}, \leq)$, as well as any totally ordered set;
2. Any filter $\mathscr{F}$ (e.g. the entire $\mathscr{P}(T)$, each system of neighborhoods $\mathscr{V}(x)$ in topological spaces, etc.) is directed by inclusion, in the sense that $A \leq B$ iff $B \subseteq A$.
3. Let us fix $x_0 \in \mathbb{R}$, and note
   $$D = \{(V, x) \in \mathscr{P}(\mathbb{R}) \times \mathbb{R} : \exists \varepsilon > 0 \text{ such that } x \in (x_0 - \varepsilon,\ x_0 + \varepsilon) \subseteq V\}.$$
   The pair $(D, \leq)$ is a directed set if the preorder $\leq$ is defined by
   $$(V, x) \leq (U, y) \Leftrightarrow U \subseteq V .$$
   The same construction is possible using neighborhoods U, V, … of a fixed point x_0 in any topological space.
4. The *partitions*, which occur in the definition of some integrals, generate directed sets (see the integral calculus). In particular, in order for us to define the Riemannian integral on $[a, b] \subset \mathbb{R}$, we consider *partitions* of the closed interval $[a, b]$, i.e. finite sets of subintervals of the form
   $$\delta = \{[x_{k-1}, x_k] : k = 1, 2, \ldots, n;\ a = x_0 < x_1 < \ldots < x_n = b\},$$
   for arbitrary $n \in \mathbb{N}^*$. In addition, for such a partition we choose different *systems of intermediate points*
   $$\xi(\delta) = \{\xi_k \in [x_{k-1}, x_k] \in \delta : k = \overline{1, n}\} .$$
   It is easy to see that set D, of all pairs $(\delta, \xi(\delta))$, is directed by relation $\leq$, where $(\delta', \xi(\delta')) \leq (\delta'', \xi(\delta''))$ iff $\delta' \subseteq \delta''$.

There is a specific terminology in preordered sets, as follows:

1.9. <u>Definition.</u> Let A be a part of T, which is (partially) ordered by $\leq$. Any element $x_0 \in T$, for which $x \leq x_0$ holds whenever $x \in A$, is said to be *an* *upper bound of A*. If $x_0 \in A$, then it is called the *greatest element of A* (if there exists one, then it is unique!), and we note $x_0 = \max A$.

If the set of all upper bounds of A has the smallest element $\bar{x}$, we say that is the *supremum of A*, and we note $\bar{x} = \sup A$.

An element $x^* \in A$ is considered *maximal* iff A does not contain elements greater than x^*(the element $\max A$, if it exists, is maximal, but the converse assertion is not generally true).

Similarly, we speak of *lower bound*, smallest element (denoted as $\min A$), infimum (noted $\inf A$), and *minimal elements*. If $\sup A$ and $\inf A$ do exist for each bounded set A, we say that $(T, \le)$ is a *complete* (in order).

Alternatively, instead of using an order $\mathcal{R}$, we can refer to the attached *strict order* $\mathcal{R} \setminus \delta$. The same, if $\mathcal{R}$ is an order on T, and $x \in T$, then $\mathcal{R}[x]$ is sometimes named *cone of vertex x* (especially because of its shape).

If a part C of T is totally (linearly) ordered by the induced order, then we say that C is a *chain* in T.

An ordered set $(T, \mathcal{R})$ is called *lattice* (or *net*) iff for any two $x, y \in T$ there exist $\inf \{x, y\} = x \wedge y$ and $\sup \{x, y\} = x \vee y$. If the infimum and the supremum exist for any bounded set in T, then the lattice is said to be *complete* (or *σ - lattice*). A remarkable example of lattice is the following:

1.10. <u>Proposition</u>. Every Boolean algebra is a lattice. In particular, $\mathscr{P}(T)$ is a (complete) lattice relative to $\subseteq$.

<u>Proof.</u> We have to show that $\subseteq$ is a (partial) order on $\mathscr{P}(T)$, and each family $\{A_i \in \mathscr{P}(T) : i \in I\}$ has an infimum and a supremum. Reasoning as for an arbitrary Boolean algebra, reflexivity of $\subseteq$ means $A \vee A = A$. In fact, according to *(iii)* and *(ii)* in proposition 1.1., we have

$$A \vee A = [A \wedge (A \vee B)] \vee [A \wedge (A \vee B)] = (A \vee A) \wedge (A \vee B) = A \vee (A \wedge B) = A,$$

From $A \le B$ and $B \le A$, we deduce that $B = A \vee B = A$, hence $\le$ is antisymmetric. For transitivity, if $A \le B$ and $B \le C$ we obtain $A \le C$ since

$$C = B \vee C = (A \vee B) \vee C = A \vee (B \vee C) = A \vee C.$$

Let us show that $\sup \{A,B\} = A \vee B$ holds for any $A,B \in \mathscr{P}(T)$. In fact, according to *(iii)*, $A \le A \vee B$ and $B \le A \vee B$. On the other hand, if $A \le X$ and $B \le X$, we have $A \wedge X = A$ and $B \wedge X = B$, so that

$$X \wedge (A \vee B) = (X \wedge A) \vee (X \wedge B) = A \vee B,$$

i.e. $A \vee B \le X$. Similarly we can reason for $\inf \{A,B\} = A \wedge B$, as well as for arbitrary families of sets in T. ◇

1.11. <u>Remark.</u> The above proof is based on the properties *(i)-(v)*, hence it is valid in arbitrary Boolean algebras. If limited to $\mathscr{P}(T)$, we could reduce it to the concrete expressions of $A \cup B$, $A \cap B$, $A \subseteq B$, etc. According to the Stone's theorem, which establishes that any Boolean algebra $\mathscr{A}$ is isomorphic to a family of parts, verifying a property in $\mathscr{A}$ as for $\mathscr{P}(T)$ is still useful.

1.12. <u>Definition.</u> Let X and Y be nonvoid sets, and $\mathcal{R} \subseteq X\times Y$ be a relation between the elements of X and Y. We say that $\mathcal{R}$ is a *function* defined on X with values in Y iff the section $\mathcal{R}$ [x] reduces to a single element of Y for any $x\in X$. Alternatively, a function is defined by X, Y and a rule f , of attaching to each $x\in X$ an element $y\in Y$. In this case we note $y = f(x)$, $x\mapsto y = f(x)$, $f : X\mapsto Y$, etc.

We say that $f : X\mapsto Y$ is *injective* (1:1, i.e. *one-to-one*) iff $f(x)\neq f(y)$ whenever $x\neq y$.

If for any $y\in Y$ there exists x in X such that $y = f(x)$, then f is called *surjective* (or *onto*). If f is both injective and surjective, it is called *bijective* (1:1 map of X on Y , or 1:1 *correspondence between X and Y*).

Any function $f : X\mapsto Y$ can be extended to $\mathscr{P}(X)$ and $\mathscr{P}(Y)$ by considering *the direct image of* $A\subseteq X$, defined by

$$f(A) = \{ f(x) : x\in A\},$$

and *the inverse image of* $B\subseteq Y$, defined by

$$\overleftarrow{f}(B) = \{x\in X : f(x) \in B\}.$$

If f is bijective, then $\overleftarrow{f}(y)$ consists of a single element, so we can speak of *the inverse function* f^{-1} ,defined by

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

If $f : X\to Y$ and $g : Y\to Z$, then $h : X\to Z$, defined by

$$h(x) = g(f(x)) \text{ for all } x\in X,$$

is called *the composition of f and g* , and we note $h = g\circ f$.

The *graph of* $f : X\to Y$ is a part of $X \times Y$, namely

$$\text{Graph}(f) = \{(x, y) \in X \times Y : y = f(x)\}.$$

On a Cartesian product $X\times Y$ we distinguish two remarkable functions, called *projections*, namely $\text{Pr}_X : X\times Y \to X$, and $\text{Pr}_Y : X\times Y \to Y$, defined by

$$\text{Pr}_X(x, y) = x \text{ , and } \text{Pr}_Y(x, y) = y .$$

In the general case of an arbitrary Cartesian product, which is defined by

$$\underset{i\in I}{\mathsf{X}} X_i = \{ f : I \to \bigcup_{i\in I} X_i \mid f(i) \in X_i\},$$

we get a *projection* $\text{Pr}_i :(\underset{i\in I}{\mathsf{X}} X_i)\to X_i$ for each $i \in I$, which has the values

$$\text{Pr}_i :(f) = f(i) .$$

Sometimes we must extend the above notion of function, and allow that $f(x)$ consists of more points; in such case we say that f is a *multivalued* (or *one to many*) *function*. For example, in the complex analysis, $f = \sqrt[n]{\ }$ is supposed to be an already known *1:n* function. Similarly, we speak of *many to one*, or *many to many* functions.

This process of extending the action of f can be continued to carry elements from $\mathscr{P}(\mathscr{P}(X))$ to $\mathscr{P}(\mathscr{P}(Y))$, e.g. if $\mathscr{V} \subseteq \mathscr{P}(X)$, then

$$f(\mathscr{V}) = \{f(A) \in \mathscr{P}(Y) : A \in \mathscr{V}\}.$$

1.13. <u>Examples</u>. Each part $A \subseteq X$ ($\neq \emptyset$) is completely determined by its *characteristic function* $f_A : X \to \{0, 1\}$, expressed by

$$f_A(x) = \begin{cases} 0 & \text{if } x \notin A \\ 1 & \text{if } x \in A. \end{cases}$$

In other terms, $\mathscr{P}(X)$ can be presented as the set of all functions defined on X and taking two values. Because we generally note the set of all functions $f : X \to Y$ by Y^X, we obtain $2^X = \mathscr{P}(X)$.

We mention that this possibility to represent sets by functions has led to the idea of *fuzzy sets*, having characteristic functions with values in the closed interval [0,1] of $\mathbb{R}$ (see [N-R], [KP], etc.). Formally, this means to replace $2^X = \mathscr{P}(X)$ by $[0,1]^X$. Of course, when we work with fuzzy sets, we have to reformulate the relations and the operations with sets in terms of functions, e.g. $f \subseteq g$ as fuzzy sets means $f \le g$ as functions, $\complement f = 1 - f$, $f \cup g = \max\{f, g\}$, $f \cap g = \min\{f, g\}$, etc.

1.14. <u>Proposition.</u> Let $f : X \to Y$ be a function, and let I, J be arbitrary families of indices. If $A_i \subseteq X$ and $B_j \subseteq Y$ hold for any $i \in I$ and $j \in J$, then:

(i) $f(\cup\{A_i : i \in I\}) = \cup\{f(A_i) : i \in I\}$;

(ii) $f(\cap\{A_i : i \in I\}) \subseteq \cap\{f(A_i) : i \in I\}$;

(iii) $\overleftarrow{f}(\cup\{B_j : j \in J\}) = \cup\{\overleftarrow{f}(B_j) : j \in J\}$;

(iv) $\overleftarrow{f}(\cap\{B_j : j \in J\}) = \cap\{\overleftarrow{f}(B_j) : j \in J\}$;

(v) $\overleftarrow{f}(\complement B) = \complement[\overleftarrow{f}(B)]$ holds for any $B \subseteq Y$,

while $f(\complement A)$ and $\complement[f(A)]$ generally cannot be compared.

The proof is left to the reader.

The following particular type of functions is frequently used in the Mathematical Analysis:

1.15. <u>Definition</u>. Let S be a nonvoid set. Any function $f : \mathbb{N} \to S$ is called *sequence in S*. Alternatively we note $f(n) = x_n$ at any $n \in \mathbb{N}$, and we mark the sequence f by mentioning the generic term (x_n).

A sequence $g : \mathbb{N} \to S$ is considered to be a *subsequence of* f iff $g = f \circ h$ for some increasing $h : \mathbb{N} \to \mathbb{N}$ (i.e. $p \le q \Rightarrow h(p) \le h(q)$). Usually we note $h(k) = n_k$, so that a subsequence of (x_n) takes the form (x_{n_k}).

More generally, if $(D, \le)$ is a directed set, then $f : D \to S$ is called *generalized sequence* (briefly *g.s.*, or *net*) in S. Instead of f, the g.s. is frequently marked by (x_d), or more exactly by $(x_d)_{d \in D}$, where $x_d = f(d)$, $\forall\, d \in D$. If $(E, \preceq)$ is another directed set, then $g : E \to S$ is named *generalized*

subsequence (g.s.s., or *subnet)* of f iff $g = f\circ h$, where $h:E\to D$ fulfils the following condition (due to Kelley, see [KJ], [DE], etc.):
[s] $\forall\, d\in D\ \exists e\in E$ *such that* $(e\preceq a\in E \Rightarrow d\leq h(a))$.

Similarly, if we note $h(a)= d_a$, then a g.s.s. can be written as (x_{d_a}).

1.16. <u>Examples</u>. a) Any sequence is a g.s., since $\mathbb{N}$ is directed.

b) If D is the directed set in the above example 1.8. *(iii)*3, then $f:D\to\mathbb{R}$, expressed by $f(V, x) = x$, is a generalized sequence.

c) Let us fix some $[a, b] \subset\mathbb{R}$, then consider the directed set $(D, \leq)$ as in the example 1.8. *(iii)*4, and define a bounded function $f : [a, b] \to\mathbb{R}$. If to each pair $(\delta, \xi) \in D$ we attach the so called *integral sum*

$$\sigma_f\ (\delta, \xi) = f(\xi_1)(x_1 - x_0) + \ldots + f(\xi_n)(x_n - x_{n-1}),$$

then the resulting function $\sigma_f : D\to\mathbb{R}$ represents a g.s. which is essential in the construction of the definite integral of f .

1.17. <u>Remark</u>. The notion of *Cartesian product* can be extended to arbitrary families of sets $\{A_i : i\in I\}$, when it is noted $\mathsf{X}\{A_i : i\in I\}$. Such a product consists of all "choice functions" $f : I\to \cup\{A_i : i\in I\}$, such that $f(i)\in A_i$ for each $i\in I$. It was shown that the existence of these *choice functions* cannot be deduced from other facts in set theory, i.e. it must be considered as an independent axiom. More exactly, we have to consider the following:

1.18. <u>The Axiom of Choice</u> (E. Zermelo). The Cartesian product of any nonvoid family of nonvoid sets is nonvoid.

We mention without proof some of the most significant relations of this axiom with other properties (for details see [HS], [KP], etc.):

1.19. <u>Theorem</u>. The axiom of choice is logically equivalent to the following properties of sets:

a) Every set can be well-ordered (Zermelo);
b) Every nonvoid partially ordered set, in which each chain has an upper bound, has a maximal element (Zorn);
c) Every nonvoid partially ordered set contains a maximal chain (Hausdorff);
d) Every nonvoid family of finite character (i.e. A is a member of the family iff each finite subset of A is) has a maximal member (Tukey).

1.20. <u>Remark.</u> The axiom of choice will be adopted throughout this book, as customarily in the treatises on Classical Analysis. Without insisting on each particular appearance during the development of the theory, we mention that the axiom of choice is essential in plenty of problems as for example the existence of a (Hamel) basis in any linear space ($\neq \{0\}$, compare to §I.3), the existence of g such that $f \circ g = I_Y$, where $f : X \to Y$, etc.

PROBLEMS § I.1.

1. Verify the De Morgan's laws:

$$\complement(A \cup B) = \complement A \cap \complement B \text{ and } \complement(A \cap B) = \complement A \cup \complement B.$$

Using them, simplify as much as possible the Boolean formulas:
(a) $\complement[A \cup (B \cap (A \cup \complement B))]$, and
(b) $\complement[(\complement X \cup Y) \cap (\complement Y \cup X)]$.

Hint. $\complement X$ is characterized in general Boolean algebras by the relations

$$X \cup \complement X = T, \text{ and } X \cap \complement X = \varnothing .$$

2. Show that $(\mathscr{P}(T), \cup)$ and $(\mathscr{P}(T), \cap)$ never form groups.
Hint. $\varnothing$, respectively T, should be the neutral elements, but the existence of the opposite elements cannot be assured anymore.

3. Verify the equalities:

(i) $A \setminus (B \cup C) = (A \setminus B) \setminus C$
(ii) $A \setminus (A \setminus B) = A \cap B$
(iii) $(A \cup B) \setminus C = (A \setminus C) \cup (B \setminus C)$
(iv) $A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C)$
(v) $(A \setminus B) \cap C = (A \cap C) \setminus (B \cap C)$
(vi) $(A \setminus B) \setminus A = \varnothing$.

Hint. Replace $X \setminus Y = X \cap \complement Y$, and use the De Morgan 's laws.

4. Prove that :
(a) $A \Delta (B \Delta C) = (A \Delta B) \Delta C$
(b) $A \cap (B \Delta C) = (A \cap B) \Delta (A \cap C)$
(c) $A \Delta A = \varnothing$
(d) $A \Delta B = A$ iff $B = \varnothing$
(e) $A \Delta B \subseteq (A \Delta C) \cup (B \Delta C)$ *and give an example when* $\subset$ *holds.*
Hint. Take $C = \varnothing$ and $A \cap B \neq \varnothing$ as an example in (e).

5. Let $\mathscr{A}$ be the set of all natural numbers that divide 30, and let us define
$x \vee y$ = *the least common multiple of x and y*
$x \wedge y$ = *the greatest common divisor of x and y*
$\rceil x = 30/x$.
Show that $\mathscr{A}$ is a Boolean algebra in which $x \Delta y = (x \vee y)/(x \wedge y)$, and represent $\mathscr{A}$ as an algebra of sets.

Hint. If $T = \{a, b, c\}$, then $\mathscr{P}(T)$ represents $\mathscr{A}$ as follows:

$\emptyset \leftrightarrow 1$, $\{a\} \leftrightarrow 2$, $\{b\} \leftrightarrow 3$, $\{c\} \leftrightarrow 5$, $\{a, b\} \leftrightarrow 6 = [2, 3] = 2 \cdot 3$, etc.,

hence T is determined by the prime divisors.

6. A filter $\mathscr{F} \subset \mathscr{P}(T)$ is said to be *tied* (*fixed* in [H-S], etc.) if $\cap \mathscr{F} \neq \emptyset$, and in the contrary case we say that it is *free*. Study whether

$$\mathscr{F} = \{\emptyset \neq A \in \mathscr{P}(T) \colon \complement A \textit{ is finite}\}$$

is a tied or free filter.

Hint. If T is finite, then A and $\complement A$ are concomitantly in $\mathscr{F}$, hence $[F_0]$ fails. $\mathscr{F}$ is free, since otherwise, if $x \in \cap \mathscr{F} \neq \emptyset$, then because $\{x\}$ is finite, we obtain $T \setminus \{x) \in \mathscr{F}$, which contradicts the hypothesis $x \in \cap \mathscr{F}$.

7. Let $(D, \leqslant)$ be a directed set. Show that

$$\mathscr{F} = \{A \subseteq D \colon \exists a \in D \textit{ such that } A \supseteq \{b \in D \colon b \geqslant a\}\}$$

is a filter in D. Moreover, if $f \colon D \longrightarrow T$ is an arbitrary net, then $f(\mathscr{F}) \subset \mathscr{P}(T)$ is a filter too (called *elementary* filter attached to the net f). Compare $f(\mathscr{F})$ by inclusion in $\mathscr{P}(\mathscr{P}(T))$ to the elementary filter attached to a subnet of f.

Hint. The elementary filter attached to a subnet is greater.

8. Let $\mathfrak{F}(T)$ be the set of all proper filters $\mathscr{F} \subset \mathscr{P}(T)$, ordered by inclusion. A filter $\mathscr{F}$, which is maximal relative to this order, is called *ultrafilter*. Show that $\mathscr{F}$ is an ultrafilter iff, for every $A \subseteq T$, either $A \in \mathscr{F}$ or $\complement A \in \mathscr{F}$ hold. Deduce that each ultrafilter in a finite set T is tied.

Hint. Let $\mathscr{F}$ be maximal. If $A \cap B = \emptyset$ for some $B \in \mathscr{F}$, then $\complement A \supseteq B$, hence $\complement A \in \mathscr{F}$. If $A \cap B \neq \emptyset$ for all $B \in \mathscr{F}$, then filter $\mathscr{F} \cup \{A\}$ is greater than $\mathscr{F}$, which contradicts the fact that $\mathscr{F}$ is maximal.

Conversely, let $\mathscr{F}$ be a filter for which either $A \in \mathscr{F}$ or $\complement A \in \mathscr{F}$ hold for all $A \subseteq T$. If filter $\mathscr{G}$ is greater than $\mathscr{F}$, and $A \in \mathscr{G} \setminus \mathscr{F}$, then $\complement A \in \mathscr{F}$. But $A \in \mathscr{G}$ and $\complement A \in \mathscr{G}$ cannot hold simultaneously in proper filters.

9. In duality to filters, the *ideals* $\mathscr{I} \subset \mathscr{P}(T)$ are defined by putting T in the place of $\emptyset$ in $[F_0]$, $\cup$ instead of $\cap$ in $[F_1]$, and $\subseteq$ instead of $\supseteq$ in $[F_2]$. Show that if $\mathscr{F} \subset \mathscr{P}(T)$ is a filter, then

$$\mathscr{I} = \{ A \subseteq T : \complement A \in \mathscr{F} \}$$

is an ideal. Reformulate and solve the above problems 6-8 for ideals.
Hint. Each ideal is dual to a filter of complementary sets.

10. If $\mathcal{R}$ is a relation on T, let us define

$$\mathcal{R}^0 = \delta,\ \mathcal{R}^{i+1} = \mathcal{R}^i \circ \mathcal{R},\ \mathcal{R}^T = \cup\{\mathcal{R}^i : i = 1,2,\ldots\},\ \text{and } \mathcal{R}^* = \mathcal{R}^0 \cup \mathcal{R}^T.$$

Show that :

(a) $\mathcal{R}^T$ is transitive (also called *transitive closure of* $\mathcal{R}$);

(b) $\mathcal{R}^*$ is a preorder (called *reflexive and transitive closure*);

(c) If $\mathcal{R} \subseteq \mathcal{S}$, then $\mathcal{R}^* \subseteq \mathcal{S}^*$;

(d) $\mathcal{R}^* \cup \mathcal{S}^* \subseteq (\mathcal{R} \cup \mathcal{S})^*$;

(e) $(\mathcal{R}^*)^* = \mathcal{R}^*$.

11. Let $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$ be functions. Show that :

1) $(h \circ g) \circ f = h \circ (g \circ f)$;
2) $f \circ I_X = f$, where I_X is the identity of X (i.e. $I_X(x) = x,\ \forall x \in X$);
3) f, g injective (surjective) $\Rightarrow$ $g \circ f$ injective (surjective);
4) $g \circ f$ injective (surjective) $\Rightarrow$ f injective (g surjective);
5) f, g bijective $\Rightarrow (g \circ f)^{-1} = f^{-1} \circ g^{-1}$;
6) $f[\overleftarrow{f}(B) \cap A] = B \cap f(A)$, but $\overleftarrow{f}[f(A) \cap B] \supseteq A \cap \overleftarrow{f}(B)$;
7) $f[\overleftarrow{f}(B)) \subseteq B$, with equality if f is surjective, and $\overleftarrow{f}(f(A)) \supseteq A$, with equality if f is injective (i.e. 1:1).

12. Let $f : X \to Y$ be a function, and suppose that there exists another function $g: Y \to X$, such that $g \circ f = I_X$ and $f \circ g = I_Y$. Prove that f must be 1:1 from X onto Y, and $g = f^{-1}$.

13. In $X = \mathbb{R}^2$ we define the relations:

$\Lambda = \{((x, y),(u, v)) :$ either $(x < u)$ or $(x = u$ *and* $y \le v)\}$;
$\Pi = \{((x, y),(u, v)) : x \le u$ *and* $y \le v\}$;
$K = \{((t, x),(s, y)) : s - t \ge |x - y|\}$.

Show that Λ is a total order (called *lexicographic*), but Π and K (called *product*, respectively *causality*) are partial orders. Find the corresponding cones of positive elements, establish the form of the order intervals, and study the order completeness.

14. In a library there are two types of books:
Class A, consisting of books cited in themselves, and

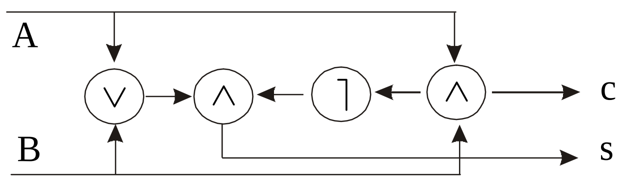
Class B, formed by the books not cited in themselves.
Classify the book in which the whole class B is cited.
Hint. Impossible. The problem reduces to decide whether class B belongs to B, which isn't solvable (for further details see [RM],[R-S], etc.).

15. Let us suppose that a room has three doors. Construct a switch net that allows to turn the light on and off at any of the doors.
Hint. Write the work function of a net depending on three switches *a, b, c,* starting for example, with $f(1,1,1) = 1$ as an initial state, and continuing with $f(1,1,0) = 0$, $f(1,0,0) = 1$,etc. ; attach a conjunction to each value 1 of the function f , e.g. $a \wedge b \wedge c$ to $f(1,1,1) = 1$, $a \wedge \rceil b \wedge \rceil c$ to $f(1,0,0) = 1$, etc.

16. Construct a logical circuit, which realizes the addition of two digits in the base 2. How is the addition to be continued by taking into account the third (*carried*) digit?
Hint. Adding two digits *A* and *B* gives a two-digit result:

A	*B*	*c*	*s*
1	*1*	*1*	*0*
0	*1*	*0*	*1*
1	*0*	*0*	*1*
0	*0*	*0*	*0*

where *s* is the *sum-digit* and *c* is the *carried-digit*. We can take $c = A \wedge B$ and $s = (A \vee B) \wedge \rceil (A \wedge B)$. The circuit has the form (called *semi-summarizer*, or *half-adder*):



Adding three digits yields a two-digits result too, as in the following table:

A	B	C	c	s
1	1	1	1	1
0	1	1	1	0
1	0	1	1	0
0	0	1	0	1
1	1	0	1	0
0	1	0	0	1
1	0	0	0	1
0	0	0	0	0

The resulting digits c and s may be obtained by connecting two semi-summarizes into a *complete summarizer* (alternatively called *full-adder,* as in [ME], etc.).

§ I.2. NUMBERS

The purpose of this paragraph is to provide the student with a unitary idea about the diagram:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \begin{array}{l} \nearrow \mathbb{C} \subset \mathbb{K} \\ \quad \ldots \\ \searrow \mathbb{D} \end{array}$$

2.1. <u>Definition.</u> We consider that two sets *A* and *B* (parts of a total set *T*) are *equivalent,* and we note *A*~*B*, iff there exists a 1:1 correspondence between the elements of *A* and *B*. Intuitively, this means that the two sets have "the same number of elements", "the same power", etc. The equivalence class generated by *A* is called *cardinal number*, and it is marked by $\overline{\overline{A}}$ = card *A*.

There are some specific *signs* to denote individual cardinals, namely:

2.2. <u>Notations</u>. card $\varnothing = 0$ (convention!)

card $A = 1$ iff *A* is equivalent to the set of natural satellites of Terra;

card $A = 2$ iff *A* is equivalent to the set of magnetic poles;

. . .

card $A = n+1$ iff for any $x \in A$ we have card $(A \setminus \{x\}) = n$;

. . .

All these cardinals are said to be *finite,* and they are named *natural number*. The set of all finite cardinals is noted by ℕ, and it is called *set of natural numbers*. If *A* ~ ℕ, then we say that *A* is *countable*, and we note card $A = \aleph_0$ (or card $A = c_0$, etc.), which is read *aleph naught*. If $A \sim \mathscr{P}(\mathbb{N})$, then *A has the power of continuum,* noted card *A* = card $(2^{\mathbb{N}}) = \aleph = 2^{\aleph_0}$ (or card $A = c$, etc.), where $2^{\mathbb{N}} \sim \mathscr{P}(\mathbb{N})$.

In order to compare and compute with cardinals, we have to specify the inequality and the operations for cardinals. If *a* = card *A*, and *b* = card *B*, then we define:

1. $a \le b$ iff there is C such that $A \sim C \subseteq B$;
2. $a + b = \text{card}(A \cup B)$, where $A \cap B = \varnothing$;
3. $a \cdot b = \text{card } (A \times B)$; and
4. $a^b = \text{card}(A^B)$.

Now we can formulate **the most significant properties of $\aleph_0$ and $\aleph$:**

2.3. <u>Theorem</u>. = is an equivalence, and $\leq$ is a total order of cardinals. In these terms, the following formulas hold:

(i) $\aleph_0 < 2^{\aleph_0} = \aleph$;

(ii) $\aleph_0 + \aleph_0 = \aleph_0$, $\aleph_0\aleph_0 = \aleph_0$;

(iii) $\aleph + \aleph = \aleph$, and $\aleph\aleph = \aleph$.

The proof can be found in [HS], etc., and will be omitted.

To complete the image, we mention that according to an axiom, known as the *Hypothesis of Continuum*, there is no cardinal between $\aleph_0$ and $\aleph$.

The *ARITHMETIC* of $\mathbb{N}$ is based on the following axioms:

2.4. <u>Peano's Axioms.</u>

[P_1] 1 is a natural number (alternatively we can start with 0);

[P_2] For each $n \in \mathbb{N}$, there exists *the next one*, noted $n' \in \mathbb{N}$;

[P_3] For every $n \in \mathbb{N}$, we have $n' \neq 1$;

[P_4] $n = m$ iff $n' = m'$;

[P_5] If $1 \in \mathbb{P}$, and [$n \in \mathbb{P}$ implies $n' \in \mathbb{P}$], then $\mathbb{P} = \mathbb{N}$.

The last axiom represents the well-known *induction principle.*

The arithmetic on $\mathbb{N}$ involves an order relation, and algebraic operations:

2.5. <u>Definition.</u> If $n, m \in \mathbb{N}$, then :

1. $n \leq m$ holds iff there exists $p \in \mathbb{N}$ such that $m = n + p$;
2. $n + 1 = n'$, and $n + m' = (n + m)'$ *(addition);*
3. $n \cdot 1 = n$, and $n \cdot m' = n \cdot m + n$ *(multiplication).*

We may precise that the *algebraical operations* are defined by induction.

2.6. <u>Remark.</u> It is easy to verify that ($\mathbb{N}$, +) and ($\mathbb{N}$, **.**) are commutative semi-groups with units, and ($\mathbb{N}$, $\leq$) is totally ordered (see [MC], [ŞG], etc.). The fact that ($\mathbb{N}$, +) is not a group expresses the impossibility of solving the equation $a + x = b$ for arbitrary $a, b \in \mathbb{N}$. In order to avoid this inconvenience, set $\mathbb{N}$ was enlarged to the so-called *set of integers*. The idea is to replace the difference $b - a$, which is not always meaningful in $\mathbb{N}$, by a pair (a, b), and to consider $(a, b) \sim (c, d)$ iff $a + c = b + d$. The *integers* will be classes $(a, b)^\wedge$ of equivalent pairs, and we note the set of all integers by

$$\mathbb{Z} = \mathbb{N} \times \mathbb{N} / \sim .$$

The operations and the order relation on $\mathbb{Z}$ are defined using arbitrary representatives of the involved classes, and we obtain:

2.7. <u>Theorem.</u> ($\mathbb{Z}$,+, **.**) is a commutative ring with unit, and ($\mathbb{Z}$, $\leq$) is totally ordered such that $\leq$ is compatible to the algebraical structure of ring.

2.8. **Remark.** Set $\mathbb{Z}$ is not a field, i.e. equation $ax = b$ is not always solvable. Therefore, similarly to $\mathbb{N}$, $\mathbb{Z}$ was enlarged, and the new numbers are called *rationals*. More exactly, instead of a quotient b/a , we speak of a pair (a, b), and we define an equivalence $(a, b) \sim (c, d)$ by $ad = bc$. The *rational numbers* are defined as equivalence classes $(a, b)^\wedge$, and the set of all these numbers is noted $\mathbb{Q} = \mathbb{Z} \times \mathbb{Z} / \sim$.

Using representatives, we can extend the algebraical operations, and the order relation, from $\mathbb{Z}$ to $\mathbb{Q}$, and we obtain:

2.9. **Theorem.** $(\mathbb{Q} , + , .)$ is a field. It is totally ordered, and the order $\leq$ is compatible with the algebraical structure.

2.10. **Remark.** Because $\mathbb{Q}$ already has convenient algebraical properties, the next extension is justified by another type of arguments. For example,

$$1;\ 1.4;\ 1.41;\ 1.414;\ 1.4142;\ \ldots$$

which are obtained by computing $\sqrt{2}$, form a bounded set in $\mathbb{Q}$, for which there is no supremum (since $\sqrt{2} \notin \mathbb{Q}$). Of course, this "lack" of elements is a weak point of $\mathbb{Q}$. Reformulated in practical terms, this means that equations of the form $x^2 - 2 = 0$ cannot be solved in $\mathbb{Q}$.

There are several methods to complete the order of $\mathbb{Q}$; the most frequent is based on the so-called Dedekind's *cuts*. By definition, a *cut* in $\mathbb{Q}$ is any pair of parts (A, B), for which the following conditions hold:

(i) $A \cup B = \mathbb{Q}$;

(ii) $a < b$ whenever $a \in A$ and $b \in B$ (hence $A \cap B = \emptyset$);

(iii) $[(a' \leq a \in A) \Rightarrow a' \in A]$, and $[(b' \geq b \in B) \Rightarrow b' \in B]$.

Every rational number $x \in \mathbb{Q}$ generates a cut, namely (A_x , B_x) , where

$$A_x = \{a \in \mathbb{Q} : a \leq x\}, \text{ and } B_x = \{b \in \mathbb{Q} : b > x\}.$$

There are still cuts which cannot be defined on this way, as for example $A = \mathbb{Q} \setminus B$, and $B = \{x \in \mathbb{Q}_+ : x^2 > 2\}$; they define the *irrational* numbers.

2.11. **Definition.** Each cut is called *real number*. The set of all real numbers is noted $\mathbb{R}$. A real number is *positive* iff the first part of the corresponding cut contains positive rational numbers. The addition and the multiplication of cuts reduce to similar operations with rational numbers in the left and right parts of these cuts.

2.12. **Theorem.** $(\mathbb{R}, +, .)$ is a field. Its order $\leq$ is compatible with the algebraical structure of $\mathbb{R}$; $(\mathbb{R}, \leq)$ is a completely and totally ordered set.

2.13. Remark. The other constructions of $\mathbb{R}$ (e.g. the Cantor's equivalence classes of Cauchy sequences, or the Weierstrass method of continuous fractions) lead to similar properties. More than this, it can be shown that the complete and totally ordered fields are all isomorphic, so we are led to the possibility of introducing real numbers in *axiomatic* manner:

2.14. Definition. We call *set of real numbers*, and we note it $\mathbb{R}$, the unique set (up to an isomorphism), for which:

1. $(\mathbb{R}, +, \cdot)$ is a field;
2. $\leq$ is a total order on $\mathbb{R}$, compatible with the structure of a field ;
3. $(\mathbb{R}, \leq)$ is complete (more exactly, every nonvoid upper bounded subset of $\mathbb{R}$ has a supremum, which is known as *Cantor's axiom*).

2.15. Remark. Taking the Cantor's axiom as a starting point of our study clearly shows that the entire *Real Analysis* is essentially based on the order completeness of $\mathbb{R}$. At the beginning, this fact is visible in the limiting process involving sequences in $\mathbb{R}$ (i.e. in *convergence* theory), and later it is extended (as in §II.2, etc.) to the general notion of *limit* of a function.

We remember that the notion of *convergence* is nowadays presented in a very general form in the lyceum textbooks, namely:

2.16. Definition. A number $l \in \mathbb{R}$ is called *limit of the sequence* (x_n) of real numbers (or x_n *tends to* l in the space $S = \mathbb{R}$, etc.), and we note

$$l = \lim_{n\to\infty} x_n ,$$

(or $x_n \to l$, etc.) iff any neighborhood $(l - \varepsilon, l + \varepsilon)$, of l, contains all the terms starting with some rank, i.e.

$$\forall\ \varepsilon > 0\ \exists\, n_0(\varepsilon) \in \mathbb{N} \text{ such that } [n > n_0(\varepsilon) \Rightarrow |x_n - l| < \varepsilon].$$

If a sequence has a limit it is said to be *convergent*, and otherwise it is considered *divergent*.

Among the most important consequences of the axioms of $\mathbb{R}$ (due to Cantor, Weierstrass, etc.) we mention the following basic theorem:

2.17. Theorem. (Cantor). If $([a_n, b_n])_{n\in\mathbb{N}}$ is a decreasing sequence of closed intervals in $\mathbb{R}$, i.e.

$$[a_0, b_0] \supseteq [a_1, b_1] \supseteq \ldots \supseteq [a_n, b_n] \supseteq \ldots$$

then:

a) The sequences (a_n) and (b_n) are convergent;

b) $\mathscr{I} \equiv \cap\{[a_n, b_n] : n \in \mathbb{N}\} \neq \varnothing$;

c) $[\inf\{b_n - a_n : n \in \mathbb{N}\} = 0] \Rightarrow [\lim_{n\to\infty} a_n = l = \lim_{n\to\infty} b_n] \Rightarrow \mathscr{I} = \{l\}$.

Proof. *a)* Sequence (a_n) is increasing and bounded by each b_m, hence there exists $\alpha = \sup a_n$. Consequently, for any $\varepsilon > 0$ there exists $n(\varepsilon) \in \mathbb{N}$ such that $\alpha - \varepsilon < a_{n(\varepsilon)} \le \alpha$, hence also $\alpha - \varepsilon < a_{n(\varepsilon)} \le a_n \le \alpha$, whenever $n > n(\varepsilon)$. This means that $\alpha = \lim_{n\to\infty} a_n$. Similarly, $\beta = \inf b_n$ is the limit of (b_n).

b) $\mathscr{I} \neq \varnothing$ because $\alpha \le \beta$, hence $\mathscr{I} \supseteq [\alpha, \beta]$. At its turn, $\alpha \le \beta$ must be accepted since the contrary, namely $\beta < \alpha$, would lead to $\beta < a_p$ and $b_q < \alpha$ for some $p, q \in \mathbb{N}$, and further $\beta \le b_s < a_p$ and $b_q < a_t \le \alpha$ for some $s, t \in \mathbb{N}$, which would contradict the very definition of α and β.

c) Of course, if $\inf \{b_n - a_n : n \in \mathbb{N}\} = 0$, then $\alpha = \beta$, because for arbitrary $n \in \mathbb{N}$ we have $b_n - a_n \ge \beta - \alpha$. If we note the common limit by l, then we finally find that $\mathscr{I} = \{l\}$. ◇

There are several more or less immediate but as for sure useful consequences of this theorem, as follows:

2.18. Corollary. The following order properties hold:

a) If $a \ge 0$ is fixed in $\mathbb{R}$, and $a \le \frac{1}{n}$ is valid for any $n \in \mathbb{N}^*$, then $a = 0$;

b) The sequence $\left(\frac{1}{n}\right)$ is convergent to 0 ;

c) Any increasing and upper bounded sequence in $\mathbb{R}$ is convergent, as well as any decreasing and lower bounded one (but do not reduce the convergence to these cases concerning monotonic sequences!);

d) If (x_n) is a sequence in $\mathbb{R}$, $x_n \in [a_n, b_n]$ for all $n \in \mathbb{N}$, and the conditions of the above theorem hold, then $x_n \to l$ (the "pincers" test).

e) If $a_n \to 0$, $l \in \mathbb{R}$, and $|x_n - l| < a_n$ holds for all $n \in \mathbb{N}$, then $x_n \to l$.

2.19. Remark. In spite of the good algebraical and order properties of $\mathbb{R}$, the necessity of solving equations like $x^2 + 1 = 0$ has led to another extension of numbers. More exactly, we are looking now for an *algebraically closed* field $\mathbb{C}$, i.e. a field such that every algebraical equation with coefficients from $\mathbb{C}$ has solutions in $\mathbb{C}$. To avoid discussions about the condition $i^2 = -1$, which makes no sense in $\mathbb{R}$, we introduce the new type of numbers in a *contradiction free fashion*, namely:

2.20. Definition. We say that $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ is the *set of complex numbers (in axiomatical form)* if it is endowed with the usual equality, and with the operations of *addition* and *multiplication* defined by:

$$(a, b) + (c, d) = (a + c, b + d),$$
$$(a, b)(c, d) = (ac - bd, ad + bc).$$

2.21. <u>Theorem.</u> $(\mathbb{C}, +, \cdot)$ is a field that contains $(\mathbb{R}, +, \cdot)$ as a subfield, in the sense that $\lambda \mapsto (\lambda, 0)$ is an embedding of $\mathbb{R}$ in $\mathbb{C}$, which preserves the algebraic operations.

The *axiomatical form* in the above definition of $\mathbb{C}$ is valuable in theory, but in practice we prefer simpler forms, like:

2.22. <u>Practical representations of ℂ.</u> By replacing $(\lambda, 0)$ by λ in the above rule of multiplying complex numbers, we see that $\mathbb{C}$ forms a linear space of dimension 2 over $\mathbb{R}$ (see §I.3, [V-P], [AE], etc.). In fact, let $\mathscr{B} = \{u, i\}$, where $u = (1,0)$ and $i = (0,1)$, be the fundamental base of this linear space. It is easy to see that u is the unit of $\mathbb{C}$ (corresponding to $1 \in \mathbb{R}$), and $i^2 = -u$. Consequently each complex number $z = (a, b)$ can be expressed as

$$z = au + bi = a + bi,$$

which is called ***algebraical*** *(traditional) form.* The components a and b of the complex number $z = (a, b)$ are called *real*, respectively *imaginary* parts of z, and they are usually noted by

$$a = \mathrm{Re}\, z, \; b = \mathrm{Im}\; z.$$

Starting with the same axiomatic form $z=(a, b)$, the complex numbers can be presented in a ***geometrical*** *form* as points in the 2-dimensional linear space $\mathbb{R}^2$, when $\mathbb{C}$ is referred to as a *complex plane.* Addition of complex numbers in this form is defined by the well-known *parallelogram's rule*, while the multiplication involves geometric constructions, which are more complicated (better explained by the *trigonometric* representation below). The geometric representation of $\mathbb{C}$ is advisable whenever some geometric images help intuition.

Replacing the Cartesian coordinates a and b of $z = (a, b)$ from the initial geometric representation by the *polar* ones (see Fig.I.2.1. below), we obtain the *modulus*

$$\rho = |z| = \sqrt{a^2 + b^2},$$

and the *argument,*

$$\theta = \arg z = \begin{cases} \alpha & \text{for } z \text{ in the quadrant I} \\ \pi + \alpha & \text{for } z \text{ in the quadrants II,III} \\ 2\pi - \alpha & \text{for } z \text{ in the quadrant IV} \end{cases}$$

where $\alpha = \operatorname{arctg} \dfrac{b}{a} \in \left(-\dfrac{\pi}{2}, +\dfrac{\pi}{2}\right)$. So we are led to the ***trigonometric*** *form* of the complex number $z = (\rho, \theta) \in \mathbb{R}_+ \times [0, 2\pi)$, namely

$$z = \rho\,(\cos\theta + i \sin\theta).$$

We mention that the complex *modulus* $|z|$ reduces to the usual *absolute value* if $z \in \mathbb{R}$, and the *argument* of z generalizes the notion of *sign* from the real case. Using the unique non-trivial (i.e. different from identity) idempotent automorphism of $\mathbb{C}$, namely $z = a + ib \mapsto \bar{z} = a - ib$, called *conjugation*, which realizes a symmetry relative to the real axis, we obtain $|z| = (z \cdot \bar{z})^{1/2}$, i.e. the norm derives from algebraical properties.

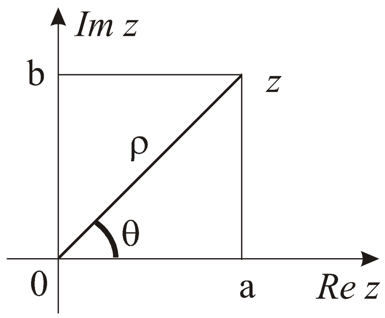


Fig.I.2.1.

The complex numbers can also be presented in the ***matrix*** *form*

$$z = \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

where a, $b \in \mathbb{R}$, based on the fact that $\mathbb{C}$ is isomorphic to that subset of $\mathcal{M}_{2,2}(\mathbb{R})$, which consists of all matrices of this form .

Finally we mention *the* ***spherical*** *form* of the complex numbers, that is obtained by the so called *stereographical projection.* Let $\mathscr{S}$ be a sphere of diameter $ON = 1$, which is tangent to the complex plane $\mathbb{C}$ at its origin. Each straight line, which passes through N and intersects $\mathbb{C}$, intersects $\mathscr{S}$ too. Consequently, every complex number $z = x + iy$, expressed in $\mathbb{R}^3$ as $z(x,y,0)$, can be represented as a point $P(\xi, \eta, \zeta) \in \mathscr{S} \setminus \{N\}$ (see Fig.I.2.2.). This correspondence of $\mathscr{S} \setminus \{N\}$ to $\mathbb{C}$ is called *stereographical projection*, and $\mathscr{S}$ is known as the Riemann's sphere (see the analytical expression of the correspondence of z to P in problem 9 at the end of this section).

The Riemann's sphere is especially useful in explaining why $\mathbb{C}$ has a single point at infinity, simply denoted by ∞ (with no sign in front!), which is the correspondent of the *North pole* $N \in \mathscr{S}$ (see §II.2.).

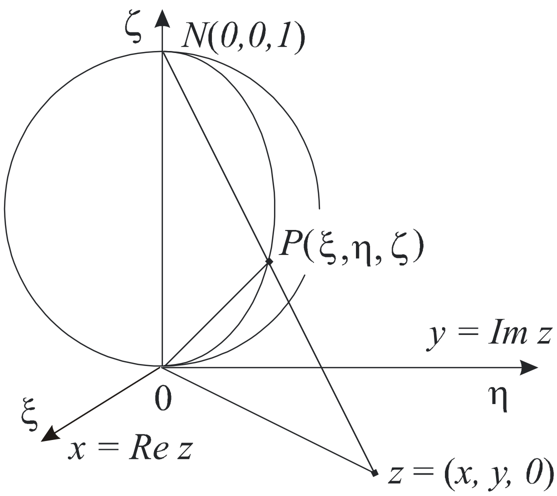


Fig.I.2.2.

The fact that $\mathbb{C}$ is algebraically closed (considered to be *the fundamental theorem of Algebra*) will be discussed later (in chapter VIII), but the special role of $\mathbb{C}$ among the other sets of numbers can be seen in the following:

2.23. Theorem. (Frobenius). The single real algebras with division (i.e. each non-null element has an inverse), of finite dimension (like a linear space), are $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{K}$ (the set of *quaternions;* see [C-E], etc.).

In other words, these theorems say that, from algebraical point of view, $\mathbb{C}$ is the best system of numbers. However, the "nice" and "powerful" order structure of the field $\mathbb{R}$ is completely lost in $\mathbb{C}$. More exactly:

2.24. Proposition. There is no order on $\mathbb{C}$, to be compatible with its algebraical structure (but different from that of $\mathbb{R}$).

Proof. By *reductio ad absurdum* (r.a.a.), let us suppose that $\leq$ is an order relation on C such that the following conditions of compatibility hold:

$$z \leq Z \text{ and } \forall\ \zeta \in C \Rightarrow z + \zeta \leq Z + \zeta;$$
$$z \leq Z \text{ and } 0 \leq \zeta \Rightarrow z\zeta \leq Z\zeta.$$

In particular, $0 \leq z$ implies $0 \leq z^n$ for all $n \in \mathbb{N}$. On the other hand from z and ζ positive in $\mathbb{C}$, and λ and μ positive in $\mathbb{R}$, it follows that $\lambda z + \mu \zeta$ is positive in $\mathbb{C}$. Consequently, if we suppose that $0 \leq z \in \mathbb{C} \setminus \mathbb{R}_+$, then all the elements of $\mathbb{C}$ should be positive, hence the order $\leq$ would be trivial. ◇

2.25. Remark. We mention that $\mathbb{R}$ can be extended to other ordered algebras, but we have to renounce several algebraical properties. Such an alternative is the algebra of *double numbers*, $\mathbb{D} = \mathbb{R} \times \mathbb{R}$, where, contrarily to $i^2 = -1$, we accept that $j^2 = +1$, i.e. $(0,1)^2 = (1,0)$.

The list of systems of numbers can be continued; in particular, the spaces of dimension 2^n can be organized as *Clifford Algebra* (see [C-E], etc.).

Besides numbers, there are other mathematical entities, called *vectors, tensors, spinors,* etc., which can adequately describe the different quantities that appear in practice (see [B-S-T], etc.).

Further refinements of the present classification are possible. For example, we may speak of *algebraic* numbers, which can be roots of an algebraical equation with coefficients from $\mathbb{Z}$, respectively *transcendent* numbers, which cannot. For example, $\sqrt{2}$ is algebraic, while $\boldsymbol{e}$ and π are transcendent (see [FG], [ŞG], etc.).

Classifying a given number is sometimes quite difficult, as for example showing that $\pi \in \mathbb{R} \setminus \mathbb{Q}$. Even if we work with π very early, proving its irrationality is still a subject of interest (e.g. [MM]). The following examples are enough sophisticated, but accessible in the lyceum framework (if necessary, see §III.3, §V.1, etc.).

2.26. <u>Proposition.</u> Let us fix $p, q, n \in \mathbb{N}^*$, and note

$$P_n(x) = \frac{x^n (qx-p)^n}{n!}, \text{ and } x_n = \int_0^{\pi} P_n(x) \sin x \, dx .$$

We claim that :

1. Both $P_n^{(s)}(0)$, and $P_n^{(s)}\left(\frac{p}{q}\right) \in \mathbb{Z}$ for any order of derivation $s \in \mathbb{N}$;
2. $\lim\limits_{n\to\infty} x_n = 0$;
3. $\pi \in \mathbb{R} \setminus \mathbb{Q}$.

<u>Proof.</u> 1. If we identify the binomial development of P_n with its Taylor formula (compare to proposition 17, §III.3), then we can write:

$$P_n(x) = \sum_{k=0}^{n} (-1)^{n-k} \frac{1}{n!} C_n^k p^{n-k} q^k x^{n+k} = \sum_{s=0}^{2n} \frac{1}{s!} P_n^{(s)}(0) x^s ,$$

where

$$P_n^{(s)}(0) = \begin{cases} 0 & \text{if } s < n \text{ or } s > 2n \\ (-1)^{n-k} \dfrac{(n+k)!}{n!} C_n^k p^{n-k} q^k & \textit{if } s = n+k, k = \overline{1, n} \end{cases}$$

Because $\frac{(n+k)!}{n!} \in \mathbb{N}$, as well as $C_n^k \in \mathbb{N}$, it follows that $P_n^{(s)}(0) \in \mathbb{Z}$ for all s in $\mathbb{N}$, including $s = 0$ (when P_n is not derived).

Changing the variable, $x \mapsto t = x - \frac{p}{q}$, we obtain

$$P_n(x) = \frac{t^n (qt+p)^n}{n!} = Q_n(t),$$

hence $P_n^{(s)}\left(\frac{p}{q}\right) \in \mathbb{Z}$ reduces to $Q_n^{(s)}(0) \in \mathbb{Z}$, which is proved like the former membership $P_n^{(s)}(0) \in \mathbb{Z}$.

2. If we note $\mu = sup\{| x(qx-p) | : x \in [0, \pi]\}$, we obtain

$$| x_n | \le \int_0^{\pi} |P_n(x)\sin x| dx \le \int_o^{\pi} \frac{\mu^n}{n!} dx = \pi \frac{\mu^n}{n!},$$

which shows that $x_n \to 0$ ($\notin \mathbb{Z}^*$!).

3. Integrating $2n+1$ times by parts, we obtain

$$x_n = - \cos x \, [P_n(x) - \ldots + (-1)^n P_n^{(2n)}(x)] \Big|_0^{\pi} .$$

If we accept that $\pi = \frac{p}{q} \in \mathbb{Q}$, then from property 1 we will deduce $x_n \in \mathbb{Z}^*$, which contradicts 2. ◇

2.27. <u>Convention.</u> Through this book we adopt the notation Γ for any one of the fields $\mathbb{R}$ or $\mathbb{C}$, especially to underline that some properties are valid in both real and complex structures (e.g. see the real and the complex linear spaces in §I.3, etc.).

A special attention will be paid to the complex analysis, which turns out to be the natural extension and even explanation of many results involving real variables. Step by step, the notion of *real function of a real variable* is extended to that of *complex function of a complex variable*:

2.28. <u>Extending functions</u> from $\mathbb{R}$ to $\mathbb{C}$ may refer to the variable, or to the values. Consequently, we have 3 types of extensions:

a) *Complex functions of a real variable*. They have the form

$$f : I \to \mathbb{C}, \text{ where } I \subseteq \mathbb{R},$$

and represent *parameterizations* of curves in $\mathbb{C}$ (compare to §VI.1). These functions are obtained by combining the real parametric equations of the curves. For example, the real equations of a <u>*straight line*</u> which passes through z_0 and has the direction ζ, lead to the complex function

$$z = z_0 + t\,\zeta, \quad t \in \mathbb{R}.$$

Similarly, the <u>*circle*</u> of center z_0 and radius r, in $\mathbb{C}$, has the parameterization

$$z = z_0 + r(\cos t + i \sin t), \quad t \in [0, 2\pi).$$

b) *Real functions of a complex variable,* which are written as

$$f : D \to \mathbb{R}, \text{ where } D \subseteq \mathbb{C}.$$

They have a complex variable, but real values, and the simplest examples are $|\cdot|$, *arg*, Re, and Im. Their graphs can be done in $\mathbb{R}^3 \approx \mathbb{C} \times \mathbb{R}$.

c) *Complex functions of one complex variable.* They represent the most important case, which is specified as

$$f : D \to \mathbb{C}\text{ , where } D \subseteq \mathbb{C}\ .$$

The assertion "*D* is the *domain* of *f* " is more sophisticated than in $\mathbb{R}$. More exactly, it means that:

- f is defined on D ;
- D is open in the Euclidean structure if $\mathbb{C}$;
- D is connected in the same structure (see §III.2. later).

The action of f is frequently noted $Z = f(z)$, which is a short form for:

$$\mathbb{C} \supseteq D \ni z \xrightarrow{\ f\ } f(z) = Z \in \mathbb{C}\ .$$

If we identify $z = x + iy \in \mathbb{C}$ with $(x, y) \in \mathbb{R}^2$, then f can be expressed by two real functions of two real variables, namely

$$f(z) = P(x, y) + i\, Q(x, y),$$

where the components P and Q are called *real part*, respectively *imaginary part* of f . This form of f is very convenient when we are looking for some geometric interpretation. Drawing graphs of such functions is impossible since $\mathbb{C} \times \mathbb{C} \sim \mathbb{R}^4$, but they can be easily represented as transformations of some plane domains (no matter if real or complex). In fact, if $Z = X + iY$, then the action of f is equivalently described by the real equations

$$\begin{cases} X = P(x, y) \\ Y = Q(x, y) \end{cases} ,\ (x, y) \in D \subseteq \mathbb{R}^2 \sim \mathbb{C}\ .$$

In other words, considering $f = (P, Q)$, we practically reduce the study of complex functions of a complex variable to that of real vector functions of two real variables. On this way, many problems of complex analysis can be reformulated and solved in real analysis. This method will be intensively used in §III.4 (see also [HD], [CG], etc.).

Alternatively, if z, the argument of f , is expressed in trigonometric form, then the image through f becomes

$$f(z) = P(\rho, \theta) + i\, Q(\rho, \theta)\ .$$

If we use the polar coordinates in the image plane to precise $f(z)$ by its modulus $| f(z)| = M(x, y)$, and its argument $\arg f(z) = A(x, y)$, then

$$f(z) = M(x, y)[\cos A(x, y) + i \sin A(x, y)]\ .$$

Finally, if both z and Z are represented in trigonometric form, then

$$f(z) = R(\rho, \theta)[\cos B(\rho, \theta) + i \sin B(\rho, \theta)].$$

PROBLEMS § I.2.

1. Let A be an infinite set (i.e. $card\ A \geq \aleph_0$), and let us fix $a \in A$. Show that

$$A \sim [A \setminus \{a\}].$$

Hint. Consider a sequence (x_n) in A, such that $x_0 = a$, and define a bijection $f : A \to [A \setminus \{a\}]$, e.g.

$$f(x) = \begin{cases} x & if \quad x \neq x_n \\ x_{n+1} & if \quad x = x_n \end{cases}.$$

2. Show that there are infinitely many prime numbers (in $\mathbb{N}$).

Hint. If $2,3,5,\ldots,p$ are the former prime numbers, then $n = 2 \cdot 3 \cdot 5 \cdot \ldots \cdot p + 1$ is another prime number, and obviously $n > p$.

3. Show that $\sqrt{2}$, e, $ln\ 2 \notin \mathbb{Q}$.

Hint. In the contrary case, we should have $\sqrt{2} = \dfrac{p}{q}$, with p and q relatively prime integers. The relation $p^2 = 2q^2$ shows that both p and q are even.

To study e , let us note a partial sum of its series by

$$s_n = \sum_{k=0}^{n} \frac{1}{k!},$$

and evaluate

$$e - s_n = \frac{1}{(n+1)!}\left[1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \ldots\right] \leq$$

$$\frac{1}{(n+1)!}\left[1 + \frac{1}{n+2} + \left(\frac{1}{n+2}\right)^2 + \ldots\right] = \frac{n+2}{(n+1)!(n+1)} < \frac{1}{nn!}$$

Thus we obtain $en! - s_n n! < \dfrac{1}{n}$, where $s_n n! \in \mathbb{N}$. If we accept that $e = \dfrac{p}{q}$ for some p, $q \in \mathbb{Z}$, then $en! \in \mathbb{Z}$ too, for enough large n, but it is impossible the difference of two integers to be under $\dfrac{1}{n}$.

Finally, $ln\ 2 = \dfrac{p}{q}$ means $e^p = 2^p$, hence e should be even (nonsense).

4. Compare the real numbers $sin\ 1$, $sin\ 2$, and $sin\ 3$.

Hint. Develop $sin\ 2\alpha$, and use $3 \cong \pi$.

5. Write in the binary system (basis 2) the following numbers (given in basis 10): $15 \in \mathbb{N}$; $-7 \in \mathbb{Z}$; $2/3$, 0.102, $-2.036 \in \mathbb{Q}$; $\pi, \sqrt{2} \in \mathbb{R} \setminus \mathbb{Q}$.
What gives the converse process?

6. Prove by induction that for any $n \in \mathbb{N}^*$ we have:

a) $\frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} \leq \frac{2n-1}{n}$,and

b) $(1+x)^n \geq 1 + nx$ if $x > -1$.

7. Verify the sub-additivity of the absolute value in $\mathbb{R}$ and $\mathbb{C}$.
Hint. It is sufficient to analyze the case of $\mathbb{C}$, where we can use the Euclidean structure of the complex plane, generated by the *scalar product*

$$\langle (x, y), (u, v) \rangle = xu + yv.$$

8. Find the formulas which correlate the coordinates of $P \in \mathscr{S}$, and $z \in \mathbb{C}$ through the stereographical projection. Use them to show that the image of any circle on the sphere is either circle or straight line in the plane.
Hint. $N(0,0,1)$, $P(\xi, \eta, \zeta)$ and $z(x, y, 0)$ are collinear, hence

$$\frac{\xi}{x} = \frac{\eta}{y} = \frac{\zeta - 1}{-1} = \frac{PN}{zN} = \frac{ON^2}{zN^2} = \frac{1}{x^2 + y^2 + 1}.$$

Point $P \in \mathscr{S}$ is on a circle if in addition it belongs to a plane

$$A\,\xi + B\,\eta + C\,\zeta + D = 0.$$

The image is a straight line iff N belongs to the circle, i.e. $C + D = 0$.

9. Write the parameterization of the following curves in the plane $\mathbb{C}$: ellipse, hyperbola, cycloid, asteroid, Archimedes's spiral, cardioid, and the Bernoulli's lemniscates.
Hint. We start with the corresponding real parameterizations in Cartesian or polar coordinates, which are based on the formulas:

Ellipse: $x = a \cos t,\ y = b \sin t,\ t \in [0, 2\pi]$;

Hyperbola: $x = a \operatorname{ch} t,\ y = b \operatorname{sh} t,\ t \in \mathbb{R}$;

Parabola: $y = ax^2 + bx + c,\ x \in \mathbb{R}$;

Cycloid: $x = a(t - \sin t),\ y = a(1 - \cos t),\ t \in [0, 2\pi]$;

Asteroid: $x = a \cos^3 t,\ y = b \sin^3 t,\ t \in [0, 2\pi]$;

Archimedes's spiral: $r = k\theta,\ \theta > 0$;

Cardioid: $r = a(1 + \cos\theta),\ \theta \in (-\pi, +\pi)$; and

Bernoulli's lemniscate: $r^2 = 2a^2 \cos 2\theta,\ \theta \in [-\frac{\pi}{4}, \frac{\pi}{4}] \cup [\frac{3\pi}{4}, \frac{5\pi}{4}]$.

If necessary, interpret the explicit equations as parameterizations. Combine these expressions to obtain $z = x + iy$, or $z = r(\cos\theta + i \sin\theta)$.

10. Using the geometrical meaning of $|\cdot|$ and *arg* in $\mathbb{C}$, find that part of $\mathbb{C}$ which is defined by the conditions $1 \le | z - i | < 2$ and $1 < \arg z \le 2$.
Show that if $| z_1 | = | z_2 | = | z_3 | > 0$, then

$$\arg \frac{z_3 - z_2}{z_3 - z_1} = \frac{1}{2} \arg \frac{z_2}{z_1} .$$

Hint. The points z_1, z_2, z_3 belong to a circle of center O, and

$$\arg \frac{\zeta}{z} = \arg \zeta - \arg z .$$

Measure the angle inscribed in this circle, which has the vertex at z_3 .

11. Let $\mathbb{D} = \{x + jy : x, y \in \mathbb{R}\}$, where $j^2 = +1$, be the algebra of double numbers, and let us note $\mathcal{K} = \{(x + jy, u + jv) : u - x \ge |v - y|\}$. Show that $\mathcal{K}$ is a partial order on $\mathbb{D}$, which extends the order of $\mathbb{R}$, and it is compatible with the algebraical structure of $\mathbb{D}$. In particular, the squares $(x + jy)^2$ are always positive.
Hint. The cone of positive double numbers is delimitated by the straight lines $y = \pm x$, and contains $\mathbb{R}_+$.

12. Solve the equation

$$2z^7 - z^6 - 4z^5 + 2z^4 - 2z^3 + z^2 + 4z - 2 = 0$$

in $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, and $\mathbb{D}$.
Hint. Use the Horner's scheme to write the equation in the form

$$(z^2 - 1)(z - \tfrac{1}{2})(z^2 - 2)(z^2 + 1) = 0 .$$

Pay attention to the fact that $z^2 \geq 0$ always holds in $\mathbb{D}$ (see problem 11 from above), so that $z^2 + 1 = 0$ has no solutions, while $z^2 - 1 = 0$ has 4 solutions in this space.

§ I.3. ELEMENTS OF LINEAR ALGEBRA

The linear structures represent the background of the Analysis, whose main purpose is to develop methods for solving problems by a local reduction to their linear approximations. Therefore, in this paragraph we summarize some results from the linear algebra, which are necessary for the later considerations. A general knowledge of the algebraic structures (like *groups, rings, fields*) is assumed, and many details are omitted on account of a parallel course on Algebra (see also [AE], [KA], [V-P], etc.).

As usually, Γ denotes one of the fields of scalars, $\mathbb{R}$ or $\mathbb{C}$.

3.1.Definition. The nonvoid set $\mathcal{L}$ is said to be a *linear space over* Γ iff it is endowed with an internal *addition* $+ : \mathcal{L} \times \mathcal{L} \to \mathcal{L}$, relative to which $(\mathcal{L},+)$ is a commutative group, and also with an external *multiplication by scalars* $\cdot : \Gamma \times \mathcal{L} \to \mathcal{L}$, such that :

[L_1] $\alpha(\beta x) = (\alpha\beta)x$ for any $\alpha, \beta \in \Gamma$ and $x \in \mathcal{L}$;

[L_2] $\alpha(x+y) = \alpha x + \alpha y$ for any $\alpha \in \Gamma$ and $x, y \in \mathcal{L}$;

[L_3] $1 \cdot x = x$ for any $x \in \mathcal{L}$.

The elements of $\mathcal{L}$ are usually called *vectors.* Whenever we have to distinguish vectors from numbers or other elements, we may note them by an arrow, or an line over, e.g. $\vec{x}$, or $\overline{x}$. In particular, the neutral element relative to the addition is noted ϑ, or simply 0 (but rarely $\vec{0}$, or $\overline{0}$), if no confusion is possible. It is called *the origin*, or *zero* of $\mathcal{L}$.

If $\Gamma = \mathbb{R}$, we say that $\mathcal{L}$ is a *real* linear space, while for $\Gamma = \mathbb{C}$ the space $\mathcal{L}$ is said to be *complex.*

Any nonvoid part $\mathcal{S}$ of $\mathcal{L}$ is called *linear subspace of* $\mathcal{L}$ iff it is closed relative to the operations of $\mathcal{L}$. In particular, $\mathcal{L}$ itself and $\{\vartheta\}$ represent *(improper)* subspaces, called *total*, respectively *null* subspaces.

3.2. Examples. ***a)*** Γ itself is a linear space over Γ. In particular, $\mathbb{C}$ can be considered a linear space over $\mathbb{R}$, or over $\mathbb{C}$. Obviously, $\mathbb{R}$ is a linear subspace of the real linear space $\mathbb{C}$, which is organized as $\mathbb{R}^2$.

b) The real spaces $\mathbb{R}^2$ or $\mathbb{R}^3$ of all physical vectors with the same origin, e.g. speeds, forces, impulses, etc., represent the most concrete examples.

Alternatively, the *position vectors* of the points in the geometrical space, or the classes of equivalent *free vectors* form linear spaces. The addition is done by the *parallelogram rule*, while the product with scalars reduces to the change of length and sense. Of course, $\mathbb{R}^2$ is a linear subspace of $\mathbb{R}^3$, more accurately *up to an isomorphism.*

c) The sets $\mathbb{R}^n$ and $\mathbb{C}^n$, where $n \in \mathbb{N}^*$, are linear spaces if the operations with vectors are reduced to components according to the following rules:

$(x_1, x_2, \ldots, x_n) + (y_1, y_2 \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots x_n + y_n)$, and

$\lambda\,(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$.

d) The set $\Gamma^{\mathbb{N}}$ of all *numerical sequences* is a linear space relative to the operations similarly reduced to components (i.e. terms of the sequences). In particular, $\mathbb{R}^n$ is a linear subspace of $\mathbb{R}^{\mathbb{N}}$, for any $n \in \mathbb{N}^*$.

e) If $\mathcal{L}$ is a linear space, and T is an arbitrary set, then the set $\mathcal{F}_{\mathcal{L}}(T)$, of all functions $f : T \to \mathcal{L}$, "borrows" the structure of linear space from $\mathcal{L}$, in the sense that, by definition, $(f + g)(t) = f(t) + g(t)$, and $(\lambda f)\,(t) = \lambda\, f(t)$ at any $t \in T$. This structure is tacitly supposed on many "function spaces" like polynomial, continuous, derivable, etc.

3.3. <u>Proposition.</u> The following formulas hold in any linear space:

(i) $0\,x = \lambda\,\vartheta = \vartheta$, and conversely,

(ii) $\lambda\,x = \vartheta$ *implies either* $\lambda = 0$, *or* $x = \vartheta$;

(iii) $(-\lambda)\,x = \lambda\,(-x) = -\lambda\,x$.

<u>Proof.</u> From $0\,x + 0\,x = (0+0)x = 0\,x + \vartheta$ we deduce $0\,x = \vartheta$; the rest of the proof is similar, and we recommend it as an exercise. ◇

A lot of notions and properties in linear spaces simply extend some intuitive facts of the usual geometry, as for example:

3.4. <u>Definition.</u> Any two distinct elements $x, y \in \mathcal{L}$ determine a *straight line* passing through these points, expressed by

$$\Delta(x, y) = \{z = (1-\lambda)x + \lambda\, y : \lambda \in \Gamma\}.$$

A set $\mathcal{A} \subseteq \mathcal{L}$ is called *linear manifold* iff $\Delta(x, y) \subseteq \mathcal{A}$ whenever $x, y \in \mathcal{A}$.

Any linear manifold $\mathcal{H} \subset \mathcal{L}$, which is maximal relative to the inclusion $\subseteq$, is called *hyper plane.*

That part (subset) of the line $\Delta(x, y)$, which is defined by

$$[x, y] = \{z = (1-\lambda)x + \lambda\, y : \lambda \in [0,1] \subset \mathbb{R}\},$$

is called *line segment* of end-points x and y. A set $C \subseteq \mathcal{L}$ is said to be *convex* iff $[x, y] \subseteq C$ whenever $x, y \in C$.

It is easy to see that any linear subspace is a linear manifold, and any linear manifold is a convex set. In this sense we have:

3.5. <u>Proposition.</u> The set $\mathcal{A} \subseteq \mathcal{L}$ is a linear manifold if and only if its *translation* to the origin, defined by

$$\mathcal{A} - x_0 = \{y = x - x_0 : x \in \mathcal{A}\},$$

where $x_0 \in \mathcal{A}$, is a linear subspace of $\mathcal{L}$.

<u>Proof.</u> If $\mathcal{A}$ is a linear manifold, then $\mathcal{S} = \mathcal{A} - x_0$ is closed relative to the addition and multiplication by scalars. In fact, if $y_1, y_2 \in \mathcal{S}$, then they have the form $y_1 = x_1 - x_0$ and $y_2 = x_2 - x_0$, for some $x_1, x_2 \in \mathcal{A}$. Consequently, $y_1 + y_2 = (x_1 + x_2 - x_0) - x_0 \in \mathcal{S}$, because

$$x_1 + x_2 - x_0 = 2\left(\frac{1}{2}x_1 + \frac{1}{2}x_2\right) - x_0 \in \mathcal{A} .$$

Similarly, if $y = x - x_0$, and $\lambda \in \Gamma$, then

$$\lambda y = ((1-\lambda)\, x_0 + \lambda x)\text{ a} - x_o \in \mathcal{S}.$$

Conversely, if $\mathcal{S} = \mathcal{A} - x_0$ is a linear subspace of $\mathcal{L}$, then $\mathcal{A} = \mathcal{S} + x_0$ is a linear manifold. In fact, for any $x_1 = y_1 + x_0$ and $x_2 = y_2 + x_0$ from $\mathcal{A}$, their *convex combination* has the form

$$(1-\lambda)\, x_1 + \lambda x_2 = ((1-\lambda)\, y_1 + \lambda y_2) + x_0 \in \mathcal{S} + x_0 ,$$

which shows that $\Delta(x_1,x_2) \subseteq \mathcal{A}$. ◇

3.6. <u>Corollary.</u> $\mathcal{H} \subset \mathcal{L}$ is a hyper plane if and only if it is the translation at some $x_0 \in \mathcal{H}$ of a maximal linear subspace $\mathcal{W}$, i.e. $\mathcal{H} = \mathcal{W} + x_0$.

The geometrical notions of *co-linearity and co-planarity* play a central role in the linear structures theory. Their generalization is expressed in terms of "linear dependence" as follows:

3.7. <u>Definition.</u> For any (finite!) set of vectors $x_1, \ldots, x_n \in \mathcal{L}$, and any system of scalars $\alpha_1,...,\alpha_n \in \Gamma$, the expression

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n ,$$

which equals another vector in $\mathcal{L}$, is called *linear combination* of these vectors. The set of all linear combinations of the elements of a subset $A \subseteq \mathcal{L}$ is called *linear span* (or *linear cover*) of A, and it is noted $\text{Lin}\, A$.

If there exists a *null* linear combination with non-null coefficients, i.e. if

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n = \vartheta$$

holds for at least one $\alpha_k \neq 0$, then the vectors $x_1, x_2, \ldots, x_n$ are said to be *linearly dependent* (or alternatively, one of them *linearly depends* on the others). In the contrary case, they are *linearly independent.*

Family $\mathscr{F} = \{x_i \in \mathcal{L} : i \in I\}$ is called *independent system* of vectors iff any of its finite subfamily is linearly independent. If such a system is maximal relative to the inclusion, i.e. any $x \in \mathcal{L}$ is a linear combination of some $x_{i_k} \in \mathscr{F}$, $i_k \in I$, $k = \overline{1,n}$, then it is called *algebraical* (or *Hamel*) *base of* $\mathcal{L}$. In other terms, we say that $\mathscr{F}$ *generates* $\mathcal{L}$, or Lin $\mathscr{F} = \mathcal{L}$.

3.8. <u>Examples.</u> a) The *canonical* base of the plane consists of the vectors $i=(1,0)$ and $j=(0,1)$; sometimes we note $\bar{i}$ and $\bar{j}$, while in the complex plane we prefer $u = (1, 0)$ and $i = (0, 1)$. Similarly, $\mathscr{B} = \{i, j, k\}$, where $i=(1,0,0)$, $j=(0,1,0)$ and $k=(0,0,1)$, represents the canonical base of $\mathbb{R}^3$. Alternatively, we frequently note $i = e_1$, $j = e_2$, and $k = e_3$ (sometimes with bars over).

b) System $\mathscr{B} = \{ (\delta_i^j)_{j=\overline{1,n}} : i = \overline{1,n} \}$, where

$$\delta_i^j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

is *the Kronecker's symbol*, is a base (named *canonical*) in Γ^n. Explicitly,

$$\mathscr{B} = \{(1,0,\ldots,0), (0,1,\ldots,0), \ldots, (0,0,\ldots,1)\}.$$

c) The space of all polynomials has a base of the form

$$\{1, t, t^2, \ldots, t^n, \ldots\},$$

which is infinite, but countable. If we ask the degree of the polynomials not to exceed some $n \in \mathbb{N}^*$, then a base of the resulting linear space consists of

$$\{1, t, t^2, \ldots, t^n\} .$$

d) Any base of $\mathbb{R}$, considered as a linear space over $\mathbb{Q}$, must contain infinitely many irrationals, hence it is uncountable.

3.9. <u>Theorem.</u> If $\mathscr{B}_1$ and $\mathscr{B}_2$ are bases of $\mathcal{L}$, and $\mathscr{B}_1$ is finite, namely card $\mathscr{B}_1 = n \in \mathbb{N}^*$, then card $\mathscr{B}_2 = n$ too.

<u>Proof.</u> In the contrary case, let us consider that

$$\mathscr{B}_1 = \{e_1, e_2, \ldots, e_n\} \text{ and } \mathscr{B}_2 = \{f_1, f_2, \ldots, f_m\}$$

are two bases of $\mathcal{L}$, and still $n<m$. We claim that in this case there exist a system of non-null numbers $\lambda_1, \lambda_2, \ldots, \lambda_m \in \Gamma$, such that

$$\lambda_1 f_1 + \lambda_2 f_2 + \ldots + \lambda_m f_m = \vartheta,$$

which contradicts the fact that $\mathscr{B}_2$ is linearly independent . In fact, since $\mathscr{B}_1$ is a base, it generates $\mathscr{B}_2$, i.e.

$$f_i = \sum_{j=1}^{n} a_{ij} e_j$$

for all $i = \overline{1, m}$. Replacing these expressions of f_i in the above combination, and taking into account the independence of $\mathscr{B}_1$, we obtain the following homogeneous system of linear equations

$$\begin{aligned} a_{11}\lambda_1 + a_{21}\lambda_2 + \ldots + a_{m1}\lambda_m &= 0 \\ a_{12}\lambda_1 + a_{22}\lambda_2 + \ldots + a_{m2}\lambda_m &= 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}\lambda_1 + a_{2n}\lambda_2 + \ldots + a_{mn}\lambda_m &= 0\,. \end{aligned}$$

This system really has non-null solutions since $n<m$. ◇

This theorem shows that the number of elements of a base is an intrinsic property of the considered linear space, i.e. it is the same for any choice of the base. In particular, if some base of $\mathcal{L}$ is finite / infinite, then any other base is also finite / infinite. In other terms, this theorem is the background of the following important notion:

3.10. <u>Definition.</u> If a linear space $\mathcal{L}$ contains infinite systems of linearly independent vectors, then $\mathcal{L}$ is said to be a space of *infinite dimension*. If $\mathcal{L}$ contains only finite systems of linearly independent vectors we say that $\mathcal{L}$ is *finite dimensional*, and the maximal cardinal of such systems (which equals the cardinal $n \in \mathbb{N}^*$ of any base) is called *the dimension of* $\mathcal{L}$, and it is noted $n = \dim \mathcal{L}$.

3.11. <u>Theorem.</u> Any change of base in a finite dimensional space is represented by a non-singular square matrix.

<u>Proof.</u> If $\mathscr{A} = \{e_1, e_2, \ldots, e_n\}$ is the "old" base of $\mathcal{L}$, and $\mathscr{B} = \{f_1, f_2, \ldots, f_n\}$ is the "new" one, then the change $\mathscr{A} \mapsto \mathscr{B}$ is explicitly given by the formulas $f_i = \sum_{j=1}^{n} t_{ij} e_j$, where $i = \overline{1, n}$. This is the exact meaning of the fact that the change of base is "represented" by the matrix $(t_{ij}) \in \mathcal{M}_{n,n}(\Gamma)$. In short, this transformation may be written in the matrix form

$$(e_1\ e_2 \ldots e_n)\,(t_{ij})^{\mathsf{T}} = (f_1\ f_2 \ldots f_n)\,,$$

where T denotes *transposition* (i.e. interchange of rows with columns); note the dimensions of the involved matrices, namely $(1, n)(n, n) = (1, n)$.

We claim that matrix (t_{ij}) is non-singular, i.e. Det $(t_{ij}) \neq 0$. In fact, in the contrary case we find a system of non-null numbers $\lambda_1, \ldots, \lambda_n$ like in the proof of theorem I.3.6. (this time $m = n$), such that $\lambda_1 e_1 + \ldots + \lambda_n e_n = \vartheta$. Such a relation is still impossible because the elements of any base are linearly independent. ◇

3.12. Remarks. ***a)*** The above representation of the change $\mathscr{A} \mapsto \mathscr{B}$ is easier expressed as a matrix relation if we introduce the so called *transition matrix* $T = (t_{ij})^{\mathsf{T}}$. More precisely, the line matrices formed with the elements of $\mathscr{A}$ and $\mathscr{B}$ are related by the equality

$$(e_1 \ e_2 \ldots e_n)\, T = (f_1 \, f_2 \ldots f_n) \, .$$

b) Continuing the idea of representing algebraical entities by matrices, we mention that any vector $x \in \mathcal{L}$ is represented in the base $\mathscr{B} = \{e_i : i = \overline{1,n}\}$ by a column *matrix of components* $X = (x_1 \, x_2 \ldots x_n)^{\mathsf{T}}$. This representation is practically equivalent to the development $x = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$, hence after the choice of some base $\mathscr{B}$ in $\mathcal{L}$, we can establish a 1:1 correspondence between vectors $x \in \mathcal{L}$ and matrices $X \in \mathcal{M}_{n,1}(\Gamma)$.

c) Using the above representation of the vectors, it is easy to see that any matrix $A \in \mathcal{M}_{n,n}(\Gamma)$ defines a function $\mathbf{U} : \mathcal{L} \to \mathcal{L}$, by identifying $y = \mathbf{U}(x)$ with $Y = A X$. A remarkable property of **U** is expressed by the relation

$$\mathbf{U}(\alpha x + \beta y) = \alpha\, \mathbf{U}(x) + \beta\, \mathbf{U}(y) \, ,$$

which holds for any $x, y \in \mathcal{L}$ and α , $\beta \in \Gamma$, i.e. **U** “respects” the linearity.

This special property of the functions, which act between linear spaces, is marked by a specific terminology:

3.13. Definition. If $\mathcal{X}$ and $\mathcal{Y}$ are linear spaces over the same field Γ, then any function $f : \mathcal{X} \to \mathcal{Y}$ is called *operator;* in the particular case $\mathcal{Y} = \Gamma$, we say that f is a *functional*, while for $\mathcal{X} = \mathcal{Y}$ we prefer the term *transformation* . The operators are noted by bold capitals **U**, **V**, etc., and the functionals by f, g, etc.

An operator $\mathbf{U} : \mathcal{X} \to \mathcal{Y}$ is said to be *linear* iff it is *additive* , i.e.

$$\mathbf{U}(x + y) = \mathbf{U}(x) + \mathbf{U}(y) \, , \ \forall x, y \in \mathcal{X} \, ,$$

and *homogeneous,* that is

$$\mathbf{U}(\alpha x) = \alpha\, \mathbf{U}(x) \, , \ \ \forall x \in \mathcal{X} \, , \text{ and } \forall \alpha \in \Gamma \, .$$

These two conditions are frequently concentrated in one, namely

$$\mathbf{U}(\alpha x + \beta y) = \alpha\, \mathbf{U}(x) + \beta\, \mathbf{U}(y),$$

which is considered valid for any $x, y \in \mathcal{X}$ and $\alpha, \beta \in \Gamma$. The set of all linear operators between $\mathcal{X}$ and $\mathcal{Y}$ is usually noted by $\mathscr{L}(\mathcal{X}, \mathcal{Y})$.

The *linear functionals* are similarly defined, and $\mathcal{X}^* = \mathscr{L}(\mathcal{X}, \Gamma)$ is the usual notation for the space of all linear functionals on $\mathcal{X}$. Alternatively, $\mathcal{X}^*$ is called *algebraical dual of* $\mathcal{X}$.

3.14. <u>Examples.</u> **1)** If $\mathcal{X} = \Gamma^n$ and $\mathcal{Y} = \Gamma^m$ for some $n, m \in \mathrm{N}^*$, then any operator $\mathbf{U} : \mathcal{X} \to \mathcal{Y}$, which is *represented* by a matrix, is linear. More exactly, there exists $A = (a_{ik}) \in \mathcal{M}_{m,n}(\Gamma)$, like in the above remark I.3.12., such that $y = \mathbf{U}(x)$ means that

$$y_i = \sum_{k=1}^{n} a_{ik} x_k, \text{ for any } i = \overline{1,m},$$

where $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_m)$.

In particular, any function $f: \Gamma^n \to \Gamma$ of values

$$f(x) = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$$

is a linear functional.

2) Let $\mathcal{X} = \mathrm{C}^1_{\mathbb{R}}(I)$ be the space of all real functions which have continuous derivatives on I, and $\mathcal{Y} = \mathbb{C}_{\mathbb{R}}(I)$ be the space of all continuous functions on I, where $I = (a, b) \subseteq \mathbb{R}$ (see also chapter IV below). Then the process of deriving, considered as an operator $\mathbf{D}: \mathcal{X} \to \mathcal{Y}$, represents a linear operator. In fact, because

$$y = \mathbf{D}(x) = x' = \frac{dx}{dt},$$

means $\mathbf{D}(x)(t) = x'(t)$ at any $t \in (a, b)$, the linearity of $\mathbf{D}$ reduces to the rules of deriving a sum and a product with a scalar.

3) Let $\mathcal{X} = \mathcal{Y} = \mathrm{C}_{\mathbb{R}}(K)$ be the function spaces from above, where this time $K = [a, b] \subset \mathbb{R}$, and let $A : K \times K \to \mathbb{R}$ be continuous on K^2 relative to each of its variables (but uniformly in respect to the other). Then *the operator of integration* $\mathbf{I} : \mathcal{X} \to \mathcal{Y}$ is also linear, where $y = \mathbf{I}(x)$ means

$$y(s) = \int_a^b A(s,t)\, x(t)\, dt.$$

In particular, the correspondence expressed by

$$\mathcal{X} \ni x \mapsto y = \int_a^b x(t)\,dt \in \Gamma$$

defines a linear functional on $\mathcal{X}$ (see also part II).

4) If $\mathcal{X} = (\mathcal{L}, <.,.>)$ is a scalar product space over Γ (see also §II.3.), then each $y \in \mathcal{L}$ generates a linear functional $f_y : \mathcal{L} \to \Gamma$, of values

$$f_y(x) = <x, y>.$$

5) Let S be an arbitrary nonvoid set, for which $\mathcal{X} = \mathcal{F}_\Gamma(S)$ denotes the space of all functions $x : S \to \Gamma$. To each $t \in S$ we may attach a linear functional $f_t : \mathcal{X} \to \Gamma$, defined by $f_t(x) = x(t)$.

The algebraical organization of $\mathscr{L}(\mathcal{X}, \mathcal{Y})$ represents the starting point in the study of the linear operators, so we mention that:

3.15. <u>Proposition.</u> ***(i)*** $\mathscr{L}(\mathcal{X}, \mathcal{Y})$ is a linear space relative to the internal *addition*, defined at any $x \in \mathcal{X}$ by

$$(\mathbf{U} + \mathbf{V})(x) = \mathbf{U}(x) + \mathbf{V}(x),$$

and the *multiplication by scalars* (from the same field Γ), given by

$$(\lambda \mathbf{U})(x) = \lambda\, \mathbf{U}(x).$$

(ii) $\mathscr{L}(\mathcal{X}, \mathcal{Y})$ is an algebra relative to the above operations of addition and multiplication by scalars, and to the *internal composition*, defined by

$$(\mathbf{V} \circ \mathbf{U})(x) = \mathbf{V}(\mathbf{U}(x))$$

at any $x \in \mathcal{X}$. This algebra has an *unit* element.

<u>Proof.</u> ***(i)*** $(\mathscr{L}(\mathcal{X}, \mathcal{Y}), +, .)$ verifies the axioms of a linear space.

(ii) $(\mathscr{L}(\mathcal{X}, \mathcal{Y}), +, \circ)$ is a ring in which

$$(\lambda \mathbf{V}) \circ (\mu \mathbf{U}) = \lambda\mu\, (\mathbf{V} \circ \mathbf{U})$$

holds for any $\lambda, \mu \in \Gamma$, and $\mathbf{U}, \mathbf{V} \in \mathscr{L}(\mathcal{X}, \mathcal{Y})$. The unit element, noted $\boldsymbol{\iota}$, and called *identity*, is defined by $\boldsymbol{\iota}(x) = x$. ◇

The proof of the following properties is also routine.

3.16. <u>Proposition.</u> If $\mathcal{X}$ and $\mathcal{Y}$ are linear spaces, and $\mathbf{U}: \mathcal{X} \to \mathcal{Y}$ is a linear operator, then:

i) The *direct image* of any linear subspace $\mathcal{L} \subseteq \mathcal{X}$, i.e.

$$\mathbf{U}(\mathcal{L}) = \{y = \mathbf{U}(x) \in \mathcal{Y}: \ x \in \mathcal{L} \},$$

is a linear subspace of $\mathcal{Y}$.

ii) The *inverse image* of any linear subspace $M \subseteq \mathcal{Y}$, i.e.

$$\mathbf{U}^{\leftarrow}(M) = \{x \in \mathcal{X} : \mathbf{U}(x) \in M \},$$

is a linear subspace of $\mathcal{X}$.

iii) **U** is invertible iff its *kernel (nucleus)*, defined by

$$\text{Ker}\,(\mathbf{U}) = \{ x \in \mathcal{X} : \mathbf{U}(x) = 0\} = \mathbf{U}^{\leftarrow}(0),$$

reduces to $\vartheta \in \mathcal{X}$, or more exactly, to the null subspace $\{ \vartheta \}$.

Several general properties of the linear functionals and operators can be formulated in "geometric" terms, being involved in the so-called *equations* of the linear manifolds, as follows:

3.17. <u>Theorem.</u> If $f \in \mathcal{X}^* \setminus \{0\}$, and $k \in \Gamma$, then

$$\mathcal{H} = \{x \in \mathcal{X} : \ f(x) = k\}$$

is a hyper plane in $\mathcal{X}$, and conversely, for any hyper plane $\mathcal{H} \subset \mathcal{X}$ there exists $f \in \mathcal{X}^* \setminus \{0\}$, and $k \in \Gamma$, such that $x \in \mathcal{H}$ iff $f(x) = k$.

<u>Proof.</u> For any $x_0 \in \mathcal{H}$ it follows that

$$\mathcal{L} = \mathcal{H} - x_0 = f^{\leftarrow}(0)$$

is a linear subspace, hence $\mathcal{H}$ is a linear manifold. We claim that $\mathcal{L}$ is maximal. In fact, we easily see that $\mathcal{L} \neq \mathcal{X}$ because $f \neq 0$, hence there exists $a \in \mathcal{X} \setminus \mathcal{L}$, where $f(a) \neq 0$. Therefore, at any $x \in \mathcal{X}$, we may define the number $\lambda = \dfrac{f(x)}{f(a)}$, and the vector $y = x - \lambda a$. Since $f(y) = 0$, it follows that $y \in \mathcal{L}$, so we may conclude that

$$\mathcal{X} = \text{Lin}\,(\mathcal{L} \cup \{a\}) ,$$

i.e. $\mathcal{L}$ is a maximal subspace of $\mathcal{X}$.

Conversely, let $\mathcal{H}$ be a hyper plane in $\mathcal{X}$, $x_0 \in \mathcal{H}$, and $\mathcal{L} = \mathcal{H} - x_0$.

Then $\mathcal{L}$ is a maximal linear subspace of $\mathcal{X}$, i.e. $\mathcal{X} \setminus \mathcal{L} \neq \emptyset$, and for any $a \in \mathcal{X} \setminus \mathcal{L}$ there is a unique decomposition of $x \in \mathcal{X}$, $x = y + \lambda a$, where $y \in \mathcal{L}$ and $\lambda \in \Gamma$. On this way we obtain a functional $f : \mathcal{X} \to \Gamma$, which attaches a number $\lambda \in \Gamma$ to each $x \in \mathcal{X}$, according to the above decomposition, i.e. $f(x) = \lambda$. It is easy to see that f is linear, so it remains to show that

$$\mathcal{H} = \{ x \in \mathcal{X} : f(x) = f(x_0) \}.$$

In fact, if $x \in \mathcal{H}$, then $x = y + x_0$ for some $y \in \mathcal{L}$, hence from $f(y) = 0$ we deduce that $f(x) = f(x_0)$. Conversely, equation $f(x) = f(x_0)$ leads to

$$y = x - x_0 \in \mathcal{L} = f^{\leftarrow}(0),$$

i.e. $x \in \mathcal{L} + x_0 = \mathcal{H}$. ◇

Going back to the problem of representing linear operators, we mention its simple solution in the case when they act on finite dimensional spaces:

3.18. <u>Theorem.</u> Let $\mathscr{B} = \{e_i : i = \overline{1,n}\}$ and $\mathscr{C} = \{f_j : j = \overline{1,m}\}$ be bases of the linear spaces $\mathcal{X}$ and $\mathcal{Y}$ over the same field Γ, such that $x \in \mathcal{X}$, and $y \in \mathcal{Y}$ are represented by the matrices X, respectively Y. If $\mathbf{U}: \mathcal{X} \to \mathcal{Y}$ is a linear operator, then there is a unique matrix $A \in \mathcal{M}_{m,n}(\Gamma)$ such that

$$y = \mathbf{U}(x)$$

is equivalent to $Y = AX$. In short, $\mathbf{U}$ is represented by A .

<u>Proof.</u> By developing each $\mathbf{U}(e_i)$ in base $\mathscr{C}$, we obtain

$$\mathbf{U}(e_i) = \sum_{j=1}^{m} a_{ij} f_j, \quad i = \overline{1,n} .$$

Evaluating $\mathbf{U}$ at an arbitrary $x = x_1 e_1 + \ldots + x_n e_n$ yields

$$\mathbf{U}(x) = \sum_{j=1}^{n} x_i \left(\sum_{j=1}^{m} a_{ij} f_j \right) = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ij} x_i \right) f_j .$$

The comparison of this expression with $y = \mathbf{U}(x) = y_1 f_1 + \ldots + y_m f_m$ leads to the relations

$$y_j = \sum_{i=1}^{n} a_{ij} x_i , \quad j = \overline{1,m} .$$

If we note $A = (a_{ji}^*)$, where $a_{ji}^* = a_{ij}$, i.e. $A = (a_{ij})^\intercal$, then the above relations between the components take the matrix form $Y = AX$. ◇

3.19. Corollary. The values of any linear functional $f: \mathcal{X} \to \Gamma$, where $\mathcal{X}$ is a finite (say n-) dimensional linear space, have the form

$$f(x) = a_1 x_1 + \ldots + a_n x_n .$$

Proof. Take $m = 1$, $A = (a_1 \ldots a_n)$, and $X = (x_1 \ldots x_n)^{\mathsf{T}}$ in the previous theorem, where $\mathscr{C} = \{1\}$. ◇

It is remarkable that algebraical operations with operators correspond to similar operations with the representative matrices. More exactly:

3.20. Proposition. Let $\mathcal{X}$ and $\mathcal{Y}$ be as in the above theorem 3.18.

a) If $\mathbf{U}, \mathbf{V} \in \mathscr{L}(\mathcal{X}, \mathcal{Y})$ are represented by $A, B \in \mathcal{M}_{m,n}(\Gamma)$, then $\mathbf{U}+\mathbf{V}$ is represented by $A+B$;

b) If $A \in \mathcal{M}_{m,n}(\Gamma)$ represents $\mathbf{U} \in \mathscr{L}(\mathcal{X}, \mathcal{Y})$, then λA represents $\lambda \mathbf{U}$ for any $\lambda \in \Gamma$;

c) Let $\mathbf{U} \in \mathscr{L}(\mathcal{X}, \mathcal{Y})$ be represented by $A \in \mathcal{M}_{m,n}(\Gamma)$ as before, and let $\mathcal{Z}$ be another liniar space over the same field Γ, where $\dim \mathcal{Z} = p \in \mathbb{N}^*$.

If $B \in \mathcal{M}_{p,n}(\Gamma)$ represents $\mathbf{V} \in \mathscr{L}(\mathcal{Y}, \mathcal{Z})$ relative to some base of $\mathcal{Z}$, then BA represents $\mathbf{V} \circ \mathbf{U}$.

Proof. The first two assertions are immediate. Even property *c)* is a direct consequence of the above theorem 3.18, since $\mathbf{V} \circ \mathbf{U}(x) = \mathbf{V}(\mathbf{U}(x))$ is equivalent to $Y = AX$, and $Z = BY$, i.e. $Z = B(AX) = (BA)X$. ◇

Because any representation (of vectors, linear operators, functionals, etc.) essentially depends on the chosen bases, it is important to see how they change by passing from a base to another.

3.21. Theorem. Let $T \in \mathcal{M}_{n,n}(\Gamma)$ represent the transition from the base $\mathscr{A}$ to $\mathscr{B}$ in the linear space $\mathcal{L}$ of finite dimension n .

a) If X represents a vector $x \in \mathcal{L}$ in the base $\mathscr{A}$, then $\overline{X} = T^{-1} X$ represents the same vector in the base $\mathscr{B}$;

b) If a linear operator $\mathbf{U}: \mathcal{L} \to \mathcal{L}$ is represented by matrix A in the base $\mathscr{A}$, and by B in $\mathscr{B}$, then the following equality holds:

$$B = T^{-1} A\, T .$$

Proof. *a)* According to theorem 11 and remark 12 from above, the change of $\mathscr{A}$ into $\mathscr{B}$ is represented by a non-singular matrix $T = (t_{ij})^{\mathsf{T}} \in \mathcal{M}_{n,n}(\Gamma)$, in the sense that the numbers t_{ij} occur in the formulas

$$f_i = \sum_{j=1}^{n} t_{ij} e_j \ , \ \ i = \overline{1,n} \ ,$$

which relates $e_j \in \mathscr{A}$ to $f_i \in \mathscr{B}$. By replacing these expressions of f_i in the developments of an arbitrary $x \in \mathcal{L}$ in these two bases, namely

$$x_1 e_1 + \ldots + x_n e_n = \overline{x}_1 f_1 + \ldots + \overline{x}_n f_n \ ,$$

we obtain $X = T\overline{X}$, hence finally $\overline{X} = T^{-1} X$.

b) Let the action of $\mathbf{U} \in \mathscr{L}(\mathcal{X}, \mathcal{Y})$, that is $y = \mathbf{U}(x)$, be alternatively expressed by $Y = AX$ in base $\mathscr{A}$, respectively by $\overline{Y} = B\overline{X}$ in base $\mathscr{B}$. According to the part *a)*, we have $\overline{X} = T^{-1} X$, and $\overline{Y} = T^{-1} Y$, wherefrom we deduce that $T^{-1} Y = B\, T^{-1} X$. Finally, it remains to equalize the two expressions of Y, namely $TBT^{-1} X = AX$, where X is arbitrary. ◇

The linear operators are useful in comparing linear spaces. In particular, it is useful to identify those spaces which present only formal differences, which are said to be *isomorphic*. More exactly:

3.22. <u>Definition.</u> Let $\mathcal{X}$ and $\mathcal{Y}$ be linear spaces over the same field Γ. We consider them *isomorphic* iff there exists a linear operator $\mathbf{U} \in \mathscr{L}(\mathcal{X}, \mathcal{Y})$, which realizes a 1:1 correspondence of their elements. In such a case we say that $\mathbf{U}$ is an *isomorphism* of $\mathcal{X}$ and $\mathcal{Y}$.

Establishing the isomorphism of the finite dimensional spaces basically reduces to the comparison of the dimensions:

3.23. <u>Theorem.</u> Two finite dimensional linear spaces $\mathcal{X}$ and $\mathcal{Y}$, over the same field Γ, are isomorphic iff $\dim\mathcal{X} = \dim\mathcal{Y}$.

<u>Proof.</u> If $\dim\mathcal{X} = \dim\mathcal{Y}$, then we put into correspondence the vectors with identical representations in some fixed bases.

Conversely, let $\mathbf{U}$ be an isomorphism of these spaces, and let us note $\dim\mathcal{X} = n$, and $\dim\mathcal{Y} = m$. Since $\mathbf{U}: \mathcal{X} \to \mathcal{Y}$ realizes an injective linear correspondence, it maps $\mathcal{X}$ into a linear subspace of $\mathcal{Y}$, and carries any base of $\mathcal{X}$ into a base of $\mathbf{U}(\mathcal{X})$. Consequently $\dim \mathbf{U}(\mathcal{X}) = n \leq m$. But $\mathbf{U}$ is also surjective, hence there exists $\mathbf{U}^{-1}$, which is linear too. Applying the same reason to $\mathbf{U}^{-1}$, we obtain the opposite inequality $m \leq n$. Consequently we obtain $m = n$. ◇

Because each particular representation of an operator depends on the chosen bases, knowing the intrinsic (i.e. independent of base) elements and properties has an extreme importance. In this sense we mention:

3.24. Definition. Number $\lambda \in \Gamma$ is called *proper* (or *characteristic*) *value* of the linear operator $\mathbf{U}: \mathcal{L} \to \mathcal{L}$ iff there exists $x \neq 0$ in $\mathcal{L}$ such that

$$\mathbf{U}x = \lambda x .$$

In this case x is named *proper vector* of $\mathbf{U}$, and

$$\mathcal{L}_\lambda = \{x \in \mathcal{L}: \mathbf{U}x = \lambda x\} = \mathrm{Ker}\,(\mathbf{U} - \lambda \mathbf{I})$$

is the *proper subspace* of $\mathbf{U}$, corresponding to λ .

3.25. Remark. The study of proper values, proper vectors, etc., leads to the so-called *spectral theory* (see [CR], [CI], etc.). One of the starting points in this theory is the notion of *spectrum* of a linear operator. More exactly, *the spectrum* of $\mathbf{U}$ consists of those $\lambda \in \Gamma$ for which operator $\mathbf{U} - \lambda \mathbf{I}$ is not invertible. If $\mathcal{L}$ has a finite dimension, the study becomes algebraic, i.e. it can be developed in terms of matrices, determinants, polynomials, etc., by virtue of the following:

3.26. Theorem. Let $\mathcal{L}$ be a n-dimensional linear space over Γ, and let $\mathbf{U}$ be a linear transformation of $\mathcal{L}$. If matrix $A \in \mathcal{M}_{n,n}(\Gamma)$ represents **U,** then $\lambda \in \Gamma$ is a proper value of $\mathbf{U}$ iff it is a root of the *characteristic* polynomial

$$P_A(\lambda) = \mathrm{Det}\,(A - \lambda I_n) ,$$

where I_n is the unit matrix of order n . The corresponding proper vectors are represented by the non-null solutions of the homogeneous system

$$(A - \lambda I_n)X = 0 .$$

The proof is direct, but useful as exercise.

Because any complex polynomial has at least one root in $\mathbb{C}$, it follows that:

3.27. Corollary. The linear operators on complex linear spaces of finite dimension have at least one proper value (respectively one proper vector).

Of course, there exist linear transformations of real spaces, which have no proper vectors, since their characteristic polynomials have no root in $\mathbb{R}$ (e.g. the rotation of the real plane). Anyway, all these facts do not depend on the base because:

3.28. Theorem. Let $\mathcal{L}$ be a linear space over Γ, and let $\mathbf{U}: \mathcal{L} \to \mathcal{L}$ be a linear transformation of $\mathcal{L}$. If two matrices $A, B \in \mathcal{M}_{n,n}(\Gamma)$ represent $\mathbf{U}$, then their characteristic polynomials coincide.

Proof. According to theorem I.3.21.b, if T represents the change of base, then $B = T^{-1} A T$. Consequently,

$$| B - \lambda I_n | = | T^{-1} A T - T^{-1} \lambda I_n T | = | T^{-1} |\cdot| A - \lambda I_n |\cdot| T | ,$$

where $|A|$ stands for Det A. Because $| T^{-1} | \cdot | T | = 1$, it follows that

$$| B - \lambda I_n | = | A - \lambda I_n | . \qquad \diamond$$

Replacing λ by A in $A - \lambda I_n$ makes no sense; however, we may evaluate $P_A(A)$, for which we have:

3.29. Theorem (Cayley-Hamilton) Each square matrix $A \in \mathcal{M}_{n,n}(\Gamma)$ vanishes its characteristic polynomial.

Proof. Let us note the characteristic polynomial of A by

$$P_A(\lambda) = a_0 + a_1 \lambda + \ldots + a_n \lambda^n ,$$

and let $A^*(\lambda)$ be the *adjoint* matrix of $A - \lambda I_n$. We remember that $A^*(\lambda)$ is obtained by transposing $A - \lambda I_n$, and replacing each element by its algebraic complement, which represent the former two of three operations in the calculus of the inverse matrix. Consequently, we have

$$(A - \lambda I_n)\, A^*(\lambda) = P_A(\lambda)\, I_n .$$

Because the matrix value of a product generally differs from the product of the values, we cannot replace here λ by A. However, a simple evaluation of the degrees shows that the adjoint matrix has the form

$$A^*(\lambda) = B_0 + B_1 \lambda + \ldots + B_{n-1} \lambda^{n-1} .$$

The identification of the matrix coefficients leads to the relations:

$$\begin{aligned} A B_0 &= a_0 I_n \\ A B_1 - B_0 &= a_1 I_n \\ &\ldots\ldots\ldots\ldots \\ A B_{n-1} - B_{n-2} &= a_{n-1} I_n \\ - B_{n-1} &= a_n I_n . \end{aligned}$$

Multiplying by appropriate powers of A, and summing up the resulting relations gives $P_A(A) = O_n$. $\diamond$

3.30. Corollary. The n^{th} power of every n-dimensional matrix linearly depends on its previous powers.

Proof. We have $a_n = (-1)^n \neq 0$ in

$$P_A(\lambda) = a_0 + a_1 \lambda + \ldots + a_n \lambda^n ,$$

and $P_A(A) = O_n$. $\diamond$

PROBLEMS **§** I.3.

1. Show that the set of all rectangular matrices, which have two rows and three columns, with elements in Γ, forms a linear space. Write a basis of this space, and identify its dimension. Generalization.
Hint. Prove the axioms of a linear space. Take matrices with a single non-null element in the base. The dimension of $\mathcal{M}_{n,m}(\Gamma)$ is nm .

2. Prove that $e_1 = (2, 1, -1)$, $e_2 = (1, 2, 0)$, and $e_3 = (1, -1, 3)$ are linearly independent vectors in $\mathbb{R}^3$, and find the coordinates of $x = (5, 0, 1)$ in the base $\{e_1, e_2, e_3\}$.
Hint. Condition $\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 = 0$ (respectively $= x$) reduces to an homogeneous (respectively non-homogeneous) system of linear equations with a non-null determinant.

3. Show that the set of solutions of any homogeneous system of m linear equations in n unknowns, of rank r, is a linear subspace of $\mathbb{R}^n$, which has the dimension $d = n - r$. Conversely, for any linear subspace $\mathcal{L}$, for which $\dim \mathcal{L} = d$, there exists a system in n unknowns, of rank $r = n - d$, whose solutions exactly fill $\mathcal{L}$.

Hint. Select r equations in r unknowns, which has non-null determinant, and construct $\mathcal{L}$ as the linear span of the resulting solutions. Conversely, if $\{e_1, e_2, \ldots, e_d\}$ is a base of $\mathcal{L}$, then noting $e_I = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in})$ for any $i = \overline{1,d}$, the equations take the form $x_1 \varepsilon_{i1} + x_2 \varepsilon_{i2} + \ldots + x_n \varepsilon_{in} = 0$.

4. Show that any hyper plane in $\mathbb{R}^3$ is defined by three of its points

$$P_k = (x_k , y_k , z_k),\ k=\overline{1,3},$$

and identify the linear functional that occurs in its equation. Generalization.
Hint. Write the equation of the plane in the form

$$\begin{vmatrix} 1 & x & y & z \\ 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \end{vmatrix} = 0 \ .$$

The generalization refers to n instead of 3, or $\mathbb{C}$ instead of $\mathbb{R}$.

5. Find the dimension and write a base of the subspace $\mathcal{L}$ of $\mathbb{R}^n$, where the equation of $\mathcal{L}$ is $x_1 + x_2 + \ldots + x_n = 0$. In particular, for $n=4$, show that the section of $\mathcal{L}$ through the 4-dimensional cube of equations $|x_i| \leq 1$, $i=\overline{1,4}$, is a 3-dimensional solid octahedron, and find its volume.
Hint. $\mathcal{B} = \{e_i = (-1, \delta_{1,i}, \ldots, \delta_{n-1,i})\ ;\ i = 1,\ldots,n-1\}$ is a base of $\mathcal{L}$, and $\dim\mathcal{L} = n-1$. The section contains some vertices of the cube, namely

$$(1,1,-1,-1), (1,-1,1,-1), (1,-1,-1,1),$$
$$(-1,1,1,-1), (-1,1,-1,1), (-1,-1,1,1),$$

which are the intersection of $\mathcal{L}$ with the edges of the cube. To see that these points are vertices of a regular octahedron it is useful to evaluate the distances between them.

6. Let $\mathcal{X}$ and $\mathcal{Y}$ be two linear subspaces of the linear space $\mathcal{L}$, and let us note $r = \dim\mathcal{X}$, $s = \dim\mathcal{Y}$, $i = \dim(X \cap \mathcal{Y})$, and $u = \dim[\mathrm{Lin}(\mathcal{X} \cup \mathcal{Y})]$. Prove that $r + s = u + i$.
Hint. In the finite dimensional case, construct some bases of $\mathcal{X}$ and $\mathcal{Y}$ by completing a base of $\mathcal{X} \cap \mathcal{Y}$. If at lest one of $\mathcal{X}$ and $\mathcal{Y}$ has an infinite dimension, then also $u = \infty$.

7. Find the intersection of the straight lines $a + \lambda x$, and $b + \mu y$, in $\mathbb{R}^5$, where $a = (2,1,1,3,-3)$, $x = (2,3,1,1,-1)$, $b = (1,1,2,1,2)$, and $y = (1,2,1,0,1)$.
Hint. Study the consistency of the system (in matrix form)

$$(2\ 1\ 1\ 3\ -3)^T + \lambda(2\ 3\ 1\ 1\ -1)^T = (1\ 1\ 2\ 1\ 2)^T + \mu(1\ 2\ 1\ 0\ 1)^T.$$

8. Precise all the mutual positions of two planes in $\mathbb{R}^n$, $n > 1$.
Hint. Write the planes in the form

$$\mathcal{X} = \{x + \alpha a + \beta b \in \mathbb{R}^n : \alpha, \beta \in \mathbb{R}\},$$

$$\mathcal{Y} = \{y + \gamma c + \delta d \in \mathbb{R}^n : \gamma, \delta \in \mathbb{R}\},$$

and study the consistency of the matrix system

$$x + \alpha a + \beta b = y + \gamma c + \delta d$$

in four unknowns $\alpha, \beta, \gamma, \delta$. If A is the principal matrix of the system, B is the completion of A, and we note $r = \mathrm{rank}\, A$, and $\underline{r} = \mathrm{rank}\, B$, then one of the following six cases is possible: ($r = \overline{2,4}$) and (either $\underline{r} = r$, or $\underline{r} = r + 1$). In some cases we can describe the intersection in geometric terms.

9. A linear operator **U**: $\mathbb{R}^4 \to \mathbb{R}^4$ is represented by the matrix

$$A = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 3 & 0 & -1 & 2 \\ 1 & 5 & 3 & 1 \\ 1 & 2 & 1 & 3 \end{pmatrix}$$

relative to some base $\mathscr{B} = \{e_1, e_2, e_3, e_4\}$. Find the matrices, which represent **U** in the bases:

(i) $\mathscr{C} = \{ e_1, e_3, e_2, e_4\}$;

(ii) $\mathscr{D} = \{ e_1, e_1+ e_2, e_1+ e_2+e_3, e_1+ e_2+e_3+e_4 \}$;

(iii) $\mathscr{E}$ = the canonical base of $\mathbb{R}^4$.

Hint. Identify the transition matrices.

10. Let the operator **U**:$\mathcal{M}_{2,2}(\Gamma) \to \mathcal{M}_{2,2}(\Gamma)$ be defined by

$$\mathbf{U}(A) = T^{-1} A\, T,$$

where T is a fixed non-singular matrix. Show that **U** is linear, and find the matrix which represents **U** in the canonical base

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Show that **U** (as a binary relation) is an equivalence, but in particular, for

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ and } B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

we have $(A,B) \notin \mathbf{U}$, even if A and B have the same *proper values.*

Hint. Introduce the components of T. Since A is the unit of $\mathcal{M}_{2,2}(\Gamma)$, we have $\mathrm{U}(A) = A \neq B$ for any nonsingular matrix T, even if

$$\mathrm{Det}(A - \lambda I) \equiv \mathrm{Det}(B - \lambda I)\ .$$

11. Show that if the matrices A and B represent the same linear operator **U**: $\Gamma^n \to \Gamma^n$ in different bases, then:

(i) $\mathrm{Det}\, A = \mathrm{Det}\, B$, and

(ii) $\mathrm{Trace}\, A = \mathrm{Trace}\, B$.

Hint. Use the fact that $B = T^{-1} A\, T$ holds for some non-singular transition matrix $T \in \mathcal{M}_{n,n}(\Gamma)$, and take Det . Express $\mathrm{Trace}\, A \overset{def.}{=} \sum_{i=1}^{n} a_{ii}$ in terms of *proper values of A,* taking into account the relations between the roots and the coefficients of the *characteristic equation,* $\mathrm{Det}\,(A - \lambda I) = 0$.

12. Analyze whether the following matrices

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \ B = \begin{pmatrix} 6 & -5 & -3 & 0 \\ 3 & -2 & -2 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

reduce to a diagonal form by a change of bases. Find the new bases and the corresponding diagonal form.

Hint. Consider that these matrices represent some linear operators on $\mathbb{R}^4$ in respect to the canonical base $\mathscr{C}$, and look for a transition matrix T, such that $T^{-1} A\, T$, respectively $T^{-1} B\, T$ are diagonal matrices. In particular, if

$$\mathscr{B} = \{ f_1 = (1,1,0,0),\ f_2 = (1,0,1,0),\ f_3 = (1,0,0,1),\ f_4 = (1,-1,-1,-1)\},$$

is the *new* base, then matrix A transforms into

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

Matrix B cannot be reduced to a diagonal form. Alternatively we may study whether relation $\mathbf{U}(x) = \lambda x$ holds for four linearly independent vectors (i.e. $\mathbf{U}$ has such *proper vectors*).

13. Establish the general form of a linear functional $f : \Gamma \longrightarrow \Gamma$, where Γ is either $\mathbb{R}$ or $\mathbb{C}$.

Hint. Take a base in Γ, e.g. $\mathscr{B} = \{1\}$, and apply theorem 3.18 from above. The form is $f(z) = k\, z$, where $k = f(1)$.

14. Reduce the rotation of angle 2α in the plane to a rotation of angle α and a central symmetry.

Hint. Apply the Cayley-Hamilton theorem to the matrix

$$A = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix},$$

which represents a rotation of angle α. The condition $P_A(A) = O_n$ becomes

$$A^2 = (2\cos\alpha)\, A - I_2,$$

where A^2 represents a rotation of angle 2α.

§ I.4. ELEMENTS OF TOPOLOGY

Mathematical Analysis is definable as a combined study of two types of structures, namely *algebraical* and *topological* (from τοπωσ = *place*). This is visible at the very beginning, when the *neighborhoods* of the points are defined in spaces already organized from the algebraic point of view, most frequently in linear spaces. The topological structures are necessary in the definition of a *limit*, which represents the main notion of the Mathematical Analysis. In the chapters of this theory, mainly dedicated to convergence, continuity, to differential and integral calculus, we study several particular cases of *limits*.

The elements of topology, which are presented in this section, will be grouped in two parts: the former will be devoted to the general topological notions and properties; the second concerns spaces where the topology derives from some particular structures (e.g. *norms* or *metrics*).

§ I. 4. Part 1. GENERAL TOPOLOGICAL STRUCTURES

We assume that elements of analysis on $\mathbb{R}$ are already known, and we take them as a starting point. We remind that the *neighborhoods* of a point $x\in\mathbb{R}$ are defined using the notion of *interval*,

$$(a, b) = \{\xi\in\mathbb{R}: a < \xi < b\},$$

which derives from the *order* of $\mathbb{R}$. More exactly, a set $V\subseteq\mathbb{R}$ is said to be a *neighborhood of* x iff there exists $a, b \in\mathbb{R}$ such that $x \in (a,b) \subseteq V$.

Alternatively, instead of (a, b) we may use a symmetric interval *centered at* x , of *radius* $\varepsilon > 0$, which is

$$I(x, \varepsilon) = \{y\in\mathbb{R}: x - \varepsilon < y < x+ \varepsilon \}.$$

If we try to introduce a similar structure in $\mathbb{C}$, which represents another very important set of numbers, we see that this technique doesn't work any more, since $\mathbb{C}$ has no proper order, compatible with its algebraic structure. The alternative definition is obtained if the symmetric intervals from above are replaced by *discs* (*balls,* or *spheres*) relative to the usual norm (*modulus*) on $\mathbb{R}$, namely

$$I(x, \varepsilon) = S(x, \varepsilon) = \{ y\in\mathbb{R} : | x - y | < \varepsilon \}.$$

Using the *modulus* in $\mathbb{C}$, the notion of *disc* (*sphere, ball*) of *center* z and *radius* $r > 0$, is similarly defined by

$$S(z, r) = \{\zeta\in\mathbb{C}: | \zeta - z | < r\} .$$

The *topology* of $\mathbb{C}$ can be specified by the following:

4.1. Definition. Set $V \subseteq \mathbb{C}$ is called *neighborhood of* $z \in \mathbb{C}$ *iff* there exists $r > 0$, such that $V \supseteq S(z, r)$.

The family of all neighborhoods of z will be noted by $\mathscr{V}(z)$. The structure of $\mathbb{C}$, realized by defining the family $\mathscr{V}(z)$ of neighborhoods for each $z \in \mathbb{C}$, is called *Euclidean topology* of $\mathbb{C}$.

4.2. Remark. So far, we have two examples of topologies, which concern two of the most usual sets of numbers. They are similar in many respects, especially in the role of the modulus, which turns out to be the Euclidean *metric*. Being derived from Euclidean metrics, the resulting topologies are called Euclidean too. The construction of a topology attached to a metric will be studied in more details in the second part of this section. For now it is important to make evident those properties of the neighborhoods, which are significant enough to be adopted in the general definition.

4.3. Proposition. The systems of neighborhoods, corresponding to the usual (Euclidean) topologies on $\mathbb{R}$ and $\mathbb{C}$, satisfy the conditions:

[N_1] $x \in V$ for each $V \in \mathscr{V}(x)$;

[N_2] If $V \in \mathscr{V}(x)$ and $U \supseteq V$, then $U \in \mathscr{V}(x)$;

[N_3] If $U, V \in \mathscr{V}(x)$, then $U \cap V \in \mathscr{V}(x)$;

[N_4] For any $V \in \mathscr{V}(x)$ there exists $W \in \mathscr{V}(x)$ such that for all $y \in W$ we have $V \in \mathscr{V}(y)$.

Proof. The former three properties are obvious. For [N_4], if $V \supseteq S(x, r)$, then we can take $W = S(x, \frac{r}{2})$, since $S(y, \frac{r}{2}) \subseteq V$ for all $y \in W$. ◇

Because properties [N_1] – [N_4] hold in plenty of examples, they are taken as axioms in the "abstract notion" of *topological structure*, namely:

4.4. Definition. Let $\mathcal{S} \neq \emptyset$ be an arbitrary set. Any function

$$\tau : \mathcal{S} \to \mathscr{P}(\mathscr{P}(\mathcal{S})) ,$$

which attaches to each $x \in \mathcal{S}$ a *system of neighborhoods* of x , noted

$$\tau(x) = \mathscr{V}(x) ,$$

is called *topology* on $\mathcal{S}$ iff $\mathscr{V}(x)$ satisfies the above conditions [N_1] – [N_4] at each $x \in \mathcal{S}$. The forthcoming structure on $\mathcal{S}$ is called *topology*, and $\mathcal{S}$, endowed with this structure, is said to be a *topological space*; it is most frequently noted as a pair $(\mathcal{S}, \tau)$.

4.5. <u>Examples</u> of topological spaces:

(i) The Euclidean Γ (i.e. ℝ or ℂ) from above.

(ii) The Euclidean Γ^n, $n \in \mathbb{N}^*$, and generally any metric space.

(iii) Let $(\mathcal{S}, \le)$ be a totally ordered nonvoid set, which has the property:

$$\forall x \in \mathcal{S} \ \ \exists a, b \in \mathcal{S} \ \ \textit{such that} \ \ a < x < b \,,$$

and let the intervals (a, b), $[a, b)$, etc. be defined like in $(\mathbb{R}, \le)$. Then the following notations make sense:

$$\mathscr{V}_E(x) = \{V \subseteq \mathcal{S} : \exists\, a, b \in \mathcal{S} \textit{ such that } x \in (a, b) \subseteq V\};$$

$$\mathscr{V}_l(x) = \{V \subseteq \mathcal{S} : \exists\, a \in \mathcal{S} \textit{ such that } (a, x] \subseteq V\};$$

$$\mathscr{V}_r(x) = \{V \subseteq \mathcal{S} : \exists\, b \in \mathcal{S} \textit{ such that } [x, b) \subseteq V\};$$

$$\mathscr{V}_{\rightarrow}(x) = \{V \subseteq \mathcal{S} : \exists\, a \in \mathcal{S} \textit{ such that } x \in (a, \rightarrow) \subseteq V\};$$

$$\mathscr{V}_{\leftarrow}(x) = \{V \subseteq \mathcal{S} : \exists\, b \in \mathcal{S} \textit{ such that } x \in (\leftarrow, b) \subseteq V\}.$$

Each of these families satisfies conditions $[N_1]$ – $[N_4]$, hence each one can be considered system of neighborhoods. The corresponding topologies are respectively called:

- *Euclidean* (or *topology of open intervals*) if $\tau(x) = \mathscr{V}_E(x)$;
- *topology of half-intervals to the left* if $\tau(x) = \mathscr{V}_l(x)$;
- *topology of half-intervals to the right* if $\tau(x) = \mathscr{V}_r(x)$;
- *topology of unbounded intervals to the right* if $\tau(x) = \mathscr{V}_{\rightarrow}(x)$;
- *topology of unbounded intervals to the left* if $\tau(x) = \mathscr{V}_{\leftarrow}(x)$.

In particular, we may consider $\mathcal{S} = \mathbb{R}$, endowed with its natural order, or $\mathcal{S} = \mathbb{R}^2$ with its lexicographic order, etc.

(iv) Let $(D, \le)$ be a directed set, and let ∞ be an element subject to the single condition $\infty \notin D$ (in addition, the order $\le$ is frequently extended by considering that ∞ is the greatest element). We note $\overline{D} = D \cup \{\infty\}$, and we claim that function $\theta : \overline{D} \to \mathscr{P}(\mathscr{P}(\overline{D}))$, expressed by

$$\theta(x) = \begin{cases} \{V \subseteq \overline{D} : x \in V\} & \textit{if } \ x \in D \\ \{V \subseteq \overline{D} : V \supseteq \{\infty\} \cup (d, \rightarrow), d \in D\} & \textit{if } \ x = \infty \end{cases}$$

represents a topology on $\overline{D}$.

Such topologies on directed sets are called *intrinsic*, and they are tacitly involved in the theory of convergence of generalized sequences (nets, in the sense of definition I.1.15). In particular, ℕ is a directed set, and we may remark that the above neighborhoods of ∞ are used to express the *convergence* of a sequence.

(v) On any nonvoid set $\mathcal{S}$ we may consider two "extreme" cases, namely:

- *the discrete topology* $\tau_d(x) = \{V \subseteq \mathcal{S} : x \in V\}$, and
- *the rough* (*anti- discrete, or trivial*) *topology* $\tau_t(x) = \{\mathcal{S}\}$.

These topologies are *extreme* in the sense that $\tau_d(x)$ is the greatest family of neighborhoods, while $\tau_t(x)$ is the smallest possible. We mention that one sense of *discreteness* in classical analysis reduces to endow the space with its discrete topology τ_d .

From the notion of neighborhood, we may derive other terms, which form the "vocabulary" of any topological study, as for example:

4.6. Definition. Point x is *interior* to a set A iff it has a neighborhood V such that $V \subseteq A$. The set of all interior points of A is called *the interior of A,* and it is noted $\mathring{A}$, A^0 , or $\iota(A)$. We say that A is *open* iff $A = \mathring{A}$.

Point y is said to be *adherent* to a set A iff $V \cap A \neq \emptyset$ holds for arbitrary $V \in \mathscr{V}(y)$. The set of all adherent points of A forms *the adherence* (or *the closure*) of A, which is noted A^{-}, $\overline{A}$, or $\alpha(A)$. In the case when $A = \overline{A}$, we say that A is *closed.*

We say that z is an *accumulation point of* A iff $A \cap (V \setminus \{z\}) \neq \emptyset$ for all $V \in \mathscr{V}(z)$. The set of all accumulation points of A is called *derivative of A,* and it is noted A' , or $\delta(A)$.

Point w is named *frontier* (or *boundary*) *point of* A iff both $A \cap V \neq \emptyset$, and $\complement A \cap V \neq \emptyset$ for any $V \in \mathscr{V}(w)$. The set of all such points forms *the frontier of A,* which is noted $A^{\sim}$, or ∂A .

Constructing the interior, adherence, derivative, and the frontier of a set is sometimes meant as the action of some *topological operators*, namely: ι = *interior*, α = *adherence*, δ = *derivation*, and ∂ = *frontier*.

4.7. Proposition. Family $\mathscr{G}$ of all open sets in the topological space $(\mathcal{S}, \tau)$ has the following properties:

[G_1] $\emptyset, \mathcal{S} \in \mathscr{G}$;

[G_2] $A, B \in \mathscr{G} \Rightarrow A \cap B \in \mathscr{G}$;

[G_3] $A_i \in \mathscr{G}$ for all $i \in I$ (arbitrary) $\Rightarrow \cup\{A_i : i \in I\} \in \mathscr{G}$.

Conversely, we can completely restore topology τ in terms of $\mathscr{G}$**.**

The proof is directly based on definitions, and therefore it is omitted, but the inexperienced reader may take it as an exercise.

A dual proposition holds for the family of closed sets, which, instead of [G_2] and [G_3], refers to finite unions and arbitrary intersections. A similar study involves the *topological operators,* as for example:

4.8. Proposition. The operator $\iota : \mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{S})$ has the properties:

$[I_1]$ $\iota(\mathcal{S}) = \mathcal{S}$;

$[I_2]$ $\iota(A) \subseteq A$, $\forall A \in \mathcal{P}(\mathcal{S})$;

$[I_3]$ $\iota(\iota(A)) = \iota(A)$, $\forall A \in \mathcal{P}(\mathcal{S})$;

$[I_4]$ $\iota(A \cap B) = \iota(A) \cap \iota(B)$, $\forall A, B \in \mathcal{P}(\mathcal{S})$.

Conversely, each ι, which satisfies $[I_1]$-$[I_4]$, uniquely determines τ .

The proof is direct, and will be omitted. The most sophisticated is $[I_3]$, which is based on $[N_4]$. We may similarly treat the *adherence*. The duality *open* / *closed*, *interior* / *adherence*, etc. can be explained by the following:

4.9. Proposition. The following relations hold in any topological space:

(i) $x \notin A^0 \Leftrightarrow x \in (\complement A)^-$, and $y \notin B^- \Leftrightarrow y \in (\complement B)^0$;

(ii) $\complement[\iota(A)] = \alpha[\complement(A)]$, and $\complement[\alpha(A)] = \iota[\complement(A)]$;

(iii) A is open $\Leftrightarrow$ $\complement A$ is closed;

4.10. Derived topologies. ***a) Topological subspaces.*** Let $(\mathcal{S}, \tau)$ be a topological space, and let $\mathcal{T} \neq \emptyset$ be a subset of $\mathcal{S}$. We say that $\mathcal{T}$ is a *topological subspace* of $\mathcal{S}$ if each $x \in \mathcal{T}$ has the neighborhoods

$$\tau_{\mathcal{T}}(x) = \{U = V \cap \mathcal{T} : V \in \tau(x) \}.$$

For example, the Euclidean $\mathbb{R}$ is a topological subspace of $\mathbb{C}$.

b) Topological products. If $(\mathcal{S}_1, \tau_1)$ and $(\mathcal{S}_2, \tau_2)$ are topological spaces, and $\mathcal{T} = \mathcal{S}_1 \times \mathcal{S}_2$, then the *product topology* of $\mathcal{T}$ is defined by

$$\tau_{\mathcal{T}}(x, y) = \{W \subseteq \mathcal{T} : \exists U \in \tau_1(x), V \in \tau_2(y) \textit{ such that } U \times V \subseteq W \}.$$

This construction can be extended to more than two topological spaces. In particular, $\mathbb{C} = \mathbb{R}^2$, and $\mathbb{R}^n$, where $n \geq 2$, are topological products.

c) Topological quotient. Let ~ be an equivalence on $\mathcal{S}$, and let $\mathcal{T} = \mathcal{S}/\sim$ be the set of equivalence classes. A topology τ on $\mathcal{S}$ is said to be *compatible* with ~ if $\tau(x) = \tau(y)$ whenever $x \sim y$. If so, we can define the *quotient topology* of $\mathcal{T}$, which attaches to each class $x^\wedge$ the neighborhoods

$$\tau_{\mathcal{T}}(x^\wedge) = \{V^\wedge \subseteq \mathcal{T} : V \in \tau(x) \}.$$

Simple examples can be done in $\mathbb{C} = \mathbb{R}^2$, and $\mathbb{R}^n$.

§ I.4. Part 2. SCALAR PRODUCTS, NORMS AND METRICS

In this part we study those particular topological spaces, which occur in the classical construction of a topology, based on some intrinsic structures of these spaces. In particular, we shall explain the scheme

Scalar product $\Rightarrow$ *Norm* $\Rightarrow$ *Metric* $\Rightarrow$ *Topology,*

where $\Rightarrow$ means "generates". Of course, we are especially interested in such a construction whenever we deal with linear spaces.

4.11. <u>Definition.</u> If $\mathcal{L}$ be a linear space over Γ, then functional

$$< . , . > : \mathcal{L} \mathrm{x} \mathcal{L} \to \Gamma$$

is called *scalar product* on $\mathcal{L}$ iff the following conditions hold:

$[SP_1]$ $<\alpha x + \beta y, z> = \alpha <x,z> + \beta <y,z>, \ \forall x, y, z \in \mathcal{L}, \ \alpha, \beta \in \Gamma$ (*linearity*);

$[SP_2]$ $<x, y> = \overline{<y, x>}, \ \forall x, y \in \mathcal{L}$ (*skew symmetry*; ‾ = conjugation);

$[SP_3]$ $<x, x> \geq 0$ at any $x \in \mathcal{L}$ (*positiveness*);

$[SP_4]$ $<x, x> = 0 \ (\in \Gamma) \Leftrightarrow x = 0 \ (\in \mathcal{L})$ (*non-degeneration*).

The pair $(\mathcal{L}, < . , . >)$ is called *scalar product space.*

If Γ = ℝ, the second condition becomes an ordinary symmetry, since the bar ‾ stands for the complex conjugation. Condition $[SP_2]$ is implicitly used in $[SP_3]$, where $<x, x> \in \mathbb{R}$ is essential.

4.12. <u>Examples.</u> If $\mathcal{L} = \mathbb{R}$, then obviously $<x, y> = xy$ is a scalar product. Similarly, if $\mathcal{L} = \mathbb{C}$, we take $<\zeta, z> = \zeta \bar{z}$. The cases $\mathcal{L} = \mathbb{R}^2$ and $\mathcal{L} = \mathbb{R}^3$ are well known from Geometry, where the scalar product of two vectors is defined as "the product of their length by the cosine of the angle between them". The analytic expression of this scalar product in $\mathbb{R}^3$ is

$$<x, y> = x_1 y_1 + x_2 y_2 + x_3 y_3 ,$$

and it can be extended to Γ^n, $\forall n \in \mathbb{N}^*$, by the formula

$$<x, y> = x_1 \overline{y_1} + x_2 \overline{y_2} + ... + x_n \overline{y_n} .$$

The finite dimensional space $\mathcal{L} = \Gamma^n$, endowed with this scalar product, or with other derived from it structures, is qualified as *Euclidean.*

A slight generalization of these products is obtained by putting some *weights* $\alpha_1 > 0, \ldots, \alpha_n > 0$ therein, namely

$$<x, y> = \alpha_1 x_1 \overline{y_1} + ... + \alpha_n x_n \overline{y_n} .$$

The Euclidean scalar product is naturally extended to sequences in Γ, if appropriate conditions of convergence are assumed. The definition is

$$<x, y> = \sum_{n=1}^{\infty} \alpha_n x_n \overline{y_n} ,$$

for any pair of sequences $x, y \in \Gamma^{\mathbb{N}}$, where sequence $\alpha = (\alpha_n)$ of strictly positive terms represents *the weight* of the respective scalar product.

If we remark that the n-dimensional vectors are functions defined on a finite set $\{1, 2, \ldots, n\}$, and the sequences are functions defined on ℕ, we may extend the Euclidean scalar products from above to functions defined on compact intervals $[a,b] \subset \mathbb{R}$. More exactly, the *scalar product of two functions* $f, g : [a, b] \to \Gamma$, is defined by

$$<f, g> = \int_a^b \alpha(t) f(t) \overline{g}(t) dt ,$$

where $\alpha : [a, b] \to \mathbb{R}^*_+$ represents the *weight function*. Of course, adequate conditions of integrability are assumed, e.g. $\alpha, f, g \in \boldsymbol{C}_\Gamma([a, b])$, etc.

The following properties are frequently used in the calculus:

4.13. <u>Proposition.</u> If $(\mathcal{L}, < . , . >)$ is a scalar product space, then:

(i) $<x, 0> = <0, x> = 0$ for any $x \in \mathcal{L}$;

(ii) $<x, y + z> = <x, y> + <x, z>$ for any $x, y, z \in \mathcal{L}$;

(iii) $<x, \alpha y> = \overline{\alpha} <x, y>$ for arbitrary $x, y \in \mathcal{L}$ and $\alpha \in \Gamma$.

The proof is directly based on the definition.

4.14. <u>The fundamental inequality</u>. (*Cauchy-Buniakowski-Schwarz*) If $(\mathcal{L}, < . , . >)$ is a scalar product space, then for all $x, y \in \mathcal{L}$ we have

$$| <x, y> |^2 \leq <x, x> <y, y> , \qquad (*)$$

with equality iff x and y are linearly dependent.

<u>Proof.</u> According to [SP_2], for any $\lambda \in \Gamma$ we have

$$T(\lambda) = <x + \lambda y, x + \lambda y> \geq 0 .$$

If $y \neq 0$, then we replace $\lambda = -\dfrac{<x, y>}{<y, y>}$ in T. Otherwise, it reduces to the obvious equality $| <x, 0> |^2 = 0 = <x, x> \cdot 0$. For $x = \lambda y$ we obviously have equality in (*) , i.e. $| \lambda <y, y> |^2 = \lambda \overline{\lambda} <y, y>^2$.

Conversely, if we suppose that (*) holds with equality for some $x \neq 0 \neq y$, then for $\lambda_0 = -\dfrac{<x, x>}{<y, x>}$ it follows that $T(\lambda_0) = 0$. Consequently, according to [SP_4] , we have $x + \lambda_0 y = 0$. ◊

Now we can show that the scalar products generate norms:

4.15. Definition. Functional $\| \cdot \| : \mathcal{L} \to \mathbb{R}_+$, where $\mathcal{L}$ is a linear space over Γ , is called *norm* iff it satisfies the conditions:

$[N_1]$ $\| x \| = 0$ iff $x = 0$ (*non degeneracy*);

$[N_2]$ $\| \lambda x \| = |\lambda| \, \|x\|$ for any $x \in \mathcal{L}$ and $\lambda \in \Gamma$ (*homogeneity*);

$[N_3]$ $\| x + y \| \leq \| x \| + \| y \|$ for all $x, y \in \mathcal{L}$ (*sub-additivity*).

Correspondingly, the pair $(\mathcal{L}, \| \cdot \|)$ is named *normed linear space.*

4.16. Corollary. Any scalar product space $(\mathcal{L}, < . , . >)$ is normed by

$$\| x \| = \sqrt{< x, x >} \ .$$

Proof. The functional $\|\cdot\|$ is well defined because of $[SP_3]$. It is no difficulty in reducing $[N_1]$ and $[N_2]$ to $[SP_4]$, $[SP_1]$ and $[SP_2]$. Finally, $[N_3]$ is a consequence of (*), because

$$\| x + y \|^2 = <x + y, x + y> = <x, x> + 2\mathrm{Re} <x, y> + <y, y> \leq$$
$$\leq \| x \|^2 + 2 \, |<x, y>| + \| y \|^2 \leq (\| x \| + \| y \|)^2.$$

The equality in $[N_3]$ holds iff the vectors are linearly dependent. ◇

4.17. Remarks. ***(i)*** There exist norms, which cannot be derived from scalar products, i.e. following the above corollary. For example:

- The *sup*-norm, acting on the space $C_\Gamma([a, b])$ of all continuous functions on the closed interval $[a, b] \subset \mathbb{R}$, defined by
$$\| f \|_{sup} = \sup \{| f(t)| : t \in [a, b]\}; \text{ and}$$
- The L^1 – norm, defined on the space $L^1_\Gamma([a, b])$ of equivalence classes of absolutely integrable functions on $[a, b] \subset \mathbb{R}$, defined by
$$\| f \|_{L^1} = \int_a^b | f(t)| dt \, .$$

(ii) There is a simple test for establishing whether a given norm derives or not from a scalar product, namely checking the formula

$$\| x + y \|^2 + \| x - y \|^2 = 2(\| x \|^2 + \| y \|^2).$$

If this equality is satisfied, and $\mathcal{L}$ is a complex linear space, then

$$<x, y> = \tfrac{1}{4}(\| x + y \|^2 - \| x - y \|^2 + i \, \| x + iy \|^2 - i \, \| x - iy \|^2)$$

is a scalar product that generates $\| \cdot \|$. For real linear spaces we have:

$$<x, y> = \tfrac{1}{4}(\| x + y \|^2 - \| x - y \|^2) \, .$$

(iii) In any real scalar product space $(\mathcal{L}, < . , . >)$, function

$$\mu_{\sphericalangle} : (\mathcal{L} \setminus\{0\}) \times (\mathcal{L} \setminus\{0\}) \to [0, \pi] \subset \mathbb{R},$$

expressed by

$$\mu_{\sphericalangle}(x, y) = arc\ cos\ \frac{< x, y >}{\|x\|.\|y\|},$$

defines the *measure of the angle* between two non-null vectors $x, y \in \mathcal{L}$.

In fact, $\mu_{\sphericalangle}$ is well defined because of (*).

The notions of *norm*, *angle*, and *orthogonality*, introduced in the next definition, represent the starting point of the Euclidean (metric) geometry on any scalar product space $(\mathcal{L}, < . , . >)$, including $\mathbb{R}^n$.

4.18. Definition. Two elements $x, y \in \mathcal{L}$, where $(\mathcal{L}, < . , . >)$ is a scalar product space, are said to be *orthogonal* iff $<x, y> = 0$. In this case we note $x \perp y$, i.e. sign $\perp$ stands for *orthogonality* as a binary relation on $\mathcal{L}$. More generally, a set of vectors $\mathscr{S} = \{x \in \mathcal{L}: i \in I\}$, where I is an arbitrary family of indices, is called *orthogonal system* iff $x_i \perp x_j$ whenever $i \neq j$.

Two sets $A, B \subseteq \mathcal{L}$ are considered *orthogonal* (to each other) iff $x \perp y$ holds for arbitrary $x \in A$ and $y \in B$. In this case we note $A \perp B$.

The orthogonal complement of A is defined by

$$A^{\perp} = \{y \in \mathcal{L}: x \perp y \text{ for all } x \in A \}.$$

4.19. Proposition. Every orthogonal system of vectors is linearly independent.

Proof. Let us consider a null linear combination of non-null vectors

$$C_1 x_1 + C_2 x_2 + \ldots + C_n x_n = 0 .$$

To show that $C_k = 0$ for all $k = \overline{1,n}$, we multiply by x_k , and we obtain $C_k \|x_k\|^2 = 0$, where $\|x_k\|^2 \neq 0$. ◇

An immediate consequence of this property establishes that every maximal orthogonal system forms a *base* of $\mathcal{L}$.

4.20. Proposition. (Pythagoras formula) If $\{x_1, x_2, \ldots, x_n\}$ is an orthogonal system, and $x = x_1 + x_2 + \ldots + x_n$, then

$$\|x\|^2 = \sum_{k=1}^{n} \|x_k\|^2 .$$

Proof. In the development of $<x, x>$ we replace $<x_i, x_k>$ by 0 if $i \neq k$, and by $\|x_i\|^2$ if $i = k$. ◇

The next level of generality refers to *metrics*:

4.21. <u>Definition.</u> Let $\mathcal{S}$ be an arbitrary non-void set (generally non-linear). A functional $\rho : \mathcal{S}\mathrm{x}\mathcal{S} \to \mathbb{R}_+$ is called *metric* on $\mathcal{S}$ iff:

[M_1] $\rho(x, y) = 0 \Leftrightarrow x = y$ (non-degeneration);
[M_2] $\rho(x, y) = \rho(y, x)$ for any $x, y \in \mathcal{S}$ (symmetry);

[M_3] $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for any $x, y, z \in \mathcal{S}$ (sub-additivity).

The pair $(\mathcal{S}, \rho)$ forms a *metric space*. The values of ρ are named *distances*.

If instead of [M_1], we have only $\rho(x, x) = 0$, then ρ is said to be a *pseudo-metric* (briefly *p-metric*), and $(\mathcal{S}, \rho)$ is called *pseudo-metric space*.

Condition [M_3] is frequently referred to as *rule of triangles*. Later we will see that the (p-) metrics are directly used to construct topologies, and this rule has an essential contribution to this construction.

4.22. <u>Examples.</u> 1. If $\mathcal{S}=\mathcal{L}$ is a linear space, and $\|\,.\,\|$ is a norm on it, then $\rho(x, y) = \|x-y\|$ is a metric. In particular, if $\mathcal{L}= \Gamma^n$, then its *Euclidean metric* is obtained on this way from the Euclidean norm, namely:

$$\rho(x,y)=\sqrt{\sum_{k=1}^{n}|x_k - y_k|^2}\ .$$

2. On non-linear spaces we cannot speak of norms, but it's still possible to define metrics, e.g. by restricting some metric of $\mathcal{L}$ (linear) to a non-linear subset $\mathcal{S}\subset\mathcal{L}$. Sometimes linear spaces are endowed with metrics that do not derive from norms, as for example s_Γ, which consists of all sequences in Γ. In fact, functional q defined for any sequence $x = (x_n)$ by

$$q(x) = \sum_{n=1}^{\infty}\frac{1}{2^n}\frac{|x_n|}{1+|x_n|}\ ,$$

isn't a norm (since $q(\lambda x)\neq|\lambda|\,q(x)$!), but $\rho(x, y) = q(x - y)$ is a metric.

3. Let $\mathcal{S}$ be an arbitrary non-void set, and let $\rho : \mathcal{S}\mathrm{x}\mathcal{S} \to \mathbb{R}_+$ be defined by

$$\rho(x, y) = \begin{cases} 0 & if \quad x = y \\ 1 & if \quad x \neq y \end{cases}\ .$$

Then $(\mathcal{S}, \rho)$ is a metric space, and even if $\mathcal{S}$ is a linear space, ρ cannot be derived from a norm. Because ρ generates the discrete topology on $\mathcal{S}$ (see the example I.4.5.*v*), it is usually named *discrete* metric.

Now we show how p-metrics generate topologies, i.e. we describe the *intrinsic topology* of a p-metric space. This is a very general construction, but in principle it repeats what we know in $\mathbb{R}$ and $\mathbb{C}$.

4.23. <u>Theorem.</u> Let ($\mathcal{S}$, ρ) be a p-metric space, and let us note *the open ball* (*sphere*) *of center x and radius* $r > 0$ by

$$S(x, r) = \{y \in \mathcal{S}: \rho(x, y) < r\}.$$

Then function $\tau : \mathcal{S} \to \mathscr{P}(\mathscr{P}(\mathcal{S}))$, expressed at any $x \in \mathcal{S}$ by

$$\tau (x) = \{V \subseteq \mathcal{S} : \exists r > 0 \text{ such that } S(x, r) \subseteq V\},$$

is a topology on $\mathcal{S}$, usually called *metric topology*.

In addition, if ρ is a metric (i.e. $\rho(x, y) = 0 \Rightarrow x = y$), then ($\mathcal{S}$, τ) is *separated*, i.e. distinct points have disjoint neighborhoods. We mention that there are plenty of *separation axioms*, and this one is known as [T_2]; see later its role in the uniqueness of the limit.

<u>Proof.</u> We have to verify conditions [N_1]-[N_4] from definition I.4.1. In particular we discuss [N_4], since [N_1]-[N_3] are obvious. In fact, let V be a neighborhood of x , and $S(x, r)$ be a sphere contained in V. We claim that the sphere $W = S(x, r/2)$ fulfils [N_4]. In fact, since for any $y \in W$, we have $S(y, r/2) \subseteq V$, we obtain $V \in \tau (y)$.

In particular, let ρ be a metric, and let $x, y \in \mathcal{S}$, $x \neq y$. Because of [M_1], we have $\rho(x, y) = r > 0$, hence $S(x, r/3)$ and $S(y, r/3)$ are samples of disjoint neighborhoods, as asked by the condition of separation. ◇

4.24. <u>Remarks.</u> 1) We may conceive Mathematical Analysis as a *two levels* theory: at a *quantitative level* it deals with numbers, vectors, and metric measurements, but at a *qualitative* one it involves limits, convergence, continuity, and other topological notions.

2) We frequently use the term *Euclidean* to qualify several things, namely:

- the natural topology of $\mathbb{R}$, $\mathbb{C}$, and more generally Γ^n ;
- the scalar product $<x, y> = \sum_{k=1}^{n} x_k \overline{y}_k$ in Γ^n , $n \in \mathbb{N}^*$;
- the norm $\| x \| = \sqrt{<x, \overline{x}>}$ of the same Γ^n ;
- the metric $\rho(x, y) = \| x - y \|$ of Γ^n.

The common feature of all these situations, which justifies the use of the same terminology, is reflected in the forthcoming topology. In other words, by an *Euclidean topological space* we understand the *n*-dimensional linear space Γ^n , $n \in \mathbb{N}^*$, endowed with the Euclidean metric topology.

PROBLEMS § I.4.

1. Compare the following topologies of $\mathbb{R}$: τ_e = Euclidean, τ_0 = rough, τ_1 =discrete, τ_l = to the left, and τ_r = to the right, where

$\tau_l(x) = \{V \subseteq \mathbb{R}: \exists \varepsilon > 0 \text{ such that } V \supseteq (x - \varepsilon, x]\}$,

$\tau_r(x) = \{V \subseteq \mathbb{R}: \exists \varepsilon > 0 \text{ such that } V \supseteq [x, x + \varepsilon)\}$.

Find and compare the corresponding families of open sets.

Hint. $\tau_0 \subset \tau_e = \inf\{\tau_l, \tau_r\} \subset \sup\{\tau_l, \tau_r\} = \tau_1$; τ_l and τ_r are not comparable. The families of open sets are: $\mathcal{G}_0 = \{\emptyset, \mathcal{S}\}$ for τ_0 , $\mathcal{G}_1 = \mathcal{P}(\mathcal{S})$ for τ_1 , and arbitrary unions of intervals of the form $(x - \varepsilon, x + \varepsilon)$, $(x - \varepsilon, x]$, and respectively $[x, x + \varepsilon)$, for the last ones.

2. Show that $\overline{\mathbb{C}}$ is homeomorphic to the Riemannian sphere (i.e. there is a 1:1 correspondence between the two sets, which carries neighborhoods from one set onto neighborhoods in the other space. Study whether $\overline{\mathbb{R}}$ has a homeomorphic copy on this sphere.

Hint. The small circular neighborhoods of points ($\neq N$) of the sphere are stereographically projected into small discs in $\mathbb{C}$. The pole N of the sphere corresponds to $\infty \in \overline{\mathbb{C}}$, and its circular neighborhoods correspond to sets of the form $\complement S(0, r)$. A similar representation of $\mathbb{R}$ is impossible since the two points $\pm\infty$ would correspond to the same N.

3. Let $\leq$ be the product order on $\mathbb{R}^n$, and let function $\omega: \mathbb{R}^n \to \mathcal{P}(\mathcal{P}(\mathbb{R}^n))$ be defined by:

$$\omega(x) = \{V \subseteq \mathbb{R}^n : \exists\ a, b \in \mathbb{R}^n \textit{ such that } x \in (a, b) \subseteq V\}.$$

Show that ω is a topology on $\mathbb{R}^n$ (called *product order topology*), which is equivalent to the product topology of $\mathbb{R}^n$. Compare this topology to the Euclidean one.

Hint. Because the order intervals have the form $(a, b) = \mathsf{X}\{(a_k, b_k): k = \overline{1, n}\}$, it follows that $V \in \omega(x)$ iff it is a neighborhood of x relative to the Euclidean topology.

4. Analyze the following sets from a topological point of view:

$A = \{n^{-1} : n \in \mathbb{N}^*\} \cup (1, 2]$ in $\mathbb{R}$;

$B = \{t + i \sin t^{-1} : t \in (0, 2/\pi)\}$ in $\mathbb{C}$;

$C = \{(z, t) \in \mathbb{C} \times \mathbb{R} : |z| \in (1, 2], \arg z \in [1, 2), t \in \mathbb{Q} \cap [1, 2]\}$.

Hint. Find the interior, adherent, accumulation, and boundary points, and establish which of the given sets are open, closed, etc.

5. Show that a topological space ($\mathcal{S}$, τ) is separated, i.e.

[T_2] $\forall x, y \in \mathcal{S},\ x \neq y$, $\exists U \in \tau(x)$, $\exists V \in \tau(y)$ *such that* $U \cap V = \emptyset$,

iff the diagonal $\delta = \{(x, x) : x \in \mathcal{S}\}$, which represents the equality on $\mathcal{S}$, is closed relative to the product topology of $\mathcal{S} \times \mathcal{S}$.

Hint. Replace the assertion "δ *is closed*" by "$x \neq y$ iff $(x, y) \notin \overline{\delta}$", and similarly, "($\mathcal{S}$, τ) *is separated*" by

"$\forall (x, y) \notin \delta\ \exists \mathrm{U} \in \mathcal{V}(x)\, and\, V \in \mathcal{V}(y)\, such\ that\ (\mathrm{U} \times V) \cap \delta = \emptyset$",
where obviously, $x \neq y$ holds if and only if $(x, y) \notin \delta$.

6. Let us note $A = \cup \{A_n : n \in \mathbb{N}\}$, where $A_n = \{1/2^n, 2/2^n, \ldots, (2^n-1)/2^n\}$. Show that $\overline{A} = [0, 1]$, i.e. A *is dense* in [0, 1], and interpret this fact in terms of binary approximation of $x \in [0, 1]$.
Hint. Divide [0, 1] in 2, 4, …, 2^n equal parts, and put either digit *0*, or *1*, according to the first or the second subinterval to which x belongs.

7. Let $\mathcal{S}$ be an arbitrary nonvoid set, $\varphi : \mathcal{S} \longrightarrow \mathbb{R}$ be 1:1 (i.e. injective), and $\rho : A \times A \longrightarrow \mathbb{R}_+$ be a metric on $A = \varphi(\mathcal{S}) \subseteq \mathbb{R}$. Show that:

a) $d_{\varphi,\rho} : \mathcal{S} \times \mathcal{S} \longrightarrow \mathbb{R}_+$, of values $d_{\varphi,\rho}(x, y) = \rho(\varphi(x), \varphi(y))$, is a metric.

b) If $\mathcal{S} = \mathbb{R}$, $\varphi(x) = \frac{x}{1+|x|}$, and $\rho(x, y) = |x - y|$, then ρ and $d_{\varphi,\rho}$ are two topologically equivalent metrics. Give an example when it is not so.
c) If card $\mathcal{S} \leq \aleph$, then for every metric d on $\mathcal{S}$, there exist φ and ρ such that $d = d_{\varphi,\rho}$. Is such a representation of d always possible ?
Hint. a) Verify the conditions in the definition I.4.21. b) Each Euclidean sphere contains some sphere relative to ρ, and conversely. Am example is

$$\varphi(x) = \begin{cases} x & if\ \ x \in \mathbb{R} \setminus \mathbb{Q} \\ n = \alpha(x) & if\ \ x \in \mathbb{Q}\ , \end{cases}$$

where α is a 1:1 correspondence of $\mathbb{Q}$ with $\mathbb{N}$.
c) An injective function φ exists iff card $\mathcal{S} \leq \aleph =$ card $\mathbb{R}$.

8. Show that in any linear normed space ($\mathcal{L}$, || . ||), the adherence $\overline{S}_{\mathrm{op}}$ of the open unit sphere centered at *0*, $S_{\mathrm{op}} = \{x \in \mathcal{L} : \|x\| < 1\}$, equals the closed unit sphere $S_{\mathrm{cl}} = \{x \in \mathcal{L} : \|x\| \leq 1\}$. What happens in general metric spaces?

Hint. Any point x from the boundary of S_{cl} , is adherent to the radius

$$\{\lambda x: \lambda \in [0,1)\} \subseteq S_{op},$$

hence $S_{cl} \subseteq \overline{S}_{op}$. Conversely, if x is adherent to S_{op} , we take $y \in S_{op}$, where $||y|| < 1$, and $||x|| \le ||x - y|| + ||y||$. This property is not generally valid in metric spaces ($\mathcal{S}$, ρ), e.g. $\mathcal{S}$ is the metric subspace of the Euclidean $\mathbb{R}^2$,

$$\mathcal{S} = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1 \text{ or } y = 0\}.$$

9. Let $f, g : [-2, 1] \times [0, 3] \to \mathbb{R}$ be functions of values $f(x, y) = x^2 + y^2$, and $g(x, y) = 2xy$. Evaluate their sup - norms and the distance $\rho(f, g)$.

Hint. Because functions f and g are bounded, it follows that the sup-norms do exist. By connecting these functions to the notions of Euclidean distance and area, we obtain the values

$$||f|| = \sup\{|f(x, y)|: (x, y) \in [-2, 1] \times [0, 3]\} = f(-2, 3), \text{ and}$$

$$||g|| = -g(-2, 3) .$$

Since $(f - g)(x, y) = (x - y)^2$, we find $\rho(f, g) = ||f - g|| = 25$.

10. Let $\mathcal{S}$, φ , ρ and $d_{\varphi,\rho}$ be defined as in the problem 7 b, and let ν be a prolongation of φ to $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, where $\nu(-\infty) = -1$ and $\nu(+\infty) = 1$. Show that $d_{\nu,\rho}$ is a bounded metric on $\mathcal{S}$, and find out the form of the open and closed spheres of center x (including $x = \pm\infty$) and radius r.

Hint. $S(+\infty, r) = (\frac{1-r}{r}, +\infty]$ for any $r \in (0, 1)$, etc.

11. Let ($\mathcal{L}$, || . ||) be a normed linear space, and let $\mathcal{F}$ be a linear finite dimensional subspace of $\mathcal{L}$. Show that for any $x \in \mathcal{L} \setminus \mathcal{F}$ there is $x^* \in \mathcal{F}$ (called *best approximation element*) such that

$$|| x - x^*|| = \min \{||x - y||: y \in \mathcal{F} \} .$$

Find the best approximation of $x = exp$ in the linear subspace $\mathcal{F}$ of all polynomials of degree 2, where $\mathcal{L} = \boldsymbol{C}_{\mathbb{R}}([0, 1])$ is normed by

$$\|f\|_{L^2} = \sqrt{\int_0^1 f^2(t)dt} .$$

Hint. Because || . || derives from a scalar product, then

$$x^* = \mathrm{Pr}_{\mathcal{F}}(x),$$

i.e. $(x - x^*) \perp \mathcal{F}$, which furnishes x^*. In particular, we have to deduce the values of a, b, and c such that $(e^t - at^2 - bt - c) \perp \{1, t, t^2\}$.

CHAPTER II. COVERGENCE

§ II.1. NETS

The nets are a very important tool of the mathematical analysis, especially when the sets where they are considered in, are endowed with a topological structure. Convergence is the main topic in this framework, which is developed in terms of limit points, accumulation points, and other topological notions. There exist particular properties of nets (e.g. boundedness, the property of being fundamental, etc.), which are studied in metric spaces. More particularly, other properties need a supplementary structure; for example, we can speak of monotony only if an order relation already exists, and we can operate with nets only if the space is endowed with an algebraic structure, etc.

From a topological point of view, the notion of *net* is the most natural extension of that of *sequence*. The study of the nets helps in understanding the principles of the convergence theory, and more than this, it is effective in the measure and integration theory. Therefore, we devote the first part of this section to general aspects involving nets.

§ II.1. Part 1. GENERAL PROPERTIES OF NETS.

1.1. <u>Definition.</u> Let $(\mathcal{S}, \tau)$ be a topological space, $(D, \leq)$ be a directed set, and $f : D \to \mathcal{S}$ be a net (generalized sequence) in $\mathcal{S}$. We say that f is *convergent* to $l \in \mathcal{S}$ iff for any $V \in \tau(l)$ there exists $d \in D$ such that $f(a) \in V$ whenever $a \geq d$.. In this case, we say that l is a *limit* of the net f, and we note

$$l = \lim_D f = \lim_{d \to \infty} f(d),$$

$$f(d) \to l,\ l \in \mathrm{Lim}\, f, \text{ etc.}$$

where sign $\mathrm{Lim}\, f$ stands for the set of all limits. If $\mathrm{Lim}\, f \neq \varnothing$ we say that f is a *convergent* net; otherwise, if $\mathrm{Lim}\, f = \varnothing$, f is said to be *divergent*.

In particular, if $D = \mathbb{N}$, we say that sequence $f : \mathbb{N} \to \mathcal{S}$ is *convergent* to $l \in \mathcal{S}$, and we note $l = \lim_{n \to \infty} x_n$, $x_n \to l$, etc. (read l is *the limit of* (x_n), or x_n *tends to* l, etc.), iff for any neighborhood $V \in \tau(l)$ there exists $n(V) \in \mathbb{N}$ such that $n \geq n(V) \Rightarrow x_n \in V$. The *convergence* and *divergence* are similarly defined for sequences, meaning $\mathrm{Lim}\, f \neq \varnothing$, respectively $\mathrm{Lim}\, f = \varnothing$.

We mention that, generally speaking, the limit points are not unique. More exactly, the sets Lim f are singletons exactly in separated (i.e. [T_2]) spaces. In addition, to each net we may attach other types of points, e.g.:

1.2. <u>Definition.</u> Let ($\mathcal{S}$, τ) be a topological space, (D, $\leq$) be a directed set, and $f : D \to \mathcal{S}$ be a net in $\mathcal{S}$. We say that $x \in \mathcal{S}$ is an *accumulation point of the net f* iff the following condition holds:

$$\forall V \in \tau(x)\ \forall a \in D\ \exists b \in D,\ b > a, \text{ such that } f(b) \in V.$$

The set of all accumulation points of a net f is noted Acc f.

In particular, the same notion makes sense for sequences.

1.3. <u>Examples.</u> ***a)*** The construction of a definite Riemannian integral is mainly based on convergent nets. To justify this assertion, let us remind this construction. We start with a bounded function $f : [a, b] \longrightarrow \mathbb{R}$, and we consider *partitions* of [a, b], which are finite sets of subintervals

$$\delta = \{[x_{k-1}, x_k] : k = 1, 2, \ldots, n;\ a = x_0 < x_1 < \ldots < x_n = b\}.$$

The *norm* of δ is defined by $\nu(\delta) = \max\{x_k - x_{k-1} : [x_{k-1}, x_k] \in \delta\}$.

Then, for each partition, we choose *systems of intermediate points*

$$\xi(\delta) = \{\xi_k \in [x_{k-1}, x_k] \in \delta : k = \overline{1,n}\}.$$

Finally, we define the *integral sums* attached to δ and $\xi(\delta)$ by

$$\sigma_f(\delta, \xi(\delta)) = \sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}).$$

We say that f is *integrable* on [a, b] iff for all sequences (δ_n) of partitions, the corresponding sequences $\sigma_f(\delta_n, \xi(\delta_n))$ of integral sums have a common limit when $\nu(\delta_n) \to 0$, not depending on the systems on intermediate points.

By definition, this limit represents the *Riemannian integral* of f on [a, b], which is usually noted

$$I = \int_a^b f(x)\,dx = \lim_{n\to\infty} \sigma_f(\delta_n, \xi(\delta_n)).$$

It is easy to see that instead of this construction, which involves a lot of sequences, we better say that the net $\sigma_f : D \to \mathbb{R}$ is convergent to I, where D is the directed set from the example I.1.8.*(iii)* 4.

b) Let sequence $f : \mathbb{N} \to \mathbb{C}$ be defined by $f(n) = i^n$. It is easy to see that f is divergent, but it has four accumulation points, namely ± 1, and $\pm i$. There are also four constant subsequences, i.e. (i^{4p}), (i^{4p+1}), (i^{4p+2}), and (i^{4p+3}), which obviously are convergent to these points. In this example we may remark that the set of values, that is $\{f(n) : n \in \mathbb{N}\} = \{1, i, -1, -i\}$ has no accumulation point, hence we have generally to distinguish between *accumulation point of a sequence* (*net*) and *accumulation point of a set* (see also the next example).

c) Let $D = \mathbb{N}\text{x}\mathbb{N}^*$ be directed by the relation

$$(m, n) \le (p, q) \Leftrightarrow m \le p ,$$

and let $f : D \to \mathbb{R}^2$ be defined by $f(m, n) = (m, \frac{1}{n})$. If we endow $\mathbb{R}^2$ with its Euclidean topology, then any point of the form $(m, 0)$ is an accumulation point of the set $f(D)$, while $\text{Acc}\, f = \varnothing$.

d) The sequence g: $\mathbb{N} \to \mathbb{R}$, expressed by $g(n) = exp((-1)^n n)$, has a single accumulation point, namely $\text{Acc}\, g = \{0\}$, but $0 \notin \text{Lim}\, g$. More exactly, the set $\text{Lim}\, g$ is void, i.e. g is a divergent sequence.

e) The sequence h: $\mathbb{N} \to \overline{\Gamma}$, of values $h(n) = (-1)^n n$, is convergent if $\Gamma = \mathbb{C}$, but it is divergent if $\Gamma = \mathbb{R}$, when $\text{Acc}\, h = \{\pm\infty\}$. Thus, we may conclude that the form of the sets Lim and Acc essentially depends on the space in which the net is considered.

The study of accumulation points naturally involves the subnets and their limit points:

1.4. <u>Theorem.</u> In the terms of the above definitions we have:

(i) Any limit of a net *f* is an accumulation point of *f;*

(ii) If the space (X, τ) is separated (i.e. $[T_2]$), and the net *f* has at least two different accumulation points, then *f* is divergent;

(iii) The element *x* is an accumulation point of the net *f* iff it is the limit of some subnet of *f*.

(iv) Every accumulation point of a subnet of *f* is an accumulation point of the initial net *f*.

(v) The set $\text{Acc}\, f$ is always closed.

<u>Proof.</u> *(i)* If $V \in \tau(x)$ determines some $d \in D$ such that $f(b) \in V$ for all $d \le b \in D$, then $f(b) \in V$ holds for some $b \ge a$, where a is arbitrary in the directed set D .Consequently $\text{Lim}\, f \subseteq \text{Acc}\, f$.

(ii) Let us suppose that $\{x, x'\} \subseteq \text{Acc}\, f$, and still (by r.a.a.) $\text{Lim}\, f = \{\ell\}$, where $x \ne x'$, but possibly $\ell = x'$, say. As for sure it remains $\ell \ne x \in \text{Acc}\, f$. Using the fact that (X, τ) is $[T_2]$, let us choose $U \in \tau(\ell)$ and $V \in \tau(x)$ such that $U \cap V = \varnothing$. Because $\ell = \lim f$, there exists $a \in D$ such that $f(b) \in U$ for all $b \ge a$, so $f(b) \in V$ is not possible for such *b*'s any more, contrarily to the supposition that $x \in \text{Acc}\, f$.

(iii) Let $(E, <<)$ be another directed set, h: $E \to D$ be a Kelley function (see condition [s] in definition I.1.15), and let the subnet $g = f \circ h$ be convergent to $x \in X$. If $V \in \tau(x)$ and $a \in D$ are fixed, then there exist e', $e'' \in E$ such that $g(e) \in V$ holds for all $e >> e'$, and $h(e) \ge a$ whenever $e >> e''$. Consequently, if e exceeds both e' and e'', then $f(b) \in V$, for $b = h(e) \ge a$, that is $x \in \text{Acc}\, f$.

Conversely, if $x \in \text{Acc}\, f$, then we may consider the set

$$E = \{(V, f(b)) \in \tau(x) \text{ x } X : f(b) \in V\} ,$$

which is directed by $<<$, in the sense that

$$(V, f(b)) << (V', f(b')) \Leftrightarrow V' \subseteq V \text{ and } b \le b'.$$

Then the function h: $E\to D$, expressed by $h((V, f(b)))= b$, satisfies the condition [s], hence $g=f\circ h$ is a subnet of f. We claim that g converges to x. In fact, if we fix $a\in D$, then for any $V\in\tau(x)$ we find some $(V, f(b)) \in E$ (where $b\geq a$ is not essential any more) such that $(V, f(b))<<(V', f(b'))$ implies $g((V', f(b'))) \in V$ since $g((V', f(b'))) = f(b') \in V' \subseteq V$.
(iv) Let E, h, g be defined as in the above proof of *(iii)* , and let $x\in \mathrm{Acc}\, g$ .To show that $x\in \mathrm{Acc}\, f$, we choose $V\in\tau(x)$ and $a\in D$, hence, using property [s] of h, we are led to some $e\in E$, such that $h(e')\geq a$ whenever $e'>>e$. Because $x\in \mathrm{Acc}\, g$, there exists some particular $e'>>e$ such that $g(e') \in V$, hence there exists $b=h(e') \geq a$ in D such that $f(b) \in V$.
(v) We have to show that $\overline{\mathrm{Acc}\, f} \subseteq \mathrm{Acc}\, f$, because the contrary is always true. In fact, if $x \in \overline{\mathrm{Acc}\, f}$, $V\in\tau(x)$, and $a\in D$, we may take W as in [N_4], hence $V\in\tau(y)$ whenever $y\in W$. Because $y\in \mathrm{Acc}\, f$, we deduce $f(b) \in W\subseteq V$ for some $b\geq a$. ◇

Some properties of the closed sets and operator "adherence" can be expressed in terms of convergent nets, as for example:

1.5. <u>Theorem.</u> If A is a subset of a topological space ($\mathcal{S}$, τ), then:

(i) $x\in \overline{A}$ iff there exists a net $f : D\to A$, convergent to x;
(ii) A is closed iff any convergent net $f : D\to A$ has the limit in A;
(iii) $\overline{A} = \mathcal{S}$ (when A is said to be *dense* in $\mathcal{S}$) iff every $x\in\mathcal{S}$ is the limit of a convergent net $f : D \to A$.

<u>Proof</u>. *(i)* If $x\in \overline{A}$, then for any $V \in \tau (x)$ there exists some $x_V\in V\cap A$, and consequently we may define the directed set $D =\{(V, x_V): V\in \tau (x)\}$ and the net f: $D\to A$ of values $f(V, x_V) = x_V$. Obviously, $f \to x$ (see also problem 1).
Conversely, if a net $f : D\to A$ (in particular a sequence, for $D = \mathbb{N}$) converges to x , then any neighborhood $V \in \tau (x)$ contains those elements of A which are the terms of f in V, i.e. $x\in \overline{A}$.
(ii) If A is closed, and $f : D\to A$ is convergent to x, then according to *(i)*, $x\in \overline{A} = A$; in particular $D = \mathbb{N}$ is possible. Conversely, using *(i)*, any $x\in \overline{A}$ is the limit of a net in A . By hypothesis $x\in A$, hence $\overline{A} = A$.
(iii) Characterization *(i)*, of the adherence, should be applied to arbitrary points of $\mathcal{S}$. ◇

1.6. <u>Theorem.</u> Let (x_n) be a sequence in a topological space ($\mathcal{S}$, τ). If none of its subsequences is convergent, then all the sets $G_k= \mathcal{S} \setminus \{x_k, x_{k+1}, \dots\}$, where $k\in\mathbb{N}$, are open.

By *reductio ad absurdum*, the proof reduces to the previous theorem, *(i)*.

§ II.1. Part 2. SEQUENCES IN METRIC SPACES

1.7. Remarks. *a)* The study of generalized sequences (nets) is redundant in metric spaces since the base $\mathscr{B}(x) = \{S(x, \frac{1}{n}): n\in\mathbb{N}^*\}$ of neighborhoods is a countable set at each point $x\in \mathcal{S}$. More exactly, each property that involves convergence of nets remains valid in metric spaces if we replace the nets by sequences.
b) The notion of fundamental sequence of numbers is essentially based on the possibility of comparing the neighborhoods of different points in $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$. Generally speaking, this comparison cannot be done in arbitrary topological spaces because of the lack of a *size* of neighborhoods. Such a *size* is still available in metric spaces, where the radii of spherical neighborhoods naturally represent it.
c) The property of boundedness is also meaningless in a general topological space, but makes sense in the metric ones.

To build a more concrete image about the convergence in metric spaces we will adapt the general definitions to this framework.

1.8. Proposition. Let $(\mathcal{S}, \rho)$ be a metric space. A sequence (x_n) in $\mathcal{S}$ is *convergent* iff there exists some $l \in \mathcal{S}$ such that

$$\forall\varepsilon > 0 \ \exists n_0(\varepsilon)\in\mathbb{N}, \text{ such that } n \geq n_0 \Rightarrow \rho(x_n, l) < \varepsilon$$

i.e. outside of any sphere centered al l there is a finite number of terms of the sequence.
Proof. If τ represents the topology generated by ρ, it follows that $V\in \tau(x)$ iff $V\supseteq S(l, \varepsilon)$ for some $\varepsilon > 0$, so it remains to reformulate the definitions. ◇

1.9. Definition. Let $(\mathcal{S}, \rho)$ be a metric space. A sequence (x_n) in $\mathcal{S}$ is said to be *fundamental* (or *Cauchy's*) iff for any $\varepsilon > 0$ there exists $n_0(\varepsilon)\in\mathbb{N}$ such that $p, q \geq n_0(\varepsilon)$ implies $\rho(x_p, x_q) < \varepsilon$. We say that $(\mathcal{S}, \rho)$ is *complete* iff "*fundamental* $\Rightarrow$ *convergent*" holds for any sequence in $\mathcal{S}$; in particular, the complete normed spaces are called *Banach* spaces, while the complete scalar product spaces are named *Hilbert* spaces.

A set $A \subseteq \mathcal{S}$ is said to be *bounded* iff it is contained in some sphere, i.e. there exist $a\in \mathcal{S}$ and $r > 0$ such that $A \subseteq S(a, r)$. We say that (x_n) is a *bounded sequence* in $\mathcal{S}$ iff the set $\{x_n\}$ of all values is bounded.

1.10. Proposition. A necessary and sufficient condition for a set $A \subseteq \mathcal{S}$ to be bounded is that for any $a' \in \mathcal{S}$ to exists $r' > 0$ such that $A \subseteq S(a', r')$.

Proof. The sufficiency is obvious. Conversely, if A is bounded and $a' \in \mathcal{S}$ is chosen, then for any $x \in A$ we have

$$\rho(x, a') \leq \rho(x, a) + \rho(a, a') \leq r + \rho(a, a'),$$

so that we may take $r' = r + \rho(a, a')$. ◇

1.11. Theorem. In every metric space $(\mathcal{S}, \rho)$ we have:

a) The limit of any convergent sequence is unique;
b) Any convergent sequence is fundamental;
c) Any fundamental sequence is bounded.

Proof. a) The property is essentially based on the fact that any metric space is separated, i.e. T_2.
b) If $\ell = \lim_{n\to\infty} x_n$, then we may compare $\rho(x_n, x_m) \leq \rho(x_n, \ell) + \rho(\ell, x_m)$.
c) If (x_n) is fundamental, then for $\varepsilon = 1$ there exists a rank $\nu \in \mathbb{N}$ such that $n > \nu$ implies $\rho(x_n, x_\nu) < 1$, i.e. the set $\{x_{\nu+1}, x_{\nu+2}, \ldots\}$ is bounded. On the other side the finite set $\{x_0, x_1, \ldots, x_\nu\}$ is bounded too. ◇

Beside the general properties involving the accumulation points and subsequences, in metric spaces we mention the following:

1.12. Theorem. If $(\mathcal{S}, \rho)$ is a metric space, and $f : \mathbb{N} \to \mathcal{S}$ is a sequence of terms $f(n) = x_n$, $n \in \mathbb{N}$, then:

a) $x \in \mathcal{S}$ is an accumulation point of the sequence f iff

$$\forall \varepsilon > 0 \; \forall \nu \in \mathbb{N} \; \exists n > \nu \text{ such that } \rho(x_n, x) < \varepsilon;$$

b) If (x_n) is fundamental and has a convergent subsequence, say $x_{n_k} \to \ell$, then $x_n \to \ell$ too;
c) $(f(\mathbb{N}))' \subseteq \mathrm{Acc}\, f$, i.e. every accumulation point of the set $\{x_n\}$, of values, is an accumulation point of the sequence.

Proof. a) We may replace $V = S(x, \varepsilon)$ and $D = \mathbb{N}$ in the definition of an accumulation point of f.
b) $\rho(x_n, \ell) \to 0$ because the distances in the greater sum

$$\rho(x_n, \ell) \leq \rho(x_n, x_{n_k}) + \rho(x_{n_k}, \ell)$$

are tending to 0 too.
c) If $x \in (f(\mathbb{N}))'$, then for any $k \in \mathbb{N}$ we find some x_{n_k} in $f(\mathbb{N})$ such that $x_{n_k} \in S(x, \frac{1}{k})$. Because the order $n_k \leq n_{k+1}$ can be easily assured, function $k \to x_{n_k}$ represents a subsequence of f. ◇

By the following theorems we show how useful is the property of completeness in scalar product, normed, and metric spaces. In particular, in the proof of the next theorem we see the role of completeness when we need to decompose Hilbert spaces into orthogonal subspaces:

1.13. <u>Theorem.</u> If $(\mathcal{L}, <.\,,.>)$ is a Hilbert space, then any closed linear subspace $\mathcal{S}$ has an orthogonal complement, i.e. $\mathcal{S} \oplus \mathcal{S}^{\perp} = \mathcal{L}$.

<u>Proof.</u> We have to show that for any $x \in \mathcal{L}$ there exist $u \in \mathcal{S}$ and $v \in \mathcal{S}^{\perp}$, such that $x = u + v$. In fact, if $x \in \mathcal{S}$, we take $x = u$, and $v = 0$. Otherwise, let $\delta = \inf\{\| x - y \| : y \in \mathcal{S}\} \overset{not.}{=} d(x, \mathcal{S})$ be the distance of x to $\mathcal{S}$, and let (y_n) be a sequence in $\mathcal{S}$, which allows the representation $\delta = \lim\limits_{n\to\infty} \|x - y_n\|$. Applying the Beppo-Levi's inequality to the terms of this sequence, namely

$$\|y_n - y_m\| \le \sqrt{\|x - y_n\|^2 - \delta^2} + \sqrt{\|x - y_m\|^2 - \delta^2}\ ,$$

it follows that (y_n) is a fundamental sequence (see [RC], etc., and some geometric interpretations). Since $\mathcal{L}$ is complete, and $\mathcal{S}$ is closed, there exists $u = \lim\limits_{n\to\infty} y_n$, $u \in \mathcal{S}$. Consequently, $\delta = \|x - u\|$, i.e. the distance δ is reached at u. It remains to show that $v = x - u \in \mathcal{S}^{\perp}$, i.e. $<v, y> = 0$ for any $y \in \mathcal{S} \setminus \{0\}$. In fact, according to the construction of u, we have

$$\| x - (u + \lambda y) \|^2 = <v - \lambda y, v - \lambda y> \ \ge \| x - u \|^2 = <v, v>$$

for arbitrary $\lambda \in \Gamma$. In particular, for $\lambda = \dfrac{<v,y>}{<y,y>}$, we obtain $-\dfrac{|<v,y>|^2}{<y,y>} \ge 0$, which obviously implies $<v, y> = 0$.

Because $\mathcal{S} \cap \mathcal{S}^{\perp} = \{0\}$, this decomposition of x is unique. ◇

The next theorem is considered a *geometric form of the fix-point principle* (see below), and represents an extension of the Cantor's theorem I.2.17.

1.14. <u>Theorem.</u> Let $(\mathcal{L}, \| . \|)$ be a Banach space, and let, for each $n \in \mathbb{N}$,

$$S_{cl}(x_n, r_n) = \{x \in \mathcal{L} : \|x - x_n\| \le r_n\}$$

denote a closed sphere in $\mathcal{L}$. If

i) the resulting sequence of spheres is decreasing, i.e.

$$S_{cl}(x_1, r_1) \supseteq S_{cl}(x_2, r_2) \supseteq \ldots \supseteq S_{cl}(x_n, r_n) \supseteq \ldots$$

ii) $r_n \to 0$,

then there exists a unique $x^* \in \mathcal{L}$ such that $\cap\{S_{cl}(x_n, r_n) : n \in \mathbb{N}^*\} = \{x^*\}$.

Proof. It is easy to see that the sequence (x_n) is fundamental in $\mathcal{L}$, which is complete, hence there exists $x^* = \lim x_n$, and $x^* \in \overline{S_{cl}}\,(x_n, r_n)$ for all $n \in \mathbb{N}^*$. Because each closed sphere is a closed set in normed spaces (see problem I.4.8), i.e. $\overline{S_{cl}}\,(x_n, r_n) = S_{cl}\,(x_n, r_n)$, we obtain $x^* \in S_{cl}\,(x_n, r_n)$ for all $n \in \mathbb{N}^*$. Consequently, $x^* \in \bigcap \{S_{cl}\,(x_n, r_n) : n \in \mathbb{N}^*\}$.

If x^{**} would be another point in the above intersection, then

$$\| x^{**} - x^* \| \leq \| x^{**} - x_n \| + \| x^* - x_n \| \to 0,$$

hence $x^* = x^{**}$. ◇

Similar results hold in complete metric spaces, i.e. linearity is not essential. The practical efficiency of these properties can be improved by thinking the transition from one sphere to the other as the action of a function. The specific terms are introduced by the following:

1.15. Definition. We say that function $f : \mathcal{S} \to \mathcal{S}$, where $(\mathcal{S}, \rho)$ is a metric space, is a *contraction* iff there exists a real number $c \in [0,1)$, called *contraction factor*, such that the inequality $\rho(f(x), f(y)) \leq c\ \rho(x, y)$ holds at arbitrary $x, y \in \mathcal{S}$. An element $x^* \in \mathcal{S}$ is called *fix point of f* iff $f(x^*) = x^*$.

1.16. Theorem. If $(\mathcal{S}, \rho)$ is a complete metric space, and $f : \mathcal{S} \to \mathcal{S}$ is a contraction, then f has an unique fix point.

Proof. Let us choose some $x_0 \in \mathcal{S}$. Generally speaking, $x_0 \neq f(x_0)$, but we may consider it like *zero order approximation of* x^**. The higher order approximations of* x^* are recurrently defined by $x_{n+1} = f(x_n)$ for all $n \in \mathbb{N}$. We claim that the sequence (x_n), of so-called *successive approximations*, is fundamental. In fact, if we note $\rho(x_0, x_1) = \delta$, then, by induction, for any $n \in \mathbb{N}$ we obtain $\rho(x_n, x_{n+1}) \leq c^n \delta$. Thus, for any $n \in \mathbb{N}$ and $p \geq 1$, we have:

$$\rho(x_n, x_{n+p}) \leq \rho(x_n, x_{n+1}) + \ldots + \rho(x_{n+p-1}, x_{n+p}) \leq c^n \delta + \ldots + c^{n+p-1} \delta =$$

$$= c^n \delta \, \frac{1-c^p}{1-c} \leq \frac{\delta}{1-c} c^n \ .$$

The case $\delta=0$ (or $c=0$) is trivial since it corresponds to constant sequences of approximations. Otherwise, since $\lim c^n = 0$, for arbitrary $\varepsilon > 0$ there exists $n_0(\varepsilon) \in \mathbb{N}$ such that $\rho(x_n, x_{n+p}) \leq \varepsilon$ whenever $n \geq n_0(\varepsilon)$ and $p \in \mathbb{N}^*$, hence (x_n) is fundamental. Because $(\mathcal{S}, \rho)$ is complete, there exists $x^* = \lim_{n\to\infty} x_n$, which turns out to be the searched fix point. In fact, since

$$\rho(x^*, f(x^*)) \leq \rho(x^*, x_n) + \rho(x_n, f(x^*)) = \rho(x^*, x_n) + \rho(f(x_{n-1}), f(x^*)) \leq$$
$$\rho(x^*, x_n) + c\, \rho(x_{n-1}, x^*) \longrightarrow 0,$$

we deduce that $\rho(x^*, f(x^*)) = 0$, hence $f(x^*) = x^*$.

Even if the above construction of x^* starts with an arbitrary x_0 , the fix point is unique in the most general sense (i.e. the same for any approximation of order zero, and for any other method, possibly different from that of successive approximations). If x^{**} would be a second fix point of f , different from x^*, then

$$0 < \rho(x^*, x^{**}) = \rho(f(x^*), f(x^{**})) \le c\, \rho(x^*, x^{**})$$

would contradict the hypothesis $c < 1$. ◇

1.17. Remark. The interest in finding fix point theorems is justified by the interpretation of the fix points as solutions of various equations. Applied to particular metric spaces, especially to function spaces, the method of successive approximations is useful in solving algebraic equations, as well as more complicated problems like systems of differential, integral or even operatorial equations (see [RI], [RC], [YK], etc.). The theoretical results are essential finding approximate solutions by digital evaluation within the desired error. In particular, this theory represents the mathematical kernel of the computer programs for solving equations. This explains why so many types of fix point theorems have been investigated, and the interest is still increasing, especially in more general than metric spaces, with the aim of developing particular techniques of approximation.

To illustrate how the method of successive approximations works to solve an equation, let us consider the following simple case:

1.18. Example. Evaluate the real root of the equation

$$x^3 + 4x - 1 = 0$$

with an error less than 10^{-4} .

The real root of this equation belongs to $\mathcal{S}$ = [0,1], and it can be considered as a fix point of $f : \mathcal{S} \to \mathcal{S}$, where

$$f(x) = (x^2 + 4)^{-1} .$$

In addition, according to Lagrange's theorem, for any $x \le y$ in $\mathcal{S}$,

$$\rho(f(x), f(y)) = | f(x) - f(y) | = | f'(\xi) | \, | x - y | = | f'(\xi) | \, \rho(x, y),$$

where $\xi \in (x, y) \subset \mathcal{S}$. Because

$$| f'(x) | = | -2x |(x^2 + 4)^{-2} \le \frac{1}{8}$$

at any $x \in \mathcal{S}$, it follows that f is a contraction of factor c = 1/8. Starting, in particular, with $x_0 = 0$, we obtain $\delta = \rho(x_0, x_1)$ = ¼, hence for $n \ge 4$ we have the error less than

$$\frac{\delta}{1-c} c^n = \frac{2}{7} 8^{-n} < 10^{-4}$$

Consequently, the searched approximation is x_4.

Because the complete metric spaces have a lot of convenient properties, we have to analyze some of the most useful examples.

The complete metric space $\mathbb{R}$. So far, we have seen that $\mathbb{R}$ is complete in order. This fact is a consequence of the Dedekind's construction (based on cuts, compare to Theorem I.2.12), and represents one of the conditions in the axiomatic definition I.2.14. In the following, we will show that $\mathbb{R}$ is a complete metric space, relative to its Euclidean metric. This result will be a consequence of some properties of $\mathbb{R}$, already discussed in § I.2.

The following theorem is based on Cantor's theorem:

1.19. Theorem. (Cesàro-Weierstrass) Every bounded sequence of real numbers contains a convergent subsequence.

Proof. Since (x_n) is bounded, let $a, b \in \mathbb{R}$ such that $a \leq x_n \leq b$ for all $n \in \mathbb{N}$. If $c = (a+b)/2$, then either $[a, c]$ or $[c, b]$ will contain infinitely many terms of the sequence. Let $[a_1, b_1]$ stand for that interval that contains infinitely many terms. Dividing it in halves, we similarly obtain $[a_2, b_2] \subseteq [a_1, b_1]$, and so on. The resulting sequence of intervals obviously satisfies the conditions in the Cantor's theorem (see I.2.17), with $\alpha = \beta$. The needed subsequence is obtained by choosing a term of the initial sequence in $[a_k, b_k]$ in the increasing order of indices. ◇

1.20. Corollary. For any infinite and bounded set $A \subseteq \mathbb{R}$ there is at least one accumulation point (α is an *accumulation* point of A iff each of its neighborhoods contains points of A, different from α).

Proof. We repeat the reason from the above theorem by considering A instead of $\{x_k\}$. ◇

1.21. Remark. There are two aspects, which concur for a sequence to be convergent, namely the relative position of the terms, and some "*richness*" of the space to provide enough limit points. An example in which these aspects can be easily distinguished is the sequence of rational approximations of $\sqrt{2}$, which looks like a convergent sequence, but in $\mathbb{Q}$ it is not so because $\sqrt{2} \notin \mathbb{Q}$. The notion of "fundamental sequence" is exactly conceived to "describe the convergence without using the limit points".

As a particular case of the definition 1.9. from above, we say that the sequence (x_n) in $\mathbb{R}$ is *fundamental* (or *Cauchy*), relative to the Euclidean metric, if for every $\varepsilon > 0$ there exists a rank $n_0(\varepsilon) \in \mathbb{N}$, such that

$$n, m > n_0(\varepsilon) \Rightarrow |x_n - x_m| < \varepsilon .$$

According to the same definition, showing that $\mathbb{R}$ is complete means to prove that each fundamental sequence in $\mathbb{R}$ is convergent, i.e. $\mathbb{R}$ contains "enough" elements, which can play the role of limit points.

The following properties of the sequences in $\mathbb{R}$ represent immediate and simple consequences of the theorems 1.11. and 1.12. from above, in the particular case $\mathcal{S} = \mathbb{R}$, and $\rho(x, y) = |x - y|$.

1.22. **Proposition.** In the metric space $\mathbb{R}$, the following implications hold:

a) Every convergent sequence is fundamental;

b) Every fundamental sequence is bounded;

c) If a fundamental sequence (x_n) contains a convergent subsequence (x_{n_k}) and $x_{n_k} \to \ell$, then also $x_n \to \ell$.

The reader is advised to produce a particular proof in $\mathbb{R}$.

Now we can formulate and prove the main result:

1.23. **Theorem.** Every fundamental sequence in $\mathbb{R}$ is convergent (or, in an equivalent formulation, $\mathbb{R}$ is *complete* relative to its Euclidean metric).

Proof. According to the above property b), (x_n) is bounded, so using the above Cesaró-Weierstrass theorem 1.19., we deduce the existence of a convergent subsequence. Being fundamental, the sequence itself has to be convergent to the same limit. ◇

The complete metric space $\mathbb{R}^p$. Considering $\mathbb{R}^p$ endowed with its usual Euclidean metric, many general properties from metric spaces will remain valid for sequences in $\mathbb{R}^p$. In particular, $\mathbb{R}^p$ is another remarkable example of complete metric space. To prove it, we have first to specify some terms and connections with sequences in $\mathbb{R}$.

Let a sequence of points in $\mathbb{R}^p$, $f : \mathbb{N} \to \mathbb{R}^p$, be defined by $f(n) = x_n$, $n \in \mathbb{N}$, where $x_n = (x_n^1, x_n^2, \ldots, x_n^p)$. For each $k = 1, 2, \ldots, p$, function f has a *component* function $f_k : \mathbb{N} \to \mathbb{R}$, defined by $f_k(n) = x_n^k$ (the p sequences of real numbers $f_1, f_2, \ldots, f_p$ are called *component sequences* of the sequence f).

The terms *convergent*, *fundamental* and *bounded* refer to the Euclidean structure of $\mathbb{R}^p$. However, the properties involving the order of $\mathbb{R}$ cannot be carried to $\mathbb{R}^p$, hence the Euclidean metric has greater importance in $\mathbb{R}^p$.

The following theorem establishes the fact that the study of sequences in $\mathbb{R}^p$, p > 1 can be reduced to a similar study of sequences in $\mathbb{R}$.

1.24. **Theorem.** If f is a sequence in $\mathbb{R}^p$, of components $f_1, f_2, \ldots, f_p$, then:

a) Sequence f is convergent and has the limit $x = (x^1, x^2, \ldots, x^p)$ iff the sequences $f_1, f_2, \ldots, f_p$ are convergent and $x^k = \lim_{n\to\infty} x_n^k$, $k = 1, 2, \ldots, p$.

b) Sequence f is fundamental iff all its components are fundamental

c) f is bounded if and only if all the component sequences are bounded.

Proof. a) Let us remark that the following double inequality takes place:

$$\left|x_n^k - x^k\right| \le \left[\sum_{k=1}^{p}(x_n^k - x^k)^2\right]^{\frac{1}{2}} \le \sum_{k=1}^{p}\left|x_n^k - x^k\right| .$$

If the sequence f is convergent to x, then for every $\varepsilon > 0$, there exists a rank $n_0(\varepsilon)\in\mathbb{N}$ such that $d(x_n, x) < \varepsilon$ holds for every $n\ge n_0(\varepsilon)$, where d is the Euclidean distance. Then, from the first inequality, it results that if $n\ge n_0(\varepsilon)$, then $|x_n^k - x^k| < \varepsilon$. Consequently, $x^k = \lim\limits_{n\to\infty} x_n^k$, for every $k = 1, 2,\dots, p$, i.e. all the component sequences are convergent.

Conversely, let us suppose that each sequence f_k converges to x^k, where $k = 1, 2,\dots, p$, and let $\varepsilon > 0$ be given. For every $\varepsilon/p > 0$, there exists $n_k(\varepsilon)\in\mathbb{N}$ such as for every $n\in\mathbb{N}$, $n \ge n_k(\varepsilon)$ we have $|x_n^k - x^k| < \varepsilon/p$. If we note $x = (x^1, x^2, \dots, x^p)$ and $n_0(\varepsilon) = \max\{n_1(\varepsilon), n_2(\varepsilon),\dots, n_p(\varepsilon)\}$, then

$$d(x_n, x) = \left[\sum_{k=1}^{p}(x_n^k - x^k)^2\right]^{\frac{1}{2}} \le \sum_{k=1}^{p}\left|x_n^k - x^k\right| < p\,\frac{\varepsilon}{p} = \varepsilon$$

holds for all $n\ge n_0(\varepsilon)$. Consequently, f is convergent and it has the limit x.

Points b) and c) of the theorem can be proved in the same way. ◇

1.25. Applications. According to point a) of this theorem, the limit of a convergent sequence from $\mathbb{R}^p$ can be calculated "on components". For example, the sequence $\left\{\left(\frac{1}{n^2}, \frac{n+1}{n+2}, (1+\frac{1}{n})^n\right)\right\}_{n\in\mathbb{N}^*}$ converges to $(0, 1, e)$ in $\mathbb{R}^3$.

Because the algebraic operations of addition and scalar multiplication are realized on components too, we may extend these operations to sequences. In addition, if f, g are two convergent sequences in $\mathbb{R}^p$, $f(n)=x_n$, $g(n)=y_n$, $n\in\mathbb{N}$, then $f+g$ is convergent and $\lim\limits_{n\to\infty}(x_n + y_n) = \lim\limits_{n\to\infty} x_n + \lim\limits_{n\to\infty} y_n$, and if $\alpha\in\mathbb{R}$, then the sequence αf is convergent and $\lim\limits_{n\to\infty}(\alpha x_n) = \alpha \lim\limits_{n\to\infty} x_n$.

Using Theorem 1.24, we can prove the completeness theorem for $\mathbb{R}^p$:

1.26. Theorem. Sequence f in $\mathbb{R}^p$ is convergent if and only if it is Cauchy, i.e. $\mathbb{R}^p$ is complete relative to its Euclidean metric.

Proof. The sequence f is convergent if and only if all his components are convergent (from Theorem 1.24 a)). Because its components are sequences of real numbers, and $\mathbb{R}$ is complete, the component sequences are convergent if and only if they are Cauchy sequences. Applying point b) this is equivalent to saying that f is a Cauchy sequence. So, f is convergent if and only if it is Cauchy, i.e. $\mathbb{R}^p$ is a complete metric space. ◇

The complete metric space $\mathbb{C}$. We consider now the complex plane $\mathbb{C}$, endowed with the Euclidean metric $d(z_1, z_2) = | z_1 - z_2 |$, $z_1, z_2 \in \mathbb{C}$. To give a sequence $\{z_n\}_{n\in\mathbb{N}}$ in $\mathbb{C}$ means to precise two sequences of real numbers, namely $\{x_n\}_{n\in\mathbb{N}}$, $\{y_n\}_{n\in\mathbb{N}}$, where $z_n = x_n + i\, y_n$. Because the Euclidean metrics on $\mathbb{C}$ and $\mathbb{R}^2$ coincide, i.e.

$$d_{\mathbb{C}}(z_1, z_2) = | z_1 - z_2 | = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} = d_{\mathbb{R}^2}((x_1, y_1), (x_2, y_2)),$$

it follows that we may treat $\mathbb{C}$ as a particular case in the theorems 1.24. and 1.26. from above. Consequently, the sequence $\{z_n\}_{n\in\mathbb{N}}$ is convergent in $\mathbb{C}$ (respectively Cauchy, or bounded) iff the real sequences $\{x_n\}_{n\in\mathbb{N}}$, $\{y_n\}_{n\in\mathbb{N}}$ have the same property. In particular, $\mathbb{C}$ is a complete metric space.

The complete metric space m_A . Let $A\subset\mathbb{R}$ be an arbitrary set and let m_A be the metric space of all bounded real functions defined on A, endowed with the *uniform* distance $d(f, g) = \sup \{| f(x) - g(x)|;\, x \in A\}$. We claim that:

1.27. Theorem. m_A is a complete metric space.

Proof. Let $\{f_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in m_A. Because

$$| f_n(x) - f_m(x) | \le d(f_n, f_m)$$

holds at each $x \in A$, it follows that $\{f_n(x)\}_{n\in\mathbb{N}}$ are Cauchy sequences of real numbers at each $x \in A$. Since $\mathbb{R}$ is complete, these sequences converge to a well-determined real number, depending on x, which we note $f(x)$. In this way we define a function $f : A\to\mathbb{R}$, called *punctual limit* of $\{f_n\}_{n\in\mathbb{N}}$. We have to prove that $\{f_n\}_{n\in\mathbb{N}}$ converges to f in the sense of d , and $f\in m_A$.

Let us take $\varepsilon > 0$. Because $\{f_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, for $\varepsilon/4 > 0$ there exists $n_0(\varepsilon)\in\mathbb{N}$, such that for every $n, m \ge n_0(\varepsilon)$ we have $d(f_n, f_m) < \varepsilon/4$, i.e. $| f_n(x) - f_m(x) | < \varepsilon/4$ holds at every $x \in A$. On the other hand, because at every $x\in A$ we have $f(x) = \lim\limits_{m\to\infty} f_m(x)$, there exists $m_0(\varepsilon, x)\in\mathbb{N}$ such that

$$| f_m(x) - f(x)| < \varepsilon/4$$

holds for all $m\ge m_0(\varepsilon, x)$.

Now, let $n\ge n_0(\varepsilon)$ and $x \in A$ be arbitrary, but fixed. If $m\in\mathbb{N}$ satisfies both $m\ge n_0(\varepsilon)$ and $m \ge m_0(\varepsilon, x)$, then:

$$| f_n(x) - f(x) | \le | f_n(x) - f_m(x)| + | f_m(x) - f(x)| < \varepsilon/4+ \varepsilon/4 = \varepsilon/2,$$

and

$$|f(x)| \le |f(x) - f_n(x)| + | f_n(x)| < \varepsilon/2 + | f_n(x)|.$$

Since $f_n \in m_A$, it follows that $f\in m_A$. In addition, $\sup\limits_{x\in A} |f_n(x) - f(x)|\le \varepsilon/2$ implies $d(f_n, f) \le \varepsilon/2 < \varepsilon$, which means that $\{f_n\}_{n\in\mathbb{N}}$ is convergent in the metric space m_A. To conclude, m_A is a complete metric space. ◇

PROBLEMS § II.1

1. In the topological space ($\mathcal{S}$, τ) we choose $z_0 \in \mathcal{S}$, and we define the set

$$D_{z_0} = \{(V,\, z) : z \in V \in \tau(z_0)\},$$

where D is directed by $(V, z) \le (U, \zeta)$ meaning $U \subseteq V$. Show that the g.s. $f : D \longrightarrow \mathcal{S}$, of terms $f((V, z)) = z$, is convergent to z_0 . Illustrate this fact by drawing two copies of $\mathcal{S} = \mathbb{C}$.

Hint. According to the definition of convergence, for every $V \in \tau(z_0)$ there is $(V, z) \in D_{z_0}$ such that for $(V, z) \le (U, \zeta)$ we have $f((U, \zeta)) \in V$ (see also the examples I.1.8(*iii*)3, and I.1.16*b*).

2. Let ($\mathcal{X}$, ξ) and ($\mathcal{Y}$, η) be topological spaces, and let $\mathcal{S} = \mathcal{X} \times \mathcal{Y}$ be endowed with the product topology τ. The *canonical projections* are noted p: $\mathcal{S} \longrightarrow \mathcal{X}$ and q: $\mathcal{S} \longrightarrow \mathcal{Y}$, where $p(x, y) = x$, and $q(x, y) = y$. Show that a g.s. $f : D \longrightarrow \mathcal{S}$ is convergent to (x_0, y_0) iff $p \circ f \to x_0$ and $q \circ f \to y_0$.

Hint. Use the form of the neighborhoods in the product topology τ (see the derived topology in I.4.10.*b*), and apply the definition of the limit.

3. Let $f : D \longrightarrow \mathbb{R}$ be a g.s. of real numbers, where $(D, \prec)$ is an arbitrary set, directed by $\prec$. Show that if

1. f is increasing (relative to $\prec$ on D and $<$ on $\mathbb{R}$), and
2. $f(D)$ has an upper bound (in $\mathbb{R}$),

then f is a convergent sequence.
Hint. According to the Cantor's axiom (see definition I.2.14), there exists the *exact* upper bound $\sup f(D) = x_0 \in \mathbb{R}$. It remains to show that $f \to x_0$.

4. Let $(D, \le)$ be a directed set, and let E be a nonvoid part of D. We say that E is *co-final* in D iff

$$\forall a \in D \quad \exists b \in E \quad \textit{such that} \quad a \le b.$$

Show that in this case the restriction $f\big|_E$ is a subnet of the g.s. $f : D \longrightarrow \mathcal{S}$. Can E consist of prime numbers if we suppose $D = \mathbb{N}$?
Hint. The *embedding* $h : E \to D$, defined by $h(e) = e$, satisfies the Kelley's condition [s] in the definition I.1.15. Consequently, $f\big|_E = f \circ h$ is a subnet of f. The set of primes is infinite, hence it is co-final in $\mathbb{N}$.

5. Study the convergence of a sequence (x_n) in the metric space $(\mathcal{S}, d)$ if:

a) The subsequences (x_{2n}), (x_{2n+1}), and (x_{3n}) are convergent;
b) The subsequences (x_{kn}) are convergent for all $k \geq 2$;
c) The subsequences (x_{n^k}) are convergent for all $k \geq 2$.

Hint. a) The sequences (x_{2n}) and (x_{3n}) have a common subsequence, e.g. (x_{6n}), hence their limits coincide. Similarly, because $(x_{3(2n+1)})$ is a common subsequence of (x_{2n+1}) and (x_{3n}), it follows that (x_{2n}) and (x_{2n+1}) have the same limit. Finally, each term x_n is either in (x_{2n}) or in (x_{2n+1}).

b) and c). Take $\mathcal{S} = \mathbb{R}$, endowed with its Euclidean metric, and remark that the subsequence (x_{p_n}), where p_n is the n^{th} prime number, may have no term in the considered subsequences, hence (x_n) may be divergent since at least two distinct accumulation points are possible.

6. On the nonvoid set $\mathcal{S}$, we consider the *discrete* metric (defined in § I.4.). Show that (x_n) is a fundamental sequence in $(\mathcal{S}, d)$ iff there exists a rank $n_0 \in \mathbb{N}$ such that $x_m = x_n$ whenever $m, n \geq n_0$. Deduce that the metric space $(\mathcal{S}, d)$ is complete.

Hint. Take $\varepsilon = \frac{1}{2} < 1$ in the definition of a fundamental sequence, and use the convergence of the constant sequences.

7. Let ρ be the Euclidean metric on $\mathcal{S} = \mathbb{R}$, and let $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ be 1:1. We note $A = \varphi(\mathbb{R})$, and we define $d_{\varphi,\rho}$ as in the problem I.4.7, i.e.

$$d_{\varphi,\rho}(x, y) = \rho(\varphi(x), \varphi(y)).$$

a) Take $\varphi(x) = x\,[1 + |x|\,]^{-1}$ and show that the sequence (n) is fundamental but not convergent relative to the corresponding metric $d_{\varphi,\rho}$.

b) Show that the real sequence (x_n), of terms

$$x_n = \begin{cases} 1 + \frac{1}{n} & if\ \ n\ is\ even \\ 1 - \frac{\pi}{n} & if\ \ n\ is\ odd \end{cases}$$

is divergent relative to the metric $d_{\varphi,\rho}$ generated by

$$\varphi(x) = \begin{cases} ax + b & if\ \ x \in \mathbb{Q} \\ -ax + b & if\ \ x \in \mathbb{R} \setminus \mathbb{Q} \end{cases},\ a > 0.$$

c) Show that, if $\mathcal{S} = \mathbb{Q}$ and $\varphi : \mathbb{Q} \longrightarrow \mathbb{N}$ is a bijection, then a sequence (x_n) is fundamental relative to the corresponding metric $d_{\varphi,\rho}$ iff it is constant except a finite number of terms.

Hint. a) φ (n) and φ (m) are arbitrarily close to 1, hence also to each other, if the values of n and m are large enough.
b) Because $1 + \frac{1}{n} \in \mathbb{Q}$ and $1 - \frac{\pi}{m} \in \mathbb{R}\setminus\mathbb{Q}$, we have $d_{\varphi,\rho}(x_n, x_m) > 2a$ whenever n is even and m is odd.
c) If $x \neq y$, then $d_{\varphi,\rho}(x, y) \geq 1$, since φ ($x$), φ ($y$) $\in \mathbb{N}$.

8. Show that in $\mathbb{C}$ we have:
a) $z_n \to 0 \Leftrightarrow |z_n| \to 0$; b) $z_n \to z \neq 0 \Leftrightarrow [|z_n| \to |z|$ *and* $\arg z_n \to \arg z]$.
c) $\lim\limits_{n\to\infty} \left(1 + i\frac{\alpha}{n}\right)^n = \cos\alpha + i\sin\alpha$.
Hint. a) Compare the definitions of a limit in $\mathbb{R}$ and $\mathbb{C}$. b) Use the formulas of $|z|$ and $\arg z$. c) Evaluate

$$\left|1 + i\frac{\alpha}{n}\right|^n = \left(1 + \frac{\alpha^2}{n^2}\right)^{\frac{n}{2}} \to 1 \text{ and } \arg\left(1 + i\frac{\alpha}{n}\right)^n = n\left(arctg\frac{\alpha}{n}\right) \to \alpha .$$

9. (Cesàro-Stolz Lemma in $\mathbb{C}$) Let (z_n) be a sequence of complex numbers, and let (r_n) be an increasing and unbounded sequence of real numbers. Show that the following implication holds

$$\exists\, l = \lim_{n\to\infty} \frac{z_{n+1} - z_n}{r_{n+1} - r_n} \Rightarrow \exists \lim_{n\to\infty} \frac{z_n}{r_n} = l$$

and use it to find the limits of the sequences:
a) $\frac{z^n}{n}$ b) $\frac{z_1 + z_2 + ... + z_n}{n}$ where $z_n \to z$ c) $\frac{n}{\frac{1}{|z_1|} + ... + \frac{1}{|z_n|}}$ where $0 \neq z_n \to z$.
Extend this problem to the linear space $\mathbb{R}^p$, where $p > 2$.
Hint. Decompose the complex numbers in real and imaginary parts and reduce the problem to $\mathbb{R}$, where we may prove the stated implication by operating with inequalities in the "ε - n_0(ε)" definition of the limit l .

10. Let (a_n) and (b_n) be sequences in the normed linear space ($\mathcal{L}$, $\|\cdot\|$), such that $a_n \to a$ and $b_n \to b$ in the topology generated by $\|\cdot\|$. Prove that:
1. $a_n + b_n \to a + b$ 2. $\lambda a_n \to \lambda a$ 3. $\|a_n\| \to \|a\|$ 4. $\langle a_n, b_m \rangle \to \langle a, b \rangle$ whenever $\|\cdot\|$ is generated by the scalar product $\langle \,.\,,\,.\, \rangle$.
Hint. Use the relations: 1. $\|(a_n + b_n) - (a + b)\| \leq \|a_n - a\| + \|b_n - b\|$,
2. $\|\lambda a_n - \lambda a\| = |\lambda|\,\|a_n - a\|$ 3. $\big|\|a_n\| - \|a\|\big| \leq \|a_n - a\|$ and 4. $|\langle a_n, b_m \rangle - \langle a, b \rangle| \leq$
$\leq |\langle a_n, b_m \rangle - \langle a_n, b \rangle| + |\langle a_n, b \rangle - \langle a, b \rangle| \leq \|a_n\|\,\|b_m - b\| + \|a_n - a\|\,\|b\|$.

§ II.2 SERIES OF REAL AND COMPLEX NUMBERS

The notion of series has appeared in practical problems, which need the addition of infinitely many numbers. A nice example is that of *Achilles and the tortoise*: Let us say that Achilles runs ten times faster than the tortoise, and between them there is an initial distance d. Trying to catch the tortoise, Achilles runs the distance d in t seconds, etc. Since this process contains infinitely many stages, it seems that Achilles will never catch the tortoise. In reality, by adding all the necessary times, we still obtain a finite time, i.e.

$$t + t/10 + t/100 + \ldots = \frac{t}{1-\frac{1}{10}} = \frac{10t}{9} \,,$$

according to the formula of the sum of a geometrical progression.

Another significant case is that of the periodical decimal numbers. For example, summing-up a geometrical progression again, we obtain

$$0.23\,23\ldots = 0.(23) = \frac{23}{100}(1 + 10^{-2} + 10^{-4} + \ldots) = \frac{23}{99}.$$

In essence, the notion of *series* is based on that of *sequence*:

2.1. <u>Definition.</u> We call *series* in Γ (which means $\mathbb{R}$ or $\mathbb{C}$) any pair (f, g) of sequences, where $f : \mathbb{N}\to\Gamma$ defines *the general terms of the series,* also noted $x_n = f(n)$, and $g : \mathbb{N}\to\Gamma$ represents *the sequence of partial sums,* i.e.

$$s_n = x_0 + x_1 + \ldots + x_n = g(n).$$

Instead of (f, g), the series is frequently marked as an "*infinite sum*"

$$x_0 + x_1 + \ldots + x_n + \ldots = \sum x_n \,.$$

More exactly, we say that the series (f, g) is *convergent to* s, respectively s is the *sum* of the series, iff the sequence (s_n), of partial sums, converges to s, and we note

$$\sum_{n=0}^{\infty} x_n = \lim_{n\to\infty} s_n = s \,.$$

2.2. <u>Remarks.</u> a) In practice, we may encounter two types of problems, which correspond to the similar problems about sequences, namely:

1^0. Establishing the nature of the given series, i.e. seeing whether the series is convergent or not, and

2^0. Searching for the value of the sum s.

Establishing the nature of a series generally involves a qualitative study. It is still essential in the practical use of the series, because the exact value of s is too rarely accessible, and we must deal with some approximations. The computer techniques are obviously efficient, but even so, we need some previous information about the behavior of the series. In other words, from a practical point of view, the two aspects are strongly connected.

b) The notion of series involves an algebraic aspect, that is the addition of n terms in s_n, and an analytical one, namely the limiting process of finding s. Consequently, a good knowledge of operating with convergent sequences from the algebraic point of view is indispensable. We recall that the case of the real sequences is well studied in high school. It is easy to see that the complex sequences have similar properties, i.e. if $(z_n)_{n\in\mathbb{N}}$ and $(\zeta_n)_{n\in\mathbb{N}}$ are convergent sequences in $\mathbb{C}$, then

- $\lim\limits_{n\to\infty} (z_n + \zeta_n) = \lim\limits_{n\to\infty} z_n + \lim\limits_{n\to\infty} \zeta_n$;
- $\lim\limits_{n\to\infty} (z_n \cdot \zeta_n) = (\lim\limits_{n\to\infty} z_n)\cdot(\lim\limits_{n\to\infty} \zeta_n)$,
- $\lim\limits_{n\to\infty} \dfrac{1}{z_n} = \dfrac{1}{\lim\limits_{n\to\infty} z_n}$, where $z_n \neq 0$ and $\lim\limits_{n\to\infty} z_n \neq 0$.

Because we realize the addition of the series "term by term", it follows that the sum of two convergent series is also convergent. The multiplication and the quotient of series is more complicated (see definition 2.27. below).
c) Adding infinitely many numbers may lead to unbounded sequences of partial sums, so we have sometimes to deal with convergence to infinity. We consider that the situation in $\mathbb{R}$ is already known, including the algebraic operations with $\pm\infty$, and the indeterminate cases. The problem of infinity in $\mathbb{C}$ is usually treated as *one point compactification*, and we sketch it later in §III.1.).

Now, we start with some *criteria* (*tests*) *of convergence*, which we need in order to answer the question about the nature of a series.

2.3. <u>Theorem.</u> (The general Cauchy's criterion) The series $\sum x_n$ in Γ is convergent iff for any $\varepsilon > 0$ we can find $n_0(\varepsilon) \in \mathbb{N}$ such that

$$| x_{n+1} + x_{n+2} + \ldots + x_{n+p} | < \varepsilon$$

holds for all $n > n_0(\varepsilon)$ and arbitrary $p \in \mathbb{N}$.

<u>Proof.</u> The assertion of the theorem reformulates in terms of ε and $n_0(\varepsilon)$ the fact that a series $\sum x_n$ is convergent iff the sequence (s_n) of partial sums is fundamental. This is valid in both $\mathbb{R}$ and $\mathbb{C}$. ◇

2.4. <u>Corollary.</u> If a series $\sum x_n$ is convergent, then $x_n \to 0$.

<u>Proof.</u> Take $p = 1$ in the above theorem. ◇

Because this corollary contains a necessary condition of convergence, namely $x_n \to 0$, it is frequently used to prove the divergence. We mention that this condition is not sufficient, i.e. $x_n \to 0$ is possible in divergent series, which is visible in plenty of examples.

2.5. <u>Examples.</u> *(i)* To get the complete answer about the convergence of the geometric series $\sum z^n$, $z\in\mathbb{C}$, we consider two cases:

- If $q = |z| < 1$, then $z^n \to 0$, and consequently $(s_n) \to s$, where

$$s_n = 1 + z + \ldots + z^{n-1} = \frac{1-z^n}{1-z} \to \frac{1}{1-z} = s.$$

- If $|z| \geq 1$, then the series is divergent because the general term is not tending to zero (as the above corollary states).

(ii) To show that condition $x_n \to 0$ is not sufficient for the convergence of a series, we may consider *the harmonic* series $\sum \frac{1}{n}$. Obviously, $x_n = \frac{1}{n} \to 0$, but the series is divergent, since grouping the terms as

$$1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \ldots$$

each bracket is greater than ½ .

From the general Cauchy's test we may derive other theoretical results:

2.6. <u>Theorem.</u> (Abel's criterion) Let $\sum z_n$ be a series of complex numbers, which has a bounded sequence of partial sums, and let (ε_n) be a decreasing sequence of real numbers, convergent to zero. Then the series $\sum (z_n \cdot \varepsilon_n)$ is convergent.

<u>Proof.</u> By hypothesis there exists $M > 0$ such that $|s_n| < M$ for all $n\in\mathbb{N}$, where s_n are the partial sums of the series $\sum z_n$, i.e. $s_n = \sum_{k=0}^{n} z_k$. We claim that the series $\sum (z_n \cdot \varepsilon_n)$ satisfies the Cauchy's criterion 2.4. from above. In fact, for arbitrary $n, p\in\mathbb{N}$ we may evaluate:

$$|\varepsilon_{n+1} z_{n+1} + \varepsilon_{n+2} z_{n+2} + \ldots + \varepsilon_{n+p-1} z_{n+p-1} + \varepsilon_{n+p} z_{n+p}| =$$
$$|\varepsilon_{n+1}(s_{n+1} - s_n) + \varepsilon_{n+2}(s_{n+2} - s_{n+1}) + \ldots + \varepsilon_{n+p}(s_{n+p} - s_{n+p-1})| =$$
$$|-\varepsilon_{n+1} s_n + (\varepsilon_{n+1} - \varepsilon_{n+2}) s_{n+1} + \ldots + (\varepsilon_{n+p-1} - \varepsilon_{n+p}) s_{n+p} + \varepsilon_{n+p} s_{n+p}| \leq$$
$$\leq M[\varepsilon_{n+1} + (\varepsilon_{n+1} - \varepsilon_{n+2}) + \ldots + (\varepsilon_{n+p-1} - \varepsilon_{n+p}) + \varepsilon_{n+p}] = 2M\varepsilon_{n+1}.$$

According to the hypothesis $\varepsilon_n \to 0$, for any $\varepsilon > 0$ we can find $n_0(\varepsilon)\in\mathbb{N}$, such that $n > n_0(\varepsilon)$ implies $2M\varepsilon_{n+1} < \varepsilon$. To conclude, for any $\varepsilon > 0$ we have

$$|\varepsilon_{n+1} z_{n+1} + \varepsilon_{n+2} z_{n+2} + \ldots + \varepsilon_{n+p-1} z_{n+p-1} + \varepsilon_{n+p} z_{n+p}| < \varepsilon$$

whenever $n > n_0(\varepsilon)$. ◇

2.7. <u>Corollary.</u> (Leibniz' test for *alternate* series). If (ε_n) is a decreasing sequence of positive real numbers, which tends to zero, then *the alternate series* $\sum (-1)^n \varepsilon_n$ is convergent.

<u>Proof.</u> The partial sums of the series $\sum (-1)^n$ are either –1 or 0, hence the sequence (s_n) is bounded. ◇

2.8. <u>Examples.</u> (*i*) The *alternate harmonic* series $\sum \frac{(-1)^n}{n}$ is convergent since we may take $\varepsilon_n = 1/n$ in the above Leibniz' criterion.

(*ii*) The condition that (ε_n) is decreasing is essential in Abel's criterion, i.e. if $\varepsilon_n \to 0$ non-monotonously, the series may be divergent. For example, let us consider the series:

$$1 - \frac{1}{5} + \frac{1}{2} - (\frac{1}{5})^2 + \frac{1}{3} - \ldots + \frac{1}{n} - (\frac{1}{5})^n + \ldots$$

where the positive terms form the harmonic series (divergent!), and the negative ones belong to a geometrical series of ratio 1/5.

(*iii*) The condition that (ε_n) is decreasing is still not necessary. As an example, the alternate series

$$1 - \frac{1}{2^2} + \frac{1}{3^3} - \frac{1}{4^2} + \ldots + \frac{1}{(2n-1)^3} - \frac{1}{(2n)^2} + \ldots$$

is convergent even if $\varepsilon_n \to 0$ in a non-monotonous fashion.

2.9. <u>Remark.</u> In the particular case of the series with real and positive terms, there are more criteria. For example, in such a case it is obvious that the convergence of a series reduces to the boundedness of its partial sums. The following theorems 2.10 – 2.20 offer other instruments in the study of convergence. These theorems can also be used for some series of complex numbers, via the series of moduli (see later the *absolute convergence*).

2.10. <u>Theorem.</u> (The integral Cauchy's criterion) Let function $f : \mathbb{R}_+ \to \mathbb{R}_+$ be continuous and decreasing, and for all $n \in \mathbb{N}$, let us note:

$$x_n = f(n),\ s_n = \sum_{k=0}^{n} x_k \text{, and } y_n = \int_0^n f(t)dt.$$

Then the series $\sum x_n$ is convergent iff the sequence (y_n) is bounded.

<u>Proof.</u> The integrals exist because f is continuous. Since f is decreasing, for all $k = 1, 2, \ldots$. and $t \in [k-1, k]$ we have $f(k-1) \geq f(t) \geq f(k)$. Integrating these inequalities on $[k-1, k]$, we obtain

$$x_{k-1} \geq \int_{k-1}^{k} f(t)dt \geq x_k.$$

Consequently, adding those relations that correspond to $k = 1, 2, \ldots, n$, we obtain $s_n - x_n \geq y_n \geq s_n - x_0$, i.e. (s_n) and (y_n) are simultaneously bounded. Since $x_n \geq 0$, it follows that (s_n) is increasing, hence its boundedness equals its convergence. ◇

2.11. <u>Example.</u> *The generalized harmonic series* $\sum 1/n^{\alpha}$, where $\alpha > 0$, is obtained by using the above *sampling process* from $f : [1, +\infty) \to \mathbb{R}_+$, where $f(t) = 1/t^{\alpha}$. It is easy to see that for any $n \in \mathbb{N}^*$ we have

$$y_n = \int_1^n \frac{dt}{t^\alpha} = \begin{cases} \frac{1}{\alpha - 1}(1 - n^{1-\alpha}) & if\ \alpha \neq 1 \\ \ln n & if\ \alpha = 1 \end{cases}.$$

So we see that (y_n) is bounded if $\alpha \in (1, +\infty)$, and unbounded if $\alpha \in (0, 1]$, hence the generalized harmonic series is convergent if and only if $\alpha > 1$.

From the already studied convergent or divergent series we can deduce the nature of some other series using *comparison criteria*:

2.12. <u>Theorem.</u> (The 1st criterion of comparison) Let $\sum x_n$ and $\sum y_n$ be series of positive real numbers, for which there exists a rank $n_0 \in \mathbb{N}$ such that $n > n_0$ implies $x_n \leq y_n$. Then the following implications hold:

a) $\sum y_n$ convergent $\Rightarrow \sum x_n$ convergent, and

b) $\sum x_n$ divergent $\Rightarrow \sum y_n$ divergent.

<u>Proof.</u> The inequalities between the general terms imply similar inequalities between the partial sums. Finally, the convergence of a series with positive terms, reduces to the boundedness of the partial sums. ◇

2.13. <u>Theorem.</u> (The 2nd criterion of comparison) Let $\sum x_n$ and $\sum y_n$ be series of positive real numbers such that

$$\frac{x_{n+1}}{x_n} \leq \frac{y_{n+1}}{y_n}$$

holds whenever n is greater than some rank $n_0 \in \mathbb{N}$. Then (as before):

a) $\sum y_n$ convergent $\Rightarrow \sum x_n$ convergent, and

b) $\sum x_n$ divergent $\Rightarrow \sum y_n$ divergent.

<u>Proof.</u> For simplicity, let us suppose $n_0 = 1$. The inequality assumed in the hypothesis leads to:

$$\frac{x_1}{y_1} \geq \frac{x_2}{y_2} \geq .... \geq \frac{x_n}{y_n} \geq .... .$$

If q denotes the first quotient, then $x_n \leq q\, y_n$ for all $n \in \mathbb{N}$, hence we can apply the former criterion of comparison. ◇

2.14. <u>Theorem.</u> (The 3rd criterion of comparison; the *limit form*) Let $\sum x_n$ and $\sum y_n$ be series of positive real numbers, $y_n > 0$, such that there exists

$$\ell = \lim_{n \to \infty} (x_n / y_n).$$

Then the following cases are possible:

a) If $0 < \ell < +\infty$, then the two series have the same nature;

b) If $\ell = 0$, then [$\sum y_n$ convergent $\Rightarrow \sum x_n$ convergent]; and

c) If $\ell = +\infty$, then [$\sum y_n$ divergent $\Rightarrow \sum x_n$ divergent].

Proof. a) By hypothesis, for any positive ε, there exists a rank $n_0(\varepsilon)\in\mathbb{N}$, such that $\ell - \varepsilon < (x_n/y_n) < \ell + \varepsilon$ holds for all $n > n_0(\varepsilon)$. In other words, we have $(\ell - \varepsilon)y_n < x_n < (\ell + \varepsilon)y_n$, hence we can use theorem 2.12.
b) Similarly, for any $\varepsilon > 0$, there exists a rank $n_0(\varepsilon)\in\mathbb{N}$, such that $n > n_0(\varepsilon)$ implies $x_n < \varepsilon\, y_n$.
c) For any $M > 0$ there exists $n_0(M)\in\mathbb{N}$, such that for all $n > n_0(M)$ we have $x_n < M \cdot y_n$. ◇

2.15. Theorem. (D'Alembert's *quotient* criterion) Let us assume that $\sum x_n$ is a series of positive real numbers, and let us note $q_n = x_{n+1}/x_n$.
a) If there exists $n_0 \in\mathbb{N}$ and $q < 1$, such that $q_n < q$ holds for all $n > n_0$, then the series is convergent;
b) If there exists $n_0\in\mathbb{N}$ such that $q_n \geq 1$ holds for all $n > n_0$, then the series is divergent.

Proof. a) For simplicity, we may suppose $n_0 = 1$. Multiplying the relations $x_{k+1} < q{\cdot}x_k$ for $k = 1, 2, \ldots, n$, we obtain that $x_{n+1} < q^n \cdot x_1$. Consequently, our series is compared with a convergent geometric series of ratio $q < 1$.
b) The sequence of general terms does not tend to zero. ◇

2.16. Corollary. (D'Alembert's criterion in the *limit form*). Let $\sum x_n$ be a series of strictly positive real numbers, and let us note $q_n = x_{n+1}/x_n$. If there exists $\ell = \lim\limits_{n\to\infty} q_n \in\mathbb{R}$, then the following implications hold:
a) $\ell < 1 \Rightarrow$ convergence, and
b) $\ell > 1 \Rightarrow$ divergence.

If $\ell = 1$ we cannot decide about the nature of the series.

Proof. By hypothesis, we have $\ell - \varepsilon < q_n < \ell + \varepsilon$ for sufficiently large n. The above theorem 2.15.a) works with $q = \ell + \varepsilon < 1$ to prove a). Similarly, in the case b), we may take $q_n > \ell - \varepsilon > 1$ in 2.15.b).

To show that the case $\ell = 1$ is undecided, we may exemplify by harmonic series $\sum \frac{1}{n^\alpha}$, which is convergent at $\alpha = 2$, and divergent at $\alpha = 1$. ◇

2.17. Theorem. (Cauchy's root criterion) If $\sum x_n$ is a series of positive real numbers, then the following implications hold:
a) If there exists $n_0 \in\mathbb{N}$ and $q \in(0, 1)$, such that $\sqrt[n]{x_n} \leq q$ for all $n > n_0$, then the series is convergent; and
b) If $\sqrt[n]{x_n} \geq 1$ holds for infinitely many indices, then the series is divergent.

Proof. a) For $n > n_0$ we have $x_n \leq q^n$, where $\sum q^n$ is a convergent geometric series. The assertion a) follows by theorem 2.12.
b) The general terms does not tend to zero. ◇

2.18. <u>Corollary.</u> (Cauchy's root criterion in *limit form*) Let $\sum x_n$ be a series of positive real numbers, for which there exists $\ell = \lim\limits_{n\to\infty} \sqrt[n]{x_n}$.

a) If $\ell < 1$, then the series is convergent; and
b) If $\ell > 1$, then the series is divergent.
The case $\ell = 1$ is undecided.

<u>Proof.</u> For any $\varepsilon > 0$ we have $\ell - \varepsilon < \sqrt[n]{x_n} < \ell + \varepsilon$ if n is large enough.

a) If $\ell < 1$, then also $q = \ell + \varepsilon < 1$ for some conveniently small ε, as in case a) of the above theorem.
b) If $\ell > 1$, then we similarly use the inequalities $1 < (\ell - \varepsilon)^n < x_n$.
If $\ell = 1$, we may reason as for corollary 2.16. ◇

2.19. <u>Theorem.</u> (Raabe-Duhamel's test) Let $\sum x_n$ be a series of strictly positive real numbers, and let us note $r_n = n\,(\frac{x_n}{x_{n+1}} - 1)$. We claim that:

a) If there exist $n_0 \in \mathbb{N}$ and $r > 1$, such that $r_n \geq r$ holds for all $n \geq n_0$, then the series is convergent;
b) If there is some $n_0 \in \mathbb{N}$ such that $r_n \leq 1$ holds for all $n \geq n_0$, then the series is divergent.

<u>Proof.</u> a) For simplicity, let us assume that $n_0 = 1$. If we note $r = 1 + \varepsilon$, for some $\varepsilon > 0$, then the inequality in the hypothesis takes the form

$$\varepsilon\, x_{k+1} \leq k\, x_k - (k+1)\, x_{k+1}\,, \quad \forall\, k \in \mathbb{N}.$$

By adding the inequalities corresponding to $k = 1, 2, \ldots, n-1$, we see that the sequence of partial sums is bounded, hence convergent.

b) If $n \geq n_0$, then the inequality from hypothesis leads to $\frac{x_{n+1}}{x_n} = \frac{\frac{1}{n+1}}{\frac{1}{n}}$, which realizes a comparison with the harmonic series. ◇

2.20. <u>Corollary.</u> (Raabe-Duhamel's criterion in *limit form*). Let $\sum x_n$ be a series of strictly positive real numbers, for which there exists

$$\ell = \lim_{n\to\infty} n(\frac{x_n}{x_{n+1}} - 1) \in \mathbb{R}_+ .$$

a) If $\ell > 1$, then the series is convergent, and
b) If $\ell < 1$, then the series is divergent.
The case $\ell = 1$ is undecided.

<u>Proof.</u> For any $\varepsilon > 0$ we have $\ell - \varepsilon < r_n < \ell + \varepsilon$ if n is large enough.

a) We take ε such that $r = \ell - \varepsilon > 1$, and we use part a) from 2.19.
b) We introduce $r_n < \ell + \varepsilon < 1$, in theorem 2.19.b).

For $\ell = 1$, see problem 7 at the end of this section. ◇

2.21. <u>Remark.</u> In practice, it is advisable to use the tests of convergence in the order of simplicity and efficiency, which has been adopted in the above presentation too, i.e. from 2.15 to 2.20. The reason of this procedure comes out from the fact that the Cauchy's test is stronger than the D'Alembert's one, and the Raabe-Duhamel's criterion is the strongest of them. In fact, beside proposition 2.22 below, which compares the quotient and the root tests, the Raabe-Duhamel's test is working in the case $x_{n+1}/x_n \to 1$, i.e. exactly when these two criteria cannot decide.

The above criteria (in the presented order) refer to slower and slower convergent series. Therefore, at least in principle, list of criteria can be infinitely enlarged, since for each series there exists another one, which is slower convergent (see problem 8 below, [BC], [SG], etc.).

The following proposition illustrates the difference of efficiency between the Cauchy and D'Alembert's tests.

2.22. <u>Proposition.</u> Let (x_n) be a sequence of strictly positive real numbers. If $\lim\limits_{n\to\infty}(x_{n+1}/x_n)$ there exists, then $\lim\limits_{n\to\infty}\sqrt[n]{x_n}$ also exists, and they are equal. The converse implication is not true.

<u>Proof.</u> If we take $z_n = \ln x_n$ and $r_n = n$ in the Cesàro-Stolz lemma (see [SG], [PM$_1$], or problem 9 in § II. 1, etc.), then the existence of

$$\lim_{n\to\infty}\frac{z_{n+1}-z_n}{r_{n+1}-r_n} = \lim_{n\to\infty}\ln\frac{x_{n+1}}{x_n} = \ln l$$

implies the existence of

$$\lim_{n\to\infty}\frac{z_n}{r_n} = \lim_{n\to\infty}\frac{\ln x_n}{n} = \ln\lim_{n\to\infty}\sqrt[n]{x_n}$$

and the equality $\lim\limits_{n\to\infty}\sqrt[n]{x_n} = l$. A direct proof, in terms of ε and $n(\varepsilon)$, with separate cases $l = 0$, $l = \infty$, and $l \in \mathbb{R}_+^*$ is recommended to the reader.

To see the invalidity of the converse, we may take as counterexample the series of general term $x_n = \dfrac{(-1)^n + 3}{2^{n+1}}$, $n = 1, 2, \dots$ The limit of x_{n+1}/x_n does not exist at all, while $\sqrt[n]{x_n} \to 1/2$. ◇

In particular, according to Corollary 2.18, but not 2.16, the series $\sum x_n$ from the above counterexample is convergent. In addition, the sum can be computed using geometric series.

2.23. <u>Remark.</u> Comparing the harmonic series (example 2.5 (ii)) with the alternate harmonic series (example 2.8 (i)), we see that taking the series of absolute values generally affects the convergence. On the other side, using criteria concerning series with positive terms, it is easier to obtain information about series of absolute values. In order to develop such a study we need more notions concerning the series in Γ; as a matter of fact, we refer to series in $\mathbb{C}$, and treat the real series as a particular case.

2.24. <u>Definition.</u> The series $\sum z_n$ is said to be *absolutely convergent* iff $\sum |z_n|$ is convergent. We say that the series $\sum z_n$ is *conditionally* (or *semi*-) *convergent* iff it is convergent, but $\sum |z_n|$ is divergent.

Besides some remarks in examples like the previously mentioned ones, the above definition makes tacitly use of the following property:

2.25. <u>Proposition.</u> The absolute convergence implies convergence.

<u>Proof.</u> The obvious inequality

$$| z_{n+1} + z_{n+2} + \dots + z_{n+p}| \leq | z_{n+1}| + | z_{n+2}| + \dots + | z_{n+p}|$$

permits us to compare the nature of the series $\sum z_n$ and $\sum |z_n|$ via the general Cauchy's criterion. ◇

Simple examples show that the converse implication is not true.

2.26. <u>Remarks.</u> *(i)* The semi-convergent series of real numbers have a remarkable property, namely by changing the order of terms, we can produce other series whose sequences of partial sums tends to a previously given number. In fact, such a series contains infinitely many positive, as well as negative terms, which tend to zero. By adding conveniently chosen terms, we can approximate any real number (see problem 9 at the end). Of course, changing the order of terms means to take another series. On the other hand, we mention that in the case of an absolutely convergent series, the sum is independent of this order (known as a Cauchy's theorem).

(ii) The absolute convergence is also important in the process of operating with convergent series. In the case of addition, there is no problem: we add term by term, and the sum of two convergent series is convergent. Doing the product is more complicated, since writing all the possible products between terms can be done in different ways, and the convergence of the product series is generally not guaranteed by that of the initial ones.

In the sequel, we present the *Cauchy's rule* of multiplying series, which is most frequently used for power series:

2.27. <u>Definition.</u> The *product* (or *convolution*) of two series

$$z_0 + z_1 + z_2 + \dots + z_n + \dots$$
$$Z_0 + Z_1 + Z_2 + \dots + Z_n + \dots$$

is defined (in the Cauchy's sense) by the series

$$\zeta_0 + \zeta_1 + \zeta_2 + \dots + \zeta_n + \dots$$

where the terms ζ_0 , ζ_1 , … are obtained by the *crossing multiplication:*

$\zeta_0 = z_0 Z_0$,

$\zeta_1 = z_0 Z_1 + z_1 Z_0$,

$\zeta_2 = z_0 Z_2 + z_1 Z_1 + z_2 Z_0$,

… … …

$\zeta_n = z_0 Z_n + z_1 Z_{n-1} + \dots + z_{n-1} Z_1 + z_n Z_0 = \sum_{k=0}^{n} z_k Z_{n-k}$,

… … …

Simple examples show how unpredictable the square of one (hence in general the product of two) semi-convergent series can be:

2.28. <u>Examples.</u> *(i) Semi-convergent series with <u>divergent</u> square.*

Let us consider the series of term $z_n = i^n n^{-1/2}$ for all $n \in \mathbb{N}^*$. Its square, obtained by taking $z_n = Z_n$ in the Cauchy's rule, has the general term

$$\zeta_n = \sum_{k=1}^{n} (i^k / \sqrt{k})(i^{n+1-k} / \sqrt{n+1-k}) = i^{n+1} \sum_{k=1}^{n} (\sqrt{k}\sqrt{n+1-k})^{-1} .$$

Because each term of the last sum is greater than $1/n$, we obtain $|\zeta_n| \geq 1$, hence the square series is divergent.

On the other hand, the absolute convergence is not necessary to the convergence of the product:

(ii) Semi-convergent series with <u>convergent</u> square.

The square of the alternate harmonic series is convergent. In fact, this square has the general term

$$\zeta_n = \sum_{k=1}^{n} \frac{(-1)^k}{k} \frac{(-1)^{n-k}}{n+1-k} = (-1)^n \sum_{k=1}^{n} \frac{1}{k(n+1-k)} = (-1)^n \frac{2}{n+1} \sum_{k=1}^{n} \frac{1}{k},$$

since according to the formula

$$\frac{1}{k} + \frac{1}{n+1-k} = \frac{n+1}{k(n+1-k)},$$

each term in ζ_n appears twice. Consequently the square series is alternating, and $|\zeta_n| \to 0$ because $1/n \to 0$ implies, via Cesàro-Stolz' theorem, that also

$$\frac{1}{n} \sum_{k=1}^{n} \frac{1}{k} \to 0 .$$

Exact information about the nature of a product can be obtained only if at least one of the series is absolutely convergent:

2.29. <u>Theorem.</u> (Mertens). The product of two (semi-)convergent series is convergent if at least one of them is absolutely convergent.

2.30. <u>Theorem.</u> (Cauchy). The product of two absolutely convergent series is absolutely convergent, and the sum of the product series equals the product of the initial sums.

<u>Proof.</u> Let us assume that the series $\sum |x_n|$ and $\sum |y_n|$ are convergent to X, respectively Y, and let $\sum z_n$ be the product of the initial series $\sum x_n$ and $\sum y_n$. Because for some $\nu, \mu \in \mathbb{N}$ we have

$$\sum_{k=0}^{n} |z_k| \leq |x_0 y_0| + ... + |x_{n_s} y_{m_s}| \leq (|x_0| + ... + |x_\nu|)(|y_0| + ... + |y_\mu|) \leq XY ,$$

it follows that $\sum z_n$ is absolutely convergent. The indices n_s and m_s are not specified because the above reason is valid for any rule of realizing the product, which takes into consideration all the pairs of terms.

Since the nature and the sum of an absolutely convergent series does not depend on the order of terms, we can arrange z_k such that:

$$\left(\sum_{k=0}^{n} x_k\right)\left(\sum_{k=0}^{n} y_k\right) = \sum_{k=0}^{n} z_k ,$$

which proves the relation between the sums. ◇

About the sum of a product series we also mention (without proof):

2.31. Theorem. (Abel). If the series $\sum u_n$, $\sum v_n$, and their product $\sum w_n$, are convergent to U, V, respectively W, then $UV=W$.

To illustrate how the above results are to be used in practice, we consider the following particular examples:

2.32. Application. Study the convergence of the series:

a) $1 + \alpha + \frac{\alpha(\alpha-1)}{2!} + \ldots + \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!} + \ldots$

b) $1 - \alpha + \frac{\alpha(\alpha-1)}{2!} - \ldots + (-1)^n \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!} + \ldots$

a) Let $a_n(\alpha)$ be the general term of the series. Looking for the absolute convergence, when the theory of series with positive terms is applicable, we see that only the Raabe-Duhamel's test is working, and it yields:

$$\lim_{n\to\infty} n\left(\frac{|a_n(\alpha)|}{|a_{n+1}(\alpha)|} - 1\right) = \lim_{n\to\infty} \frac{n(\alpha+1)}{n-\alpha} = \alpha + 1.$$

Consequently, if $\alpha > 0$, then we have $\alpha + 1 > 1$, and according to the corollary 2.20, the series is absolutely convergent. As a matter of fact, this case includes $\alpha = 0$, when $a_n(\alpha) = 0$ holds for all $n \geq 1$, i.e. the partial sums of the series are constantly equal to 1.

In the sub-case $\alpha < 0$, let us note $\alpha = -\beta$, and remark that

$$a_n(\alpha) = (-1)^n \frac{\beta(1+\beta)\ldots(n+\beta-1)}{n!} = (-1)^n b_n(\beta),$$

where the last equality represents a notation. Consequently, the series is alternate, but because

$$\frac{b_{n+1}(\beta)}{b_n(\beta)} = \frac{n-\alpha}{n+1},$$

it follows that, for $\alpha \leq -1$, the general term doesn't tend to zero any more, hence the series diverges.

The remaining case corresponds to $\alpha \in (-1, 0)$, when the sequence $(b_n(\beta))$ is decreasing. Using 2.20 again, the series $\sum b_n(\beta)$ diverges since

$$\lim_{n\to\infty} n\left(\frac{b_n(\beta)}{b_{n+1}(\beta)} - 1\right) = \lim_{n\to\infty} \frac{n(1-\beta)}{n+\beta} = 1 - \beta < 1.$$

In other terms, the initial series is not absolutely convergent, so we have to analyze it semi-convergence. In this respect, let $c_n(\beta)$ be the general term of a sequence for which:

$$b_n(\beta) = \frac{1}{n}[c_1(\beta) + c_2(\beta) + \ldots + c_n(\beta)].$$

Using a similar expression of $b_{n-1}(\beta)$, we obtain

$$c_n(\beta) = n\, b_n(\beta) - (n-1)b_{n-1}(\beta) = \beta\, b_{n-1}(\beta).$$

Because $(b_n(\beta))$ is a decreasing sequence of positive numbers, it is convergent, hence $(c_n(\beta))$ is convergent too. Let us note its limit by $\ell = \lim_{n\to\infty} c_n(\beta) \in \mathbb{R}_+$. According to the Cesàro-Stolz' theorem, we have also $\ell = \lim_{n\to\infty} b_n(\beta)$, hence $\ell = \beta\ell$. Because $\beta > 0$, it follows that $\ell = 0$. Consequently, the condition of the Leibniz' test (corollary 2.7) are fulfilled, and we can conclude that the initial series is semi-convergent.

In conclusion, the complete answer in the case a) is:

- absolute convergence if $\alpha \geq 0$;
- semi-convergence if $\alpha \in (-1, 0)$;
- divergence if $\alpha \leq -1$.

b) Similarly to the case a), the absolute convergence holds if $\alpha \geq 0$.

On the other hand, if $\alpha < 0$, then the same substitution, namely $\alpha = -\beta$, reduces the series to $\sum b_n(\beta)$. As we have already seen, this series is divergent according to the Raabe-Duhamel's test.

The conclusion relative to case b) is:

- absolute convergence if $\alpha \geq 0$, and
- divergence if $\alpha < 0$.

Using some theoretical results from the next sections, we will be able to get information about the corresponding sums (namely 2^α, respectively 0). So far we can decide only for $\alpha = n \in \mathbb{N}$, when the series represent finite (binomial) sums, namely 2^n in the first case, and 0 in the second one.

PROBLEMS §II.2.

1. Test the following series for convergence:

$$\text{(a) } \sum \frac{1}{n}2^{-n}, \quad \text{(b) } \sum \frac{\ln n}{n}, \quad \text{(c) } \sum (n-2^n)^{-1}, \quad \text{(d) } \sum \frac{1}{2n-1}$$

Hint. Compare to the series $\sum 1/2^n$, and $\sum 1/n$.

2. Give examples of convergent series $\sum a_n$ and divergent $\sum b_n$ such that one of the following inequalities be valid for all $n = 1, 2, \ldots$

(i) $a_n \geq b_n$ (in $\mathbb{R}$), *(ii)* $|a_n| \geq |b_n|$ (in $\mathbb{R}$, and in $\mathbb{C}$).

Hint. *(i)* $a_n = 1/n^2$, $b_n = -1/n$;

(ii) $a_n = (-1)^n / n$, and $b_n = \frac{1}{2n}\left[\cos(1-(-1)^n)\frac{\pi}{4} + i\sin(1-(-1)^n)\frac{\pi}{4}\right]$.

3. Test the following series for convergence:

$$\text{(i) } \sum (2n-1)2^{-n/2}, \quad \text{(ii) } \sum \frac{2\cdot 5\cdot 8\cdot \ldots\cdot (3n-1)}{1\cdot 5\cdot 9\cdot \ldots\cdot (4n-3)}.$$

Hint. Use the D'Alembert 's test.

4. Using the Cauchy's criterion, study the convergence of the series:

$$\text{(i) } \sum \left(\frac{n+1}{2n-1}\right)^n, \quad \text{(ii) } \sum \left(\frac{n}{3n-1}\right)^{2n-1}.$$

Hint. In the second case, the coefficients of the even powers are null.

5. Test for convergence the generalized harmonic series $\sum \frac{1}{n^\alpha}$, $\alpha \in \mathbb{R}$.

Hint. If $\alpha \leq 0$, the series diverges. If $\alpha > 0$, we evaluate r_n in the Raabe-Duhamel's criterion, and we obtain

$$r_n = n\left(\frac{(n+1)^\alpha}{n^\alpha} - 1\right) = \frac{\left[1+(1/n)^\alpha\right]-1}{1/n}.$$

Because $\lim_{n\to\infty} r_n = \lim_{x\to 0} \frac{(1+x)^\alpha - 1}{x} = \alpha$, the series is convergent for $\alpha > 1$, and divergent for $\alpha \leq 1$ (the case $\alpha = 1$ has to be studied separately).

6. Decide about the nature of the series $\sum_{n=1}^{\infty} \frac{1\cdot 3\cdot 5\cdot \ldots\cdot (2n-1)}{2\cdot 4\cdot 6\cdot \ldots\cdot (2n)}$.

Hint. The Raabe-Duhamel's criterion gives $r_n = n/(2n+1) \to \frac{1}{2}$.

7. Use the Raabe-Duhamel test to establish the nature of the series:

(a) $\sum_{n=1}^{\infty}(2n+1)\left[\frac{\alpha(\alpha-1)...(\alpha-n+1)}{(\alpha+1)(\alpha+2)...(\alpha+n)}\right]^2$, where $\alpha\in\mathbb{R}$, and

(b) $\sum_{n=1}^{\infty}\frac{n!}{\beta(\beta+1)...(\beta+n-1)}$, where $\beta>0$.

Hint. (a) If x_n denotes the general term of the series, then

$$n\left(\frac{x_n}{x_{n+1}}-1\right)=\frac{n}{2n+3}\cdot\frac{(8\alpha+2)n^2+(12\alpha+4)+(1+2\alpha-2\alpha^2)}{(\alpha-n)^2}\to 4\alpha+1.$$

Consequently, the series is convergent for $\alpha > 0$, and divergent for $\alpha < 0$. If $\alpha = 0$, then $x_n = 0$, hence the series is convergent.

(b) If y_n represents the general term of the series, then

$$n\left(\frac{y_n}{y_{n+1}}-1\right)=(\beta-1)\frac{n}{n+1}\to\beta-1.$$

So we deduce that the series converges for $\beta > 2$, and it diverges for $\beta < 2$. In the case $\beta = 2$, we have $y_n = \frac{1}{n+1}$, hence the series is divergent.

8. Let $a_n\to a$ and $b_n\to b$ in Γ, such that $a_n\neq a$ and $b_n\neq b$ for all $n\in\mathbb{N}$. We say that (a_n) *faster* converges than (b_n), respectively (b_n) *slower* converges than (a_n), iff $\lim_{n\to\infty}\frac{a_n-a}{b_n-b}=0$. Similarly, if $\sum_{n=0}^{\infty}a_n=a$ and $\sum_{n=0}^{\infty}b_n=b$, we say that $\sum a_n$ is *faster* than $\sum b_n$, respectively $\sum b_n$ is *slower* than $\sum a_n$, iff $\lim_{n\to\infty}\frac{\alpha_k}{\beta_k}=0$, where $\alpha_n=a-\sum_{k=0}^{n}a_k$ and $\beta_n=b-\sum_{k=0}^{n}b_k$ are the remainders of the respective series. Show that:

(a) If $\sum a_n$ and $\sum b_n$ are convergent series of strictly positive numbers, and $a_n\to 0$ faster than $b_n\to 0$, then $\sum a_n$ is *faster* than $\sum b_n$.

(b) For each convergent series $\sum b_n$ there exist other convergent series, $\sum a_n$ faster, and $\sum c_n$ slower than $\sum b_n$.

(c) If we rewrite the geometric series of terms $x_n = 2^{-n}$, where $n\geq 1$, by decomposing the general term x_n into n terms $\frac{1}{n}x_n+...+\frac{1}{n}x_n$, then the resulting series $\sum y_n$ is slower than $\sum x_n$.

Hint. (a) $\lim_{n\to\infty}\frac{a_n}{b_n}=0$ means that to arbitrary $\varepsilon > 0$ there correspond a rank $n(\varepsilon)\in\mathbb{N}$ such that $0<\frac{a_k}{b_k}<\varepsilon$ holds for all $k\geq n(\varepsilon)$. By multiplying the inequalities $0 < a_k < \varepsilon\, b_k$, where $k\geq n\geq n(\varepsilon)$, we obtain

$$0 < \alpha_n = \sum_{k=n+1}^{\infty} a_k < \varepsilon\, \beta_n = \varepsilon \sum_{k=n+1}^{\infty} b_k \,.$$

(b) Using (a) we may take $a_n = b_n^{1+\nu}$ and $c_n = b_n^{1-\nu}$, where $\nu \in (0, 1)$.

(c) To express y_n, let us note $\sigma_n = 1 + 2 + \ldots + (n-1) = \frac{n(n-1)}{2}$. It is easy to see that for each $k \in \mathbb{N}^*$, there exists an unique rank $n(k) \in \mathbb{N}^*$, such that $\sigma_{n(k)} \le k < \sigma_{n(k)+1}$. Consequently, we obtain $y_k = \frac{1}{n(k)} x_{n(k)}$, and finally

$$\frac{x_k}{y_k} = \frac{x_k}{\frac{1}{n(k)} x_{n(k)}} \le \frac{x_{\sigma_{n(k)}}}{\frac{1}{n(k)} x_{n(k)}} = n(k) \cdot 2^{\frac{3n(k) - n^2(k)}{2}} \to 0 \,.$$

9. Approximate the numbers 5/3, π, $-e$, and ln 2 with four exact decimals using terms of the alternate harmonic series.

Hint. We have $1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} < 5/3 < 1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9}$, hence we start to use negative terms and we obtain $1 + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9} - \frac{1}{2} < 5/3$. Then we add positive terms until we overpass 5/3 and so on. We similarly treat the numbers π, $-e$. Finally, $\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} = ln\, 2$.

10. Test for convergence the following series:

(a) $\sum_{n=2}^{\infty} \frac{1}{\ln n}$, (b) $\sum_{n=2}^{\infty} \frac{1}{n \ln n}$, (c) $\sum_{n=2}^{\infty} \frac{1}{n(\ln n)(\ln \ln n)}$, and (d) $\sum_{n=2}^{\infty} (n \ln^2 n)^{-1}$.

Hint. (a) can be compared to the harmonic series; The others can be studied by the integral test, using the primitives ln(ln x), ln(ln(ln x)), and $-\ln^{-1} x$. Consequently, the single convergent series in this exercise is that of (d).

11. Let us note $\varepsilon_n = \left[\sqrt{n} + (-1)^n\right]^{-1}$, where $n \in \mathbb{N}$, $n \ge 2$. Give an explanation why $\sum (-1)^{n+1} \varepsilon_n$ is divergent, while $\varepsilon_n \to 0$.

Hint. (ε_n) is not monotonic. To justify the divergence we may use the inequality $v(n) = 1 + \sqrt{2n} - \sqrt{2n+1} > 0$ in evaluating

$$\frac{1}{\sqrt{2n+1} - 1} - \frac{1}{\sqrt{2n} + 1} = \frac{1 + v(n)}{\sqrt{2n+1}\sqrt{2n} - v(n)} > \frac{1}{2n+1},$$

which offers clear information about the odd partial sums, namely:

$$s_3 = \frac{-1}{\sqrt{2} + 1} + \frac{1}{\sqrt{3} - 1} > \frac{1}{3}; \ldots; s_{2n+1} > \frac{1}{3} + \ldots + \frac{1}{2n+1}; \ \ldots \ .$$

12. Let $(a_n)_{n\in\mathbb{N}^*}$ be a sequence of positive real numbers such that $\sum a_n^2$ is convergent. Show that the series $\sum \frac{1}{n} a_n$ converges too.

Hint. From $(a_n - \frac{1}{n})^2 > 0$, we deduce that $\frac{2}{n} a_n < a_n^2 + \left(\frac{1}{n}\right)^2$.

13. Let s be the sum of the alternating harmonic series (s = ln2 will be obtained in the next section, using function series). Find the sum:

$$\left(1-\frac{1}{2}-\frac{1}{4}\right)+\left(\frac{1}{3}-\frac{1}{6}-\frac{1}{8}\right)+\ldots+\left(\frac{1}{2k-1}-\frac{1}{4k-2}-\frac{1}{4k}\right)+\ldots$$

Hint. Express the partial sums σ_n of the rearranged series by the partial sums s_n of the initial (alternate harmonic) series. For example,

$$\sigma_{3m} = \sum_{k=1}^{m}\left(\frac{1}{2k-1}-\frac{1}{4k-2}-\frac{1}{4k}\right) = \sum_{k=1}^{m}\left(\frac{1}{4k-2}-\frac{1}{4k}\right) = \frac{1}{2} s_{2m},$$

which leads to $\sigma_{3m} \to \frac{1}{2}$ s. We may similarly treat σ_{3m+1} and σ_{3m+2}.

14. Evaluate the following sums:

(a) $\sum_{n=1}^{\infty}(\sqrt{n+\alpha+1}-2\sqrt{n+\alpha}+\sqrt{n+\alpha-1})$, where $\alpha > 0$;

(b) $\sum_{n=1}^{\infty}\frac{1}{(\alpha+n)(\alpha+n+1)}$, where $\alpha \in \mathbb{R} \setminus \mathbb{Z}$;

(c) $\sum_{n=1}^{\infty}(16n^2-8n-3)^{-1}$; (d) $\sum_{n=1}^{\infty}\ln\frac{n+1}{n}$; (e) $\sum_{n=1}^{\infty}\frac{2^n-(-1)^n}{5^n}$.

Hint. (a) Note $\sqrt{n+\alpha}-\sqrt{n+\alpha-1} = a_n$, and find $s_n = a_{n+1} - a_1$. For (b) and (c), decompose into elementary fractions, and compute s_n.

15. Show that the divergent series

$$2 + 2 + 2^2 + 2^3 + \ldots + 2^n + \ldots \quad \text{and} \quad -1 + 1 + 1^2 + 1^3 + \ldots + 1^n + \ldots$$

have an absolutely convergent product. Extend this property to an arbitrary pair of series of the form $a_0 + \sum_{n\ge1} a^n$, and $b_0 + \sum_{n\ge1} b^n$, where $a \neq b$.

Hint. The general terms of the product series are $c_0 = -2$, and $c_n = 0$ at the remaining $n = 1, 2, \ldots$. In general, we have

$$c_n = a_0\, b^n + b_0\, a^n - a^n - b^n + \frac{a^{n+1}-b^{n+1}}{a-b} = \frac{a^n A - b^n B}{a-b},$$

where $A = a + (b_0 - 1)(a - b)$, and $B = b + (1 - a_0)(a - b)$. In particular, we can realize $A = B = 0$, even if $a - b = 1$.

§ II.3. SEQUENCES AND SERIES OF FUNCTIONS

In this section, we consider sequences and series whose terms are real or complex functions. From lyceum, we already know some simple cases of real functions, and we can easily extend them to complex variables. In principle, each property in $\mathbb{R}$ has a valid extension in $\mathbb{C}$, so that the notation Γ for both $\mathbb{R}$ and $\mathbb{C}$ will be very useful.

3.1. <u>Examples</u>. a) Rising $x \le -1$ to an increasing power leads to no limit, but applying the same process to $x > -1$ yields

$$\lim_{n\to\infty} x^n = \begin{cases} 0 & if \ \ |x|<1 \\ 1 & if \ \ x=1 \\ +\infty & if \ \ x>1 \end{cases} .$$

Obviously, this formula describes the behavior of the general term of a geometric progression when $n \to \infty$.

Using the trigonometric form of a complex number, we can easily pass from $x \in \mathbb{R}$ to $z \in \mathbb{C}$. In fact, because $|z^n| = |z|^n$, we have

$$\lim_{n\to\infty} z^n = \begin{cases} 0 & if \ |z|<1 \\ 1 & if \ z=1 \\ \infty & if \ |z|>1 \end{cases}$$

(where understanding ∞ on the Riemann's sphere is advisable). In the remaining cases, when $|z|=1$ but $z \ne 1$, from $\arg z^n = n \arg z (\mathrm{mod}\, 2\pi)$ we deduce that $\lim\limits_{n\to\infty} z^n$ doesn't exist.

b) Adding the former terms of the geometric progression leads to the *geometric series* $\sum x^n$. According to the above result, it is convergent iff $|x|<1$, when its sum is

$$\sum_{n=0}^{\infty} x^n = \lim_{n\to\infty} \frac{1-x^n}{1-x} = \frac{1}{1-x}.$$

Otherwise, this series is divergent.

Similarly, in the complex framework, we have

$$\sum_{n=0}^{\infty} z^n = \lim_{n\to\infty} \frac{1-z^n}{1-z} = \frac{1}{1-z}$$

if and only if $|z|<1$.

c) All the sequences and series with parameters, frequently met in lyceum, represent sequences and series of functions. For example, it should be well known that $\lim_{n\to\infty} \frac{1}{n} \sin n\, x = 0$ at any $x \in \mathbb{R}$, etc. The complex analogue of such results needs a thorough knowledge of the complex functions of complex variables, like sin $n\, z$, with $z \in \mathbb{C}$, in this case.

Similar theoretical aspects, as well as plenty of practical problems, lead us to investigate general methods of defining complex functions, including the elementary ones (see § II.4). To anticipate, we mention the utility of the complex functions in tracing conformal maps, calculating real (sometimes improper) integrals, and solving differential equations (see [HD], etc.).

In other words, our primary interest in studying sequences and series of functions is their role in the construction of other functions.

Beside the analytical method of defining functions, which is based on power series, considering sequences of functions may offer significant information about plenty of numerical sequences and series. The advantage consists in the use of derivation, integration, and other operations based on a limit process that involves functions.

As a consequence of these purposes, the present section has two parts:

(i) Types of convergence and properties of the limit function, and

(ii) Developments in (real) Taylor series.

3.2. <u>Definition</u>. Let $D \subseteq \mathbb{R}$ be a fixed domain, and let $\mathcal{F}(D, \mathbb{R}) = \mathbb{R}^D$ be the set of all functions $f : D \to \mathbb{R}$. Any function $F : \mathbb{N} \to \mathcal{F}(D, \mathbb{R})$ is called *sequence of* (*real*) *functions*. Most frequently it is marked by mentioning the terms (f_n), where $f_n = F(n)$, and n is an arbitrary natural number.

We say that a number $x \in D$ is a *point of convergence* of (f_n) if the numerical sequence $(f_n(x))$ is convergent. The set of all such points forms *the set* (or *domain*) *of convergence*, denoted D_c . The resulting function, say $\varphi : D_c \to \mathbb{R}$, expressed at any $x \in D_c$ by

$$\varphi(x) = \lim_{n\to\infty} f_n(x),$$

is called *limit* of the given sequences of functions. Alternatively we say that φ is the (point-wise) limit of (f_n), (f_n) *p-tends to* φ, etc., and we note $\varphi \overset{p}{=} \lim_{n\to\infty} f_n$, $f_n \xrightarrow{p} \varphi$, etc.

The notions of *series of functions, partial sums, infinite sum, domain of convergence,* etc., are similarly defined in $\mathcal{F}(D, \mathbb{R})$. For this reason, in the former part of the present paragraph we mainly refer to sequences (not series) of functions.

In addition, these notions have the same form in the case of complex functions, i.e. in $\mathcal{F}(D, \mathbb{C})$, where also $D \subseteq \mathbb{C}$.

3.3. Proposition. A sequence (f_n) of functions is *point-wise* convergent to function φ, (i.e. $\varphi = \overset{p}{\lim\limits_{n\to\infty}} f_n$ on D_c) iff

$$\forall x \in D_c\ \forall \varepsilon > 0\ \exists n_0(x,\varepsilon) \in \mathbb{N} \text{ such that } n > n_0(x,\varepsilon) \Rightarrow | f_n(x) - \varphi(x)| < \varepsilon .$$

Proof. This condition expresses the convergence of the numerical sequence $(f_n(x))$ at any $x \in D_c$. Of course, in the complex case we preferably replace x by $z \in D_c \subseteq \mathbb{C}$, and the above condition takes the form

$$\forall z \in D_c\ \forall \varepsilon > 0\ \exists n_0(z,\varepsilon) \in \mathbb{N} \text{ such that } n > n_0(z,\varepsilon) \Rightarrow | f_n(z) - \varphi(z)| < \varepsilon ,$$

where the modulus $|\cdot|$ corresponds to either $\mathbb{R}$ or $\mathbb{C}$. ◇

This condition of point-wise convergence can be formulated without explicit mention of the limit function:

3.4. Theorem. (The Cauchy's general test of point-wise convergence) The sequence (f_n) is point-wise convergent on D_c iff

$$\forall x \in D_c\ \forall \varepsilon > 0\ \exists n_0(x,\varepsilon) \in \mathbb{N} \text{ such that } m, n > n_0(x, \varepsilon) \Rightarrow | f_m(x) - f_n(x) | < \varepsilon.$$

Proof. The numerical sequence $(f_n(x))$ is convergent iff it is a Cauchy sequence at any $x \in D_c$. A similar condition holds in $\mathbb{C}$. ◇

3.5. Remark. We have to mention x in $n_0(x, \varepsilon)$ because generally speaking, this rank depends on x. As for example, we may consider $f_n : [0,1] \to \mathbb{R}$, $f_n(x) = x^n$; $g_n : [-1, 1] \to \mathbb{R}$, $g_n(x) = x^n(1 - x^{2n})$; etc. In fact, if we evaluate f_n (or g_n) at different points $x_n = 2^{-1/n} \to 1$, $n \in \mathbb{N}^*$, then $| f_n(x_n) - 0 | = \frac{1}{2}$, hence the condition in proposition 3.3. can not be satisfied with the same rank $n_0(\varepsilon)$ at all $x \in D_c$. A similar behavior takes place in the complex case, when D denotes the closed unit disk in $\mathbb{C}$.

On the other hand, the restrictions $f_n : [0,\delta] \to \mathbb{R}$, and $g_n : [-\delta, \delta] \to \mathbb{R}$, of the above examples, where $0 < \delta < 1$, show that the rank $n_0(\varepsilon)$ may happen to be valid for all $x \in D_c$. It is easy to see that any other restrictions to compact subsets of D_c have similar properties. In the complex case, the functions f_n and g_n shall be restricted to closed disks $D_\delta = \{z \in \mathbb{C}: |z| \le \delta\}$, or to other compact subsets of D_c.

In order for us to distinguish such cases, we will introduce other types of convergence as follows:

3.6. Definition. Let (f_n) be a sequence of functions $f_n : D \to \Gamma$, which is point-wise convergent to $\varphi : D_c \to \Gamma$ (remember that Γ means either $\mathbb{R}$ or $\mathbb{C}$). We say that (f_n) is *uniformly* (*u*-) *convergent to* φ iff

$$\forall \varepsilon > 0\ \exists n_0(\varepsilon) \in \mathbb{N} \text{ such that } n > n_0(\varepsilon) \Rightarrow | f_n(x) - \varphi(x)| < \varepsilon \text{ at any } x \in D_c.$$

In this case we note $f_n \xrightarrow[D_c]{u} \varphi$, $\varphi = \overset{u}{\lim\limits_{n\to\infty}} f_n$, etc.

If $f_n \xrightarrow{u} \varphi$ only on arbitrary compact sets $K \subseteq D_c$, but not necessarily on D_c , then we say that (f_n) is *almost uniformly* (briefly *a.u.*-) convergent to function φ . In this case, we note $\varphi \overset{a.u.}{\underset{D_c}{=}} \lim_{n\to\infty} f_n$, $f_n \xrightarrow[D_c]{a.u.} \varphi$, etc.

3.7. <u>Remarks</u>. a) In the definition of the uniform (and the almost uniform) convergence we already assume the p-convergence, since it furnishes the limit function φ. More than this, any u-convergent sequence is also a.u.-convergent. In fact, if a rank $n_0(\varepsilon)$ is good for all the points of D_c, then it is good for all $x \in K \subseteq D_c$ too. Therefore, we say that the uniform convergence is *stronger than* the a.u. one, which, at its turn, is *stronger* than the point-wise one, i.e. the following implications hold:

u-convergence $\Rightarrow$ *a.u.-convergence* $\Rightarrow$ *p-convergence.*

b) In the case of a series of functions, besides the point-wise, uniform, and almost uniform convergence, other nuances of convergence are frequently taken into consideration, e.g. the *absolute* and the *semi*-convergence. These details are omitted here because of their strong analogy with the numerical series in both $\mathbb{R}$ and $\mathbb{C}$.

c) The condition of uniform convergence may be formulated without special reference to the variable x, as for example if the involved functions are bounded (continuous, etc.), and the sup-norm makes sense. More exactly, the sequence (f_n) is u-convergent to φ on D_c iff

$$\forall \varepsilon > 0\ \exists n_0(\varepsilon) \in \mathbb{N} \text{ such that } n > n_0(\varepsilon) \Rightarrow \|f_n - \varphi\| < \varepsilon .$$

For this reason, $\|f\| = \sup\{|f(x)| : x \in D_c\}$ is said to be *norm of the uniform convergence*, or simply *u-norm*. In a similar manner, we describe the a.u.-convergence in terms of family of semi-norms $p_K(f) = \sup\{|f(x)| : x \in K\}$, where K denotes a compact subset of D_c.

The completeness of $\mathbb{R}$ and $\mathbb{C}$ allows us to formulate also the uniform convergence with no reference to the limit function, by analogy to theorem 3.4, concerning the point-wise convergence, namely:

3.8. <u>Theorem</u>. (The general Cauchy's test of uniform convergence) For a sequence (f_n) to be uniformly convergent to φ on D_c it is necessary and sufficient that

$$\forall \varepsilon > 0\ \exists n_0(\varepsilon) \in \mathbb{N} \text{ such that } m, n > n_0(\varepsilon) \Rightarrow | f_m(x) - f_n(x) | < \varepsilon$$

at any $x \in D_c$ (when we say that (f_n) is *uniformly fundamental*, or *Cauchy*).

<u>Proof.</u> If $\varphi \overset{u}{=} \lim_{n\to\infty} f_n$, then the u-Cauchy condition follows from

$$|f_m(x) - f_n(x)| \le |f_m(x) - \varphi(x)| + |\varphi(x) - f_n(x)| .$$

Conversely, if (f_n) is uniformly Cauchy on D_c , then it also is point-wise Cauchy (as in theorem 3.4). Consequently, there exists $\varphi \overset{p}{=} \lim_{n\to\infty} f_n$ on D_c .

We claim that $\varphi = \overset{u}{\lim\limits_{n\to\infty}} f_n$ too. In fact, for any $\varepsilon > 0$, let $n_0(\varepsilon) \in \mathbb{N}$ be such that at any $x \in D_c$ we have

$$m, n > n_0(\varepsilon) \Rightarrow | f_m(x) - f_n(x) | < \varepsilon / 2 .$$

On the other hand, each $x \in D_c$ determines a rank $m_0(x, \varepsilon) \in \mathbb{N}$, such that (because of the point-wise convergence)

$$m > m_0(x, \varepsilon) \Rightarrow | f_m(x) - \varphi(x) | < \varepsilon / 2$$

at any $x \in D_c$. So we may conclude that for any $\varepsilon > 0$ there exists $n_0(\varepsilon) \in \mathbb{N}$ such that (eventually adjusting m up to x) we have

$$|f_n(x) - \varphi(x)| \leq |f_n(x) - f_m(x)| + |f_m(x) - \varphi(x)| < \varepsilon ,$$

whenever $m > \max \{n_0(\varepsilon), m_0(x, \varepsilon)\}$. Consequently, (f_n) is u-convergent to function φ on D_c . ◇

In spite of its generality, it is quite difficult to apply the Cauchy's test. However, it has many practical consequences, as for example the following corollaries 3.9, 3.10 and 3.11. More than this, we are especially interested in criteria for u-convergence, since the p-convergence immediately reduces to numerical series.

3.9. Corollary. (The Weierstrass' test) Let $\sum f_n$ be a series of real or complex functions $f_n : D \to \Gamma$, and let $\sum a_n$ be a series of real numbers. If

1. $\sum a_n$ is convergent, and
2. $| f_n(x) | \leq a_n$ holds at any $x \in D$, and for all $n \in \mathbb{N}$,

then $\sum f_n$ is uniformly (and absolutely) convergent on D .

Proof. The sequence of partial sums is fundamental, because

$$| f_n(x) + f_{n+1}(x) + \ldots + f_{n+m}(x) | \leq a_n + a_{n+1} + \ldots + a_{n+m}$$

holds at any $x \in D$. ◇

3.10. Corollary. (The Abel's test) Let $\sum f_n \cdot g_n$ be a series of functions defined on $D \subseteq \mathbb{R}$, where $f_n : D \to \Gamma$ and $g_n : D \to \mathbb{R}$ for all $n \in \mathbb{N}$. If

1. the series $\sum f_n$ is uniformly convergent on D , and
2. the sequence (g_n) is bounded and monotonic on D,

then also the initial series is uniformly convergent on D .

Proof. If we note $\sigma_0 = f_n , \ldots, \sigma_m = f_n + \ldots + f_{n+m}$, etc., then according to 1, it follows that $| \sigma_m | < \varepsilon$ holds for all m, whenever n is sufficiently large. In addition, we have:

$$f_n = \sigma_0$$
$$f_{n+1} = \sigma_1 - \sigma_0$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$f_{n+m} = \sigma_m - \sigma_{m-1} \text{ , etc.,}$$

so that the sum in the Cauchy's general test becomes:

$$f_n g_n + \ldots + f_{n+m} g_{n+m} =$$
$$= \sigma_0 g_n + (\sigma_1 - \sigma_0) g_{n+1} + \ldots + (\sigma_m - \sigma_{m-1}) g_{n+m} =$$
$$= \sigma_0 (g_n - g_{n+1}) + \ldots + \sigma_{m-1}(g_{n+m-1} - g_{n+m}) + \sigma_m g_{n+m}.$$

Because all the differences $(g_n - g_{n+1}), \ldots, (g_{n+m-1} - g_{n+m})$ have the same sign (take separately +1 and –1, if easier), it follows that the inequalities

$$| f_n(x) g_n(x) + \ldots + f_{n+m}(x) g_{n+m}(x) | \le$$
$$\le \ \varepsilon | g_n(x) - g_{n+m}(x) | + | \sigma_m(x) g_{n+m}(x) | \le$$
$$\le \varepsilon |g_n(x)| + \varepsilon |g_{n+m}(x)| + |\sigma_m(x) g_{n+m}(x)| \le \varepsilon (| g_n(x) | + 2| g_{n+m}(x) |)$$

hold at any $x \in D$. Thus it remains to use the boundedness of the sequence (g_n), which assures the existence of $M > 0$ such that $| g_k(x) | < M$ holds at any $x \in D$, for sufficiently large n, and arbitrary $m \in \mathbb{N}$. ◇

3.11. <u>Corollary</u>. (The Dirichlet's test). Let $\sum f_n \cdot g_n$ be a series of functions defined on $D \subseteq \mathbb{R}$, where $f_n : D \to \Gamma$ and $g_n : D \to \mathbb{R}$ for all $n \in \mathbb{N}$, and let s_n be the partial sums of the series $\sum f_n$. If

1. $\exists K > 0$ such that $| s_n(x) | \le K$ at any $x \in D$, and for all k in $\mathbb{N}$ (i.e. the sequence (s_n) is *equally bounded* on D), and

2. the sequence (g_n) is monotonic and u-convergent to 0 on D ,

then $\sum f_n \cdot g_n$ is u-convergent on D too.

<u>Proof.</u> Similarly to the proof of the above corollary, replacing

$$f_n = s_n - s_{n-1}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$f_{n+m} = s_{n+m} - s_{n+m-1}$$

in the sum involved in the general Cauchy's test, we obtain:

$$f_n g_n + \ldots + f_{n+m} g_{n+m} =$$
$$= (s_n - s_{n-1}) g_n + (s_{n+1} - s_n) g_{n+1} + \ldots + (s_{n+m} - s_{n+m-1}) g_{n+m} = - s_{n-1} g_n +$$
$$+ s_n(g_n - g_{n+1}) + s_{n+1}(g_{n+1} - g_{n+2}) + \ldots + s_{n+m-1} (g_{n+m-1} - g_{n+m}) + s_{n+m} g_{n+m}.$$

Now, let $\varepsilon > 0$ be arbitrary, and $n_0(\varepsilon) \in \mathbb{N}$ be the rank after which (i.e. for all $n \ge n_0(\varepsilon)$) we have $| g_n(x) | < \varepsilon / 4K$ at any $x \in D$. Because $n + m \ge n_0(\varepsilon)$ also holds for all $m \in \mathbb{N}$, it follows that

$$| f_n(x) g_n(x) + \ldots + f_{n+m}(x) g_{n+m}(x)| \le$$
$$\le | s_{n-1}(x) g_n(x) | + K | g_n(x) - g_{n+m}(x) | + | s_{n+m}(x) g_{n+m}(x) | \le$$
$$\le 2K (| g_n(x) | + | g_{n+m}(x) |) \le \varepsilon ,$$

hence $\sum f_n \cdot g_n$ is uniformly Cauchy. ◇

3.12. <u>Remark</u>. As a general scheme, the notion of convergence shall be based on some topology of the space wherefrom the terms of the sequences are taken. In particular, the fact that x is a point of convergence can be expressed in terms of a semi-norm on $\mathscr{F}(D, \mathbb{R})$, namely $p_x(f) = | f(x) |$. More exactly, (f_n) is convergent to φ at $x \in D$ iff $\lim\limits_{n \to \infty} p_x (f_n - \varphi) = 0$.

Similarly, the u-convergence frequently makes use of sup – norm, and the a.u. convergence is described by the family of semi-norms p_K , already mentioned in remark 3.7.c. Naturally, it is quite difficult to identify such structures for other types of convergences. For example, the problem of introducing a topology on $\mathscr{F}$ $(D, \mathbb{R})$ such that the corresponding convergence *carries* (*preserves*, or *transports*) the property of continuity from the terms f_n of the sequence to the limit φ , is (we hope completely) solved in [PM$_2$] and [PM$_3$], but many other cases remain open.

Sequences like (f_n) in the above remark 3.5 show that the point-wise convergence is too weak for carrying the continuity from the terms to the limit (i.e. each f_n is continuous while φ isn't). The following theorem points out that the uniform convergence assures the transportation of the continuity from term functions to the limit function. Because the notions of continuity, derivative and integral are not yet analyzed in the complex case, the rest of this section refers to real functions of real variables. Later on, we will see that these properties remain valid in the complex framework.

3.13. <u>Theorem</u>. If $\varphi = \overset{u}{\lim\limits_{n\to\infty}} f_n$ on $D_c \subseteq \mathbb{R}$, and the terms $f_n : D \to \mathbb{R}$ are continuous on D_c , then φ is also continuous on D_c .

<u>Proof.</u> Let us fix $x_0 \in D_c$, and $\varepsilon > 0$. To prove the continuity of φ at x_0 , we have to find $\delta > 0$ such that $|\ \varphi(x) - \varphi(x_0)\ | < \varepsilon$ be valid whenever $x \in D_c$ and $|\ x - x_0\ | < \delta$. For this purpose we primarily consider the rank $n_0(\varepsilon)$, furnished by the u-convergence of (f_n), and choose some $n > n_0(\varepsilon)$ such that $|\ f_n(x) - \varphi(x)\ | < \varepsilon/3$ holds at any $x \in D_c$, including x_0 . We claim that the continuity of f_n at x_0 yields the desired $\delta > 0$. In fact, because the inequality

$$|f_n(x) - f_n(x_0)| < \varepsilon/3$$

holds at any $x \in D_c$ whenever $|\ x - x_0\ | < \delta$, it follows that

$$|\ \varphi(x) - \varphi(x_0)\ | < |\ \varphi(x) - f_n(x)\ | + |\ f_n(x) - f_n(x_0)\ | + |\ f_n(x_0) - \varphi(x_0)\ | < \varepsilon ,$$

i.e. φ is continuous at x_0 . ◇

A similar result refers to the sequence of derivatives:

3.14. <u>Theorem</u>. If (f_n) is a sequence of functions $f_n : D \to \mathbb{R}$, such that :

1. $f_n \xrightarrow{p} f$ on D_c ,
2. each f_n is derivable on D_c ,
3. there exists $g = \overset{u}{\lim} f_n^{/}$ on D_c ,

then f is derivable on D_c , and $f^{/} = g$.

<u>Proof.</u> For any x_0 , $x \in D_c$, we obviously have:

$$\left| \frac{f(x) - f(x_0)}{x - x_0} - g(x_0) \right| \leq$$

$$\leq \left| \frac{f(x)-f(x_0)}{x-x_0} - \frac{f_n(x)-f_n(x_0)}{x-x_0} \right| + \left| \frac{f_n(x)-f_n(x_0)}{x-x_0} - f_n'(x_0) \right| + \left| f_n'(x_0) - g(x_0) \right|.$$

Because the last two moduli in the above inequality can be easily made arbitrary small by acting on n and $|x - x_0|$, the key problem is to show that

$$\frac{f_n(x)-f_n(x_0)}{x-x_0} \xrightarrow{u} \frac{f(x)-f(x_0)}{x-x_0}.$$

In fact, this sequence is u-Cauchy, since according to the Lagrange's theorem we have

$$\frac{f_m(x)-f_m(x_0)}{x-x_0} - \frac{f_n(x)-f_n(x_0)}{x-x_0} = \frac{(f_m-f_n)(x)-(f_m-f_n)(x_0)}{x-x_0} =$$

$$= (f_m - f_n)'(\xi_x) = f_m'(\xi_x) - f_n'(\xi_x),$$

where ξ_x is lying between x_0 and x. On this way we reduce the problem to the uniform convergence of the derivatives.

The rest of the proof is routine. ◇

Finally, we have a *rule of integrating term by term*:

3.15. <u>Theorem</u>. Let (f_n) be a sequence of functions $f_n : D \to \mathbb{R}$, $D \subseteq \mathbb{R}$, and let D_c be its domain of convergence. If

1. each f_n is continuous on D_c, and
2. $f_n \xrightarrow{u} f$,

then f is integrable on any interval $[a, b] \subseteq D_c$, and

$$\int_a^b f(x)dx = \lim_{n\to\infty} \int_a^b f_n(x)dx.$$

<u>Proof.</u> According to theorem 3.13, f is continuous on D_c, hence it is also integrable on any interval [a, b] $\subseteq D_c$. If $\varepsilon > 0$ is given, then hypothesis 2 assures the existence of $n_0(\varepsilon) \in \mathbb{N}$ such that $| f_n(x) - f(x) | < \varepsilon / (b - a)$ at any $x \in [a, b]$, whenever $n > n_0(\varepsilon)$. Consequently, we have

$$\left| \int_a^b f_n(x)dx - \int_a^b f(x)dx \right| = \left| \int_a^b [f_n(x) - f(x)]dx \right| \leq$$

$$\leq \int_a^b |f_n(x) - f(x)| dx < \int_a^b \frac{\varepsilon}{b-a} dx = \varepsilon,$$

which proves the last assertion of the theorem. ◇

3.16. <u>Remarks</u>. a) The above theorems 3.13, 3.14 and 3.15 lead to similar properties of the series of functions. They are omitted here because usually, formulating the corresponding statements and proving them shouldn't raise problems (however recommended exercises).

b) Theorems 3.13, 3.14 and 3.15 from above remain valid under the hypothesis of almost uniform convergence. The proofs shall be slightly modified by putting forward some compact sets. Most simply, in theorem 3.15, $[a, b]$ already is a compact set. Similarly, sequences like (g_n) from the above remark 3.5. show that the uniform convergence is not necessary to assign continuous limits to sequences of continuous functions.

In addition, other hypotheses can be replaced by weaker conditions without affecting the validity of these theorems. For example, theorem 3.15, concerning the integrability, remains true if we replace the continuity (hypothesis 1) by the weaker condition of integrability. Generally speaking, one of the most important problems in studying the convergence in spaces of functions is that of identifying the type of convergence, which is both necessary and sufficient to carry some property from the terms to the limit.

A contribution to this problem, which concerns the property of continuity, can be find in [PM_1] and [PM_2]). The key step consists in formulating the adequate type of convergence, namely:

3.17. <u>Definition</u>. The sequence (f_n) of functions $f_n : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}$ and $n \in \mathbb{N}$, is said to be *quasi-uniformly* (briefly *q.u.*) *convergent* on $A \subseteq D$ to a function $f : A \to \mathbb{R}$, if

$$\underset{\varepsilon>0}{\forall}\ \underset{x_0\in A}{\forall}\ \underset{n_0\in\mathbb{N}}{\exists}\ \textit{such that}\ \underset{\mathbb{N}\ni n\geq n_0}{\forall}\ \underset{V_n\in\mathscr{V}(x_0)}{\exists}\ \textit{such that}$$
$$[(\forall)\, x \in V_n \Rightarrow \left| f_n(x) - f(x) \right| < \varepsilon].$$

If so, we note $f \overset{q.u.}{\underset{A}{=}} \lim f_n$, $f_n \overset{q.u.}{\underset{A}{\to}} f$, etc.

We mention that the q.u. convergence is a topological one, i.e. it corresponds to a particular topology on $\mathscr{F}(D, \mathbb{R})$, in the sense of [KJ], [PM_2], etc. To place the q.u. convergence among other convergences, we may easily remark that

u. convergence $\Rightarrow$ *q.u. convergence* $\Rightarrow$ *p. convergence*.

Simple examples show that the converse implications fail to be generally valid (see also [PM_1], etc.):

3.18. <u>Examples</u>. a) The sequence (f_n), where $f_n : [0, 1] \to \mathbb{R}$, for all $n \in \mathbb{N}$, $f_n(x) = \exp(-n x^2)$, is point-wise but not q.u. convergent on [0,1] to

$$f(x) = \begin{cases} 1 & at\ x = 0 \\ 0 & at\ x \in (0,1] \end{cases} .$$

b) The sequence (g_n), where $g_n : [0, 1] \to \mathbb{R}$, for all $n \in \mathbb{N}$, $g_n(x) = x^n(1 - x^n)$, is q.u. but not u. convergent to the null function on [0, 1].

3.19. <u>Theorem</u>. Let the sequence (f_n) of functions $f_n : D \to \mathbb{R}$, where $D \subseteq \mathbb{R}$ and $n \in \mathbb{N}$, be point-wise convergent on $A \subseteq D$ to a function $f : A \to \mathbb{R}$. In addition, we suppose that each function f_n is continuous on A. Then the limit function f is continuous on A if and only if the convergence of the sequence (f_n) is quasi-uniform on A.

Proof. Let us suppose that $f_n \underset{A}{\overset{q.u.}{\to}} f$, and let us choose some $x_0 \in A$. To show that f is continuous at x_0, let $\varepsilon > 0$ be given. According to definition 3.17, there is some $n_0 \in \mathbb{N}$, such that for each $n > n_0$ we can find a neighborhood $V_n \in \mathcal{V}(x_0)$ such that $|f_n(x) - f(x)| < \varepsilon/3$ holds at each $x \in V_n$ (in particular at x_0 too). Because f_n is continuous at x_0, there is some $U_n \in \mathcal{V}(x_0)$ such that $|f_n(x_0) - f_n(x)| < \varepsilon/3$ holds at each $x \in U_n$. If we correspondingly fix a rank $n > n_0$, then $W = U_n \cap V_n \in \mathcal{V}(x_0)$ doesn't depend on n. In addition, all these inequalities hold at each $x \in W$. By putting them together, we obtain

$$|f(x) - f(x_0)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < \varepsilon,$$

which proves the continuity of f at $x_0 \in A$.

Conversely, let us say that f is continuous at each $x_0 \in A$. More exactly,

$$\underset{\varepsilon>0}{\forall} \ \underset{U \in \mathcal{V}(x_0)}{\exists} \text{ such that } [(\forall) x \in U \Rightarrow |f(x) - f(x_0)| < \varepsilon/3 \ .$$

A similar condition holds for f_n, i.e.

$$\underset{\varepsilon>0}{\forall} \ \underset{U_n \in \mathcal{V}(x_0)}{\exists} \text{ such that } [(\forall) x \in U_n \Rightarrow |f_n(x) - f_n(x_0)| < \varepsilon/3 \ .$$

Finally, the convergence of numerical sequence $(f_n(x_0))$ to $f(x_0)$ means that

$$\underset{\varepsilon>0}{\forall} \ \underset{n_0 \in \mathbb{N}}{\exists} \text{ such that } [n > n_0 \Rightarrow |f_n(x_0) - f(x_0)| < \varepsilon/3 \ .$$

If we note $V_n = U_n \cap U \in \mathcal{V}(x_0)$, then, at any $x \in V_n$, we have

$$|f_n(x) - f(x)| \le |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| + |f(x_0) - f(x)| < \varepsilon \ .$$

Consequently, $f_n \underset{A}{\overset{q.u.}{\to}} f$. ◇

3.20. Remark. a) So far we have investigated how particular types of convergences may carry some *good* properties (like continuity, derivability, and integrability) into similar *good* properties. To complete the image, we mention that the limiting process in sequences and series of functions may transform *bad* properties into *worse*, although the uniform and absolute (i.e. the strongest) convergence is assured. This fact is visible in the examples of continuous but nowhere derivable functions, which are limits of continuous or piece-wise derivable functions. One of the first examples of this type (due to K.W.T. Weierstrass) is the sum w of the series

$$w(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x) \ ,$$

where $a \in (0, 1)$ and $b \in \mathbb{N}$ is odd, such that $ab > 1 + (3\pi/2)$.

Another, perhaps more "popular" example (due to B.L. van der Waerden), starts with the periodical prolongation, noted $f : \mathbb{R} \to \mathbb{R}$, of the modulus $|\cdot|: [-1/2, +1/2] \to \mathbb{R}$, and realizes the so-called *condensation of the singularities* (see [FG], [G-O], etc.) by the summation of the series

$$\sum_{n=0}^{\infty} 4^{-n} f(4^n x) \ .$$

b) From another point of view, we have only considered that a sequence or series has been given, and the task was to study the existence of the limit functions, and their properties. The converse process is also very useful in practice, namely starting with some function (which for example is to be evaluated or approximated) we need a sequence (or series) that converges somehow to this function. When the process involves series we use to say that we *develop* the given function in a series. An important type of such developments consists of Taylor series, which will be discussed during the rest of this section. This type of series is intensively used in the concrete evaluation of the functions, including the elementary ones. The upcoming values are usually put in trigonometric, logarithmic, and other tables, or, most frequently in our days, worked by computer techniques. These series are equally important from a theoretical point of view, since they play the role of *definitions* of the complex functions, operator functions, etc.

The simplest case of Taylor series involves polynomials.

3.21. <u>Proposition</u>. If $P(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$ is a polynomial function, and $x_0 \in \mathbb{R}$ is fixed, then the equality

$$P(x) = P(x_0) + \frac{P'(x_0)}{1!}(x - x_0) + \dots + \frac{P^{(n)}(x_0)}{n!}(x - x_0)^n$$

holds at each $x \in \mathbb{R}$. In particular, the coefficients have the expressions

$$a_0 = P(0),\ a_1 = \frac{1}{1!}P^{/}(0),\ \dots,\ a_n = \frac{1}{n!}P^{(n)}(0),$$

where $P^{/}$ up to $P^{(n)}$ represent the derivatives of P.

<u>Proof</u>. We write the polynomial in the form

$$P(x) = b_0 + b_1 (x - x_0) + \dots + b_n (x - x_0)^n ,$$

and we identify the coefficients. By repeated derivation in respect to x, and the replacement of $x = x_0$, we obtain

$$b_0 = P(x_0),\ b_1 = P^{/}(x_0),\ \dots,\ n!\, b_n = P^{(n)}(x_0) ,$$

which lead to the announced relations. ◇

Of course, the above equality is no longer valid for other than polynomial functions. However, in certain circumstances we can give approximating polynomials of this form, according to the following result:

3.22. <u>Theorem</u>. Let I be an interval of $\mathbb{R}$, and let function $f : I \to \mathbb{R}$ be n+1 times derivable on I . If x_0 is fixed in I, then the equality

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \ldots +$$

$$+ \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}dt$$

holds at any $x \in I$.

Proof. We may reason by induction on $n \in \mathbb{N}$. In fact, the case $n = 0$ reduces to the obvious relation

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(t)dt \; .$$

If the formula is supposed to be valid up to $n - 1$, then verifying it for n, means to prove the equality:

$$\int_{x_0}^{x} f^{(n)}(t)\frac{(x-t)^{n-1}}{(n-1)!}dt = \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}dt$$

By evaluating the difference of these two integrals we obtain:

$$\frac{1}{n!}\int_{x_0}^{x}\left[nf^{(n)}(t) - (x-t)f^{(n+1)}(t)\right](x-t)^{n-1}dt =$$

$$= -\frac{1}{n!}\int_{x_0}^{x}\frac{d}{dt}\left[(x-t)^n f^{(n)}(t)\right]dt = \frac{1}{n!}(x-x_0)^n f^{(n)}(x_0) \; ,$$

which achieves the poof. ◇

3.23. Definition. If a function $f : I \to \mathbb{R}$ is n times derivable, and $x_0 \in I$, then the polynomial function $T_n : \mathbb{R} \to \mathbb{R}$, expressed by

$$T_n(x) = f(x_0) + \frac{f'(x_0)}{1!}(x-x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$$

is called *Taylor polynomial* of degree n, attached to f at x_0 .

The expression of $f(x)$, established in theorem 3.22, is called *Taylor formula*. It is easy to see that these formulas represent extensions of the Lagrange’s theorem on finite increments.

The difference $R_n = f - T_n$ is called *Taylor remainder* of order n, of f at x_0.

In particular, if f is n+1 times derivable, then the remainder expressed by the integral in theorem 3.22, i.e.

$$R_n(x) = \int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}dt$$

is called *remainder in integral form*.

Because sometimes we need other forms of the remainder, it is useful to know more results similar to theorem 3.22:

3.24. Theorem. Let function $f : I \to \mathbb{R}$ be n+1 times derivable on I, and let $x_0 \in I$ be fixed. For each $x \in I$ there exists at least one point ξ_x (depending on x), between x and x_0, such that the following equality holds

$$f(x) = f(x_0) +$$

$$\frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n +$$

$$+ \frac{f^{(n+1)}(\xi_x)}{(n+1)!}(x-x_0)^{n+1} .$$

Proof. Let $E : I \to \mathbb{R}$ be a function for which the equality

$$f(x) = f(x_0) +$$

$$\frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \ldots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n +$$

$$+ \frac{E(x)}{(n+1)!}(x-x_0)^{n+1}$$

holds at any $x \in I$. Since f is n+1 times derivable, it follows that E is simply derivable on I. Let us suppose that $x_0 < x$, and define $\varphi : [x_0, x] \to \mathbb{R}$ by

$$\varphi(t) = f(t) +$$

$$\frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}(x-t)^2 + \ldots + \frac{f^{(n)}(t)}{n!}(x-t)^n + \frac{E(x)}{(n+1)!}(x-t)^{n+1}.$$

We claim that φ is a *Rolle* function on $[x_0, x]$ (i.e. it satisfies the conditions in the Rolle's theorem, well-known from lyceum), namely:

1. φ is derivable on (x_0, x) because f is n+1 times derivable on $I \supset (x_0, x)$,
2. $\varphi(x) = f(x) = \varphi(x_0)$.

According to the conclusion of the Rolle's theorem, there exists $\xi_x \in (x_0, x)$ such that $\varphi'(\xi_x) = 0$. Taking into account that

$$\varphi'(t) = \frac{f^{(n+1)}(t)}{n!}(x-t)^n - \frac{E(x)}{n!}(x-t)^n ,$$

it follows that $E(x) = f^{(n+1)}(\xi_x)$. ◇

3.25. Remark. The remainder from theorem 3. 24, which involves the derivative $f^{(n+1)}(\xi_x)$, i.e.

$$L_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x)(x-x_0)^{n+1}$$

is referred to as *the Lagrange's remainder* of f at x_0.

Other types of remainders are possible , e.g. the Cauchy's one

$$C_n(x) = \frac{f^{(n+1)}(\vartheta x)}{n!}(1-\vartheta)^n x^{n+1},$$

where $\vartheta \in (0, 1)$ depends on x and n, and the list continues.

All these remainders, and particularly R_n, and L_n and C_n represent different forms of the same quantity. In particular, L_n results from R_n by the following *generalized mean formula* concerning the integral of a product:

3.26. Lemma. If φ, $\psi : [a, b] \to \mathbb{R}$ are continuous functions, and ψ does not change the sign on $[a, b]$, then there exists $\xi \in [a, b]$ such that

$$\int_a^b \varphi(t)\psi(t)\,dt = \varphi(\xi)\int_a^b \psi(t)\,dt .$$

Proof. Since φ is continuous, and $[a, b]$ is compact, there exist

$$m = \inf \{\varphi(x) : x \in [a, b]\} \text{ and } M = \sup \{\varphi(x) : x \in [a, b]\}.$$

To make a choice, let us say that $\psi \geq 0$. In this case, the inequalities

$$m\,\psi(t) \leq \varphi(t)\,\psi(t) \leq M\,\psi(t)$$

hold at any $t \in [a, b]$, and the monotony of the integral gives

$$m\int_a^b \psi(t)\,dt \leq \int_a^b \varphi(t)\psi(t)\,dt \leq M\int_a^b \psi(t)\,dt .$$

The searched ξ is one of the points where φ takes the intermediate value represented by the quotient of these integrals, i.e.

$$\varphi(\xi) = \left(\int_a^b \varphi(t)\psi(t)\,dt\right) \cdot \left(\int_a^b \psi(t)\,dt\right)^{-1} .$$

Similarly, we treat the case $\psi \leq 0$. ◇

3.27. Proposition. Let function $f : I \to \mathbb{R}$ be $n+1$ times derivable on I.

1. If $x_0 \in I$ is fixed, then for any $x \in I$ there exists (at least one) point ξ_x between x_0 and x such that $R_n(x) = L_n(x)$, and
2. If $x_0 = 0$, then there exists $\vartheta \in (0, 1)$ such that $R_n(x) = C_n(x)$.

Proof. In order to obtain the first equality, namely

$$\int_{x_0}^{x} f^{(n+1)}(t)\frac{(x-t)^n}{n!}\,dt = \frac{f^{(n+1)}(\xi_x)}{(n+1)!}(x-x_0)^{n+1} ,$$

we may apply the previous lemma to the particular pair of functions

$$\varphi(t) = f^{(n+1)}(t) \text{ and } \psi(t) = \frac{1}{n!}(x-t)^n$$

on $[x_0, x]$ if $x_0 \leq x$, respectively on $[x, x_0]$ if $x \leq x_0$.

The other equality, i.e. $R_n(x) = C_n(x)$, follows by changing the variable in the integral remainder. In fact, if we put $t = \tau x$, then $R_n(x)$ becomes

$$\int_0^x f^{(n+1)}(t)\frac{(x-t)^n}{n!}\,dt = \frac{x^{n+1}}{n!}\int_0^1 f^{(n+1)}(\tau x)(1-\tau)^n\,d\tau .$$

The usual mean formula furnishes the searched $\vartheta \in (0, 1)$. ◇

It is useful to know several variants of remainder because it may happen that only one of them be adequate to the concrete problem. In particular, the Cauchy's form is well fitting to the following *binomial development*:

3.28. Example. The equality

$$(1+x)^{\alpha} = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n + \dots$$

holds at each $x\in(-1, 1)$ in the sense of the absolute and a.u. convergence, for arbitrary $\alpha\in\mathbb{R}$.

In fact, function $f : (-1,\infty)\to\mathbb{R}$, of values $f(x) = (1+x)^{\alpha}$, is infinitely derivable, and for any $n\in\mathbb{N}$ we have:

$$f^{(n)}(x) = \alpha(\alpha-1)\dots(\alpha-n+1)(1+x)^{\alpha-n}.$$

Consequently, the following Taylor formula makes sense

$$(1+x)^{\alpha} = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n + C_n(x),$$

where $C_n(x)$ represents the Cauchy's remainder, i.e.

$$C_n(x) = \frac{\alpha(\alpha-1)\dots(\alpha-n)(1+\vartheta x)^{\alpha-n-1}}{n!}(1-\vartheta)^n x^{n+1}.$$

On the other hand, the D'Alembert test in limiting form, shows that

$$1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n + \dots$$

is an absolutely and point-wise convergent series on $(-1, 1)$, because

$$\lim_{n\to\infty}\frac{\left|\frac{\alpha(\alpha-1)\dots(\alpha-n+1)(\alpha-n)}{(n+1)!}x^{n+1}\right|}{\left|\frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n\right|} = \lim_{n\to\infty}\frac{n-\alpha}{n+1}|x| < 1$$

whenever $|x| < 1$. This convergence suggests that in the Taylor formula we have $C_n \xrightarrow{a.u.} 0$ on $(-1, 1)$. To prove this fact, we write

$$C_n(x) = \underbrace{\frac{(\alpha-1)\dots(\alpha-1-n+1)}{n!}x^n}_{a_n}\underbrace{\alpha\, x(1+\vartheta x)^{\alpha-1}}_{b}\underbrace{\left(\frac{1-\vartheta}{1+\vartheta x}\right)^n}_{c_n},$$

and we remark that $a_n \to 0$ as the general term of a similar series for $\alpha - 1$; b is bounded, namely $|\alpha x|(1-|x|)^{\alpha-1} \le |b| \le |\alpha x|(1+|x|)^{\alpha-1}$; and $0 < c_n < 1$, since $0 < 1-\vartheta < 1+\vartheta x$ reduces to $\vartheta(1+x) > 0$. More precisely, these properties of a_n, b, and c_n hold uniformly on any compact set $K\subset(-1, 1)$ since they are implicitly valid at that point $x_0 \in K$, where we have

$$|x_0| = \max\ \{|x| : x\in K\}.$$

We mention that some particular values of $\alpha \in \mathbb{R}$ correspond to important developments. For example, integrating and deriving in the developments of $(1+x)^{-1}$, $\sqrt[m]{1+x}$, etc., we may obtain many other formulas, like the developments of $\ln(1+x)$, $(1+x)^{-2}$, etc. (see also the final list).

In the particular case $\alpha \in \mathbb{N}$, the above binomial developments reduce to the finite sums of the Newton's formulas, i.e. $C_n(x) \equiv 0$ for enough large n, and the convergence is obvious.

The convergence at $x = \pm 1$ has been analyzed in II.2.32.

3.29. <u>Application</u>. Besides approximation problems, the Taylor formulas are useful in the study of the local extremes. In fact, if $f \in \mathbf{C}_{\mathbb{R}}^2(I)$ takes an extreme value at x_0, then $f'(x_0) = 0$, hence

$$f(x) - f(x_0) = \frac{1}{2} f''(\xi_x)(x - x_0)^2 .$$

Consequently, we have to distinguish two cases, namely:

a) If $f''(x_0) \neq 0$, then the increment $f(x) - f(x_0)$ preserves the sign on some neighborhood of x_0, hence x_0 is an extreme point, and respectively

b) If $f''(x_0) = 0$, but there exists $f'''(x_0) \neq 0$, then x_0 is not extreme point any more (and we call it *inflexion point*).

In the more general case, when several derivatives vanish at x_0, the result depends on the parity of the first non-null derivative, namely:

1. If $f'(x_0) = f''(x_0) = \ldots = f^{(2p-1)}(x_0) = 0$ and $f^{(2p)}(x_0) \neq 0$, then x_0 really is an extreme point;
2. If $f'(x_0) = f''(x_0) = \ldots = f^{(2p)}(x_0) = 0$ and $f^{(2p+1)}(x_0) \neq 0$,then x_0 is not an extreme point (but only inflexion point).

If f has derivatives of any order n, on D, and the remainder tends to zero when $n \to \infty$ (as in example 3.23), it is more advisable to speak of Taylor *series* instead of Taylor *formulas*, according to the following:

3.30. <u>Definition</u>. Let function $f : I \to \mathbb{R}$ be infinitely derivable on I, i.e. the derivatives $f^{(n)}(x)$ exist at any $x \in I$ and for any $n \in \mathbb{N}$, (which is briefly noted $f \in \mathbf{C}_{\mathbb{R}}^{\infty}(I)$). If $x_0 \in I$ is fixed, then the series

$$f(x_0) + \frac{f'(x_0)}{1!}(x - x_0) + \ldots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \ldots$$

is called *Taylor series* attached to f at x_0. If this series is convergent to f, we say that f can be *developed* in Taylor series *around* x_0. In the particular case when $x_0 = 0$, some people call it *Mac Laurin series*.

3.31. <u>Remark</u>. The terms of Taylor series are monomials, powers of $x - x_0$, and the partial sums are polynomials. As usually, the main problem about Taylor series concerns the convergence, generally expressed by $R_n \to 0$. This time it is completed by another question, namely: "*Is the Taylor series attached to f at x_0 convergent to the same f on a neighborhood of x_0 ?*". The answer is generally negative, like in the following case.

3.32. <u>Example</u>. Let function $f: \mathbb{R} \to \mathbb{R}$ be expressed by

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}.$$

This function is infinitely derivable on $\mathbb{R}$ and $f^{(n)}(0) = 0$ for all $n \in \mathbb{N}$, so that the attached Mac Laurin series is identically null, hence u-convergent. However, excepting $x = 0$, we have $f(x) \neq 0$.

The equality holds if the derivatives are equally bounded on I, i.e.

3.33. <u>Theorem</u>. Let $f \in \mathbf{C}^{\infty}_{\mathbb{R}}(I)$, and let $x_0 \in I$. If there exist a neighborhood $V \subseteq I$ of x_0, and a constant $M > 0$ such that the inequalities

$$| f^{(n)}(x) | < M$$

hold at any $x \in V$ and for any $n \in \mathbb{N}$, then the Taylor series attached to f at x_0 is u-convergent to f on V.

<u>Proof</u>. Without losing generality, we may suppose that

$$V = \{x \in I : | x - x_0 | < \varepsilon\}$$

for some $\varepsilon > 0$. By maximizing the Lagrange's remainder on V, we obtain the inequality

$$| R_n(x) | = \frac{\left| f^{(n+1)}(\xi_x) \right|}{(n+1)!} \left| x - x_0 \right|^{n+1} < M \frac{\varepsilon^{n+1}}{(n+1)!},$$

which realizes the comparison of $| R_n(x)|$ with the general term of the a.u.-convergent series Me^{ε}. Consequently, $R_n \xrightarrow{u} 0$ on this neighborhood.

We mention the following direct consequence of this theorem:

3.34. <u>Corollary</u>. The Mac Laurin (and generally Taylor) series attached to the functions *exp, sin, cos, sinh,* and *cosh* are absolutely and almost uniformly convergent on $\mathbb{R}$ to the same functions.

<u>Proof</u>. All these functions have equally bounded derivatives on the set λK, where $K \subset \mathbb{R}$ is compact, and $\lambda \in [-1,1]$. More exactly,

$$\underset{\text{compact } K \subset \mathbb{R}}{\forall} \quad \underset{\lambda \in [-1,1]}{\forall} \quad \underset{M > 0}{\exists} \quad \textit{such that}$$

$$[(\forall) x \in \lambda K, (\forall) n \in \mathbb{N} \Rightarrow \left| f^{(n)}(x) \right| < M]$$

Consequently, we may apply theorem 3.33 at $x_0 = 0$. ◇

Because the developments of the real functions will be starting points (i.e. definitions) in the complex analysis, we end this section by mentioning the most remarkable ones, which refer to some elementary functions. The reader is kindly advised to learn them by heart.

3.35. <u>List of developments</u>. The following real functions have a.u. and absolutely convergent Mac Laurin series on the mentioned domains:

1) The exponential function, on $\mathbb{R}$:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots$$

2) The circular trigonometric sine, on $\mathbb{R}$:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots + (-1)^p \frac{x^{2p+1}}{(2p+1)!} + \ldots$$

3) The circular trigonometric cosine, on $\mathbb{R}$:

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots + (-1)^q \frac{x^{2q}}{(2q)!} + \ldots$$

4) The hyperbolic sine, on $\mathbb{R}$:

$$\sinh x = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots + \frac{x^{2p+1}}{(2p+1)!} + \ldots$$

5) The hyperbolic cosine, on $\mathbb{R}$:

$$\cosh x = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + \frac{x^{2q}}{(2q)!} + \ldots$$

6) The binomial function, on (–1, 1):

$$(1 + x)^\alpha = 1 + \frac{\alpha}{1!} x + \frac{\alpha(\alpha - 1)}{2!} x^2 + \ldots + \frac{\alpha(\alpha - 1)\ldots(\alpha - n + 1)}{n!} x^n + \ldots$$

where α is arbitrary in $\mathbb{R}$.

7) The elementary fraction, on (–1, 1):

$$\frac{1}{1 - x} = 1 + x + x^2 + \ldots + x^n + \ldots$$

8) The natural logarithm, on (–1, 1):

$$\ln (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots + (-1)^{n+1} \frac{x^n}{n} + \ldots$$

9) The m^{th} root, on (–1, 1):

$$\sqrt[m]{1 + x} = 1 + \frac{x}{m} + \frac{1 - m}{2! m^2} x^2 + \ldots + \frac{(1 - m)\ldots[1 - (n - 1)m]}{n! m^n} x^n + \ldots$$

10) The inverse trigonometric function *arctg*, on (–1, 1):

$$\operatorname{arctg} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + (-1)^n \frac{x^{2n+1}}{2n+1} + \ldots$$

Of course, the list is to be completed by many other expansions depending on the concrete searched problem.

PROBLEMS § II.3.

1. Find the domains of convergence, the limit functions, and the types of convergence (namely point-wise, uniform or almost uniform) of the following sequences of functions:

a) $f_n(x) = arcsin\ n\ x$; b) $g_n(x) = \frac{2}{\pi}\ arctg\ n\ x$;

c) $u_n(x) = \frac{x^n}{1+x^{2n}}$; d) $v_n(x) = \frac{1+x+x^2+...+x^n}{1+x^n}$.

Replace $x \in \mathbb{R}$ by $z \in \mathbb{C}$ in the last two examples, and analyze the same aspects concerning convergence.

Hint. a) Because $f_n : [-1/n, 1/n] \to \mathbb{R}$, the sequence (f_n) makes sense only on $D = \{0\}$, where it reduces to a convergent numerical sequence.

b) $D_c = \mathbb{R}$, and $\lim\limits_{n\to\infty} g_n(x) \overset{p}{=} sign\ x$. The convergence is not almost uniform (hence also not uniform) since the limit is not continuous.

c) $D_c = \mathbb{R} \setminus \{-1\}$, and the limit function is (only point-wise)

$$\lim_{n\to\infty} u_n(x) \overset{p}{=} \begin{cases} 1/2 & \text{if } x = 1 \\ 0 & \text{if } |x| \neq 1 \end{cases}.$$

d) $D_c = \mathbb{R} \setminus \{-1\}$, and the limit is (only point-wise)

$$\lim_{n\to\infty} v_n(x) \overset{p}{=} \begin{cases} (1-x)^{-1} & if \quad |x| < 1 \\ +\infty & if \quad x = 1 \\ x(x-1)^{-1} & if \quad |x| > 1. \end{cases}$$

In the complex case we have $D_c = \{z \in \mathbb{C} : |\ z\ | \neq 1\} \cup \{1\}$, and we shall replace $+\infty \in \overline{\mathbb{R}}$ by $\infty \in \overline{\mathbb{C}}$.

2. Justify the uniform convergence of the following series of real functions (of real variables):

a) $\sum\limits_{n=1}^{\infty} \frac{\cos nx}{n^2}$; b) $\sum\limits_{n=1}^{\infty} \frac{\sin nx}{n^\alpha}$, $\alpha > 1$;

c) $\sum\limits_{n=1}^{\infty} \frac{1}{n} \cos nx$; d) $\sum\limits_{n=1}^{\infty} \frac{1}{\sqrt{n}} \sin nx$;

e) $\sum\limits_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$, where the numerical sequences (a_n) and (b_n) decreasingly tend to 0.

Hint. The problems a) and b) are solved by the Weierstrass' test since the series $\sum 1/n^{\alpha}$ is convergent whenever $\alpha > 1$. The other examples can be studied by the Dirichlet's test, because a finite sum of sines and cosines is bounded, e.g.

$$\left|\sum_{k=1}^{n} \sin kx\right| = \left|\frac{\cos\frac{x}{2} - \cos(n+\frac{1}{2})x}{2\sin\frac{x}{2}}\right| \leq \frac{1}{\left|\sin\frac{x}{2}\right|} .$$

3. Show that the sequence of functions $f_n : D \to \mathbb{R}$, where

$$f_n(x) = x^n - 1 + \frac{n(1-x)}{1+n(1-x)} ,$$

is convergent to 0 in the following manner:

a) point-wise but not almost uniformly if $D = [0, 1]$, and

b) almost uniformly but not uniformly if $D = [0, 1)$.

Analyze the same problem in a complex framework, by taking

$$D = \overline{S}(0,1) \overset{def.}{=} \{z \in \mathbb{C} : |z| \leq 1\} .$$

Hint. Take $x = 1$ and $x \neq 1$ separately. Evaluate $| f_n(x) - 0 |$ at $x = 1 - \frac{1}{n}$,

$$| f_n(x) | = | (1 - \tfrac{1}{n})^n - \tfrac{1}{2}| > \tfrac{1}{2} - \tfrac{1}{e} > 0 .$$

The difference between the two cases rises because [0, 1] is compact, while [0, 1) isn't. At the same time, for any compact set $K \subset [0, 1)$, the nearest point to 1 is $x_0 = \sup K \in K$.

The complex case is similar, but a clear distinction between $| z | = 1$ and $z = 1$ is necessary.

4. Let the functions $f_n : \mathbb{R} \to \mathbb{R}$ be expressed by:

$$f_n(x) = \begin{cases} \frac{1}{n} & if \;\; x \in \mathbb{Q} \\ 0 & if \;\; x \in \mathbb{R} \setminus \mathbb{Q} . \end{cases}$$

Show that the sequence (f_n) of everywhere discontinuous functions is uniformly convergent to a continuous limit (i.e. the u-convergence preserves the *good behavior* of the terms, not the *bad* one!).

Construct a similar example of complex functions.

5. Prove that a uniform limit of a sequence of bounded functions is bounded too. Using the functions $f_n : (0, 1] \to \mathbb{R}$, of values

$$f_n(x) = \min \; \{n, \frac{1}{x}\} ,$$

show that the a.u.-convergence is not strong enough to transport the property of *boundedness* from terms to the limit.

Consider similar functions of a complex variable.

Hint. If $f_n \xrightarrow{u} \varphi$, we may use the inequality $\|\varphi\| \le \|\varphi - f_n\| + \|f_n\|$, where $\|f\| = \sup\{|f(x)| : x \in D\}$. The particularly considered f_n are bounded functions, but the a.u. limit $\varphi(x) = \frac{1}{x}$ is unbounded .

Take into account that $\mathbb{C}$ is not ordered. However, $|z| \in \mathbb{R}_+$, so we may consider $f_n : \{z \in \mathbb{C} : 0 < |z| \le 1\} \to \mathbb{R}$, of values

$$f_n(z) = \min\ \{n, \frac{1}{|z|}\} .$$

6. Derive term by term in the sequence (f_n), where $f_n(x) = \frac{\sin nx}{\sqrt{n}}$, at any $x \in \mathbb{R}$, and $n \ge 1$. Similarly discuss the series $\sum [f_{n+1} - f_n]$.

Hint. $f_n \xrightarrow{u} 0$ on $\mathbb{R}$, but $(f_n^{/})$ is not convergent, neither point-wise. The series is u-convergent to – *sin x*, while the series of derivatives is divergent (1.e. the hypotheses of theorem 3.14 are not fulfilled).

7. For any $n \ge 2$ we define $f_n : [0, 1] \to \mathbb{R}$ by the formula

$$f_n(x) = \begin{cases} n^2 x & \text{if } x \in \left[0, \frac{1}{n}\right) \\ -n^2(x - \frac{2}{n}) & \text{if } x \in \left[\frac{1}{n}, \frac{2}{n}\right) \\ 0 & \text{if } x \in \left[\frac{2}{n}, 1\right] . \end{cases}$$

Show that $f_n \xrightarrow{u} 0$, but $\int_0^1 f_n(x)\,dx = 1$ for all $n \ge 2$.

Hint. $f_n(0) = 0$, and for any $x > 0$ there exists $n \in \mathbb{N}$ such that $x > 2/n$. Disregarding theorem 3.15, the different results

$$\lim_{n\to\infty} \int_0^1 f_n(x)dx = 1 \ne 0 = \int_0^1 (\lim_{n\to\infty} f_n(x))dx$$

are possible because the convergence is not almost uniform.

8. For each $n \in \mathbb{N} \setminus \{0, 1\}$ we note

$$X_n = \{\frac{p}{n} : p \in \mathbb{N}, 0 < p < n, (p, n) = 1\},$$

where (p, n) means the greatest common divisor of p and n. Show that each function $f_n : [0, 1] \to \mathbb{R}$, of values

$$f_n(x) = \begin{cases} 1 & \textit{if } x \in X_n \\ 0 & \textit{otherwise} \end{cases},$$

is integrable on [0, 1] , but the series $\sum f_n$ is point-wise convergent to a non integrable function.

Hint. Each f_n , $n \geq 2$ (and consequently each partial sum s_n of the considered series) has only a finite number of discontinuities at points of the form

$$\underbrace{\frac{1}{2}}_{f_2};\ \underbrace{\frac{1}{3},\frac{2}{3}}_{f_3};\ \underbrace{\frac{1}{4},\frac{3}{4}}_{f_4};\ \underbrace{\frac{1}{5},\frac{2}{5},\frac{3}{5},\frac{4}{5}}_{f_5};\ \dots;\ \underbrace{\frac{1}{n},\dots,\frac{n-1}{n}}_{f_n};\dots$$

where they equal 1, hence these functions are integrable on [0, 1]. The series is point-wise convergent to the function

$$\varphi(x) = \begin{cases} 1 & if \quad x \in \mathbb{Q} \cap (0,1) \\ 0 & otherwise \end{cases} .$$

9. Using adequate Taylor developments evaluate the sums:

a) $1 - \frac{1}{2} + \frac{1}{3} - \dots + \frac{(-1)^{n-1}}{n} + \dots$

b) $1 - \frac{1}{5} + \frac{1}{9} - \dots + \frac{(-1)^n}{4n+1} + \dots$

c) $1 - \frac{1}{3 \cdot 3} + \frac{1}{5 \cdot 3^2} - \frac{1}{7 \cdot 3^3} + \dots + \frac{(-1)^{n-1}}{(2n-1) \cdot 3^{n-1}} + \dots$

d) $3 - \frac{3}{4} + \frac{9}{4^2} - \dots + \frac{1-(-2)^{n+1}}{4^n} + \dots$

Hint. a) The development of $ln(1 + x)$ is convergent at $x = 1$, so take $x \to 1$ in theorem 3.13, applied to example 8 on the list 3.35.

c) Take $x = 1$ in the function $f : (-1, +1] \to \mathbb{R}$, defined by the formula

$$f(x) \underset{(-1,+1]}{\overset{a.u.}{=}} x - \frac{x^5}{5} + \frac{x^9}{9} - \dots + \frac{(-1)^n x^{4n+1}}{4n+1} + \dots$$

According to theorem 3.14, the derivative of f is

$$f'(x) = 1 - x^4 + x^8 - \dots + (-1)^n x^{4n} + \dots = \frac{1}{1+x^4} .$$

Because $f(0) = 0$, it follows that

$$f(x) = \int_0^x \frac{1}{1+t^4} dt .$$

d) Replace $x^2 = \frac{1}{3}$ in

$$f(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \dots + (-1)^n \frac{x^{2n+1}}{2n+1} + \dots$$

e) Use the function $f : (-1/2, 1/2) \to \mathbb{R}$, defined by

$$f(x) = \sum_{n=0}^{\infty} \left[1-(-2)^{n+1}\right] x^n = \sum_{n=0}^{\infty} x^n + 2\sum_{n=0}^{\infty} (-2x)^n = \frac{1}{(1-x)(1+2x)}$$

10. Use adequate Taylor formula to approximate:
a) *sin* 33º to five exact decimals;
b) $e^{-0.01}$ to three decimals;
c) *ln* 2 to two decimals;
d) π to two decimals;
e) $\int_0^1 e^{-t^2} dt$ to four decimals.

Hint. a) Take $x_0 = \pi/6$ and $h = x - x_0 = \pi/60$ in the development of *sin*. Since for $n = 3$ the Lagrange's remainder is maximized to

$$|R_n| = \left|\frac{h^4}{4!}\sin\xi_x\right| \le 10^{-6},$$

the searched approximation is

$$\sin 33^\circ \cong \frac{1}{2} + \frac{\pi}{60}\frac{\sqrt{3}}{2} - \frac{1}{2}\left(\frac{\pi}{60}\right)^2\frac{1}{2} - \frac{1}{6}\left(\frac{\pi}{60}\right)^3\frac{\sqrt{3}}{2} \cong 0.54464.$$

b) According to the example 3.32, the development of *exp* $(-x^{-2})$ is not useful. Alternatively, we may develop e^x around $x_0 = 0$. Because the third term at $x = -0.01$ is 0.00005, and the series is alternate, it follows that the remainder has a smaller value.
c) Use the series of *ln* $(1 + x)$ on the list; since it is slowly convergent, other expansions are recommended.
d) Develop *arctg* x around $x = \pi/6$.
e) Integrate in the development of $exp(-x^2)$

$$f(x) \overset{def}{=} \int_0^x \exp(-t^2)dt \underset{R}{\overset{a.u.}{=}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!}\frac{x^{2n+1}}{2n+1}.$$

Because the eighth term of the alternate series at $x = 1$ in less than the imposed error, we obtain $0.74681 < f(1) < 0.74685$.

11. Find the Mac Laurin developments of the functions:

a) $Si(x) = \int_0^x \frac{\sin t}{t} dt$ (called *integral sine*);

b) $u(x) = \int_0^x \frac{1-\cos t}{t} dt$ (part of the *integral cosine*).

Hint. a) Integrating term by term in the series

$$\frac{\sin t}{t} = 1 - \frac{t^2}{3!} + \ldots + (-1)^p \frac{t^{2p}}{(2p+1)!} + \ldots$$

we obtain

$$Si(x) = \frac{x}{1\cdot 1!} - \frac{x^3}{3\cdot 3!} + \ldots + (-1)^p \frac{x^{2p+1}}{(2p+1)\cdot(2p+1)!} + \ldots$$

b) Similarly, integrating the development of $\frac{1-\cos t}{t}$ leads to

$$u(x) = \frac{x^2}{2\cdot 2!} - \frac{x^4}{4\cdot 4!} + \ldots + (-1)^q \frac{x^{2q}}{2q\cdot(2q)!} + \ldots$$

The *integral cosine* is defined by the improper integral

$$Ci(x) = \int_0^x \frac{\cos t}{t} dt .$$

We mention the relation $Ci(x) = \ln x + \gamma - u(x)$, where γ is the Euler's constant $\gamma = \lim_{n\to\infty}\left(1+\frac{1}{2}+\ldots+\frac{1}{n}-\ln n\right)$.

12. Search the following functions for extreme values:

a) $f(x) = 2x + 3\sqrt[3]{x^2}$; b) $g(x) = \sin x - x$;
c) $u(x) = \cosh 2x - 2x^2$; d) $v(x) = x^5 e^x$.

Hint. a) $f(-1) = 1$ is a local maximum, and $f(0) = 0$ is a local minimum, even $f'(0)$ does not exist. b) g has infinitely many stationary points $x_k = 2k\pi$, $k\in\mathbb{Z}$, but no local extreme. c) We evaluate $u(0) = u'(0) = u''(0) = u'''(0) = 0$ and $u^{(4)}(0) > 0$, hence 0 is a point of minimum. d) $x_1 = -5$ is a local minimum, but $x_2 = 0$ is an inflexion point since $v(0) = \ldots = v^{(4)}(0) = 0$ and $v^{(5)}(0) \neq 0$.

13. Identify the type of convergence of the sequence (f_n), where the functions $f_n : \mathbb{R}\to\mathbb{C}$ take the values

$$f_n(\theta) = \left(1+i\frac{\theta}{n}\right)^n .$$

Hint. According to problem 3.1.8.c, we have

$$\lim_{n\to\infty} f_n(\theta) \overset{p.}{\underset{\mathbb{R}}{=}} \cos\theta + i\sin\theta .$$

The uniform convergence on $[0, 2\pi]$ and the periodicity of *sin* and *cos* do not assure the uniform convergence of (f_n). The analysis of the mentioned problem shows that the convergence is almost uniform.

§ II.4. POWER SERIES

It is easy to remark that the partial sums in Taylor series are polynomials, whose coefficients are determined by the developed function. The *power series* generalize this feature, i.e. they are series of monomials. The main idea is to reverse the roles: Taylor series are attached to a given function, while power series are used to define functions.

In the first part of this section we study real power series, which provide a direct connection with the well-known differential and integral calculus involving real functions of a real variable. In the second part we extend this study to complex power series, which open the way to a complex analysis. Finally, we use the method of power series to introduce several elementary complex functions of a complex variable.

4.1. <u>Definition</u>. Let (a_n) be a sequence of real numbers, and let $x_0 \in \mathbb{R}$ be fixed. The functions series $\left(a_n (x - x_0)^n ; \sum_{k=1}^{n} a_k (x - x_0)^k \right)$, briefly noted

$$\sum a_n (x - x_0)^n , \tag{1}$$

is called (real, since x, $a_n \in \mathbb{R}$) *power series*. The point x_0 is the *center*, and the numbers a_n are the *coefficients* of the series. The series (1) is said to be *centered* at x_0 .

Obviously, the entire information about function series, contained in the previous section, remains valid for power series. In particular, the terms of a power series, namely the monomial functions $f_n(x) = a_n (x - x_0)^n$, are defined for all $n \in \mathbb{N}$, and at all $x \in \mathbb{R}$. However, the domain of convergence generally differs from $\mathbb{R}$, as we cam see in several particular cases.

4.2. <u>Examples</u>. a) If $a_n = 1$ for all $n \in \mathbb{N}$, then (1) becomes $\sum (x - x_0)^n$, and we call it *geometric series* of ratio $q = x - x_0$. Because

$$s_n(x) = \sum_{k=0}^{n} (x - x_0)^k = \frac{1 - q^{n+1}}{1 - q} ,$$

and $\lim_{n \to \infty} q^n = 0$ if and only if $|q| < 1$, it follows that the geometric series is convergent exactly in the interval $I = (x_0 - 1, x_0 + 1)$. The divergence in the case $|q| \geq 1$ is based on the fact that the general term of a convergent series necessarily tends to zero.

To conclude, the domain of convergence in the case of a geometric series is the interval I , centered at x_0, of radius 1.

b) The Taylor series of the exponential function, i.e. $\sum_{n=0}^{\infty} \frac{1}{n!} x^n = e^x$, is a power series of coefficients $a_n = \frac{1}{n!}$, which is convergent at any $x \in \mathbb{R}$. We may interpret $\mathbb{R}$ like an interval centered at x_0, of infinite radius.

c) The power series $\sum n^n (x - x_0)^n$ is convergent only at $x = x_0$. In fact, if $x \neq x_0$, then we can find some $n_0(x) \in \mathbb{N}$, such that $n|x - x_0| > 1$ holds for all $n > n_0(x)$, hence the general term of the series doesn't tend to zero at this point. Consequently, the set of convergence reduces to $\{x_0\}$, which can also be viewed as an interval centered at x_0, of null radius.

These examples lead us to the conjecture that the set of convergence of any real power series is an interval, including $\mathbb{R}$ and $\{x_0\}$ in this notion. To prove the validity of this supposition in the most general case, let us note:

$$\omega = \overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} \,, \tag{2}$$

and

$$R = \begin{cases} 1/\omega & if \;\; \omega \in \mathbb{R}_+^* \\ \infty & if \;\; \omega = 0 \\ 0 & if \;\; \omega = \infty \end{cases} . \tag{3}$$

The following theorem explains why R is called *radius of convergence*.

4.3. <u>Theorem</u>. (Cauchy – Hadamard) Each power series (1) is absolutely and almost uniformly convergent in the interval $I = (x_0 - R, x_0 + R)$, and divergent outside of its closure $\overline{I}$.

<u>Proof</u>. First we remind that (2) concentrates the following two conditions:

(I) $\forall_{\varepsilon > 0} \; \exists_{n_0 \in \mathbb{N}} \;$ *such that* $[(\forall) n > n_0 \Rightarrow \sqrt[n]{|a_n|} < \omega + \varepsilon]$, and

(II) $\forall_{\varepsilon > 0} \; \exists_{m \in \mathbb{N}} \;$ *such that* $\sqrt[m]{|a_m|} > \omega - \varepsilon$.

<u>Case</u> a) $0 < \omega < \infty$. From (I) we immediately deduce that

$$\left| a_n (x - x_0)^n \right| < (\omega + \varepsilon)^n |x - x_0|^n \overset{not.}{=} q^n$$

holds for all $n > n_0$. If

$$|x - x_0| < \frac{1}{\omega + \varepsilon} \,,$$

then $q < 1$, hence the general term of the given series is less than the term q^n of a convergent geometric series. Using the comparison test II.2.12, it follows that (1) is absolutely and a.u. convergent in the interval

$$I_\varepsilon = (x_0 - \frac{1}{\omega + \varepsilon}, x_0 + \frac{1}{\omega + \varepsilon}) .$$

Because $\varepsilon > 0$ is arbitrary, this convergence holds in I too.

To prove the assertion concerning divergence, we use condition (II). Let us take $\varepsilon > 0$ such that $\omega - \varepsilon > 0$, and $m \in \mathbb{N}$ like in (II), so that

$$\left|a_m (x - x_0)^m\right| > (\omega - \varepsilon)^m \left|x - x_0\right|^m \overset{not.}{=} p^m .$$

Consequently, if $\left|x - x_0\right| > \dfrac{1}{\omega - \varepsilon}$, then the geometric series of ratio p is divergent, and according to the same comparison test, so is the given power series. Taking into account that ε is arbitrary, divergence holds whenever $\left|x - x_0\right| \geq 1/\omega = R$.

<u>Case</u> b) $\omega = 0$. Using (I), to each $\varepsilon > 0$ and $x \in \mathbb{R}$ there corresponds a rank $n_0(\varepsilon, x) \in \mathbb{N}$, such that $\left|a_n (x - x_0)^n\right| < \varepsilon^n \left|x - x_0\right|^n$ holds for all $n > n_0(\varepsilon, x)$. At any $x \in \mathbb{R}$ we may take $\varepsilon > 0$ such that

$$\varepsilon \left|x - x_0\right| \overset{not.}{=} q < 1,$$

hence the geometric series of ratio q is convergent. As before, it remains to use the comparison test.

Divergence is impossible, since (II) is trivial at $\omega = 0$.

<u>Case</u> c) $\omega = \infty$. Instead of (I) and (II), we express $\omega = \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \infty$ by

(III) $\quad \underset{M>0}{\forall} \quad \underset{m\in\mathbb{N}}{\exists} \quad$ *such that* $\sqrt[m]{|a_m|} > M$.

Consequently, $\left|a_m (x - x_0)^m\right| > M^m \left|x - x_0\right|^m$ is possible for arbitrary $M > 0$, which shows that series (1) is divergent at any $x \neq x_0$. ◇

4.4. <u>Remarks</u>. a) Apart from its theoretical significance, we have to take the superior limit in (2) whenever infinitely many coefficients of the series are vanishing. For example, the series $\sum_{n=1}^{\infty} \frac{1}{n} x^{2n}$ has the coefficients

$$a_p = \begin{cases} 1/n & \textit{if} \quad p = 2n \\ 0 & \textit{if} \quad p = 2n + 1 \ . \end{cases}$$

Because $\lim_{n\to\infty} \sqrt[n]{n} = 1$, we have $R = 1$, but $\lim_{n\to\infty} \sqrt[n]{|a_n|}$ does not exist.

b) In the case $0 < \omega < \infty$, theorem 4.3. solves the problem of convergence at any $x \in \mathbb{R}$, except the endpoints of I, say $x_1 = x_0 - R$, and $x_2 = x_0 + R$. To get the complete answer of the convergence problem, it remains to study the two numerical series: $\sum_{n=0}^{\infty} a_n x_1^n$ and $\sum_{n=0}^{\infty} a_n x_2^n$.

This behavior at the endpoints x_1 and x_2 is unpredictable, i.e. all the combinations convergence / divergence are possible. For example,

- Divergence at both x_1 and x_2 holds for the geometric series $\sum_{n=0}^{\infty} x^n$;
- Convergence at x_1 and divergence at x_2 hold for $\sum_{n=1}^{\infty} \frac{1}{n} x^n$;
- Divergence at x_1 and convergence at x_2 hold for $\sum_{n=1}^{\infty} \frac{(-1)^n}{n} x^n$;
- Convergence at both x_1 and x_2 holds for the series $\sum_{n=1}^{\infty} \frac{1}{n^2} x^n$.

c) Instead of (2), we may evaluate ω using the D'Alembert formula

$$\omega = \overline{\lim_{n\to\infty}} \left| \frac{a_{n+1}}{a_n} \right| , \tag{4}$$

because the existence of the limit in (4) assures the existence of that in (2), and these limits are equal. Formula (4) is sometimes very useful in practice, especially when the evaluation of the limit in (2) is difficult. As for example, we may compare the efficiency of these formulas in the case of the exponential series

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n .$$

The theory of the real power series gas several "weak points", i.e. there exist some phenomena that cannot be explained within the frame of real variables. The simplest ones concern the difference between the domain of existence (convergence) of the series and that of the sum function.

4.5. <u>Examples</u>. a) The geometric series $\sum_{n=0}^{\infty} x^{2n} = \frac{1}{1-x^2}$ has the radius of convergence $R = 1$. This is perfectly explained by the fact that it represents the Taylor series of the function $f(x) = \frac{1}{1-x^2}$, which is not defined at ± 1. However, f makes sense outside of (– 1, + 1), where the series diverges.

b) The alternating geometric series $\sum_{n=0}^{\infty} (-1)^n x^{2n} = \frac{1}{1+x^2}$ has the same domain of convergence as before, while its sum $g(x) = \frac{1}{1+x^2}$ is defined on the whole $\mathbb{R}$. In addition, the restriction of the convergence to (– 1, + 1) is no longer explained since g has no singularities. Drawing the graphs of f and g is recommended to visualize the situation.

The explanation is immediate in complex variables, since the complex function $h(z) = \dfrac{1}{1+z^2}$ cannot be defined at $\pm i$, and $|\pm i| = 1 = R$ too.

In the second part of this section we'll see that the study of the complex power series offers such explanations, and in addition it is similar to the real case in many respects (pay attention to differences!).

4.6. <u>Definition</u>. Let (a_n) be a sequence of complex numbers, and let us fix $z_0 \in \mathbb{C}$. The function series $\left(a_n (z - z_0), \sum_{k=0}^{n} a_k (z - z_0)^k \right)$, briefly noted

$$\sum a_n (z - z_0)^n , \tag{5}$$

is called *complex power series* centered at z_0.

The coefficients a_n, as well as the center z_0, may be real numbers, but the series (5) is essentially *complex* since $z \in \mathbb{C}$. To get a quick view of some domains of convergence we may consider several particular cases.

4.7. <u>Examples</u>. a) The *complex geometric* series $\sum z^n$ is convergent in the open unit disc (centered at zero), $D(0,1) = \{z \in \mathbb{C} : |z| < 1\}$.

b) The "exponential like" series $\sum \frac{1}{n!} z^n$ is absolutely and a.u. convergent on the entire complex plane (since $|z| \in \mathbb{R}$!).

c) The series $\sum n^n (z - z_0)^n$ is convergent only at $z = z_0$.

The general result concerning the domain of convergence of a complex series makes use of the following simple fact:

4.8. <u>Lemma</u>. If the power series (5) is convergent at some $z_1 \in \mathbb{C}$, then it is absolutely convergent at any other $z \in \mathbb{C}$, which is *closer* to z_0 than z_1, i.e.

$$|z - z_0| < |z_1 - z_0| .$$

<u>Proof</u>. If $\sum a_n (z_1 - z_0)^n$ is convergent, then $\lim\limits_{n \to \infty} a_n (z_1 - z_0)^n = 0$, hence $\{|a_n (z_1 - z_0)^n| : n \in \mathbb{N}\}$ is a bounded set of real numbers. Because we may reformulate the hypothesis $|z - z_0| < |z_1 - z_0|$ by $\left| \dfrac{z - z_0}{z_1 - z_0} \right| \overset{not.}{=} q < 1$, we have $|a_n (z - z_0)^n| = |a_n (z_1 - z_0)^n| \cdot \left| \dfrac{z - z_0}{z_1 - z_0} \right|^n \leq M q^n$, where M satisfies the only condition $M \geq \sup \{|a_n (z_1 - z_0)^n| : n \in \mathbb{N}\}$. The comparison of the given series to the geometric series of (positive) ratio $q < 1$ shows that (5) is absolutely convergent at z. ◇

By analogy to $\mathbb{R}$, if we accept the entire $\mathbb{C}$ and the singletons $\{z_0\}$ to be called *discs* too, then the problem concerning the form of the domain of convergence of a complex power series is simply solved by the following:

4.9. <u>Proposition</u>. Every complex power series (5) is absolutely and a.u. convergent in a disc centered at z_0.

<u>Proof</u>. First of all, let us identify the two extreme situations, namely:

a) The series (5) is convergent at each point of a sequence $(z_n)_{n\in\mathbb{N}^*}$, where $z_n \in \mathbb{C}$ and $\lim\limits_{n\to\infty}|z_n - z_0| = \infty$. According to the above lemma, the series (5) is absolutely and a.u. convergent on $\mathbb{C}$.

b) (5) is divergent at any point $z_n \neq z_0$ of a sequence $(z_n)_{n\in\mathbb{N}^*}$, where $z_n \in \mathbb{C}$ and $z_n \to z_0$. According to the same lemma, z_0 is the unique point of convergence.

In the remaining cases, there exist $z_1, \zeta_1 \in \mathbb{C}$ such that (5) is convergent at z_1 and divergent at ζ_1. Using lemma 4.8. again, we establish the behavior of the series in the interior of the circle

$$C_{conv.}(z_0, |z_1 - z_0|) = \{z \in \mathbb{C} : |z - z_0| = |z_1 - z_0|\},$$

where it is convergent, and in the exterior of the circle

$$C_{div.}(z_0, |\varsigma_1 - z_0|) = \{z \in \mathbb{C} : |z - z_0| = |\varsigma_1 - z_0|\},$$

where it diverges. To find out the nature of the series (5) in the remaining cases, we consider "testing points" in the annulus

$$A(z_0, |z_1 - z_0|, |\varsigma_1 - z_0|) = \{z \in \mathbb{C} : |z_1 - z_0| < |z - z_0| < |\varsigma_1 - z_0|\}.$$

If the series converges at some $z \in A$, we note it z_2, and we increase the disc of convergence. By contrary, if the series diverges at this point, we note it ζ_2, and we use it to decrease the radius of the circle $C_{div.}$ (as sketched in Fig.II.4.1. below).

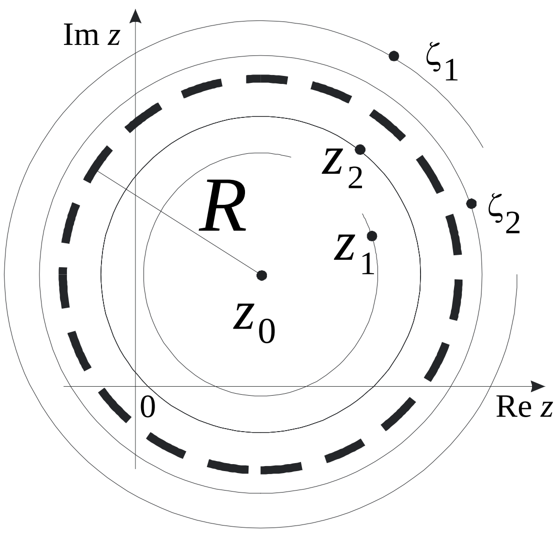


Fig. II.4.1.

By repeating this construction, we obtain an increasing sequence of real numbers ($|z_n - z_0|$), and a decreasing sequence ($|\zeta_n - z_0|$). In addition, we can chose them such that $\lim_{n\to\infty}\left(|\varsigma_n - z_0| - |z_n - z_0|\right) = 0$, hence the number

$$R \overset{not.}{=} \lim_{n\to\infty}|z_n - z_0| = \lim_{n\to\infty}|\varsigma_n - z_0|$$

is uniquely determined according to Cantor's theorem (e.g. I.2.17). It is easy to see that (5) is absolutely and a.u. convergent in the open disc

$$D(z_0,R) = \{z \in \mathbb{C} : |z - z_0| < R\},$$

and divergent in its exterior $\complement\,\overline{D}(z_0,R)$. ◇

Like in $\mathbb{R}$, the above number R is called *radius of convergence*, and it can be evaluated by formulas similar to (2), (3), and (4). More exactly:

4.10.Theorem. (Cauchy – Hadamard) If (5) is a complex power series, for which we note

$$\omega = \overline{\lim_{n\to\infty}}\sqrt[n]{|a_n|}\ , \tag{2'}$$

then its radius of convergence has the value

$$R = \begin{cases} 1/\omega & if \quad \omega \in \mathbb{R}_+^* \\ \infty & if \quad \omega = 0 \\ 0 & if \quad \omega = \infty \end{cases} . \tag{3'}$$

The proof is similar to that of theorem 4.3, and will be omitted. ◇

4.11.Remarks. a) The intervals of convergence $I = (x_0 - R,\ x_0 + R)$, in the case of real power series, also represent discs, in the sense of the intrinsic metric of $\mathbb{R}$. Consequently, theorem 4.10 extends theorem 4.3 in the same way as the norm of $\mathbb{C}$ extends the norm of $\mathbb{R}$. In particular, if in (5) we have $a_n \in \mathbb{R}$, and $z_0 = x_0 \in \mathbb{R}$, then the interval of convergence for the resulting real power series is a "trace on $\mathbb{R}$" of the plane disc of convergence, i.e.

$$I(x_0,R) = D(x_0,R) \cap \mathbb{R}\ .$$

b) Following D'Alembert, instead of (2') we may use the formula

$$\omega = \overline{\lim_{n\to\infty}}\left|\frac{a_{n+1}}{a_n}\right| , \tag{4'}$$

to evaluate the radius of convergence. The only difference between (4) and (4') concerns the domain of the modulus, which is $\mathbb{R}$ in (4), and $\mathbb{C}$ in (4').

c) Theorem 4.10 gives no information about the nature of (5) on the circle

$$C(z_0,R) = \{z \in \mathbb{C} : |z - z_0| = R\} = Fr\,D(z_0,R)\,.$$

Simple examples put forward a large variety of situations between the two extreme cases:

- Divergence overall $C(z_0, R)$, as for the geometric series, and

- Convergence on the entire $C(z_0, R)$, as for the series $\sum_{n=1}^{\infty} \frac{1}{n^2} z^n$.

To obtain more information, we need additional hypotheses, e.g.:

4.12. <u>Theorem</u>. (Abel) Let $\sum_{n=0}^{\infty} a_n z^n$ be a complex power series with real coefficients, which has the radius of convergence R, such that $0 < R < \infty$. If

$$a_0 > a_1 R > a_2 R^2 > \ldots > a_n R^n > \ldots \to 0, \tag{6}$$

then the series is (point wise) convergent at any $z \in C(0, R) \setminus \{R\}$.

<u>Proof</u>. Let us first remember another Abel's theorem (see II.2.6 in complex form): "If $\sum z_n$ is a numerical series with bounded partial sums, and (ε_n) is a decreasing sequence of positive (real) numbers, which converges to 0, then $\sum \varepsilon_n z_n$ is a convergent series." Now, for any $n \in \mathbb{N}$ we may write

$$a_n z^n = a_n R^n \left(\frac{z}{R} \right)^n = \varepsilon_n z_n ,$$

where $\varepsilon_n = a_n R^n$ and $z_n = \left(\frac{z}{R} \right)^n$ fulfill the conditions of the cited theorem. In fact, if $|z| = R$, but $z \neq R$, then the partial sums

$$s_n(z) = 1 + \frac{z}{R} + \ldots + \left(\frac{z}{R} \right)^n = \frac{1 - \left(z/R \right)^{n+1}}{1 - \left(z/R \right)}$$

are equally bounded, more exactly, for any $n \in \mathbb{N}$ we have

$$|s_n(z)| \le \frac{2}{\left| 1 - \left(z/R \right) \right|} .$$

Consequently, except $z = R$, the gives series is convergent on $C(0, R)$. ◇

4.13. <u>Remarks</u>. a) Convergence at $z = R$ is neither affirmed nor denied by theorem 4.12, so it remains to be studied separately. On this way we may complete the answer concerning the behavior of the series on the whole $\mathbb{C}$.

b) Taking $z_0 = 0$ in theorem 4.12 is not essential. More generally, we can reduce every qualitative problem concerning power series, including convergence, to the case, via the translation of z_0 to 0.

c) We may use theorem 4.12 to identify more points of divergence on the frontier of the disc $D(z_0, R)$. For example, if we replace $z = \zeta^k$, where $k \in \mathbb{N}^*$, in the power series $\sum_{n=1}^{\infty} \frac{1}{n} z^n$, then we find out that the series $\sum_{n=1}^{\infty} \frac{1}{n} \varsigma^{kn}$ is divergent at the k^{th} roots of 1.

4.14. **<u>Classes of Functions</u>.** It is essential for us to distinguish between the senses in which we may speak of functions and classes of functions. An abstract class of functions, e.g. continuous, derivable, etc., is defined by a specific property, while the concrete employment of a particular class of functions asks an effective knowledge of the values. For example, we can uniquely define a function by specifying its derivative $t^{\alpha} e^{-t}$, its value at 0, and $\alpha > -1$, but whenever we want to use this function in practice, we need its values at different points. For this reason mathematicians have written plenty of books containing *tables of values* of some particular functions, as for example tables of logarithms, sine, cosine, Bessel functions, etc.

During the former stage of the study, we deal with the class of *elementary functions*, which is generated by the *algebraic functions* and function *exp.* Any non elementary function is said to be *special*. More exactly, function f is called *algebraic* iff to obtain its values $f(x)$ we have to perform a finite number of algebraic operations (sum, product, difference, quotient, power, root). If the calculation of the values $f(x)$ involves infinitely many algebraic operations, and consequently some limit processes (like in series!), then f is named *transcendental* function. In particular, the exponential is the only transcendental function in the class of elementary functions, since its values are obtained by summing up the series

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} \overset{def.}{=} \exp(x) = e^x .$$

To be more specific, we mention that the assertion "class X is generated by the functions $f, g, \ldots$ " means that besides $f, g, \ldots$, this class contains restrictions, compositions, inverses and algebraic operations with them. For example, due to the Euler's formulas

$$\cos x = \frac{e^{ix} + e^{-ix}}{2} \ , \ \sin x = \frac{e^{ix} - e^{-ix}}{2i} ,$$

it follows that the trigonometric functions are elementary. This simple case already shows that the real framework is not sufficient to study functions, since the trigonometric functions are expressed by complex exponentials.

The initial way of learning functions is usually called *geometric*, because of the strong connections to trigonometry, graphs, etc. Whenever we think of a function as a set of numerical values, we need some rule of computing these values. Most frequently, the method of introducing the functions by this computation is called *analytic*. Simple cases of analytical definitions of some functions, known from lyceum, involve the primitives that cannot be expressed by elementary functions (e.g. $\int_0^x e^{-t^2}\, dt$, $\int_0^{\pi/4} \ln(1+\tan x) dx$, etc.). Generally speaking, a function is known if we can approximate its values. Because we naturally prefer to approximate by polynomials, the analytic method reduces to define functions by power series.

To conclude this analysis, we have to define the elementary functions in complex variables, using the analytic method, i.e. by power series.

4.15. <u>Definition</u>. The *complex exponential* function is defined by

$$e^z \overset{not.}{=} \exp z \overset{def.}{=} \sum_{n=0}^{\infty} \frac{1}{n!} z^n . \qquad (7)$$

We similarly define (compare to II.3.35 in $\mathbb{R}$) the *circular* and *hyperbolic* complex trigonometric functions *sine* and *cosine* by the power series:

$$\sin z \overset{def.}{=} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1} ;$$

$$\cos z \overset{def.}{=} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n} ;$$

$$\sinh z \overset{def.}{=} \sum_{n=0}^{\infty} \frac{1}{(2n+1)!} z^{2n+1} ;$$

$$\cosh z \overset{def.}{=} \sum_{n=0}^{\infty} \frac{1}{(2n)!} z^{2n} .$$

To give a model of how to study the complex elementary functions, we'll analyze the exponential in more details. In particular we'll see that also the complex trigonometric functions can be expressed by the exponential.

4.16. <u>Theorem</u>. Function *exp* has the following properties:

1. Its domain of definition is $\mathbb{C}$;
2. $e^{z_1} \cdot e^{z_2} = e^{z_1+z_2}$ at all $z_1, z_2 \in \mathbb{C}$ (fundamental algebraic property);
3. $e^{iz} = \cos z + i \sin z$ at all $z \in \mathbb{C}$;
4. $\sin z = \dfrac{e^{iz} - e^{-iz}}{2i}$ and $\cos z = \dfrac{e^{iz} + e^{-iz}}{2}$ at all $z \in \mathbb{C}$ (Euler);
5. $e^{x+iy} = e^x (\cos y + i \sin y)$ at all $z = x + iy \in \mathbb{C}$;
6. $\left| e^z \right| = e^x$ and $\arg e^z = y + 2k\pi$, where $z = x + iy \in \mathbb{C}$ and $k \in \mathbb{Z}$;
7. $T = 2\pi i$ is a period of the function *exp*.

<u>Proof</u>. 1) The domain of definition for *exp* is the disc of convergence of the power series (7). Using (4'), we easily obtain $R = \infty$, hence (7) converges on the whole $\mathbb{C}$.

2) We have to multiply the series of e^{z_1} and e^{z_2}, which are

$$e^{z_1} = 1 + \frac{z_1}{1!} + \frac{z_1^2}{2!} + \ldots + \frac{z_1^n}{n!} + \ldots \text{ and}$$

$$e^{z_2} = 1 + \frac{z_2}{1!} + \frac{z_2^2}{2!} + \ldots + \frac{z_2^n}{n!} + \ldots .$$

According to the Cauchy's rule, the product power series has the terms $\varsigma_0 = 1$, $\varsigma_1 = \dfrac{z_1 + z_2}{1!}$, $\varsigma_2 = \dfrac{(z_1 + z_2)^2}{2!}$, and by induction, for all $n \in \mathbb{N}$,

$$\varsigma_n = \sum_{k=0}^{n} \frac{z_1^k}{k!} \frac{z_2^{n-k}}{(n-k)!} = \frac{(z_1 + z_2)^n}{n!} .$$

3) We may replace $i^{4k} = 1$, $i^{4k+1} = i$, $i^{4k+2} = -1$, and $i^{4k+3} = -i$, where $k \in \mathbb{N}$, in the power series of e^{iz}.

4) Add and subtract the previously established relations

$$e^{iz} = \cos z + i \sin z, \text{ and}$$
$$e^{-iz} = \cos z - i \sin z .$$

5) Combining properties 2 and 3 from above, we obtain the asked relation

$$e^{x+iy} = e^x \cdot e^{iy} = e^x (\cos y + i \sin y) .$$

6) Interpret the formula from 5 as trigonometric form of $Z = e^z$.

7) The functions *sin* and *cos* in 5 have the period 2π. ◇

4.17. Geometric Interpretation. The complex exponential is a complex function of one complex variable, hence its graph is a part of $\mathbb{C} \times \mathbb{C} \cong \mathbb{R}^4$. Because we cannot draw the subsets of $\mathbb{R}^4$, the method of visualize the properties of a function on its graph, so useful for real functions, now is not helpful any more. However, we can give geometric interpretations to the properties of the complex functions like *exp*, if we conceive these functions as transformations of the complex plane into itself. Fig. II.4.2 illustrates this method in the case of the complex exponential.

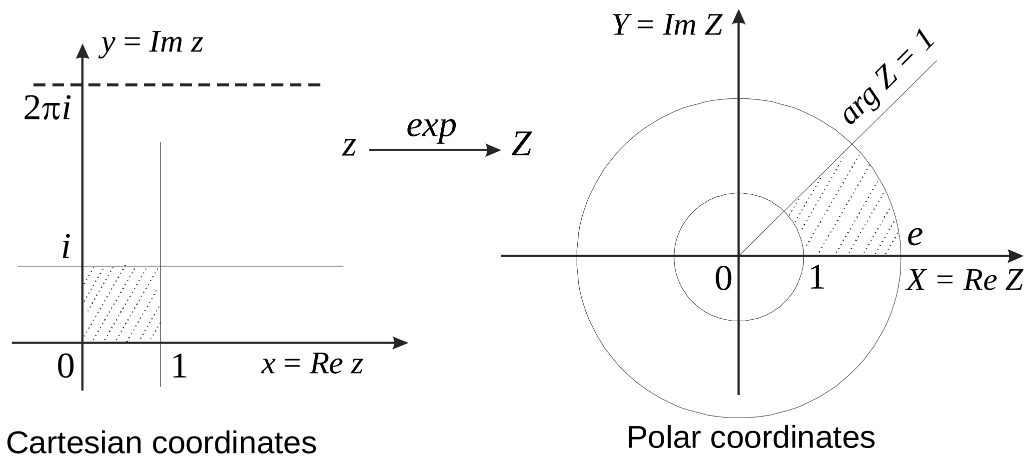


Fig. II.4.2

In fact, according to the property 6 in theorem 4.16, the correspondence $\mathbb{C} \ni z \mapsto Z = e^z \in \mathbb{C}$ takes the real form $\mathbb{R}^2 \ni (x, y) \mapsto (X, Y) \in \mathbb{R}^2$, where

$$\begin{cases} |Z| = \sqrt{X^2 + Y^2} = e^x \\ \arg Z = y + 2k\pi,\ k \in \mathbb{Z}. \end{cases}$$

The property of periodicity shows that the band

$$\{(x,y) \in \mathbb{R}^2 : y \in [0,2\pi)\} = \mathbb{R} \times [0,2\pi)$$

is bijectively carried onto $\mathbb{C} \setminus \{0\}$. The other bands, which are parallel to this one and have the same breadth 2π, have the same image.

Another remarkable property is $e^z \neq 0$. It is based on the inequality

$$|e^z| = e^x > 0,$$

which holds at all $z = x + iy \in \mathbb{C}$, respectively at all $x \in \mathbb{R}$.

The inverse of *exp* is defined as usually, by reversing the correspondence:

4.18. <u>Definition</u>. The inverse of the complex exponential is called *complex logarithm*, and we note it *Ln*. More exactly, if $Z = e^z$, then $z = Ln\ Z$.

The main properties of *Ln* naturally extend those of *ln*.

4.19. <u>Theorem</u>. The complex logarithm has the properties:

1. The domain of definition is $\mathbb{C} \setminus \{0\}$;
2. *Ln* is a multi-valued function (of type *one to many*), i.e.
$$Ln\, Z = \{\ln|Z| + i(\arg Z + 2k\pi) : k \in \mathbb{Z}\} \tag{8}$$
3. $Ln\ Z_1 Z_2 = Ln\ Z_1 + Ln\ Z_2$ for any $Z_1, Z_2 \in \mathbb{C} \setminus \{0\}$.

<u>Proof</u>. 1. The domain of *Ln* is the image of *exp*.

2. *Ln* is multi-valued because *exp* is periodical. If we note $z = x + iy = Ln\ Z$, then from the relation $Z = e^{x+iy} = e^x(\cos y + i \sin y)$ we deduce that

$$|Z| = e^x \text{ and } \arg Z = y + 2k\pi$$

for some $k \in \mathbb{Z}$. Consequently, $x = \ln|Z|$ and $y = \arg Z + 2k\pi$.

3. The equality refers to sets. If $z_1 \in Ln\, Z_1$ and $z_2 \in Ln\, Z_2$, then $Z_1 = e^{z_1}$ and $Z_2 = e^{z_2}$. According to the fundamental property of the exponential, we have $Z_1 Z_2 = e^{z_1 + z_2}$, hence $z_1 + z_2 \in Ln\, Z_1 Z_2$.

Conversely, if $z \in Ln\, Z_1 Z_2$, then $Z_1 Z_2 = e^z$. Let us take $z_1 \in Ln\, Z_1$, and note $z_2 = z - z_1$. Since $Z_1 = e^{z_1} \neq 0$, relation $Z_1 Z_2 = e^{z_1 + z_2} = e^{z_1} e^{z_2}$ gives $Z_2 = e^{z_2}$. Consequently, $z_2 \in Ln\, Z_2$ and $z \in Ln\, Z_1 + Ln Z_2$. ◇

4.19. <u>Remarks</u>. a) Theorem 4.18 suggests we better to write $z \in Ln\, Z$ than $z = Ln\, Z$, which remains specific to the real logarithm. According to (8), the set *Ln Z* is infinite, but countable.

b) Using the complex logarithm we can define the *complex power*

$$Z^{\varsigma} = e^{\varsigma\, Ln\, Z} \tag{9}$$

where $\varsigma \in \mathbb{C}$ and $Z \in \mathbb{C} \setminus \{0\}$. Obviously, this function is multi-valued too. In particular, if $\varsigma = \frac{1}{n}$ for some $n \in \mathbb{N} \setminus \{0, 1\}$, then the complex power $Z^{\frac{1}{n}}$ reduces to the n^{th} root, which is a set of n numbers. More exactly,

$$Z^{\frac{1}{n}} = e^{\frac{1}{n} Ln\, Z} = e^{\frac{1}{n}[\ln|Z| + i(\arg Z + 2k\pi)]} = e^{\frac{1}{n}[\ln|Z| + i \arg Z]} e^{\frac{2k\pi}{n} i}, \; k \in \mathbb{Z}.$$

Because $2\pi i$ is a period of the exponential function, it follows that $e^{\frac{2k\pi}{n} i}$ takes only n distinct values, which correspond to $k = 0, 1, \ldots, n - 1$.

c) A similar study of the complex trigonometric functions is left to the reader. We just mention that the formulas of the real trigonometry remain valid. In addition, it's useful to retain the formulas

$$\sin iz = i \sinh z; \qquad \sinh iz = i \sin z;$$
$$\cos iz = \cosh z; \qquad \cosh iz = \cos z,$$

which connect the circular and hyperbolic trigonometric functions.

Of course, not all properties of the real trigonometric functions are valid in complex variables, e.g. *sin* and *cos* are not bounded any longer.

A typical problem about numerical series, which can be solved by means of power series, concerns the evaluation of the sum. The following example refers to real series, but later on we'll see that similar techniques, based on the operations of derivation and integration of the terms, remains valid in the complex framework.

4.20. <u>Application</u>. Evaluate the sum of the alternating harmonic series

$$s \overset{not.}{=} \sum_{n=0}^{\infty} \frac{(-1)^n}{n+1} = 1 - \frac{1}{2} + \frac{1}{3} - \ldots + \frac{(-1)^n}{n+1} + \ldots$$

The solution is based on the fact that s is the particular value, at $x_0 = 1$, of the power series

$$f(x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}.$$

Deriving f, we obtain the geometric series

$$f'(x) = \sum_{n=0}^{\infty} (-1)^n x^n = \frac{1}{1+x}.$$

By integrating f', we find $f(x) = \ln(1+x) + C$, where $C = f(0) - \ln 1 = 0$. So we may conclude that $s = f(1) = \ln 2$.

PROBLEMS §II.4.

1. Find the radius of convergence and the sum of the following real power series:

a) $\sum_{n=0}^{\infty}\left(\frac{x}{\alpha}\right)^n$, $\alpha>0$; b) $\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{2n+1}$; c) $\sum_{n=0}^{\infty}(-1)^n\frac{x^{3n+1}}{3n+1}$,

then deduce the sum of the numerical series:

a') $\sum_{n=0}^{\infty}\frac{n}{\alpha^n}$, α > 1 ; b') $\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}$; c') $\sum_{n=0}^{\infty}\frac{(-1)^n}{3n+1}$.

Hint. a) The geometric power series converges at $|x|<\alpha$, hence the radius of convergence is $R=\alpha$. The sum is

$$f(x)=\sum_{n=0}^{\infty}\left(\frac{x}{\alpha}\right)^n=\frac{\alpha}{\alpha-x} .$$

Deriving term by term in this series we obtain

$$f^{/}(x)=\sum_{n=0}^{\infty}\frac{n}{\alpha}\left(\frac{x}{\alpha}\right)^{n-1}=\frac{\alpha}{(\alpha-x)^2} .$$

In particular, if $x = 1 < \alpha$, we find the answer to a'), which is:

$$\sum_{n=0}^{\infty}\frac{n}{\alpha^n}=f^{/}(1)=\frac{\alpha}{(\alpha-1)^2} .$$

b) $R = 1$. To find the sum

$$g(x)=\sum_{n=0}^{\infty}(-1)^n\frac{x^{2n+1}}{2n+1} ,$$

we may first evaluate the derivative

$$g^{/}(x)=\sum_{n=0}^{\infty}(-1)^n x^{2n}=\frac{1}{1+x^2} ,$$

then integrate, so that $g(x)=\arctan x+C$. Using the value at $x = 0$ we can identify $C=g(0)-\arctan 0=0$, and at $x = 1$ we find the answer to b'):

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{2n+1}=g(1)=\arctan 1=\frac{\pi}{4} .$$

c) $R = 1$. Deriving the function $h(x) = \sum_{n=0}^{\infty}(-1)^n\frac{x^{3n+1}}{3n+1}$, we obtain

$$h^{/}(x)=\sum_{n=0}^{\infty}(-1)^n x^{3n}=\frac{1}{1+x^3}=\frac{1}{3}\,\frac{1}{1+x}+\frac{1}{3}\,\frac{2-x}{1-x+x^2} .$$

By integration we go back to

$$h(x)=\frac{1}{3}\ln(1+x)-\frac{1}{6}\ln(x^2-x+1)+\frac{1}{2}\arctan\frac{2x-1}{\sqrt{3}}+C,$$

where $C=\frac{1}{2}\arctan\frac{1}{\sqrt{3}}=\frac{\pi}{12}$ follows by taking $x = 0$. At $x = 1$ we obtain

$$\sum_{n=0}^{\infty}\frac{(-1)^n}{3n+1}h(1)=\frac{1}{3}\ln 2+\frac{\pi}{6}\ .$$

2. Find the discs of convergence and study the convergence at the frontier points of these discs for the following complex power series:

a) $\sum_{n=0}^{\infty}\left(\frac{z}{\alpha}\right)^n,\ \alpha\neq 0$; b) $\sum_{n=0}^{\infty}(-1)^n\frac{z^{2n+1}}{2n+1}$; c) $\sum_{n=0}^{\infty}(-1)^n\frac{z^{3n+1}}{3n+1}$.

Hint. a) $R = |\alpha|$, and divergence holds whenever $|z|\geq R$.

b) $R = 1$. If we note $\varsigma=-z^2$, then the series becomes $\sum_{n=0}^{\infty}\frac{\varsigma^n}{2n+1}$. Using theorem 4.12, it follows that this series converges at any ζ on the circle of equation $|\zeta| = 1$, except $\zeta = 1$. Consequently, the series in z is convergent if $|z|\leq 1$, except the points $z_{1,2}=\pm i$.

c) $R = 1$. A reason similar to b) leads to divergence at the cubic roots of -1.

3. Establish the domain of convergence for the series:

a) $\sum_{n=1}^{\infty}\frac{n+i}{n^n}z^n$; b) $\sum_{n=1}^{\infty}\frac{z^n}{\sqrt{n!}}$; c) $\sum_{n=1}^{\infty}\frac{(nz)^n}{n!}$; d) $\sum_{n=1}^{\infty}\frac{z^n}{\sqrt[n]{n!}}$.

Hint. a) $\omega=\lim_{n\to\infty}\frac{\sqrt[n]{n+i}}{n}=0$, hence $R=\infty$, i.e. the series converges on $\mathbb{C}$.

b) Similarly to a), $D_{conv.}=\mathbb{C}$, since $\omega=\lim_{n\to\infty}\frac{\sqrt{n!}}{\sqrt{(n+1)!}}=0$.

c) Applying (4) or (4') to $a_n = n^n / n!$, we obtain $\omega=\lim_{n\to\infty}\frac{n}{\sqrt[n]{n!}}=e$, hence $R=1/e$. We may use theorem 4.12 to find out the behavior at the frontier points. In fact, because $(1+\frac{1}{n})<e$, the sequence of terms $a_nR^n=\frac{n^n}{n!e^n}$ is decreasing, while the Stirling's formula $\lim_{n\to\infty}\frac{n^n}{n!e^n}\sqrt{2\pi n}=1$ shows that $\lim_{n\to\infty}a_nR^n=0$. Consequently, possibly except the point $z=1/e$, this power series is convergent at the other points where $|z|=1/e$.

To clarify the nature of the series at the remaining point $z=1/e$, we have to study the convergence of the numerical series

$$\sum_{n=1}^{\infty} \frac{n^n}{e^n\, n!} .$$

With this purpose we reformulate the Stirling's result by saying that "for every $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that the inequalities

$$1-\varepsilon < \frac{n^n}{n!\, e^n}\sqrt{2\pi n} < 1+\varepsilon$$

hold whenever $n > n_0$ ". Using the resulted relation

$$\frac{n^n}{n!\, e^n} > \frac{1-\varepsilon}{\sqrt{2\pi n}}$$

in a comparison test, we conclude that the series is divergent at $z = 1/e$.

d) From $\lim_{n\to\infty} \frac{n}{\sqrt[n]{n!}} = e$ we deduce that $\omega = R = 1$. The sequence of terms

$$a_n = \frac{1}{\sqrt[n]{n!}}$$

is decreasing, and $\lim_{n\to\infty} a_n = 0$, hence the convergence of the power series holds when $|\, z \,| = 1$, possibly except $z = 1$. At this point we have divergence on account of the relation

$$\frac{1}{\sqrt[n]{n!}} > \frac{1}{\sqrt[n]{n^n}} = \frac{1}{n},$$

where $1/n$ is the general term of a divergent series.

4. Function $f : (-R, R) \to \mathbb{R}$, where $R > 0$, is defined as the sum of the series

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots$$

where $a_0 = a_1 = 1$, and $a_n = \sum_{k=0}^{n-1} a_k\, a_{n-k-1}$ for all $n > 1$. Show that:

a) $f(x) = \frac{1-\sqrt{1-4x}}{2x}$;

b) $a_n = \frac{1}{n+1} C_{2n}^{n}$;

c) The radius of convergence is $R = 1/4$;

d) The power series converges to 2 at $x = 1/4$.

<u>Hint</u>. a) If we identify the coefficients of the series

$$f^{\,2}(x) = b_0 + b_1 x + b_2 x^2 + \ldots + b_n x^n + \ldots ,$$

which represents the product of the series of f by itself, then we obtain the relation $b_{n-1} = a_n$. In other words, we have

$$f^{\,2}(x) = a_1 + a_2 x + a_2 x^2 + \ldots + a_n x^{n-1} + \ldots ,$$

i.e. $x f^{\,2}(x) = f(x) - 1$. This equation has two solutions, namely

$$f_{1,2}(x) = \frac{1 \pm \sqrt{1-4x}}{2x} \ ,$$

but $f\,(0) = 1$ holds (in limit form) only for

$$f_2(x) = \frac{1-\sqrt{1-4x}}{2x} = \frac{2}{1+\sqrt{1-4x}} \ .$$

b) Develop $\sqrt{1-4x}$ as a binomial series

$$(1-4x)^{\frac{1}{2}} = \sum_{n=0}^{\infty} (4x)^n \frac{1\cdot 3\cdot \ldots \cdot(2n-3)}{n!\, 2^n} \ ,$$

and identify the coefficients in the series of f, which becomes

$$f\,(x) = 1 + x + \ldots + \frac{1}{2^{n+1}} \frac{1\cdot 3\cdot \ldots \cdot(2n-1)}{(n+1)!} 4^{n+1} \frac{1}{2} x^n + \ldots \ .$$

A simple calculation shows that

$$a_n = \frac{2^n \cdot 1\cdot 3\cdot \ldots \cdot(2n-1)}{(n+1)!} = \frac{(2n)!}{n!(n+1)!} = \frac{(2n)!}{[n!]^2(n+1)} = \frac{1}{n+1} C_{2n}^{n} \ .$$

c) $\omega = \lim\limits_{n\to\infty} \dfrac{a_{n+1}}{a_n} = 4$. D) $\lim\limits_{n\uparrow\frac{1}{4}} f\,(x) = 2$.

8. Evaluate: a) *Ln* 1 ; b) *Ln* (-1) ; c) *Ln i* ; d) *Ln* 1 + *i* ; and e) i^i .
Hint. Using (8) for *Ln*, and (9) for the complex power we obtain:

$$Ln\,1 = \{2k\pi\, i : k \in \mathbb{Z}\};$$

$$Ln\,(-1) = \{(\pi + 2k\pi) i : k \in \mathbb{Z}\}$$

$$Ln\, i = \{(\tfrac{\pi}{2} + 2k\pi) i : k \in \mathbb{Z}\}$$

$$Ln\,(1+i) = \{\tfrac{1}{2}\ln 2 + (\tfrac{\pi}{4} + 2k\pi) i : k \in \mathbb{Z}\}$$

$$i^i = e^{i\cdot Ln\, i} = \{e^{-\frac{\pi}{2}+2k\pi} : k \in \mathbb{Z}\}.$$

6. Show that the complex *sin* and *cos* are not bounded. In particular, evaluate the modulus of *sin*[π + *i ln*(2+$\sqrt{5}$)].
Hint. Find the real and imaginary parts of the functions, e.g.

$$\sin z = \sin(x+iy) = \sin x \cos iy + \sin iy \cos x = \sin x \cosh y + i \sinh y \cos x \ .$$

Consequently, we have $|\sin z| = \sqrt{\sin^2 x + \sinh^2 y}$. It remains to remember that the (real!) *sinh* is unbounded. In particular, | *sin*[π + *i ln*(2+$\sqrt{5}$)] | = 2.

9. Establish the formulas

$$Arc\sin Z = -iLn(iZ + \sqrt{1-Z^2}), \ Arc\cos Z = -iLn(Z + \sqrt{Z^2-1}),$$

$$Arc\tan Z = -\frac{i}{2} Ln \frac{1+iZ}{1-iZ}, \ Arc\cot Z = -\frac{i}{2} Ln \frac{Z+i}{Z-i} \ ,$$

and solve the equation *sin z* = 2. Find similar formulas for hyperbolic functions.

Hint. $Arc\sin Z = \{z \in \mathbb{C} : \sin z = Z\}$. If we replace $e^{iz} = \varsigma$ in the Euler's formula $\sin z = \dfrac{e^{iz} - e^{-iz}}{2i}$, then we find $\varsigma = iZ + \sqrt{1 - Z^2}$, and $z = \dfrac{1}{i} Ln\varsigma$.
In particular, the solutions of the equation *sin z* = 2 have the form

$$z \in Arc\sin 2 = -iLn\left[i(2 \pm \sqrt{3})\right] = \{\frac{\pi}{2} - i\ln(2 \pm \sqrt{3}) + 2k\pi : k \in \mathbb{Z}\}.$$

If we introduce *the first determination* by $\arcsin 2 = \frac{\pi}{2} - i\ln(2 + \sqrt{3})$, then we obtain the "old" formula

$$Arc\sin 2 = \begin{cases} \arcsin 2 + 2k\pi \\ \pi - \arcsin 2 + 2k\pi \end{cases}, \ k \in \mathbb{Z}.$$

CHAPTER III. CONTINUITY

§ III.1. LIMITS AND CONTINUITY IN $\mathbb{R}$

This section is a synthesis about the notions of "limit" and "continuity" as they are learned in the high school (also referred to as *lyceum*). Such a recapitulation naturally extracts the essential aspects concerning the notions of limit and continuity for real functions depending of one real variable.

As usually, the high school textbooks are thought like introduction to the field of interest, and they are merely based on description of practical facts, constructions, and direct applications. Their purpose is to offer some ideas and models, which hold out enough motivation for a rigorous and extensive study. Now, we suppose that these starting points are already known.

In particular, we consider that the following aspects are significant in the textbooks on Mathematical Analysis, at the standard level in high school:

1.1. Remarks. a) The real numbers are not rigorously constructed, but only described by their decimal approximations from $\mathbb{Q}$; most frequently, the fundamental algebraic and order properties are mentioned as axioms. The representation of $\mathbb{R}$ as *a real line* is currently used to help intuition.

b) The notion of continuity is studied with no reference to the specific structure; its qualitative feature is obvious if compared with Algebra or Geometry, but a strong relation to measurements is predominant. More exactly, the *neighborhoods* $V \in \mathscr{V}(x_0)$, where $x_0 \in \mathbb{R}$, are described in terms of order and absolute value, using the condition to contain *open symmetric intervals*, namely

$$V \supseteq (x_0 - \varepsilon, x_0 + \varepsilon) = \{x \in \mathbb{R} : |x - x_0| < \varepsilon\}$$

for some $\varepsilon > 0$. The *topological* forthcoming structure is not mentioned at all. In addition, almost all proofs are based on some particular algebraic and geometric properties of $\mathbb{R}$, but not on the topological ones.

c) The symbols $\pm\infty$ are introduced before speaking about limits and convergence, so they are directly related to the order structure of $\mathbb{R}$. In particular, $\pm\infty$ are involved in the study of *boundedness*, as well as in the notations $(-\infty, a) = \{x \in \mathbb{R}: x < a\}$, $[b, +\infty)$, $\overline{\mathbb{R}}$, etc. (compare later to the property of being *compact*). Similarly, the sign ∞ makes a purely formal sense when we express a limit, e.g. $\ell = \lim\limits_{n\to\infty} x_n$.

Other remarks refer to the manner in which different classes of functions are dealt with. For example, there is no mention that solving the practical problems, or giving some examples, strongly depends on some previously constructed and "known" functions, which, at the beginning, are the so-called *elementary functions* (as already said in § II.4.)

Besides the algebraic and order properties, which include the rules of operating with inequalities, the following properties are essential starting points for the development of the mathematical analysis on $\mathbb{R}$:

1.2. Fundamental Properties. a) (Cantor) For any bounded set in $\mathbb{R}$ there exist the infinimum and the supremum in $\mathbb{R}$.

b) (Archimedes) For every $x\in\mathbb{R}$ there exists a unique integer $n\in\mathbb{Z}$ such that $n\leq x < n + 1$ (called *entire part* of x, and noted $[x]$).

Proof. a) The problem of proving this assertion rises only for particular constructions of $\mathbb{R}$. In the axiomatic descriptions of $\mathbb{R}$ it is known as the *Cantor's axiom.*

b) The Archimedes' property is sometimes considered as an *axiom* too. However, in a framework like the present one, it is a consequence of a). In fact, if we suppose the contrary, then $n\leq x$ would hold for all $n\in\mathbb{Z}$. This means that $\mathbb{Z}$ is bounded, hence according to the Cantor's axiom, there would exist $\xi = \sup \mathbb{Z}\in\mathbb{R}$. Consequently, $\xi - 1 < p \leq \xi$ must hold for some p in $\mathbb{Z}$, hence $\xi < p + 1$. Because $p + 1 \in\mathbb{Z}$ too, this is in contradiction to the very definition $\xi = \sup \mathbb{Z}$. ◇

1.3. Remarks. a) Taking the Cantor's axiom as a starting point of our study clearly shows that the Real Analysis is essentially based on the order *completeness* of $\mathbb{R}$. At the beginning, this fact is visible in the limiting process involving sequences in $\mathbb{R}$, which is later extended to the general notion of *limit* of a real function.

b) We remember that the notion of *convergence* is presented in a very general form in the actual high school textbooks. The limiting process is essential in approximation problems, which naturally involve convergent sequences and series. For example, in practice we frequently approximate the irrationals, i.e. we operate with 1.4142 instead of $\sqrt{2}$, or with 3.14 instead of π, etc. In particular the Euler's number

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

is carefully introduced and studied in the most textbooks, including the presentation in the form of a series

$$e = 1 + \frac{1}{1!}+\frac{1}{2!}+\ldots+\frac{1}{n!}+\ldots .$$

c) Excepting the algebraic functions, whose values are obtained after a finite number of algebraic operations, the evaluation of any other function (generally being *transcendent*) requires a limiting process. For example, the values of e^x, *log x*, *sin x*, etc., as well as simpler expressions like $\sqrt{x}$, $\frac{1}{1-x}$, etc., represent sums of some series. More particularly, let us say that we have to evaluate $f(x) = x^{\sqrt{3}}$ at the point $x_0 = 2$. Primarily, we must reduce the problem to rational powers, because $f(2)$ is the limit of the sequence

$$2, 2^{1.7}, 2^{1.73}, \ldots$$

Finally, it remains to approximate the roots of indices 10, 100, etc.

Such problems clearly show that we need to extend the limiting process from sequences, which are functions on ℕ, to arbitrary functions. On this way, $n \to \infty$ is naturally generalized to $x \to a \in \mathbb{R}$.

1.4. <u>Definition</u>. Let $f : D \to \mathbb{R}$ be an arbitrary function of a real variable $x \in D \subseteq \mathbb{R}$, and let $a \in \overline{\mathbb{R}}$ be an accumulation point of D (i.e. $a \in D'$). We say that f *has the limit* $\ell \in \overline{\mathbb{R}}$ at a iff for any neighborhood V of ℓ there exists a neighborhood U of a such that at any $x \in D \cap (U \setminus a)$ we have $f(x) \in V$. In this case we note

$$\ell = \lim_{x \to a} f(x).$$

If, in addition, $a \in D$ and $f(a) = \ell$, we say that f is *continuous* at this point. If f is continuous at any point $a \in D$, then we say that f is *continuous on D.*

In practice it is useful to describe the existence of the limit in other terms, as follows:

1.5. <u>Theorem</u>. The following assertions are equivalent:

a) There exists $\ell = \lim_{x \to a} f(x)$;

b) For any sequence (x_n) in $D \setminus \{a\}$, we have [$x_n \to a$ implies $f(x_n) \to \ell$];

c) The *lateral limits* exist, and $f(a - 0) = f(a + 0)$.

In addition, if a, $\ell \in \mathbb{R}$ (i.e. they differ from $\pm\infty$), then these conditions are equivalent to the following:

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \textit{such that} \ [\, x \in D \setminus \{a\} \ \& \ |x - a| < \delta \,] \Rightarrow \left| f(x) - \ell \right| < \varepsilon .$$

<u>Proof</u>. b) $\Rightarrow$ a), more exactly $\rceil$ a) $\Rightarrow \rceil$ b) : If f has no limit at a, then for any $\ell \in \mathbb{R}$ there exists some $\varepsilon > 0$ such that for arbitrary $\delta > 0$, there exists some $x \in (D \setminus \{a\}) \cap (a - \delta, a + \delta)$ for which $|f(x) - \ell| > \varepsilon$. In particular, let us take $\delta = \frac{1}{n}$, where $n \in \mathbb{N}$, and note by x_n the corresponding point. It is easy to see that $x_n \to a$, but $f(x_n) \not\to \ell$.

The rest of the proof is recommended as exercise. ◇

1.6. <u>Remarks</u>. a) Many properties, which are well known for sequences, remain valid for the general notion of limit. As a model, we mention the following rule of adding limits:

Let us consider *f*, $g : D \to \overline{\mathbb{R}}$, where $D \subseteq \overline{\mathbb{R}}$, and let us fix $a \in D'$. If there exist $l = \lim_{x \to a} f(x)$, and $k = \lim_{x \to a} g(x)$, then there exists $\lim_{x \to a} (f+g)(x)$ too, and it equals $l + k$ (respecting the rules of operating with $\pm\infty$).

b) A lot of limits in the textbooks express either continuity of particular elementary functions, or their behavior at $\pm\infty$. However, there are some remarkable cases, called *undeterminable*, which cam be establish by using the derivatives (e.g. l'Hôpital rules). We recall some of the most important:

$$\lim_{x \to \infty} \frac{a_0 x^n + a_1 x^{n-1} + \ldots + a_n}{b_0 x^n + b_1 x^{n-1} + \ldots + b_n} = \frac{a_0}{b_0}$$

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

$$\lim_{x \to \pm\infty} \left(1 + \frac{1}{x}\right)^x = \lim_{y \to 0} (1+y)^{1/y} = e$$

$$\lim_{x \to \infty} \frac{x^n}{a^x} = 0 \ (\forall n \geq 1,\ a > 1)$$

$$\lim_{x \to 0} x \ln|x| = 0$$

$$\lim_{x \to \infty} \frac{\ln x}{x} = 0$$

$$\lim_{x \to 0} \frac{\ln(1+x)}{x} = 1$$

$$\lim_{x \to 0} \frac{a^x - 1}{x} = \ln a \ (a > 0)$$

$$\lim_{x \to 0} \frac{(1+x)^r - 1}{x} = r \ (\forall r \in \mathbb{R}).$$

The most part of practical problems combine such "fundamental limits", in the sense that *composing continuous functions gives rise to continuous functions.*

c) The evaluation of a limit is deeply involved in the notion of *asymptote* of a graph. There are three types of asymptotes:

- The graph has a *horizontal* asymptote $y = l$ at $+\infty$ if
$$\exists\, l = \lim_{x \to \infty} f(x) \in \mathbb{R};$$
- The straight line $x = a$ is a *vertical* asymptote from the left, upwards (respectively downwards) the graph, if
$$\lim_{\substack{x \to a \\ x < a}} f(x) = \pm\infty \ ;$$
- The straight line $y = m\,x + n$ is an *oblique* asymptote at $+\infty$ if

$$\exists\, m = \lim_{x\to\infty} \frac{f(x)}{x} \in \mathbb{R}^* \textit{and}\ \exists\, n = \lim_{x\to\infty} \left[f(x) - mx \right] \in \mathbb{R}.$$

Similarly, we define the horizontal and the oblique asymptotes at $-\infty$, as well as the vertical asymptotes from the right.

The most significant properties of the continuous functions are expressed in terms of *boundedness* and *intermediate values* (see later *compactness, respectively connectedness*). The following two results refer to the behavior of the continuous functions on compact intervals.

1.7. <u>Definition</u>. We say that a set $D \subset \mathbb{R}$ is *bounded* iff there exist $a, b \in \mathbb{R}$ such that $D \subseteq [a, b]$. In particular, $[a, b]$ is a *compact interval* (i.e. bounded and closed). We say that a function $f : D \to \mathbb{R}$ is *bounded* iff $f(D)$ is a bounded set in ℝ, i.e. there exist the *lower* and *upper bounds* :

$$m = \inf \{f(x) : x \in D\} \in \mathbb{R} \text{ and } M = \sup \{f(x) : x \in D\} \in \mathbb{R}.$$

If $m = f(x^*)$ and $M = f(x^{**})$ at some x^*, $x^{**} \in D$, then we say that f *attains* (*touches*) its extreme values (bounds).

1.8. <u>Theorem</u>. If a function $f: D \to \mathbb{R}$ is continuous on a compact interval $[a, b] \subseteq D$, then it is bounded function that attains its bounds on $[a, b]$.

<u>Proof</u>. Let us suppose that function f is continuous but not upper bounded on $[a, b]$. Then there exists a sequence (x_n) in $[a, b]$, such that the sequence $(f(x_n))$ tends to $+\infty$. According to the Weierstrass theorem, this sequence has a <u>convergent</u> subsequence, say (x_{n_k}), for which $f(x_{n_k}) \to +\infty$ too. Let us note $l = \lim_{k\to\infty} x_{n_k}$, and remark that $l \in [a, b]$. Since f is continuous at l, it follows that $f(x_{n_k}) \to f(l) \neq +\infty$. The contradiction shows that f must have an upper bound. We similarly treat the lower boundedness.

To show that f attains its lower bound, let (ξ_n) be a sequence in $[a, b]$, such that $f(\xi_n) \overset{not.}{=} \eta_n \to m = \inf \{f(x) : x \in [a, b]\}$. Using the Cesàro's theorem, let us construct a convergent subsequence $(\xi_{n_k}) \to \underline{x} \in [a, b]$. Since f is continuous, we deduce that $m = \lim_{k\to\infty} f(\xi_{n_k}) = f(\underline{x})$.

Similarly, we show that $M = f(\overline{x})$ at some $\overline{x} \in [a, b]$. ◇

1.9. <u>Definition</u>. Let $f: D \to \mathbb{R}$ be a continuous function on $D \subseteq$R, i.e.

$$\forall_{x\in D}\ \forall_{\varepsilon>0}\ \exists_{\delta(x,\varepsilon)>0} \textit{ such that } \forall_{y\in D} \left[|x-y| < \delta(x,\varepsilon) \Rightarrow |f(x) - f(y)| < \varepsilon \right].$$

If in this condition we can use some $\delta(\varepsilon)$ for all $x \in D$, i.e.

$$\forall_{\varepsilon>0}\ \exists_{\delta(\varepsilon)>0} \textit{ such that } \forall_{x,y\in D} \left[|x-y| < \delta(\varepsilon) \Rightarrow |f(x) - f(y)| < \varepsilon \right],$$

then we say that f is *uniformly* (briefly *u.* -) *continuous on D.*

1.10. <u>Examples</u>. a) Function $f: D\to\mathbb{R}$ is said to be *Lipschitzean* iff there exists $L > 0$ (called *Lipschitz constant*) such that the Lipschitz' condition

$$| f(x) - f(y) | \le L \,| x - y |$$

holds at any $x, y\in D$. It is easy to see that the Lipschitzean functions are u.-continuous (e.g. functions with bounded derivatives, like *sin*, *cos*, etc.).

b) The polynomial functions P_n of degree $n\ge 2$ are not u.-continuous on $\mathbb{R}$, but they are u.-continuous on compact intervals.

c) The function $\frac{1}{x}$, defined on $\mathbb{R}^*$ (or on (0, 1], etc.) is not u.-continuous.

1.11. <u>Theorem</u>. If a function is continuous on a compact interval, then it is uniformly continuous on that interval.

<u>Proof</u>. Suppose by *reductio ad absurdum* that $f: [a, b]\to\mathbb{R}$ is continuous, but not uniformly continuous on $[a, b]$. Then there exists $\varepsilon_0 > 0$ such that for any $\delta_n = \frac{1}{n}$, where $n\in\mathbb{N}^*$, we can find some points x_n , $y_n \in [a, b]$, for which $| x_n - y_n | < \delta_n$, but $| f(x_n) - f(y_n) | \ge \varepsilon_0$. Making use of the same Cesàro's theorem (in two dimensions, since $(x_n , y_n) \in [a, b]^2 \subset\mathbb{R}^2$), let us construct a convergent subsequence of (x_n, y_n), say (x_{n_k}, y_{n_k}), for which we have $\xi = \lim\limits_{k\to\infty} x_{n_k} \in [a,b]$ and $\eta = \lim\limits_{k\to\infty} y_{n_k} \in [a,b]$. In addition, we claim that $\xi = \eta$. In fact, the second term in the inequality

$$| \xi - \eta | < | \xi - x_{n_k} | + | x_{n_k} - y_{n_k} | + | y_{n_k} - \eta |$$

can be made arbitrarily small for sufficiently large n.

The proof is achieved if we remark that the equality $\xi = \eta$ contradicts the hypothesis $\left| f(x_{n_k}) - f(y_{n_k}) \right| \ge \varepsilon_0$. ◇

Finally, we remind the property of intermediate values, namely:

1.12. <u>Definition</u>. Let $f: I\to\mathbb{R}$ be a function, where I is an interval of $\mathbb{R}$. We say that f has the *property of intermediate values on I*, iff for any $x_1, x_2\in I$, and any $c\in(f(x_1), f(x_2))$, there exists some $\xi \in (x_1, x_2)$ such that $f(\xi) = c$.

In this case, we also say that f is a *Darboux function* (or, it has the *Darboux' property*), where c is called *intermediate* value.

It is easy to see that f has the property of intermediate values if and only if it transforms any interval from I into another interval.

1.13. <u>Example</u>. Function $f: \mathbb{R}\to\mathbb{R}$, defined (using $a\in\mathbb{R}$) by

$$f(x) = \begin{cases} \sin\frac{1}{x} & \text{if } x \ne 0 \\ a & \text{if } x = 0 \end{cases}$$

is not continuous, but it is a Darboux function iff $a\in[-1, +1]$.

1.14. <u>Theorem</u>. Each continuous function is a Darboux function.
<u>Proof</u>. The assertion of the theorem can be reduced to the fact that for any continuous function $f: [a, b] \to \mathbb{R}$ which changes the sign at the endpoints (i.e. $f(a) \cdot f(b) \leq 0$), there is some $\xi \in [a, b]$ where $f(\xi) = 0$. To prove it in this last form, we may divide $[a, b]$ into equal parts and choose that half interval on which f changes the sign, which we note I_1. Dividing I_1, we similarly obtain I_2, and so on. To conclude, we may apply the principle of included intervals to find ξ. ◇

1.15. <u>Remark</u>. The above theorem assures the existence of at least one root for the equation $f(x) = 0$, where f is continuous. More than this, following the above proof, we can concretely solve such equations. More exactly, we can approximate the solution up to the desired degree of accuracy (actually done by the computer algorithms).

The same theorem is used to establish the intervals of constant sign of a continuous function. In fact, according to this theorem, any continuous function preserves its sign on the intervals where it is not vanishing.

1.16. <u>Theorem</u>. Let $f: I \to J$ be a continuous function, where $I \subseteq \mathbb{R}$ denotes an interval, and $J = f(I)$. This function is 1:1 if and only if it is strictly monotonous. If so, $f^{-1}: J \to I$ is continuous and strictly monotonous too.
<u>Proof</u>. It is easy to see that every strictly monotonous function is 1:1. Conversely, because any continuous function is Darboux, it follows that the property of being injective implies the strict monotony.

The inverse of any strictly monotonous function obviously is of the same type. It remains to show that f^{-1} is continuous on J. In fact, if we remember that the neighborhoods of any point in $\mathbb{R}$ contain intervals, then according to the Darboux' theorem, it follows that the continuous functions are carrying intervals into intervals. ◇

Later we will see how such properties concerning the limiting process can be extended from the case of real sequences, and real functions of a single real variable, to more general situations.

PROBLEMS §III.1.

1. Study the continuity and find the asymptotes of the function

$$f(x) = \begin{cases} x\left(1+e^{1/x}\right)^{-1} & if \quad x<0 \\ 1 & if \quad x=0 \\ e^{1/x} & if \quad x>0 \end{cases}$$

Hint. *f* is discontinuous at 0.

2. Find the following undetermined limits:

a) $\lim\limits_{x\to\infty} \dfrac{\sqrt{x^2+1}+\sqrt{x}}{\sqrt[4]{x^3+x}-x}$; b) $\lim\limits_{x\to\infty} \left(\dfrac{1+x}{2+x}\right)^{\frac{1-\sqrt{x}}{1+x}}$;

c) $\lim\limits_{x\to\infty} \dfrac{\ln(\sin x+2)}{x}$; d) $\lim\limits_{x\to 0} \dfrac{tgx-\sin x}{x^3}$.

Hint. a) Put forward the factor *x*. b) Use a fundamental limit that leads to *e* . d) Function *sin* is bounded and *ln* is increasing. d) Reduce to fundamental undetermined limits using trigonometric formulas.

3. The *rational function* is defined as a quotient of irreducible polynomials. Pick up the rational functions from the following:

a) *cos* (*n arccos x*); b) [*x*] = entire part of *x*; c) *x* – [*x*] ; d) $\sqrt{x}$;

e) | *x* | / *x* ; f) $\sqrt{x^2+1}$; g) e^x ; h) *sin x*.

Hint. a) Use the Moivre's formula to show that this function is polynomial, hence rational function (the single on the list). b) and c) have infinitely many discontinuities (of the first type!). d) is not defined on $\mathbb{R}_-$. e) The quotient of irrational functions may be rational. However, if we suppose that | *x* | / *x* = *P*(*x*) / *Q*(*x*), where *P* and *Q* have the same degree (since the limit of *P* / *Q* at ∞ is finite), then we are led to the contradiction

$$\lim_{x\to\pm\infty} \frac{|x|}{x} = \pm 1 \neq \lim_{x\to\pm\infty} \frac{P(x)}{Q(x)} = k .$$

f) and g) idem. h) has infinitely many zeros.

4. A real function *f* of one real variable is said to be *algebraic* iff there exists a polynomial (of degree *n*, with a parameter *x*),

$$P(u) = \sum_{k=0}^{n} a_k(x)\cdot u^k ,$$

where a_k are real polynomials, such that $P\circ f \equiv 0$. If not, *f* is called *transcendent.* Show that the rational functions as well as the roots (with arbitrary index) of polynomials (in particular $\sqrt{x}$ and $|x| = \sqrt{x^2}$) are algebraic functions while e^x and *sin x* are transcendent.

Hint. If $f = A/B$, we take $P(u) = A - Bu$; if $f = \sqrt[m]{A}$, we use $P = A - u^m$, etc. On the other hand, the identity

$$a_0(x) + a_1(x)e^x + \ldots + a_n(x)e^{nx} \equiv 0$$

is not acceptable. In fact, it leads to $\lim\limits_{x\to-\infty} a_0(x) = 0$, hence $a_0 \equiv 0$, and similarly (after dividing by e^x), $a_1 \equiv 0$, etc.

In the case of *sin* we may reason by referring to the finite set of zeros, namely $\{k\pi : k\in\mathbb{Z}\}$, and we similarly show that $a_0 \equiv 0, \ldots, a_n \equiv 0$.

5. Give examples of real functions having the properties:

a) They are defined on ℝ but continuous at a single point;

b) They are defined and discontinuous at each $x\in\mathbb{R}$;

c) They are continuous at each irrational, and discontinuous in rest.

Hint. a) $\pm x$, depending on $x\in\mathbb{Q}$ or not; b) ± 1, similarly depending on the rationality of x; c) Analyze the function:

$$f(x) = \begin{cases} \dfrac{1}{n} & if \quad x = \dfrac{m}{n} \in \mathbb{Q} \\ 0 & if \quad x \in \mathbb{R}\setminus\mathbb{Q}. \end{cases}$$

6. Let us note $\overset{(n)}{f} = f \circ f \circ \ldots \circ f$, where $f : \mathbb{R}\to\mathbb{R}$. Show that $\overset{(n)}{f}$ is continuous, periodical, respectively bounded for arbitrary $n\in\mathbb{N}^*$, if f is so. In particular, evaluate $\lim\limits_{n\to\infty} \overset{(n)}{f}(x)$ if $f(x) = \sin x$, at several points $x \in\mathbb{R}$.

Hint. Function $\overset{(n)}{\sin}$ has the same intervals of monotony as *sin*. The limit is *0* everywhere, since $\lim\limits_{n\to\infty} \overset{(n)}{\sin}\left(\dfrac{\pi}{2}\right) = 0$.

§ III.2. LIMITS AND CONTINUITY IN TOPOLOGICAL SPACES

In this section we aim to present the notions of limit and continuity in the most general framework, namely for functions acting between topological spaces. Therefore we expect the reader to be well acquainted with general topological structures, in the sense of § I.4 at least. A good knowledge of the high school handbooks, briefly sketched in the previous § III.1, will be also very useful.

To explain why this general theory is necessary, we mention that it offers the advantage of taking a long view over many particular cases that involve mathematical analysis, and other forms of the idea of continuity. On the other hand, the general theory is easily accessible since it naturally extends the case of a real function of one real variable.

2.1. <u>Definition</u>. Let $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \theta)$ be topological spaces, and let us consider a set $A \subseteq \mathscr{X}$, a point $a \in A'$, and a function $f: A \to \mathscr{Y}$. We say that an element $l \in \mathscr{Y}$ *is the limit of f at the point a*, iff for any $V \in \theta(l)$ there exists $U \in \tau(a)$ such that $f(x) \in V$ whenever $x \in \dot{U} \cap A$, where $\dot{U} = U \setminus \{a\}$. If $a \in A$, and $f(a) = l$, we say that f is *continuous at a*. If f is continuous at each $a \in A$, we say that f is *continuous on the set A.*

If f is continuous on $\mathscr{X}$, is 1:1, and $f^{-1}: \mathscr{Y} \to \mathscr{X}$ is continuous on $\mathscr{Y}$, then f is called *homeomorphism* between the topological spaces $\mathscr{X}$ and $\mathscr{Y}$. In other words, $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \theta)$ are said to be *homeomorphic* iff there exists a homeomorphism between them.

2.2. <u>Remark</u>. The use of the same notation, namely *lim*, for more notions, namely for the *convergence* of a sequences, as well as for that of *limit* and continuity of a function acting between topological spaces, is naturally explained by the existence of some intrinsic topology on any directed set. More exactly, if $(D,\leq)$ is a directed set, and $\infty \notin D$, then $\overline{D} = D \cup \{\infty\}$ is naturally endowed with its intrinsic topology in the sense of I.4.5.(*iv*). In addition, let $(\mathscr{S}, \theta)$ be a topological space, and let $f: D \to \mathscr{S}$ be a net in $\mathscr{S}$. It is easy to see that $l = \lim_D f$ iff the *prolongation* $\overline{f}: \overline{D} \to \mathscr{S}$, defined by

$$\overline{f}(x) = \begin{cases} f(x) & if \quad x \in D \\ l & if \quad x = \infty \end{cases}$$

is continuous at ∞ relative to the intrinsic topology θ on D, and to the initially considered topology τ on $\mathscr{S}$.

In particular, the limit of a sequence $f: \mathbb{N} \to \mathscr{S}$ can be viewed as the limit of its *prolongation* $\overline{f}$, relative to an *intrinsic* topology θ of $\overline{N} = \mathbb{N} \cup \{\infty\}$, where $\infty \notin \mathbb{N}$, and θ has the values

$$\theta(x) = \begin{cases} \{V \subseteq \overline{\mathbb{N}} : x \in V\} & if \quad x \in \mathbb{N} \\ \{V \subseteq \overline{\mathbb{N}} : \exists n \in \mathbb{N} \text{ such that } V \supseteq (n,\infty]\} & if \quad x = \infty \ . \end{cases}$$

Similarly, defining (if $a \notin A$), or modifying (if $l \neq f(a)$) the value of f at a, such that $f(a) = l$, the existence of $\lim_{x \to a} f(x)$ can be reformulated in terms of continuity. Such a connection between convergence and continuity is involved in the following property concerning the composed functions:

2.3. <u>Theorem</u>. Let $(\mathscr{X}, \tau)$, $(\mathscr{Y}, \theta)$, $(\mathscr{Z}, \zeta)$ be topological spaces, in which we take the points $x \in \mathscr{X}$, $y \in \mathscr{Y}$, and $z \in \mathscr{Z}$.

(*i*) We note $\mathscr{X}_0 = \mathscr{X} \setminus \{x\}$, $\mathscr{Y}_0 = \mathscr{Y} \setminus \{y\}$, and we consider that $x \in (\mathscr{X}_0)'$, and $y \in (\mathscr{Y}_0)'$. If the functions $f : \mathscr{X}_0 \to \mathscr{Y}_0$ and $g : \mathscr{Y}_0 \to \mathscr{Z}$ have the limits $y = \lim_{u \to x} f(u)$ and $z = \lim_{v \to y} g(v)$, then $z = \lim_{u \to x} (g \circ f)(u)$.

(*ii*) If the function $F : \mathscr{X} \to \mathscr{Y}$ is continuous at x, and the function $G : \mathscr{Y} \to \mathscr{Z}$ is continuous at $y = F(x)$, then the function $G \circ F$ is continuous at x.

(*iii*) Each net (in particular sequence) is convergent iff all of its subnets (subsequences) are convergent to the same limit.

<u>Proof</u>. (*i*) To any $V \in \zeta(z)$ there corresponds $U \in \theta(y)$, hence $W \in \tau(x)$, such that $u \in \dot{W} \cap \mathscr{X}_0$ implies $v = f(u) \in \dot{U} \cap \mathscr{Y}_0$, and finally $(g \circ f)(u) \in V$.

(*ii*) Similarly to (*i*), for each $V \in \zeta(z)$ there exists $U \in \theta(y)$, hence $W \in \tau(x)$, such that $u \in W$ implies $v = F(u) \in U$, and finally $(G \circ F)(u) \in V$. In addition, we have $y = F(x)$ and $z = G(y)$, hence $z = (G \circ F)(x)$.

(*iii*) Let $(E, <<)$ be a directed set, and let τ be the intrinsic topology of the space $\mathscr{X} = \overline{E} = E \cup \{\$\}$, where \$ plays the role of infinity to E. In a similar manner, let $(D, \leq)$ be another directed set, and let θ be the intrinsic topology of $\mathscr{Y} = \overline{D} = D \cup \{\infty\}$. Finally, let the net $f_0 : E \to D$ be extended to $f : \mathscr{X} \to \mathscr{Y}$ by

$$f(u) = \begin{cases} f_0(u) & if \quad u \in E \\ \infty & if \quad u = \$ \notin E \ , \end{cases}$$

and let $g_0 : D \to \mathscr{Z}$ be extended to $g : \mathscr{Y} \to \mathscr{Z}$ by

$$g(v) = \begin{cases} g_0(v) & if \quad v \in D \\ z & if \quad v = \infty \notin D \ . \end{cases}$$

Obviously, the net g_0 is convergent to z in the topological space $(\mathscr{Z}, \zeta)$, if and only if $z = \lim_{v \to \infty} g(v)$. In addition, it is easy to see that f_0 is subject to the Kelley's condition [*s*] of subnets if and only if $\infty = \lim_{u \to \$} f(u)$. Using the above result on composed functions, it follows that $z = \lim_{u \to \$} (g \circ f)(u)$, which shows that the subnet $g_0 \circ f_0$ is convergent to z too.

The converse implication is obvious, because in particular, each net is a subnet of itself, for which $E = D$, and f_0 the identity on D.

Of course, taking $E = D = \mathbb{N}$, and $\$ = \infty$, we obtain the similar property for sequences. ◇

Simple examples (see problem 4 at the end of this section) show that the rule (*i*) of composed limits cannot be formulated as simply as the rule (*ii*) of composed continuous functions in theorem 2.3 from above. Of course, other hypotheses are possible in (*i*) (see [DE], [FG], [G-O], etc.).

In order for us to complete the list of relations between convergence and continuity, we introduce the following:

2.4. <u>Theorem</u>. (Heine). Let $(\mathscr{Y}, \theta)$, $(\mathscr{Z}, \zeta)$ be topological spaces, and let $(D, \leq)$ be a directed set. Function $g: \mathscr{Y} \to \mathscr{Z}$ has the limit $z = \lim_{v \to y} g(v)$ if and only if for any net $f: D \to \mathscr{Y}$, we have

$$y = \lim_D f \Rightarrow z = \lim_D (g \circ f). \tag{H}$$

<u>Proof</u>. According to theorem 2.3 about composed functions, from $y = \lim_D f$ and $z = \lim_{v \to y} g(v)$ it follows that $z = \lim_D (g \circ f)$.

Conversely, let us suppose that the implication (H), usually called *Heine condition*, is fulfilled. Let us remark that the particular set

$$D = \{(V, v) \in \theta(y) \times \mathscr{Y} : v \in V\}$$

is directed by the order relation defined by

$$(U, u) \leq (V, v) \Leftrightarrow V \subseteq U.$$

The main use of $(D, \leq)$ is that the net $f: D \to \mathscr{Y}$, defined by $f(V, v) = v$, always converges to y. Finally, according to theorem 2.3, $z = \lim_D (g \circ f)$ is nothing but $z = \lim_{v \to y} g(v)$. ◇

For practical reasons (e.g. approximation problems, modeling continuous and deterministic systems, etc.), it is desirable to ensure the uniqueness of the limit, whenever there exists one. This property of the limit turns out to depend on the topological structure of the target space of the considered function (in particular net, or sequence). More exactly:

2.5. <u>Definition</u>. (*Hausdorff axiom*) We say that a topological space $(\mathscr{Y}, \theta)$ is *separated* (*Hausdorff*, or T_2) iff

[T_2] For each pair of points $y', y'' \in \mathscr{Y}$, where $y' \neq y''$, there exist some neighborhoods $V' \in \theta(y')$ and $V'' \in \theta(y'')$, such that $V' \cap V'' = \emptyset$.

2.6. <u>Theorem</u>. A topological space $(\mathscr{Y}, \theta)$ is separated iff for any other topological space $(\mathscr{X}, \tau)$, and any function $f: \mathscr{X} \to \mathscr{Y}$, which has a limit at an arbitrary $x \in \mathscr{X}$, this limit is unique.

Proof. Let us suppose that ($\mathscr{Y}$, τ) is separated, and still there exists a function $f: \mathscr{X} \to \mathscr{Y}$, such that both $y' = \lim_{u \to x} f(u)$ and $y'' = \lim_{u \to x} f(u)$. Then there exist $U' \in \tau(x)$ and $U'' \in \tau(x)$ such that $f(U') \subseteq V'$ and $f(U'') \subseteq V''$. Because of [N$_1$] and [N$_3$], we have $U' \cap U'' \neq \emptyset$, contrarily to $V' \cap V'' = \emptyset$.

Conversely, let us suppose that ($\mathscr{Y}$, θ) is not separated, and let $y', y'' \in \mathscr{Y}$ be a pair of different points for which $V' \cap V'' \neq \emptyset$ holds for all $V' \in \theta(y')$ and $V'' \in \theta(y'')$. If so, we may define the set

$$D = \{(V', V'', y) \in \theta(y') \times \theta(y'') \times \mathscr{Y} : y \in V' \cap V''\},$$

which is directed by the product relation of inclusion

$$(U', U'', u) \leq (V', V'', v) \Leftrightarrow V' \subseteq U' \textit{ and } V'' \subseteq U''.$$

As usually, we construct $\mathscr{X} = D \cup \{\infty\} = \overline{D}$ and endow it with its natural topology τ. Consequently, the function (more exactly the net) $f: D \to \mathscr{Y}$, of values $f(V', V'', y) = y$, has two limits at ∞, namely y' and y''. ◇

By extending the function (and its inverse) from points to sets, and to families of sets, we obtain other forms for the notion of "continuity":

2.7. Theorem. Let ($\mathscr{X}$, τ) and ($\mathscr{Y}$, ϑ) be topological spaces. If $f: \mathscr{X} \to \mathscr{Y}$, then:

(a) f is continuous at $a \in \mathscr{X}$ iff $f^{\leftarrow}(\vartheta(f(a))) \subseteq \tau(a)$,

(b) f is continuous on $\mathscr{X}$ iff [$f^{\leftarrow}(A)$ is open (closed) in $\mathscr{X}$ whenever A is open(closed) in $\mathscr{Y}$],

(c) f is continuous on $\mathscr{X}$ iff [for any $A \subseteq \mathscr{X}$, $a \in \overline{A}$ implies $f(a) \in \overline{f(A)}$], i.e.

$$f(\overline{A}) \subseteq \overline{f(A)}.$$

Continuity is useful when we need to compare topologies, or to obtain new topologies (e.g. on a subset, product space, quotient space, etc.):

2.8. Standard constructions. 1) Let τ and ϑ be two topologies on the same set $\mathscr{X}$. We say that τ is *coarser* (*smaller*, etc.) than ϑ (which is *finer*, *greater*, etc. than τ) iff the identity $\iota: \mathscr{X}_\vartheta \to \mathscr{X}_\tau$ is continuous, where the indices τ and ϑ represent the topologies considered on $\mathscr{X}$. Since this means that $\tau(x) \subseteq \vartheta(x)$ holds at each $x \in \mathscr{X}$, we may note $\tau \subseteq \vartheta$.

2) Let ($\mathscr{X}$, τ) be a topological space, and let a subset $\mathscr{Y} \subseteq \mathscr{X}$ be endowed with a topology ϑ. We say that ($\mathscr{Y}$, ϑ) is a *topological subspace of* ($\mathscr{X}$, τ) iff ϑ is the coarsest topology for which the *canonical embedding*

$$\varepsilon : \mathscr{Y} \to \mathscr{X},$$

defined by $\varepsilon(y) = y \in \mathscr{X}$ at each $y \in \mathscr{Y}$, is continuous on $\mathscr{Y}$. Alternatively, $H \subseteq \mathscr{Y}$ is ϑ-open iff $H = G \cap \mathscr{Y}$ holds for some τ-open set $G \subseteq \mathscr{X}$.

3) Let $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \vartheta)$ be topological spaces. The *product topology* ζ on $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ is defined as the coarsest topology on $\mathscr{Z}$ for which the *projections* $P_x : \mathscr{Z} \to \mathscr{X}$, and $P_y : \mathscr{Z} \to \mathscr{Y}$ are continuous (remember that the projections are defined by $P_x(x, y) = x$ and $P_y(x, y) = y$).

In order to construct the product topology, we mention that for each pair of neighborhoods $V \in \tau(x)$, and $U \in \vartheta(y)$, the *cylinders* $P_x^{-1}(V)$ and $P_y^{-1}(U)$ represent ζ – neighborhoods of $(x, y) \in \mathscr{Z}$, hence $W \in \zeta(x, y)$ holds exactly when W contains a *rectangle* of the form $P_x^{-1}(V) \cap P_y^{-1}(U)$.

4) Let $(\mathscr{X}, \tau)$ be a topological space , let ~ be an equivalence on $\mathscr{X}$, and let $\hat{\mathscr{X}} = \mathscr{X}/\sim$ be the quotient space. The finest topology on $\hat{\mathscr{X}}$, for which the *canonical application* $\varphi : \mathscr{X} \to \hat{\mathscr{X}}$ is continuous on $\mathscr{X}$, is called *quotient topology*. It is frequently noted $\hat{\tau}$. We remind that that the canonical application in the construction of the quotient space is defined by

$$\varphi(x) = \hat{x} = \{y \in \mathscr{X} : y \sim x\}$$

at all $x \in \mathscr{X}$. In other words, if $V \in \tau(x)$, then $\hat{V} \in \hat{\tau}(\hat{x})$.

These "standard" constructions of a subspace, a product space, and a quotient space, confer special roles to the specific functions of embedding, projection and quotient, as shown in the following propositions 2.9 to 2.11:

2.9. <u>Proposition</u>. Let $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \vartheta)$ be topological spaces. The function $f : A \to \mathscr{Y}$, where $\emptyset \neq A \subseteq \mathscr{X}$, is $\tau - \vartheta$ continuous on A iff it is continuous on A relative to that topology σ, which makes it a topological subspace of $\mathscr{X}$.

<u>Proof</u>. For any $a \in A$ we have $V \in \sigma(a)$ iff $V = W \cap A$ for some $W \in \tau(a)$, hence f and $f \circ \iota$ are simultaneously continuous. ◇

2.10. <u>Proposition</u>. Let $(\mathscr{S}, \sigma)$, $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \vartheta)$ be topological spaces, and let $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ be endowed with the product topology ζ. Then function $f : \mathscr{S} \to \mathscr{Z}$ is continuous at $s \in \mathscr{S}$ (respectively on $\mathscr{S}$) iff their components $f_x = P_x \circ f$, and $f_y = P_y \circ f$ are continuous at s (respectively on $\mathscr{S}$).

<u>Proof</u>. Since P_x and P_y are continuous on $\mathscr{Z}$, according to the previous proposition, f_x and f_y will be continuous.

Conversely, let the functions f_x and f_y be continuous at $s \in \mathscr{S}$, and let W be a neighborhood of $f(s) = (f_x(s), f_y(s)) \in \mathscr{Z}$. According to the construction of the product space, there exist $U \in \tau(f_x(s))$, and $V \in \vartheta(f_y(s))$ such that $W \supseteq U \times V$. The continuity of f_x and f_y shows that $f_x^{\leftarrow}(U) \in \sigma(s)$ and $f_y^{\leftarrow}(V) \in \sigma(s)$. It is easy to see that for $L = f_x^{\leftarrow}(U) \cap f_y^{\leftarrow}(V) \in \sigma(s)$ we have $f(L) \subseteq W$. ◇

2.11. <u>Proposition</u>. Let $(\mathscr{X}, \tau)$ be a topological space, and let $(\hat{\mathscr{X}}, \hat{\tau})$ be the quotient topological space corresponding to the equivalence relation ~ on $\mathscr{X}$. If $(\mathscr{Y}, \vartheta)$ is another topological space, then $f: \hat{\mathscr{X}} \to \mathscr{Y}$ is continuous on $\hat{\mathscr{X}}$ iff $f \circ \varphi$ is continuous on $\mathscr{X}$, where φ is the canonical application of the quotient space.

<u>Proof</u>. We may directly refer to construction 2.8.4). The topology $\hat{\tau}$ is so defined such that the function $\varphi : \mathscr{X} \to \hat{\mathscr{X}}$ is continuous, hence the assertion "f continuous" obviously implies "$f \circ \varphi$ continuous".

Conversely, if $f \circ \varphi$ is continuous on $\mathscr{X}$, then at each point $x \in \mathscr{X}$, we have

$$V \in \vartheta(f \circ \varphi(x)) \Rightarrow (f \circ \varphi)^{\leftarrow}(V) \in \tau(x).$$

Obviously, $(f \circ \varphi)^{\leftarrow}(V) = \varphi^{\leftarrow}(f^{\leftarrow}(V))$. Because $f^{\leftarrow}(V)$ is a neighborhood of $\hat{x}$ in some topology of $\hat{\mathscr{X}}$, and $\hat{\tau}$ is the finest topology for which φ is continuous, it follows that $f^{\leftarrow}(V) \in \hat{\tau}(\hat{x})$. Consequently, the function f is continuous at x, which is arbitrary in $\mathscr{X}$. ◇

In the remaining part of this section we study two of the most important topological properties of sets, namely *connectedness and compactness*. We remind that these notions have been partially studied in lyceum, because in $\mathbb{R}$, *connected* means *interval*, and *compact* means *closed and bounded*.

2.12. <u>Definition</u>. The *junction* of two subsets A and B of a topological space $(\mathscr{S}, \tau)$ is defined by $\mathscr{J}(A, B) = (A \cap \overline{B}) \cup (\overline{A} \cap B)$.

If $\mathscr{J}(A, B) = \varnothing$, we say that A and B are *separated*. A set $M \subseteq \mathscr{S}$ is said to be *disconnected* iff $M = A \cup B$, where $A \neq \varnothing \neq B$, and A and B are separated. In the contrary case, we say that M is *connected*.

It is useful to recognize some particular connected sets:

2.13. <u>Theorem</u>. In the Euclidean topology of $\mathbb{R}$, M is connected iff it is an interval (no matter how, closed or open).

<u>Proof</u>. Let us assume that M is connected. The fact that M is an interval means that for any $x, y \in M$, $x < y$, we have $[x, y] \subseteq M$. If we suppose the contrary, i.e. M isn't interval, there exists $c \in (x, y) \setminus M$. Using c, we can construct the sets $A = \{x \in M : x < c\}$, and $B = \{x \in M : x > c\}$. Obviously, $M = A \cup B$, $A \neq \varnothing \neq B$ and $\mathscr{J}(A, B) = \varnothing$, hence M should be disconnected. The contradiction shows that M must be an interval.

Conversely, let us show that any interval $I \subseteq \mathbb{R}$ is connected. Assuming the contrary, we can decompose I into separate parts, i.e. $I = A \cup B$, such that $A \neq \varnothing \neq B$, and $\mathscr{J}(A, B) = \varnothing$. If so, let us fix $a \in A$ and $b \in B$, say in the relation $a < b$. Because I is an interval, we have $[a, b] \subseteq I$. In particular, also $c = sup\ (A \cap [a, b]) \in I$. Two cases are possible, namely either $c \in A$, or $c \in B$. Finally, we show that each one is contradictory.

Case c∈A. Since $\mathscr{J}(A, B) = \emptyset$, we necessarily have $c < b$. On the other side, $(c, b] \cap A = \emptyset$. Because I is an interval, it follows that $(c, b] \subseteq I$, hence $(c, b] \subseteq B$. In conclusion, $c \in \mathscr{J}(A, B) = \emptyset$, which is absurd.

Case c∈B. By its construction, $c \in \overline{A}$, hence $c \in \mathscr{J}(A, B) = \emptyset$ again, so this case is also impossible.

To conclude, the assumption that I is not connected is false. ◇

2.14. <u>Theorem</u>. Let $\{M_\alpha : \alpha \in \mathscr{I}\}$ be a family of connected sets. If the sets M_α are pair wise non-separated, i.e. $\mathscr{J}(M_\alpha, M_\beta) \neq \emptyset$ whenever $\alpha \neq \beta$, then the union $M = \cup\{M_\alpha : \alpha \in \mathscr{I}\}$ is a connected set too.

<u>Proof</u>. Let us suppose the contrary, i.e. $M = A \cup B$, where A and B are non void and separated. Because each M_α is connected, we have either $M_\alpha \subseteq A$, or $M_\alpha \subseteq B$ for all $\alpha \in \mathscr{I}$ (otherwise the sets $M_\alpha \cap A$ and $M_\alpha \cap B$ would be non void and separated components of M_α, which contradicts the connectedness of M_α). Now, using the monotony of the junction relative to the relation of inclusion, we obtain $\emptyset \neq \mathscr{J}(M_\alpha, M_\beta) \subseteq \mathscr{J}(A, B) = \emptyset$, whenever $M_\alpha \subseteq A$ and $M_\alpha \subseteq B$, contrarily to the hypothesis. ◇

2.15. <u>Theorem</u>. Let $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \vartheta)$ be topological spaces, and let the set $M \subseteq \mathscr{X}$ be connected. If $f : \mathscr{X} \to \mathscr{Y}$ is continuous on $\mathscr{X}$, then $f(M)$ is connected in $\mathscr{Y}$.

<u>Proof</u>. If we suppose the contrary, then we may decompose $f(M) = \mathfrak{A} \cup \mathfrak{B}$, such that $\mathfrak{A} \neq \emptyset \neq \mathfrak{B}$, and $\mathscr{J}(\mathfrak{A}, \mathfrak{B}) = \emptyset$. Consequently, the inverse images of $\mathfrak{A}$ and $\mathfrak{B}$, namely $A = M \cap f^{\leftarrow}(\mathfrak{A}) \neq \emptyset$, and $B = M \cap f^{\leftarrow}(\mathfrak{B}) \neq \emptyset$, realize the decomposition $M = A \cup B$. Since f is continuous, we have

$$\mathscr{J}(A, B) \subseteq f^{\leftarrow}(\mathscr{J}(\mathfrak{A}, \mathfrak{B})) = \emptyset,$$

which contradicts the hypothesis that M is connected. ◇

2.16. <u>Corollary</u>. Every real continuous function on $\mathbb{R}$ (endowed with the Euclidean topology) has the Darboux property.

<u>Proof</u>. The Darboux property claims that the direct image of any interval is also an interval. According to theorem 2.13, we may replace the term *interval* by *connected set*, and then apply theorem 2.15. ◇

2.17. <u>Remark</u>. We may use the above results to construct connected sets, e.g. continuous arcs in the complex plane, open or closed discs, sets obtained by taking the adherence of connected sets, unions, etc.

Another useful notion in this respect is that of *connectedness by arcs*. More exactly, M is *connected by arcs* iff for any two points $x, y \in M$ there exists a continuous arc γ, of end-points x and y, which is entirely contained in M (a *continuous arc* is the image through some continuous function of an interval of the real axis). We mention without proof that, for open sets, the conditions *connected* and *connected by arcs* are equivalent.

2.18. Definition. Let K be a set in a topological space $(\mathscr{S}, \tau)$, and let $\mathscr{G}$ be the family of all open sets in $\mathscr{S}$. By *open cover* of a set $K \subseteq \mathscr{S}$ we understand any family $\mathscr{A} \subseteq \mathscr{G}$ whose union covers K, i.e.

$$K \subseteq \cup\{G : G \in \mathscr{A}\}.$$

We say that K is a *compact* set iff from every open cover $\mathscr{A}$ of K we can extract a finite sub-cover, i.e. there exists a finite sub-family $\mathscr{B} \subseteq \mathscr{A}$, for which a similar inclusion holds, namely

$$K \subseteq \cup\{G : G \in \mathscr{B}\}.$$

2.19. Examples. a) We may easily obtain open covers of a set starting with open sets conceived to cover only one point. On this way, in $\mathbb{R}$, we find out that the set $A = \{1/n : n \in \mathbb{N}^*\}$ is not compact. However, the set $A \cup \{0\}$ is compact, because covering 0, we cover infinitely many terms of A.

b) The finite sets are compact in any topology of an arbitrary space.

c) Each closed interval $[a, b] \subset \mathbb{R}$ is compact. More generally, any closed and bounded set in $\mathbb{R}^n$ is compact.

d) The Riemann sphere is not compact because it isn't closed (see I.2.22). If we add the "North Pole" N ($\leftrightarrow \infty$), then $\mathscr{S} \cup \{N\} (\leftrightarrow \overline{\mathbb{C}})$ is compact.

e) The spaces $\mathbb{R}$, $\mathbb{C}$, and generally $\mathbb{R}^n$ are not compact, but $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ and $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ are compact.

The main property of a compact set refers to the transfer of this property to the image through a continuous function. In particular, if $f : I \to \mathbb{R}$ is continuous and $K \subseteq I \subseteq \mathbb{R}$ is compact, then f is bounded on K and it attains its extreme values. This property turns out to be generally valid, i.e. it holds if we replace $\mathbb{R}$ by arbitrary topological spaces, namely:

2.20. Theorem. Let $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \vartheta)$ be topological spaces, and let K be a compact set in $\mathscr{X}$. If the function $f : \mathscr{X} \to \mathscr{Y}$ is continuous on $\mathscr{X}$, then the image $f(K)$ is a compact set in $\mathscr{Y}$.

Proof. Let $\mathscr{A}$ be an open cover of $f(K)$, and let us consider

$$\mathscr{U} = \{f^{\leftarrow}(G) : G \in \mathscr{A}\}.$$

Since f is continuous, $\mathscr{U}$ represents an open cover of K. Let $\mathscr{V} \subseteq \mathscr{U}$ be a finite sub-cover of K, which exists because K is compact. It is easy to see that the corresponding subfamily of images

$$\mathscr{B} = \{f(X) : X \in \mathscr{V}\} \subseteq \mathscr{A}$$

is the finite open cover of $f(K)$, which we are looking for. ◇

We can express many general properties of the compact sets in terms of convergence. The analysis of such aspects is based on some properties of closed sets and operator “adherence” expressed in terms of convergence, as partially contained in Theorem II.1.5. In addition we mention:

2.21. <u>Lemma</u>. Let (x_n) be a sequence in a topological space $(\mathscr{S}, \tau)$. If none of its subsequences is convergent, then all the sets $G_k = \mathscr{S} \setminus \{x_k , x_{k+1} , \ldots\}$, where $k\in\mathbb{N}$, are open.

<u>Proof</u>. If we suppose the contrary, then some point $x\in G_k$ is not interior to G_k , i.e. any neighborhood $V\in \tau(x)$ contains terms x_n of the sequence, of rank $n\geq k$. It follows that x is adherent to $\{x_n\}$, and by Theorem II.1.5.(*i*), x should be the limit of some sequence in the set $\{x_n\}$. Obviously, this sequence can be arranged as a subsequence of (x_n), contrarily to the hypothesis that such subsequences do not exist at all. ◇

Now, from compactness we may deduce properties of convergence:

2.22. <u>Theorem</u>. If K is a compact set in the separated topological space $(\mathscr{S}, \tau)$, then the following properties hold:

a) K is closed, and
b) For any sequence (x_n) in K (i.e. $x_n\in K$ for all $n\in\mathbb{N}$), there exists a subsequence (x_{n_k}), convergent to some $x\in K$ (when we say that K is *sequentially compact*).

<u>Proof</u>. A) If we suppose the contrary, i.e. $K \neq \overline{K}$, it follows that there exists some $x\in \overline{K} \setminus K$ (the converse inclusion, $K\subseteq \overline{K}$, always holds). According to condition $[T_2]$, for every $y\in K$, $y \neq x$, there exists a pair of neighborhoods $V_y\in\tau(x)$ and $U_y\in\tau(y)$ such that $V_y\cap U_y = \emptyset$. Since K is compact, there exists a finite set of points, say $\{y_1, y_2, ..., y_n\}\subseteq K$, such that $K\subseteq \bigcup\{U_{y_k} : k=\overline{1,n}\}$. On the other hand, the neighborhood $V = \bigcap\{V_{y_k} : k=\overline{1,n}\}\in\tau(x)$ has no point in K , contrarily to the hypothesis $x\in \overline{K}$. To avoid this contradiction, we have to accept that $\overline{K} =K$, i.e. K is closed.

b) Supposing the contrary again, let (x_n) be a sequence in K, such that no subsequence is convergent to some $x\in K$. Because K is compact, hence just proved closed, the subsequences of (x_n) cannot be convergent in $\mathscr{S}$ (see Theorem II.1.5(*ii*)). Now, let us construct the sets G_k as in the lemma 2.21, which form an increasing sequence, i.e. $G_k\subseteq G_{k+1}$ holds for all $k\in\mathbb{N}$. On the other hand, $\{G_k\}$ forms an open cover of K. Using the compactness of K, let G_n be the greatest element of a finite sub-cover of K, hence $G_n \supseteq K$. This relation contradicts the fact that $x_n\in K$, but $x_n\notin G_n$. Consequently, the initial supposition is impossible, i.e. sequence (x_n) cannot ever exist. ◇

It is easy to see that finite unions and arbitrary intersections of compact sets are compact too. The problem of compactness of an arbitrary Cartesian product is more difficult, except the case of a finite family of compact sets, which is relatively simple:

2.23. <u>Theorem.</u> Let $(\mathscr{X}, \tau)$ and $(\mathscr{Y}, \vartheta)$ be topological spaces, and let their Cartesian product $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ be endowed with product topology ζ. If the sets $K \subseteq \mathscr{X}$ and $L \subseteq \mathscr{Y}$ are compact, then $K \times L$ is a compact set in $\mathscr{Z}$.

<u>Proof</u>. Because any open set in $(\mathscr{Z}, \zeta)$ is a union of sets of the form $G \times F$, where G is open in $(\mathscr{X}, \tau)$ and F is open in $(\mathscr{Y}, \vartheta)$, it follows that each open cover $\mathscr{A}$ of $K \times L$ generates another open cover, say

$$\mathscr{A}^* = \{G_i \times F_j : (i, j) \in P\},$$

where $P \subseteq I \times J$, and I, J are certain families of indices. It is clear that family $\{G_i : (i, j) \in P\}$ is an open cover of K, while $\{F_j : (i, j) \in P\}$ is an open cover of L. Consequently, using the hypothesis concerning the compactness of K and L, we find finite subfamilies $I_0 \subseteq I$ and $J_0 \subseteq J$, such that the sub-families $\{G_i : i \in I_0\}$ and $\{F_j : j \in J_0\}$ are open sub-covers of K, respectively of L. In conclusion we see that the family

$$\mathscr{A}_0 = \{A \in \mathscr{A} : A \supseteq G_i \times F_j \textit{ for some } (i, j) \in I_0 \times J_0\}$$

is a finite open sub-cover of $\mathscr{A}$, hence $K \times L$ is compact. ◇

2.24. <u>Remarks</u>. (*i*) Using the above results on compactness we can easily construct particular compact sets. For example, the compact sets in $\mathbb{R}$ are finite unions of closed intervals; the continuous arcs, which contain their end-points, are compact sets in $\mathbb{C} \sim \mathbb{R}^2$, or generally in $\mathbb{R}^n$; the closed balls, and closed parallelograms, etc. For more examples and details in *Euclidean spaces*, we recommend the reader to see the next section.

(*ii*) A lot of assertions reveal properties (e.g. "point *a* is adherent to *A*", "*M* is connected", "*K* is compact"), which are *invariant* under continuous transformations, i.e. they remain valid for images through continuous functions. In general, such properties are said to be *topological*, and *topology* itself is defined as their study (compare to geometries!).

(*iii*) It is easy to see that important topological spaces like $\mathbb{R}$, $\mathbb{C}$ ($\sim \mathbb{R}^2$), and generally $\mathbb{R}^n$ (thoroughly discussed in the next section), are not compact. Because of many convenient properties of the compact spaces, especially involving continuity and convergence, a natural tendency of transforming such spaces into compact ones has risen. This process is frequently called *compaction*, and usually it consists of adding some new elements (called *points at infinity*) to the initial space (e.g. $+\infty$ and $-\infty$ to $\mathbb{R}$, ∞ to $\mathbb{C}$, etc.) such that the forthcoming space becomes compact.

2.25. <u>Comment</u>. The above topological structures represent (in the present framework) the most general *structures of continuity*. There are many extensions of these structures, but according to the main purpose of this book, later on we pay more attention to the metric, normed, Euclidean, and other particular spaces, where the concrete (i.e. numerical) measurements are eloquent.

From a general point of view, we can say that *topology is a qualitative theory* if compared to geometry, algebra or other branches of mathematics, physics, etc. The specific topological concepts like *convergence*, *continuity*, *compactness*, etc., express properties that cannot be measured, or described by numbers, because in fact, they represent conclusions of infinitely many measurements, and infinite sets of elements (e.g. numbers).

Choosing some topology as a mathematical instrument of studying a particular phenomenon is an evidence of the investigator's belief in the *continuous nature* of that problem. The comparison of the theoretical results with the practical experience decides how inspired is the continuous vision of the problem. Obviously, this point of view is not always adequate, i.e. there exist many non-continuous aspects in nature, which need other kind of mathematical structures to be modeled. By philosophical duality, the essential feature of these problems is *discreteness*. Nowadays, discrete phenomena represent the object of the discrete system theory, including the computer engineering.

If we limit ourselves to topological structures, then the sense of *discrete* reduces to "space endowed with the *discrete topology*", that is an extreme case when any subset of the space is open. From this point of view, each set may be considered discrete, which however is not always the case. On the other hand, the *continuous sets* are thought as "compact and connected", which have no discrete counterpart. So, we may conclude that it is difficult enough to develop the great idea of a "continuous-discrete dualism of the world" exclusively using topological structures. Therefore we need a larger framework, where some structures of discreteness are justified to be dual to topologies, but not particular topologies. Without going into details, we mention that such structures have been proposed in [BT_3]. In brief, the idea is that, instead of defining a topology τ by filters $\tau(x)$ of neighborhoods at each $x \in V \in \tau(x)$, to consider a dual structure, called *horistology*, which is specified by *ideals of perspectives* $\chi(x)$, such that $x \notin P$ whenever $P \in \chi(x)$. The terminology is naturally inspired from relativity theory, where super-additivity is accepted as a real physical fact.

The coexistence of continuity and discreteness in the real world, which is reflected in the topology – horistology dualism, is also met at many other particular levels, defined by metrics, norms, or inner products. Respecting the traditional framework of the classical Analysis, we shall not discuss about horistologies any further, and we let the reader to appreciate whether such qualitative structures are useful to study discreteness.

PROBLEMS § III. 2.

1. Let $(D, \geq)$ be a directed set, and let $f, g : D \to \mathbb{R}$ be nets. Show that:
(i) If the net f is monotonic (i.e. $f(d) \leq f(e)$ whenever $\alpha \leq d \leq e$ for some $\alpha \in D$), and bounded (i.e. $\exists M \in \mathbb{R}$ such that $f(d) \leq M$ whenever $d \geq \alpha$), then f is convergent (the limit being always unique in $\mathbb{R}$);
(ii) If f and g are convergent nets, and $f(d) \leq g(d)$ holds at any $d \geq \alpha$ for some $\alpha \in D$, then also $\lim f \leq \lim g$;
(iii) If the nets f and g are convergent to the same limit l, and $h : D \to \mathbb{R}$ is another net for which $f(d) \leq h(d) \leq g(d)$ holds at any $d \geq \alpha$ for some $\alpha \in D$, then h is convergent to l too.
Hint. Repeat the proof of the similar properties for real sequences.

2. Show that any closed part F, of a compact set K, is compact. Analyze the case $K \subset \mathbb{R}$ and $F = K \cap \mathbb{Q}$, where F, generally not closed any more, is alternatively referred to as a part of $\mathbb{R}$ and $\mathbb{Q}$.
Hint. If $\mathscr{A}$ is an open cover of F, then $\mathscr{A} \cup \{\complement F\}$ is an open cover of K. The open sets in topology of $\mathbb{Q}$ are intersections of the form $\mathbb{Q} \cap G$, where G is open in $\mathbb{R}$.

3. Let $f : (a, b) \to \mathbb{R}$ be a continuous function. Show that:

$$\left(\exists \lim_{x \to a} f(x) \in \mathbb{R},\ \exists \lim_{x \to b} f(x) \in \mathbb{R} \right) \Rightarrow \text{ (f is bounded).}$$

Is the converse implication generally valid?
Hint. f can be continuously prolonged to the compact $[a, b]$. Consider the example $g : (0, 1) \to \mathbb{R}$, where $g(x) = \sin x^{-1}$.

4. Compare $\lim_{x \to 0} g(x)$ to $\lim_{x \to 0} (g \circ f)(x)$ if the functions $f, g : [0, 1] \to \mathbb{R}$ have the values $f(0) = g(0) = 1$ and $f(x) = g(x) = 0$ at each $x \in (0, 1]$.
Hint. $\lim_{x \to 0} g(x) = 0$, while $\lim_{x \to 0} (g \circ f)(x) = 1$.

5. Show that for each pair of compact (connected) sets $A, B \subset \Gamma$ (i.e. $\mathbb{R}$ or $\mathbb{C}$), the sets $A \pm B$, and AB are also compact (connected). Comment $A : B$.
Hint. Use the fact that the operations of addition and multiplication are defined by continuous functions on the product space.

§ III.3. LIMITS AND CONTINUITY IN METRIC SPACES

It is easy to see that each contraction of a metric space $\mathscr{S}$ is continuous on this space relative to the intrinsic topology. More than this, because in metric spaces we can "correlate" the neighborhoods of different points by considering that the spheres of equal radiuses $S(x, r)$ and $S(y, r)$ have the *same size*, we deduce a possibility to compare the continuity at different points. Based on this feature of the metric spaces, we remark a uniform behavior of the contractions from the continuity point of view. These ones, and many similar cases lead us to consider the following type of continuity in the metric space framework:

3.1. Definition. Let $(\mathscr{X}, \rho)$ and $(\mathscr{Y}, \sigma)$ be metric spaces, and let A be a (non-void) subset of $\mathscr{X}$. We say that the function $f : A \to \mathscr{Y}$ is *uniformly continuous on A* iff for any $\varepsilon > 0$ there exists $\delta > 0$ such that the inequality $\sigma(f(x), f(y)) < \varepsilon$ holds whenever $\rho(x, y) < \delta$. In particular, a function can be uniformly continuous on the entire space $\mathscr{X}$.

The following theorem is frequently used to establish that some functions are uniformly continuous.

3.2. Theorem. Let $(\mathscr{X}, \rho)$ and $(\mathscr{Y}, \sigma)$ be metric spaces, and let $K \subseteq \mathscr{X}$ be a compact set. If the function $f : K \to \mathscr{Y}$ is continuous on K, then it is also uniformly continuous on K.

Proof. The continuity of f at x allows us to assign some $\delta_x > 0$ to each $\varepsilon > 0$, such that $\rho(x, y) < \delta_x$ implies $\sigma(f(x), f(y)) < \varepsilon / 3$. Since K is compact, and the family $\mathscr{A} = \{S(x, \frac{1}{3}\delta_x): x \in K\}$ is an open cover of K, it follows that there exists a finite set $\{x_1, x_2, \dots, x_n\} \subseteq K$ (hence a finite sub-family of $\mathscr{A}$) such that $K \subseteq \bigcup \{S(x_i, \frac{1}{3}\delta_x): i=1, 2, \dots, n\}$. Consequently, for each $x \in K$ we can find some $i \in \{1, 2, \dots, n\}$ such that $\rho(x, x_i) < \frac{1}{3}\delta_{x_i} < \delta_{x_i}$, and so we secure the inequality $\sigma(f(x), f(x_i)) < \varepsilon / 3$.

We claim that $\delta = \frac{1}{3}\min\{\delta_{x_i} : i=1, 2, \dots, n\}$ is right to fulfill the condition of uniform continuity of f. In fact, let us arbitrarily chose $x, y \in K$ such that $\rho(x, y) < \delta$, and let $i, j \in \{1, 2, \dots, n\}$ be indices for which the inequalities $\rho(x, x_i) < \frac{1}{3}\delta_{x_i}$ and $\rho(x, x_j) < \frac{1}{3}\delta_{x_j}$ are valid. It follows that

$$\rho(x_i, x_j) \le \rho(x, x_i) + \rho(x, y) + \rho(y, x_j) < \max\{\delta_{x_i}, \delta_{x_j}\},$$

which shows that $f(x_i, x_j) < \varepsilon / 3$. Finally, using the inequality

$$\sigma(f(x), f(y)) \le \sigma(f(x), f(x_i)) + \sigma(f(x_i), f(x_j)) + \sigma(f(x_j), f(y))$$

we deduce that $\sigma(f(x), f(y)) < \varepsilon$ whenever $\rho(x, y) < \delta$. ◇

3.3. <u>Remark</u>. So far we referred to the *size* of the neighborhoods in metric spaces when we discussed about fundamental sequences and about uniform continuity. We mention that such aspects are specific to the so called *uniform topological spaces*. In this framework, theorem 3.2 from above takes a very general form, as follows: *If $\mathscr{X}$ and $\mathscr{Y}$ are uniform topological spaces, $A\subseteq\mathscr{X}$ is compact, and $f : A\to\mathscr{Y}$ is continuous on A, then f is uniformly continuous on A.*

Another example of specific topological properties, which rise in metric spaces, refers to compactness. In fact, compact sets are always closed, but in addition, in metric spaces they are also bounded. More exactly:

3.4. <u>Proposition</u>. If $(\mathscr{S}, \rho)$ is a metric space, and $K\subseteq\mathscr{S}$ is a compact set relative to the corresponding metric topology τ, then K is bounded.

<u>Proof</u>. Let us consider the following family of open spheres

$$\mathscr{A} = \{S(x, n): n\in\mathbb{N}\},$$

where x is fixed (in K, say). Because $\mathscr{A}$ covers any subset of $\mathscr{S}$, and K is compact, it follows that K has a finite sub-cover $\mathscr{A}^*$. In addition, $n\leq m$ implies $S(x, n)\subseteq S(x, m)$, hence the greatest sphere from $\mathscr{A}^*$ contains K. The existence of such a sphere means that K is bounded. ◇

The possibility of expressing compactness in terms of convergence is an important facility in metric spaces. To develop this idea, we will consider other types of compactness, namely:

3.5. <u>Definition</u>. A set K in a metric space $(\mathscr{S}, \rho)$ is said to be *sequentially compact* (briefly *s.c.*) iff each sequence (x_n) from K contains a sub-sequence (x_{n_k}), which is convergent to some $x_0\in K$.

We say that the set $K\subseteq\mathscr{S}$ is ε*– compact* (briefly ε – c.) iff for any $\varepsilon > 0$ there exists a finite family of open spheres of radiuses ε, which covers K.

3.6. <u>Examples</u>. (*i*) Every compact set is s.c. as well as ε – c. (obviously). There are still ε – c. sets, which are not compact, as for example

$$K = \{\tfrac{1}{n}: n\in\mathbb{N}^*\},$$

in $\mathbb{R}$, relative to the Euclidean metric.

(*ii*) The ε – c. sets are bounded, since the union of two spheres is contained in a greater one. The converse is generally false, i.e. boundedness is not enough for ε– compactness, as in the case of the balls in $\mathbf{C}_{\mathbb{R}}([a, b])$, endowed with the *sup* norm.

(*iii*) The same set K from the former example (*i*) shows that the ε – c. sets are not necessarily s.c. However, the converse is valid, namely:

3.7. <u>Proposition</u>. Let $(\mathscr{S}, \rho)$ be a metric space. If $K\subseteq\mathscr{S}$ is a sequentially compact set, then it is also ε – compact.

Proof. Let us suppose that, by contrary, there exists some $\varepsilon > 0$ such that K cannot be covered by finite families of open spheres of radius ε. Then, starting with an arbitrary $x_0 \in K$, and this ε, we can find the elements

$$x_1 \in K \setminus S(x_0, \varepsilon),\ x_2 \in K \setminus [S(x_0, \varepsilon) \cup S(x_1, \varepsilon)],\ \ldots$$

The resulting sequence (x_n) has the property that $\rho(x_n, x_m) \geq \varepsilon$ holds for all $n, m \in \mathbb{N}$, which makes it unable to contain convergent subsequences. ◇

To establish the main result concerning the compact sets in metric spaces, namely the equivalence between compactness and sequential compactness, we need the following:

3.8. Lemma. If K is a s.c. set in a metric space $(\mathscr{S}, \rho)$, and $\mathscr{A}$ is an open cover of K, then there exists $\varepsilon > 0$ such that for all $x \in K$ we have $S(x, \varepsilon) \subseteq A$ for some $A \in \mathscr{A}$.

Proof. Let $\mathscr{A} = \{A_i \in \mathscr{G} : i \in I\}$ be a cover of K, and let us suppose that the assertion isn't true, i.e. for any $\varepsilon > 0$ there exists $x \in K$ such that $S(x, \varepsilon) \not\subset A_i$ holds for all $i \in I$. In particular, taking $\varepsilon = \frac{1}{n}$, where $n \in \mathbb{N}^*$, we find $x_n \in K$ such that $S(x_n, \frac{1}{n}) \not\subset A_i$ for all $i \in I$. Because K is supposed to be s.c, the resulting sequence (x_n) contains a subsequence, say (x_{n_k}), convergent to some $\xi \in K$. Let $j \in I$ be the index for which $\xi \in A_j$. More that this, because A_j is open, we have $S(\xi, r) \subseteq A_j$ for some $r > 0$. On the other hand, from the convergence of (x_{n_k}) to ξ it follows that $x_{n_k} \in S(\xi, \frac{r}{2})$ holds if k overpasses certain value k_0. If we take k great enough to obtain $\frac{1}{n_k} < \frac{r}{2}$, then finally $S(x_{n_k}, \frac{1}{n_k}) \subseteq S(\xi, r) \subseteq A_j$, contrarily to the initial hypothesis. ◇

3.9. Theorem. A set K in a metric space $(\mathscr{S}, \rho)$ is compact if and only if it is sequentially compact.

Proof. Each compact set in $\mathscr{S}$ is s.c. since the metric spaces are separated.

Conversely, let us suppose that $K \subseteq \mathscr{S}$ is sequentially compact, and let $\mathscr{A} = \{A_i : i \in I\}$ be an open cover of K. Using the above lemma, let $\varepsilon > 0$ be the number for which to each point $x \in K$ there corresponds $j \in I$ such that $S(x, \varepsilon) \subseteq A_j$. Finally, to obtain the necessary finite sub-cover of K we may proceed as in proposition 3.7. ◇

3.10. Corollary. Every closed and bounded interval $[a, b] \subset \mathbb{R}$ is a compact set (relative to the Euclidean topology of $\mathbb{R}$).

Proof. In terms of compactness, the Cesàro-Weierstrass theorem (e.g. see II.1.19) says that $[a, b]$ is sequentially compact.

Because the Euclidean topological spaces are particular metric spaces, there are specific properties in addition to the metric ones, as for example:

3.11 Theorem. A set $K \subset \Gamma^n$, where $n \in \mathbb{N}^*$, is compact if and only if it is closed and bounded.

Proof. The compact sets are closed in any separated topological space. They are bounded in each metric space (see proposition 3.4. from above), hence in Γ^n they are closed and bounded.

Conversely, if K is bounded, then $K \subseteq \overline{S}(0, r)$ holds for some $r > 0$. Because $\overline{S}(0, r)$ is sequentially compact, and K is a closed part, it follows that K itself is s.c., hence, according to theorem 3.9, it is compact. ◇

The following theorem shows that the compactness of the closed and bounded sets is specific to finite dimensional spaces:

3.12. Theorem. Let $(\mathscr{L}, \| \cdot \|)$ be a linear normed space, and let

$$K = \overline{S}(0, r) = \{x \in \mathscr{L} : ||x|| \leq 1\}$$

be the closed unit ball. If K is compact, then dim $\mathscr{L}$ is finite.

Proof. Because $V = S(0, \frac{1}{2})$ is open, and $\mathscr{A} = \{x + V : x \in K\}$ covers K, it follows that there exist $x_1, x_2, \ldots, x_n \in K$ such that

$$K \subseteq (x_1 + V) \cup (x_2 + V) \cup \ldots \cup (x_n + V) .$$

If $\mathscr{L}_0 = \mathrm{Lin}\{x_1, x_2, \ldots, x_n\}$ denotes the linear space spanned by these vectors, then dim $\mathscr{L}_0 \leq n$, hence it is closed in $\mathscr{L}$ as a finite dimensional subspace. Because $2V \subset K \subset \mathscr{L}_0 + V$, and $\lambda \mathscr{L}_0 \subseteq \mathscr{L}_0$ for every $\lambda \neq 0$, we obtain the inclusion $V \subset \mathscr{L}_0 + \frac{1}{2} V$, and successively

$$K \subset \mathscr{L}_0 + \frac{1}{2} V \subset \mathscr{L}_0 + \frac{1}{4} V \subset \ldots \subset \mathscr{L}_0 + 2^{-n} V \subset \ldots .$$

Consequently, $K \subset \overline{\mathscr{L}_0} = \mathscr{L}_0$. Since $\lambda \mathscr{L}_0 \subseteq \mathscr{L}_0$ for each $\lambda \in \Gamma$, and $K \subset \mathscr{L}_0$, we have $\Gamma \cdot K \subset \mathscr{L}_0$. On the other hand, $\Gamma \cdot K = \mathscr{L}$, i.e. for every $x \in \mathscr{L}$ there exist $\lambda \in \Gamma$ (e.g. $\lambda = \|x\|$) and $x^* \in K$ such that $\lambda x^* = x$. Consequently, we deduce that $\mathscr{L} \subseteq \mathscr{L}_0$, hence dim $\mathscr{L} \leq$ dim $\mathscr{L}_0 \leq n$. ◇

3.13. Remarks. a) The above theorem remains valid in the more general case of topological vector spaces, i.e. *any locally compact topological vector space* (that has a neighborhood of the origin with compact closure) *has finite dimensions.*

b) Using the above results we can find more examples of compact sets. In particular, if $a = (a_1, a_2, \ldots, a_n)$, $b = (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$, then the closed n-dimensional interval $[a, b] = [a_1, b_1] \times \ldots \times [a_n, b_n]$ is a compact set in $\mathbb{R}^n$. Since the notion of *interval* is based on *order*, this construction doesn't work in the complex space $\mathbb{C}^n$. However, using the above theorem 3.12 in the case $\mathscr{L} = \Gamma$, we obtain a similar examples by replacing the closed intervals by closed spheres.

c) Other specific properties in Euclidean spaces, besides those involving compactness, derive from the fact that Γ^n represents a Cartesian product. In addition, the Euclidean topology of Γ^n can be considered a product topology of Γ, n times by itself. Therefore, plenty of properties concerning the limiting process in Γ^n, e.g. continuity and convergence, naturally reduce to similar properties in Γ. In order to make more explicit this fact we have to precise some terms and notation:

3.14. <u>Definition</u>. Let $X \neq \emptyset$ be an arbitrary set, and $f : X \to \Gamma^n$, be a vector function for some $n \in \mathbb{N}^*$. As usually, we define the *projections* $Pr_k : \Gamma^n \to \Gamma$ by $Pr_k(y_1, \ldots, y_n) = y_k$ for all $k = \overline{1,n}$. The functions $f_k = Pr_k \circ f$, where $k = \overline{1,n}$, are called *components* of f, and for every $x \in X$ we note

$$f(x) = (f_1(x), f_2(x), \ldots, f_n(x)).$$

If f is a (generalized) sequence, i.e. $X = \mathbb{N}$ (or a directed set D), then the components of f are called *component sequences*, and the general term of the vector sequence is noted $f(p) = y_p = (y_p^{\,1}, \ldots, y_p^{\,n})$ for each $p \in \mathbb{N}$. Alternatively, the sequence is written as the set of n component sequences, namely

$$(y_p) = ((y_p^{\,1}), \ldots, (y_p^{\,n})).$$

In particular, if $n = 1$, and $\Gamma = \mathbb{C} = \mathbb{R}^2$, the components of f are named *real* and *imaginary parts* of f. Most frequently, f is defined on a domain (i.e. open and connected set) $D \subseteq \mathbb{C}$, and we note $f : D \to \mathbb{C}$, where

$$f(z) = P(x, y) + i\,Q(x, y)$$

at any $z = x + iy \in D$. In brief, $f = P + iQ$, where $P = \mathrm{Re}\, f$ and $Q = \mathrm{Im}\, f$ are the components of f.

If f is a sequence of complex numbers, i.e. $f(n) = z_n = x_n + iy_n \in \mathbb{C}$ is defined at any $n \in \mathbb{N}$, then (x_n) is the *sequence of real parts*, and (y_n) is the *sequence of imaginary parts* of f (alternatively noted (z_n)).

3.15. <u>Theorem</u>. Let (X, τ) be a topological space, and let $f : X \to \Gamma^n$ be a function of components $f_1, \ldots, f_n$. If $x_0 \in X$ is fixed, then:

a) $\ell = (\ell_1, \ldots, \ell_n) \in \Gamma^n$ is the limit of f at x_0 iff each component ℓ_k is the limit of f_k at x_0, and

b) f is continuous at x_0 iff each f_k is continuous at this point.

<u>Proof</u>. a) We may either extend the proposition III.2.10 by induction upon $n \in \mathbb{N}^*$, or directly involve the Euclidean metric ρ of Γ^n, namely

$$\rho(\ell, y) = \sqrt{\sum_{k=1}^{n} |\ell_k - y_k|^2}\,,$$

where $\ell = (\ell_1, \ldots, \ell_n) \in \Gamma^n$ and $y = (y_1, \ldots, y_n) \in \Gamma^n$. Following the second way, we put forward the double inequality

$$|\ell_m - y_m| \le \rho(\ell, y) \le \sum_{k=1}^{n} |\ell_k - y_k| ,$$

which holds for each $m = \overline{1,n}$ (the second inequality is the *triangle's rule*!).

If we consider the vector $y = f(x)$ of components $y_k = f_k(x)$, where x is arbitrary in a neighborhood $V \in \tau(x_0)$, then we may express the existence of the *limit* ℓ by the condition:

$$\forall \varepsilon > 0 \;\; \exists V \in \tau(x_0) \;\; \forall x \in V \;\; \Rightarrow \rho(\ell, f(x)) < \varepsilon .$$

Similarly, the existence of the limit ℓ_k for each $k = \overline{1,n}$ means that

$$\forall \varepsilon_k > 0 \;\; \exists V \in \tau(x_0) \;\; \textit{such that} \; \forall x \in V \;\; \Rightarrow \; |\ell_k - y_k| < \varepsilon_k .$$

The first inequality from above shows that

$$\rho(\ell, y) < \varepsilon \;\; \Rightarrow \;\; |\ell_k - y_k| < \varepsilon ,$$

hence the existence of ℓ assures the existence of ℓ_k for each $k = \overline{1,n}$.

Conversely, let the limits ℓ_k exist for all $k = \overline{1,n}$, and let $\varepsilon > 0$ be given. If we introduce $\varepsilon_k = \varepsilon / n$ in the conditions concerning each ℓ_k, we obtain a set of neighborhoods V_k, so we may construct $V = \cap\{V_k : k = \overline{1,n}\} \in \tau(x_0)$. It is easy to see that $x \in V$ leads to $\rho(\ell, y) < \varepsilon$, hence ℓ exists.

b) In addition to a) we take $\ell = f(x_0)$, which is equivalent to $\ell_k = f_k(x_0)$ for all $k = \overline{1,n}$. ◇

3.16 Corollary. If D is a domain in the complex plane $\mathbb{C}$, and the function $f : D \to \mathbb{C}$ has the components $P = \mathrm{Re}\, f$ and $Q = \mathrm{Im}\, f$, then the limits of f, P and Q at any $z_0 = x_0 + iy_0 \in D$ are in the relations:

a) $\zeta = \xi + i\eta = \lim\limits_{z \to z_0} f(z)$ iff

$$\xi = \lim_{(x,y) \to (x_0, y_0)} P(x, y) \;\text{ and }\; \eta = \lim_{(x,y) \to (x_0, y_0)} Q(x, y)$$

at the corresponding point $(x_0, y_0) \in \mathbb{R}^2$;

b) f is continuous at z_0 iff both P and Q are continuous at (x_0, y_0).

Proof. We may identify $\mathbb{C}$ to $\mathbb{R}^2$ from topological point of view, and reduce the limit of f to those of its components. ◇

3.17. Theorem. A vector sequence $(y_p) = ((y_p^1), \ldots, (y_p^n))$, $p \in \mathbb{N}$, in Γ^n, is bounded (convergent, or fundamental) iff all its component sequences (y_p^k), $k = \overline{1,n}$, are so.

Proof. As in the proof of theorem 3.15, we may use the inequalities

$$| x^m - y_p^m | \le \rho(x, y_p) \le \sum_{k=1}^{n} \left| x^k - y_p^k \right| ,$$

which hold for any $x = (x^1, \ldots, x^n)$, $p \in \mathbb{N}$, and $m = \overline{1,n}$. The analysis of the boundedness involves a fixed x. To establish the property of convergence, we take $x = \lim\limits_{p\to\infty} y_p$. Similarly, to study the Cauchy's property, we may replace x by y_q . ◇

3.18. Corollary. A sequence of complex numbers is bounded (convergent, fundamental) iff both sequences of real and imaginary parts are bounded (convergent, respectively fundamental).
Proof. We consider $\mathbb{C} = \mathbb{R}^2$ in the above theorem. ◇

3.19. Corollary. The Euclidean spaces Γ^n are complete for any $n \in \mathbb{N}^*$.
Proof. According to the theorem in III.2, $\mathbb{R}$ is complete, i.e. we have

$$convergent \equiv fundamental$$

for any sequence of real numbers. Consequently, this identity of properties holds in $\mathbb{C} = \mathbb{R}^2$, as well as in any Cartesian product Γ^n. ◇

The above properties concerning the sequences in Γ^n can be easily extended to nets (i.e. generalized sequences). The detailed analysis of this possibility is left to the reader.

3.20. Remark. In practice we frequently face the problem of comparing the convergence of the series (a) $\sum x_n$, (b) $\sum \|x_n\|$, and (c) $\sum \|x_n\|^2$ in a scalar product space, particularly when the system $\{x_n : n \in \mathbb{N}\}$ is orthogonal. As a general rule, the convergence of the series (b) implies that of (a) and (c). The real series $\sum \frac{(-1)^n}{n}$ shows that none of the converse implications is generally true. Using the series of terms $y_n = 1/n$, and $z_n = \frac{(-1)^n}{\sqrt{n}}$, we see that the series of types (a) and (c) do converge independently. However, the equivalence (a)$\Leftrightarrow$(c) holds for series of orthogonal elements. In fact, using the continuity of the scalar product, if $x = \sum x_n$, then the following extended Pythagoras' formula holds

$$<x, x> = <\sum x_n , x> = \sum \ <x_n, x> = \sum \|x_n\|^2 .$$

Consequently, if $\sum \|x_n\|^2$ is convergent, and s_p, s_q are some partial sums of the series $\sum x_n$, then (assuming $p < q$), we have

$$\|s_q - s_p\|^2 = \sum_{n=p+1}^{q} \|x_n\|^2 .$$

Even so, the convergence of (b) is not generally implied by the others, as in the case of $x_n = \left(\frac{1}{n}\delta_n^i\right)_{i\in N}$, where δ_n^i is the Kronecker delta.

PROBLEMS §III.3.

1. Formulate the main topological notions (e.g. adherence, interior, limit, convergence, continuity, etc.) in terms of metrics.
Hint. For example, $\ell = \lim_{x \to x_0} f(x)$ takes the form

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \textit{such that} \ \rho(x, x_0) < \delta \Rightarrow \sigma(f(x), \ell) < \varepsilon .$$

2. Let (x_n) be a sequence in a metric space, such that (x_{2n}) and (x_{2n+1}) are convergent subsequences. Show that the sequences (x_n) and (x_{3n}) are simultaneously convergent.
Hint. Any subsequence of a convergent sequence is convergent to the same limit. Conversely, the convergence of (x_{3n}) implies

$$\lim x_{2n} = \lim x_{2n+1} = \lim x_n.$$

3. Show that the function $f:[1,\infty) \to [1,\infty)$, expressed by $f(x) = x + \frac{1}{x}$, has no fixed point, even if $|f(x) - f(y)| < |x - y|$ at any $x \neq y$.
Hint. $x = f(x)$ is impossible because $1/x \neq 0$. Because of the relation

$$|f(x) - f(y)| = (1 - \frac{1}{xy})|x - y|,$$

the claimed inequality is obvious, but f is not contraction.

4. Let $\mathscr{S}$ denote the set $\mathbb{R}$ or any interval $(-\infty, a]$, $[a, b]$, $[b, +\infty)$ of $\mathbb{R}$, and let $f: \mathscr{S} \to \mathscr{S}$ be a derivable function on $\mathscr{S}$. Show that f is a contraction if there exists $c < 1$ such that $|f'(x)| \leq c$ at any $x \in \mathscr{S}$. In particular, analyze the possibility of approximating a solution of the equation

$$x = \alpha \sin x + \beta \cos x + \gamma$$

making discussion upon $\alpha, \beta, \gamma \in \mathbb{R}$.
Hint. Use the Lagrange's theorem about finite increments, which assures the existence of $\xi \in (x, y)$ such that $f(x) - f(y) = f'(\xi)(x - y)$ holds whenever $x, y \in \mathscr{S}$, with $x < y$. Write the particular equation using the function $f(x) = \sqrt{\alpha^2 + \beta^2} \sin(x + \varphi) + \gamma$, where f is a contraction if $\sqrt{\alpha^2 + \beta^2} < 1$; otherwise the method of successive approximation doesn't work, even if the equation always has at least one solution.

5. Let $(\mathscr{X}, \rho)$, $(\mathscr{Y}, \sigma)$ and $(\mathscr{Z}, \zeta)$ be metric spaces. Show that if the functions $f: \mathscr{X} \to \mathscr{Y}$ and $g: \mathscr{Y} \to \mathscr{Z}$ are uniformly continuous, then $g \circ f$ is uniformly continuous too. Is the converse implication true?

Hint. The property is directly based on definitions. The converse assertion is generally false, as for $\mathscr{X} = \mathscr{Y} = \mathscr{Z} = \mathbb{R}$, $g = |\cdot|$ and

$$f(x) = \begin{cases} x & if\, x \in \mathbb{Q} \\ -x & if\, x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

6. Let us fix a set $A \subseteq \mathbb{R}$ (or $A \subseteq \mathbb{C}$). Show that if any continuous function $f: A \to \mathbb{R}$ is bounded, then A is compact.

Hint. Taking $f(x) = x$, it follows that A is bounded. For any $\xi \in \overline{A} \setminus A$, the function $f_\xi(x) = |x - \xi|^{-1}$ is continuous but unbounded, hence we necessarily have $A = \overline{A}$, i.e. A is closed.

7. Let $(\mathscr{X}, \rho)$ and $(\mathscr{Y}, \sigma)$ be metric spaces. We say that $f: \mathscr{X} \to \mathscr{Y}$ is a *Lipschitzean* function iff there exists $L > 0$, called Lipschitz constant, such that the inequality

$$\sigma(f(x), f(y)) \leq L\, \rho(x, y)$$

holds for arbitrary $x, y \in \mathscr{X}$. Show that, in the case $(\mathscr{X}, \rho) \equiv (\mathscr{Y}, \sigma)$, the following implications hold:

f is a contraction $\Rightarrow$ *f is Lipschitzean* $\Rightarrow$
$\Rightarrow$ *f is uniformly continuous* $\Rightarrow$ *f is continuous*,

but none of their converses is generally true. More particularly, if $\mathscr{X}$ is an interval of $\mathbb{R}$, place the property "*f has (bounded) derivative on* $\mathscr{X}$" between the above properties.

Hint. Use functions like ax^2 or $x \sin\frac{1}{x}$ on intervals of $\mathbb{R}$. Any function with bounded derivative is Lipschitzean. The function $\sqrt{x}$ is a counter-example for the converse implication.

8. Let $X \neq \emptyset$ be an arbitrary set. Find a metric on X such that all the real functions $f: X \to \mathbb{R}$ (where $\mathbb{R}$ is Euclidean) are continuous on X.

Hint. The continuity of the characteristic functions attached to the point-wise sets $\{x\}$ shows that the only possible topology on X is discrete.

§ III.4. CONTINUOUS LINEAR OPERATORS

The *linear* functions play an important role in Analysis because of their simple form and convenient properties. For example, they are used in local approximations of a function, in *integral* calculus, in dynamical systems theory, etc. When the functions act between linear spaces and they are linear, we use to call them *operators*. In particular, the term *functional* is preferred whenever the target space is Γ, i.e. it takes scalar values.

In this section we mainly study the *continuous* linear functionals and operators acting between *normed* linear spaces, when continuity and other topological properties can be considered relative to the intrinsic topologies of the involved spaces. For example, the operator $U:(\mathscr{X}, \| \cdot \|_{\mathscr{X}}) \to (\mathscr{Y}, \| \cdot \|_{\mathscr{Y}})$ is *continuous* at the point $x_0 \in \mathscr{X}$ iff for each $\varepsilon > 0$ there exists $\delta > 0$ such that $\|U(x) - U(x_0)\| < \varepsilon$ holds whenever $\| x - x_0 \| < \delta$. Genarally speaking, if there is no danger of confusing the norms on $\mathscr{X}$ and $\mathscr{Y}$, we may renounce the distinctive notation $\|\cdot\|_{\mathscr{X}}$, $\|\cdot\|_{\mathscr{Y}}$, and mark all of then by $\| \cdot \|$.

According to the following property we may simply speak of *continuous linear operators* without mentioning the particular point x_0 any more.

4.1. Proposition. Every linear operator U: $(\mathscr{X}, \| \cdot \|) \to (\mathscr{Y}, \| \cdot \|)$ is continuous on $\mathscr{X}$ (i.e. at each point of $\mathscr{X}$) if and only if it is continuous at the origin $0 \in \mathscr{X}$.

Proof. The essential part is the "if" implication, so let U be continuous at 0, and let $x_0 \in \mathscr{X}$ and $\varepsilon > 0$ be arbitrary. Using the continuity of U at 0, we find $\delta > 0$ such that $\|y\| < \delta$ implies $\|U(y)\| < \varepsilon$. If we note $y = x - x_0$, then $U(x) - U(x_0) = U(y)$ follows from the linearity of U. Consequently, the condition $\|x - x_0\| < \delta$ implies $\|U(x) - U(x_0)\| < \varepsilon$. ◇

4.2. Corollary. The continuity of the linear operator is *uniform.*

Proof. In the proof of the above proposition, δ depends only on ε, i.e. it is the same for all $x_0 \in \mathscr{X}$. ◇

4.3. Definition. Let $(\mathscr{X}, \| \cdot \|)$ and $(\mathscr{Y}, \| \cdot \|)$ be normed linear spaces, and let U: $\mathscr{X} \to \mathscr{Y}$ be a linear operator. We say that U is *bounded* on $\mathscr{X}$ iff there exists a real number $\mu > 0$ such that the inequality

$$\|U(x)\| \leq \mu\|x\|$$

holds at any $x \in \mathscr{X}$. The set of all bounded operators between $\mathscr{X}$ and $\mathscr{Y}$ is noted by $\mathscr{B}(\mathscr{X}, \mathscr{Y})$. In particular, the set $\mathscr{B}(\mathscr{X}, \Gamma)$ of bounded functionals on $\mathscr{X}$ is called *topological dual of* $\mathscr{X}$, and noted $\mathscr{X}'$.

The following theorem explains why for linear operators we may identify the notions of boundedness and continuity.

4.4. Theorem. If $U: (\mathscr{X}, \| \cdot \|) \to (\mathscr{Y}, \| \cdot \|)$ is a linear operator, then it is continuous if and only if it is bounded.

Proof. We easily see that a bounded operator is continuous at the origin 0. Consequently, according to proposition 4.1, it is continuous on $\mathscr{X}$.

Conversely, if U is continuous at the origin $0\in\mathscr{X}$, then for arbitrary $\varepsilon > 0$, hence also for $\varepsilon = 1$, there exists $\delta > 0$ such that $\|x\| < \delta$ implies $\|U(x)\| < 1$. Except 0, where the condition of boundedness obviously holds, for any other $x\in\mathscr{X}$ we have $\|(\delta/2\|x\|)x\| < \delta$, hence $\|U((\delta/2\|x\|)x)\| < 1$. Since U is linear, the inequality $\|U(x)\| < (2/\delta)\|x\|$ holds at all $x\in\mathscr{X}$. This shows that the condition of boundedness is verified with $\mu = 2/\delta$. ◇

4.5. Theorem. The functional $\|\cdot\|^* : \mathscr{B}(\mathscr{X}, \mathscr{Y}) \to \mathbb{R}_+$, defined by

$$\|U\|^* = \inf \{ \mu \in \mathbb{R}_+ : \|U(x)\| < \mu \|x\| \textit{ for all } x\in \mathscr{X}\},$$

is a norm. In addition, this norm also allows the following expressions:

$$\|U\|^* = \sup \{\|U(x)\|: \|x\|\le 1\} =$$

$$= \sup \{\|U(x)\|: \|x\| = 1\} = \sup \left\{\frac{\|U(x)\|}{\|x\|} : x \neq 0\right\}.$$

Proof. Obviously, $\mathscr{B}(\mathscr{X}, \mathscr{Y})$ is a linear sub-space of $(\mathscr{L}(\mathscr{X}, \mathscr{Y}), +, \cdot)$. To establish that $\|\cdot\|^*$ is a norm, we have to prove the conditions $[N_1]$, $[N_2]$ and $[N_3]$ from the definition I.4.15. For example $[N_1]$, i.e. $\|U\|^* = 0 \Leftrightarrow U = 0$, is directly based on the definition of $\|\cdot\|^*$.

To prove $[N_2]$ it is enough to remark that for any $\lambda \neq 0$ we have

$$\|(\lambda U)(x)\|\le \mu \|x\| \Leftrightarrow \|U(|\lambda| x) \le \frac{\mu}{|\lambda|} \| |\lambda| x \|,$$

because it leads to the following expression of $\|\lambda U\|^*$:

$$\|\lambda U\|^* = \inf \{ \mu > 0 : \|(\lambda U)(x)\| < \mu \|x\| \textit{ for all } x\in \mathscr{X}\} =$$

$$= |\lambda| \inf \left\{\frac{\mu}{|\lambda|} > 0 : \|U(y)\| < \frac{\mu}{|\lambda|}\|y\| \textit{ for all } y\in\mathscr{X}\right\}.$$

The case $\lambda = 0$ in $[N_2]$ is trivial.

Finally, for $[N_3]$, let λ and μ be positive numbers showing that U and V are bounded operators. Because $\|U(x)\| < \mu\|x\|$ and $\|V(x)\| < \lambda\|x\|$ imply

$$\|(U + V)(x)\| \le (\lambda + \mu)\|x\|,$$

it follows that $\|U + V\|^* \le \lambda + \mu$. It remains to take here inf.

To conclude, $\|\cdot\|^*$ is a norm.

To obtain the other forms of $\|U\|^*$, let us note $\nu = \sup \left\{\frac{\|U(x)\|}{\|x\|} : x \neq 0\right\}$. In other words, for arbitrary $\varepsilon > 0$, and $x \neq 0$, we have $\|U(x)\| < (\nu + \varepsilon)\|x\|$, i.e. the number $\nu + \varepsilon$ verifies the condition of boundedness. Consequently, $\|U\|^* \leq \nu + \varepsilon$ holds with arbitrary $\varepsilon > 0$, hence also $\|U\|^* \leq \nu$.

Conversely, if μ is a number for which $\frac{\|U(x)\|}{\|x\|} < \mu$, then we can show that $\nu \leq \mu$ holds too. First, because $\|U\|^*$ is the infimum of such μ's, it follows that $\nu \leq \|U\|^*$. Thus we may conclude that $\|U\|^* = \sup \left\{\frac{\|U(x)\|}{\|x\|} : x \neq 0\right\}$.

Because $\left\| \frac{1}{\|x\|} x \right\| = 1$, and $\frac{\|U(x)\|}{\|x\|} = \left\|U\left(\frac{1}{\|x\|} x\right)\right\|$, we obtain

$$\sup \{\|U(x)\| : \|x\| = 1\} = \sup \left\{\frac{\|U(x)\|}{\|x\|} : x \neq 0\right\}.$$

Finally, from $\{x \in \mathscr{X} : \|x\| = 1\} \subseteq \{x \in \mathscr{X} : \|x\| \leq 1\}$, we deduce that

$$\sup \{\|U(x)\| : \|x\| = 1\} \leq \sup \{\|U(x)\| : \|x\| \leq 1\},$$

while from $\|U(x)\| \leq \left\|U\left(\frac{1}{\|x\|} x\right)\right\|$, which is valid whenever $0 < \|x\| \leq 1$, we obtain the converse inequality. ◇

4.6. <u>Corollary</u>. For every $U \in \mathscr{B}(\mathscr{X}, \mathscr{Y})$ and $x \in \mathscr{X}$ we have

$$\|U(x)\| \leq \|U\|^* \|x\|.$$

<u>Proof</u>. According to the formula $\|U\|^* = \sup \left\{\frac{\|U(x)\|}{\|x\|} : x \neq 0\right\}$, it follows that $\frac{\|U(x)\|}{\|x\|} \leq \|U\|^*$ holds at any $x \in \mathscr{X} \setminus \{0\}$. ◇

4.7. <u>Corollary</u>. If $U : \mathscr{X} \to \mathscr{X}$ is a linear operator in the scalar product space $(\mathscr{X}, \langle \cdot , \cdot \rangle)$, then $U \in \mathscr{B}(\mathscr{X}, \mathscr{X})$ if and only if the inequality

$$| \langle U(x), y \rangle | \leq \mu \|x\| \|y\|$$

holds for some $\mu > 0$, at all $x, y \in \mathscr{X}$.

<u>Proof</u>. According to theorem 4.4, if U is continuous, then it is bounded, hence there exists $\mu > 0$ such that $\|U(x)\| \leq \mu \|x\|$ holds at each $x \in \mathscr{X}$. Using the fundamental inequality of a scalar product space, we obtain

$$| \langle U(x), y \rangle | \leq \|U(x)\| \|y\| \leq \mu \|x\| \|y\|.$$

Conversely, let us suppose that the inequality $| <U(x), y> | \leq \mu \|x\| \|y\|$ holds at each $x, y \in \mathscr{X}$. In particular, for $y = U(x)$, we obtain:

$$\|U(x)\|^2 \leq \mu \|x\| \|U(x)\|.$$

Considering the essential case $U(x) \neq 0$, this inequality shows that U is a bounded operator. ◇

4.8. Remark. In many practical problems (e.g. finding the differential of a given function) it is important to know the *form* of the linear functionals on a particular linear space. So far we can say that each linear functional on Γ has the form $f(x) = c\, x$ for some $c \in \Gamma$. More generally, the general form of the linear functionals on Γ^n is

$$f(x) = a_1 x_1 + a_2 x_2 + \dots + a_n x_n = < a, x >,$$

where $a = (a_1, a_2, \dots, a_n) \in \Gamma^n$ depends on f. It is a remarkable fact that a similar form is kept up in all Hilbert spaces:

4.9. Theorem. (F. Riesz) For every linear and continuous functional f in a Hilbert space $(\mathscr{X}, < . , . >)$ there exists a vector $y \in \mathscr{X}$ such that:

a) $f(x) = < x, y >$ at any $x \in \mathscr{X}$, and

b) $\| f \| = \|y\|$.

In addition, this vector is uniquely determined by f.

Proof. If $f = 0$, the assertion is proved by $y = 0$, hence we shall essentially analyze the case $f \neq 0$. Let $L = f^{\leftarrow}(0)$ be the null subspace of f. Since f is continuous, L is a closed linear subspace. Because $L \neq \mathscr{X}$ holds in this case, we can decompose the space as a direct sum $\mathscr{X} = L \oplus L^{\perp}$, where $L^{\perp} \neq \{0\}$. If we fix some $z \in L^{\perp} \setminus \{0\}$, then $f(z) \neq 0$, and to each $x \in \mathscr{X}$ we can attach the element $u = x - \dfrac{f(x)}{f(z)} z$. Because $f(u) = 0$, we deduce that $u \in L$, hence $< u, z > = 0$, and finally $< x, z > - \dfrac{f(x)}{f(z)} < z, z > = 0$. So we can evaluate

$$f(x) = \frac{f(z) < x, z >}{< z, z >} = < x, \frac{\overline{f(z)}}{\|z\|^2} z > ,$$

hence the asked element is $y = \overline{f(z)} \|z\|^{-2} z$.

In order to evaluate $\| f \|$, we may start with the fundamental inequality

$$| f(x)| = |< x, y >| \leq \|x\| \|y\|,$$

which shows that $\| f \| \leq \| y\|$.

On the other hand, if $x = y \neq 0$, the fundamental inequality becomes equality, i.e. $\dfrac{|f(y)|}{\|y\|} = \| y\|$. Because $\| f \| = \sup \left\{ \dfrac{|f(y)|}{\|y\|} : y \neq 0 \right\}$, it follows that $\| f \| \geq \| y\|$. The two contrary inequalities show that $\| f \| = \| y\|$.

Let us suppose that y is not unique, and let $y^{/} \in \mathscr{X}$ be another element for which the representation $f(x) = \langle x, y^{/} \rangle$ holds at any $x \in \mathscr{X}$. If compared to the initial form $f(x) = \langle x, y \rangle$, it shows that $\langle x, y - y^{/} \rangle = 0$ holds at each $x \in \mathscr{X}$. In particular, taking $x = y - y^{/}$, we obtain $\| y - y^{/} \| = 0$, which shows that $y = y^{/}$. Because this contradicts the supposition $y \neq y^{/}$, we may conclude that y is the single element in $\mathscr{X}$ for which a) holds. ◇

4.10. <u>Remark</u>. The above theorem shows that all the Hilbert spaces can be identified with their duals, i.e. $\mathscr{X} = \mathscr{X}^{/}$. This result has many important consequences in the theory of the adjoint and self-adjoint operators, in spectral theory, etc., as well as in practice (e.g. in Quantum Physics).

Because we mainly use linear functionals in order to develop the classical *Differential Calculus*, we continue their study from some other point of view, namely we will extend the above results to *multi-linear* continuous functionals. In this framework the starting space has the form

$$\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \ldots \times \mathscr{X}_n ,$$

where $(\mathscr{X}_k, \| \cdot \|_k)$, $k \in \{1, 2, \ldots, n\}$, are normed linear spaces over the same field Γ. It is easy to see that $(\mathscr{X}, \| \cdot \|)$ is a normed linear space too, if the norm $\| \cdot \|: \mathscr{X} \to \mathbb{R}_+$ is expressed at any $x = (x_1, x_2, \ldots, x_n) \in \mathscr{X}$ by

$$\|x\| = \max \{\|x_k\|_k : k = \overline{1,n}\}.$$

4.11. <u>Definition</u>. Let $(\mathscr{X}, \| \cdot \|)$ and $(\mathscr{Y}, \| \cdot \|)$ be normed linear spaces, where $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \ldots \times \mathscr{X}_n$ is organized as in the previous remark. The function $U: \mathscr{X} \to \mathscr{Y}$ is called *multi-linear* (more exactly *n-linear*) *operator* (*functional* in the case $\mathscr{Y} = \Gamma$) iff it is linear relative to each of its variables $x_k \in \mathscr{X}_k$, i.e. for all $k = \overline{1,n}$.

The *2* - linear functions are frequently called *bilinear*. Of course, it makes sense to speak of "*multi-* …" iff $n > 1$, but the results known for $n = 1$ should be recovered from this more general framework.

We say that the n-linear operator U is *bounded* iff there exists $\mu > 0$ such that the following inequality holds at any $x = (x_1, x_2, \ldots, x_n)$:

$$\|U(x)\| \leq \mu \|x_1\|_1 \|x_2\|_2 \ldots \|x_n\|_n .$$

For the sake of shortness, we write $\|U(x)\| \leq \mu \, \pi(x)$, where $\pi(x) = \prod_{k=1}^{n} \|x_k\|_k$.

The *continuity* of a multi-linear function refers to the natural (uniform) topologies of $(\mathscr{X}, \| \cdot \|)$ and $(\mathscr{Y}, \| \cdot \|)$.

4.12. <u>Examples</u>. If we take $\mathscr{X}_1 = \mathscr{X}_2 = \mathscr{Y} = \mathbb{R}$, then the function $f(x, y) = xy$ is bilinear, but not linear. In the same framework, $g(x, y) = x + y$ is linear but not bilinear, hence the notions *linear* and *n-linear* are independent. In particular, *1-linear* coincides with *linear*, but the term *n-linear* essentially refers to $n > 1$. Both f and g from above are continuous functions, while the function $h:\mathbf{C}_{\mathbb{R}}^1(K) \times \mathbf{C}_{\mathbb{R}}^1(K) \to \mathbb{R}$, of values $h(x, y) = x'(t_0)\, y'(t_0)$, where t_0 is fixed in the interior of the compact interval $K \subseteq \mathbb{R}$, is not continuous relative to the norms $\|x_k\|_k = \sup\ \{|x_k(t)|: t \in K\}$, $k \in \{1, 2\}$ (see also problem 3 at the end of the section).

The above theorem 4.4 can be extended to n-linear operators.

4.13. <u>Theorem</u>. In the terms of the above definition, the n-linear operator U is continuous on $\mathscr{X}$ if and only if it is bounded.

<u>Proof</u>. If U is continuous on $\mathscr{X}$, then it is continuous at the origin of $\mathscr{X}$ too. Let $\delta > 0$ be the number that corresponds to $\varepsilon = 1$ in this condition of continuity, i.e. $\|x\| < \delta$ implies $\|U(x)\| < 1$. Let x be a vector of components $x_k \neq 0_k \in \mathscr{X}_k$ for all $k = \overline{1,n}$, and let us note $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$, where $x_k^* = \dfrac{\delta}{2\sqrt{n}\|x_k\|_k}\, x_k$ for all $k = \overline{1,n}$. It is easy to see that $\| x^*\| < \delta$, hence $\|U(x^*)\| < 1$. Because U is n-linear, this inequality takes the form

$$\frac{\delta^n}{2^n(\sqrt{n})^n}\cdot\frac{1}{\|x_1\|_1\cdot\ldots\cdot\|x_n\|_n}\cdot\|U(x_1, \ldots, x_n)\| < 1,$$

or, equivalently, $\|U(x)\| < \left[\dfrac{2\sqrt{n}}{\delta}\right]^n \pi(x)$. In this case, $\mu = \left[\dfrac{2\sqrt{n}}{\delta}\right]^n$ is the constant in the condition of boundedness for U.

Otherwise, if $x_k = 0_k$ for some $k = \overline{1,n}$, then $U(x) = 0$, hence the condition of boundedness reduces to a trivial equality. So we conclude that U is a bounded operator.

Conversely, let the n-linear operator U be bounded. In order to prove its continuity at an arbitrary point $x_0 = (x_1^0, x_2^0, \ldots, x_n^0) \in \mathscr{X}$, we may evaluate

$$\|U(x) - U(x_0)\| =$$

$$= \left\| \sum_{k=1}^{n} \left[U(x_1^0, \ldots, x_{k-1}^0, x_k, x_{k+1}, \ldots, x_n) - U(x_1^0, \ldots, x_{k-1}^0, x_k^0, x_{k+1}, \ldots, x_n) \right] \right\| \leq$$

$$\leq \sum_{k=1}^{n} \left\| U(x_1^0, \ldots, x_{k-1}^0, x_k - x_k^0, x_{k+1}, \ldots, x_n) \right\| \leq$$

$$\leq \mu \sum_{k=1}^{n} \|x_1^0\|_1 \cdot\ldots\cdot \|x_{k-1}^0\|_{k-1} \cdot \|x_k - x_k^0\|_k \cdot \|x_{k+1}\|_{k+1} \cdot\ldots\cdot \|x_n\|_n .$$

Now let us remark that $\|x - x_0\| < 1$ implies

$$\|x_k\|_k \leq \|x_k - x_k^0\| + \|x_k^0\|_k < 1 + \|x_k^0\|_k .$$

If we note $\nu(x_0) = \prod_{k=1}^{n} (1 + \|x_k^0\|_k)$, then we may write

$$\|U(x) - U(x_0)\| \leq \mu\nu(x_0)(\sum_{k=1}^{n} \|x_k - x_k^0\|) \leq n\ \mu\nu(x_0)\|x - x_0\|.$$

To conclude, for each $\varepsilon > 0$ there exists $\delta = \min\{1, \dfrac{\varepsilon}{n \cdot \mu \cdot \nu(x_0)}\}$, such that $\|x - x_0\| < \delta$ implies $\|U(x) - U(x_0)\| < \varepsilon$, i.e. U is continuous at x_0 . ◇

4.14. Remark. It is easy to see that the set $\mathscr{L}(\mathscr{X}, \mathscr{Y})$, of all n–linear operators on $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2 \times \ldots \times \mathscr{X}_n$, is a linear space too. More than this, the set $\mathscr{B}_n(\mathscr{X}, \mathscr{Y})$, consisting of all continuous n-linear operators (which are also bounded, according to the above theorem), forms a linear space over Γ, and for each $U \in \mathscr{B}_n(\mathscr{X}, \mathscr{Y})$ it makes sense to consider

$$\|U\|^* = \sup\{\|U(x)\|: \|x\| \leq 1\}.$$

Following the same steps as in theorem 4.5 from above, we obtain a similar result relative to continuous n-linear operators, namely:

4.15. Theorem. The functional $\|\cdot\|^* : \mathscr{B}_n(\mathscr{X}, \mathscr{Y}) \to \mathbb{R}_+$ is a norm.

Finally, we analyze some problems of isomorphism between spaces of operators, which will be necessary to study the higher order differentials.

4.16. Definition. Let $(\mathscr{X}, \|\cdot\|)$ and $(\mathscr{Y}, \|\cdot\|)$ be normed linear on the same field Γ. These spaces are said to be *metrically isomorphic* iff there exists a function $\Phi : \mathscr{X} \to \mathscr{Y}$, called *metric isomorphism*, such that:

[I_1] Φ is a 1:1 correspondence between $\mathscr{X}$ and $\mathscr{Y}$;

[I_2] Φ is linear, i.e. $\Phi(\alpha x + \beta y) = \alpha\, \Phi(x) + \beta\, \Phi(x)$ holds at each $x, y \in \mathscr{X}$ and $\alpha, \beta \in \Gamma$

[I_3] Φ preserves the norm, i.e. we have $\|\Phi(x)\| = \|x\|$ at each $x \in \mathscr{X}$.

If only conditions [I_1] and [I_2] hold, we say that these spaces are *linearly isomorphic*, and Φ is called *linear isomorphism*.

4.17. Examples. The finite dimensional linear spaces over the same Γ are linearly isomorphic iff they have the same dimension. Even so, they are not metrically isomorphic, as for example the Euclidean Γ^n, and the space of all polynomial functions of degree strictly smaller than n, defined on some compact set, and endowed with the sup-norm.

The above theorem 4.9 (due to Riesz) shows that each Hilbert space is metrically isomorphic with its topological dual.

The following theorem establishes an isomorphic representation of the operators whose target space consists of operators.

4.18. Theorem. Let $(\mathscr{X}_1, \|.\|_1)$, $(\mathscr{X}_2, \|.\|_2)$ and $(\mathscr{Y}, \|.\|)$ be normed spaces, and let the space $\mathscr{X} = \mathscr{X}_1 \times \mathscr{X}_2$ be endowed with the above product norm $\|.\|$. If the spaces $(\mathscr{B}_2(\mathscr{X}, \mathscr{Y}), \|\cdot\|^*)$ and $(\mathscr{B}(\mathscr{X}_1, \mathscr{B}(\mathscr{X}_2, \mathscr{Y})), \|\cdot\|^{**})$ are normed according to theorem 4.5 and 4.15, then they are metrically isomorphic.

Proof. We start with the construction of the isomorphism. More exactly, to each $U\in \mathcal{B}_2(\mathcal{X}, \mathcal{Y})$, we have to attach a bounded linear operator defined at each $x\in \mathcal{X}_1$. Primarily we consider an operator $U_x : \mathcal{X}_2\to \mathcal{Y}$, of values $U_x(t) = U(x, t)$ at any $t\in \mathcal{X}_2$. Because U is a continuous bilinear operator, it follows that U_x is continuous and linear, hence $U_x\in \mathcal{B}(\mathcal{X}_2, \mathcal{Y})$.

Now we may construct $\underline{U}: \mathcal{X}_1\to \mathcal{B}(\mathcal{X}_2, \mathcal{Y})$, by the formula $\underline{U}(x) = U_x$. Obviously, at each $x, y \in \mathcal{X}_1$, $t\in \mathcal{X}_2$, and $\alpha, \beta\in \Gamma$, we have:

$$\underline{U}(\alpha x + \beta y)(t) = U_{\alpha x+\beta y}(t) = U(\alpha x + \beta y, t)= \alpha U(x, t) + \beta U(y, t) =$$
$$= \alpha U_x(t) + \beta U_y(t) = (\alpha U_x + \beta U_y)(t) = (\alpha \underline{U}(x) + \beta \underline{U}(y))(t) .$$

Consequently, $\underline{U}(\alpha x + \beta y) = \alpha \underline{U}(x) + \beta \underline{U}(y)$, i.e. $\underline{U}$ is linear.

In order for us to show that $\underline{U}$ is continuous, we may note the norm of $\mathcal{B}(\mathcal{X}_2,\mathcal{Y})$ by $\|\cdot\|_2^*$, and we evaluate

$$\|\underline{U}(x)\|_2^* = \|U_x\|_2^* = \sup\{\|U_x(t)\| : \|t\|_2 \leq 1\} = \sup\{\|U(x, t)\| : \|t\|_2 \leq 1\} \leq$$
$$\underset{\text{cont of } U}{\leq} \sup\{\mu \|x\|_1 \|t\|_2 : \|t\|_2 \leq 1\} \leq \mu \|x\|_1.$$

Consequently, $\underline{U} \in \mathcal{B}(\mathcal{X}_1, \mathcal{B}(\mathcal{X}_2, \mathcal{Y}))$.

On the other hand, we may interpret this construction as a description of the function $\Phi : \mathcal{B}_2(\mathcal{X}, \mathcal{Y}) \to \mathcal{B}(\mathcal{X}_1 , \mathcal{B}(\mathcal{X}_2 , \mathcal{Y}))$, of values

$$\Phi (U) = \underline{U} .$$

The rest of the proof is a study of Φ .

***Φ* is injective**. In fact, if $\Phi (U) = \Phi (V)$, then $\underline{U}(x) = \underline{V}(x)$ must hold at each $x\in\mathcal{X}_1$. From $U_x = V_x$ it follows that $U_x(t) = V_x(t)$ at each $t\in\mathcal{X}_2$. Consequently, we have $U(x, t) = V(x, t)$ at any $x\in \mathcal{X}_1$ and $t\in \mathcal{X}_2$, i.e. $U = V$.

***Φ* is surjective.** If $\mathfrak{U}\in \mathcal{B}(\mathcal{X}_1 , \mathcal{B}(\mathcal{X}_2, \mathcal{Y}))$, then at each $x\in \mathcal{X}_1$ we have $\mathfrak{U}(x)\in \mathcal{B}(\mathcal{X}_2, \mathcal{Y}))$, hence the operator $U: \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}$ is well defined by considering $U(x, t) = (\mathfrak{U}(x))(t)$. It is easy to see that U is a continuous operator, and $\Phi (U) =\mathfrak{U}$.

***Φ* is linear.** If $U, V \in \mathcal{B}_2(\mathcal{X}, \mathcal{Y})$, and $\alpha, \beta \in \Gamma$, then

$$\alpha \Phi (U) + \beta \Phi (V) = \alpha \underline{U} + \beta \underline{V} ,$$

and at every $x\in \mathcal{X}_1$ we have

$$(\alpha \underline{U} + \beta \underline{V})(x) = \alpha \underline{U}(x) + \beta \underline{V}(x) = \alpha U_x + \beta V_x.$$

Furthermore, at each $t \in\mathcal{X}_2$, we have

$$(\alpha U_x + \beta V_x)(t) = \alpha U_x(t) + \beta V_x(t) =$$
$$= \alpha U(x, t) + \beta V(x, t) = (\alpha U + \beta V)(x, t) = (\alpha U + \beta V)_x(t),$$

which shows that $\alpha U_x + \beta V_x = (\alpha U + \beta V)_x$.

In other terms, the equality

$$\alpha \underline{U} (x) + \beta \underline{V} (x) = (\underline{\alpha U + \beta V})(x)$$

holds at any $x\in \mathcal{X}_1$, i.e. Φ fulfils the condition

$$(\alpha \Phi (U) + \beta \Phi (V))(x) = \Phi (\alpha U + \beta V)(x).$$

Because x is arbitrary, we obtain

$$\alpha\, \Phi(U) + \beta\, \Phi(V) = \Phi(\alpha\, U + \beta\, V).$$

Φ preserves the norm. In fact, for each $U\in \mathscr{B}_2(\mathscr{X}, \mathscr{Y})$, we have :

$$\|\Phi(U)\|^{**} = \|\underline{U}\|^{**} = \sup\ \{\|\underline{U}(x)\|^{*} : \|x\|_1 \le 1\} = \sup\ \{\|U_x\|^{*} : \|x\|_1 \le 1\} =$$
$$= \sup\{[\ \sup\{\|U_x(t)\| : \|t\|_2 \le 1\}] : \|x\|_1 \le 1\} =$$
$$= \sup\ \{\|U(x,t)\| : \|x\|_1 \le 1,\ and\ \|t\|_2 \le 1\} =$$
$$= \sup\ \{\|U(x,t)\| : \|(x,t)\| \le 1\} = \|U\|^{*}.$$

To conclude, $\|\Phi(U)\|^{**} = \|U\|^{*}$. ◇

4.19. Remark. For practical purposes it is useful to know more particular bilinear functions. For example, in real Hilbert spaces, we may generate bilinear functions by linear operators, according to the formula

$$f(x, y) = \langle x, Uy\rangle.$$

In particular, each square matrix A with real elements generates a bilinear function on $\mathbb{R}^n$, according to a similar formula, $f(x, y) = \langle x, Ay\rangle$.

These examples make use of the fact that the scalar product itself is a bilinear function on real spaces. For the case of a complex space, there is another theory that takes into consideration the so called *skew symmetry* of the scalar product, and involves Hermitian operators and matrices, self-adjoint operators, general inner products, etc. Some elements of this sort will be discussed later.

The following proposition shows that the bilinear functions in the above examples have the most general form (in that frame).

4.20. Proposition. If $(\mathscr{X}, \langle\ .\ ,\ .\ \rangle)$ is a real Hilbert space, then to each continuous bilinear function $f : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ there corresponds a continuous linear operator $U : \mathscr{X} \to \mathscr{X}$ such that the equality

$$f(x, y) = \langle x, Uy\rangle$$

holds at each $x, y \in \mathscr{X}$.

Proof. Whenever we take $y\in\mathscr{X}$, it follows that the function $f(\ .\ , y) : \mathscr{X} \to \mathbb{R}$ is continuous and linear. Then, according to the Riesz' theorem, there exists $y^*\in\mathscr{X}$ such that $f(x, y) = \langle x, y^*\rangle$ at any $x\in\mathscr{X}$. Let us define U by $U(y) = y^*$. This operator is linear because for all $U(y) = y^*$, $U(z) = z^*$, and $\alpha, \beta\in\mathbb{R}$, we deduce that the following equalities are valid at each $x\in\mathscr{X}$:

$$\langle x, U(\alpha y + \beta z)\rangle = f(x, \alpha y + \beta z) = \alpha f(x, y) + \beta f(x, z) =$$
$$= \alpha \langle x, U(y)\rangle + \beta \langle x, U(z)\rangle = \langle x, \alpha y^* + \beta z^*\rangle .$$

According to the above theorem 4.13, f is bounded, hence there exists a real positive number μ, such that $|f(x, y)| \le \mu\, \|x\|\, \|y\|$ holds at any $x, y \in\mathscr{X}$. In particular, at $x = U(y)$, this inequality becomes

$$\|U(y)\|^2 \le \mu\, \|U(y)\|\, \|y\| ,$$

which proves the boundedness of U. ◇

4.21. Corollary. For each bilinear function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ there exists a matrix $A \in \mathcal{M}_n(\mathbb{R})$ which represents f in the sense that $f(x, y) = < x, Ay >$

holds at any $x, y \in \mathbb{R}^n$. More exactly, if we note the *transposed* matrix by the superscript ^{T}, so that $(x_1, x_2, \ldots, x_n) = X^T$, and $(y_1, y_2, \ldots, y_n) = Y^T$ represent the vectors x and y in some base $\mathfrak{B} = \{e_1, e_2, \ldots, e_n\}$ of $\mathbb{R}^n$, then

$$f(x, y) = X^T A\, Y = \sum_{i,j=1}^{n} a_{ij}\, x_i\, y_j \ ,$$

where A consists of the elements $a_{ij} = f(e_i, e_j)$, for all $i, j = \overline{1, n}$.

Proof. We represent the above operator U in the base $\mathfrak{B}$, as well as x, y and $< x, Uy >$. The continuity of f is implicitly assured since $\mathbb{R}^n$ has the finite dimension $n \in \mathbb{N}^*$ (see also problem 1). ◇

An important type of functions (also called *forms* in $\mathbb{R}^n$) derives from the n-linear functions by identifying the variables. This technique will be later used to connect the higher order differentials with the terms of a Taylor's development (see the next chapter).

4.22. Definition. Let $\mathscr{X}$ be a real linear space. If $f : \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is a bilinear function, then the function $\varphi : \mathscr{X} \to \mathbb{R}$, expressed by $\varphi(x) = f(x, x)$ at any $x \in \mathscr{X}$, is called *quadric* function. Similarly, if g is a three-linear function, then the function of values $\psi(x) = g(x, x, x)$ is called *cubic*, etc.

A bilinear function f is said to be *symmetric* iff $f(x, y) = f(y, x)$ holds at arbitrary $x, y \in \mathscr{X}$.

If $\mathscr{X} = \mathbb{R}^n$, then f, g, etc. (respectively φ, ψ, etc.) are called *forms.*

4.23. Remark. Sometimes, we prefer to define the quadratic functions by starting with symmetric bilinear functions. The advantage of this variant is that the forthcoming correspondence of f to φ is 1:1 (otherwise $f \neq g$ may yield the same quadratic function φ, e.g. $x_1 y_2$ and $x_2 y_1$ in $\mathbb{R}^2$). Going back, from the quadratic form φ to the generating bilinear symmetric function f (frequently called *the polar* of φ), is done by the formula

$$f(x, y) = [\, \varphi(x + y) - \varphi(x) - \varphi(y)] / 2.$$

The quadratic forms, i.e. the functions $\varphi : \mathbb{R}^n \to \mathbb{R}$, are always continuous. Their simplest description is that of *homogeneous polynomial* functions of the second degree, which has strong connections with the geometric theory of the conics and quadrics.

4.24. Example. The function $\varphi : \mathbb{R}^2 \to \mathbb{R}$, expressed at each $x = (x_1, x_2)$ by

$$\varphi(x) = x_1^2 + 4x_1 x_2 - 3x_2^2 \ ,$$

is generated by the (unique) symmetric bilinear form

$$f(x, y) = x_1 y_1 + 2(x_1 y_2 + x_2 y_1) - 3x_2 y_2 \, ,$$

which represents the *polar form* of φ. From the geometrical point of view, $\varphi(x) = k$ is the equation of a centered quadratic curve (conic) of matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & -3 \end{pmatrix}.$$

In addition, doing a convenient rotation, i.e. replacing

$$\begin{cases} x_1 = x_1^{/} \cos\alpha - x_2^{/} \sin\alpha \\ x_2 = x_1^{/} \sin\alpha + x_2^{/} \cos\alpha \end{cases}$$

we find another system of coordinates, in which the equation of this conic becomes *canonical*, namely $\lambda_1 x_1^{/\,2} + \lambda_2 x_2^{/\,2} = k^{/}$. The corresponding *canonical* form of φ in the new base has the simpler diagonal matrix

$$B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

The question is whether such a reduction to a form that contains only squares (represented by a diagonal matrix) is always possible and if "yes", then how is it concretely realizable? In order for us to get the answer, which will be positive, we need several results about the complex spaces (as mentioned in the above remark 4.19).

4.25. Definition. Let $\mathscr{X}$ be a (real or complex) linear space. We say that a function $f: \mathscr{X} \times \mathscr{X} \to \Gamma$ is *Hermitian* iff it fulfils the conditions:

$[H_1]$ $f(\alpha x + \beta y, z) = \alpha f(x, z) + \beta f(y, z)$ at all $x, y, z \in \mathscr{X}$, and $\alpha, \beta \in \Gamma$ (called *linearity relative to the first variable*);

$[H_2]$ $f(x, y) = \overline{f(y, x)}$ at all $x, y \in \mathscr{X}$ (called *skew symmetry*).

4.26. Remarks. a) If $\Gamma = \mathbb{R}$, then the condition $[H_2]$ reduces to the usual symmetry, hence the real Hermitian functions are symmetric (speaking of symmetric complex functions is also possible, but not very fruitful).

b) A property similar to proposition 4.20 holds, i.e. every Hermitian continuous function f on Hilbert space $(\mathscr{X}, <\cdot,\cdot>)$ has the form

$$f(x, y) = < x, Uy >,$$

where $U : \mathscr{X} \to \mathscr{X}$ is a continuous linear operator.

c) By analogy to Corollary 4.21, the Hermitian functions on Γ^n (where they all are continuous) are represented by matrices. More exactly, if the matrix $\overline{A} \in \mathcal{M}_n(\Gamma)$ represents the operator U in a particular base $\mathfrak{B}$ of $\mathscr{X} = \Gamma^n$, and the matrices $X, Y \in \mathcal{M}_{n\,1}(\Gamma)$ represent the vectors $x, y \in \mathscr{X}$, then

$$f(x, y) = X^T A \overline{Y} = \sum_{i,j=1}^{n} a_{ij} x_i \overline{y_j},$$

where $a_{ij} = f(e_i, e_j)$, with $e_i, e_j \in \mathfrak{B}$, for all $i, j = \overline{1,n}$.

d) Each matrix A, which represents a Hermitian function f as before, is said to be *Hermitian*. Its specific property, i.e. $A = \overline{A}^T$, is a reformulation of the relations $a_{ij} = \overline{a_{ji}}$, for all $i, j = \overline{1,n}$.

Operators, which are represented by Hermitian matrices, allow a general theory based on the notion of *adjoint operator*, as follows:

4.27. Theorem. Let $(\mathscr{X}, <\cdot,\cdot>)$ be a Hilbert space over the field Γ. If U is a continuous linear operator on $\mathscr{X}$, then there exists another (unique) linear and continuous operator U^* on $\mathscr{X}$, such that $< Ux, y > = < x, U^* y >$ holds at all $x, y \in \mathscr{X}$.

Proof. Obviously, if $U : \mathscr{X} \to \mathscr{X}$ is a continuous linear operator, then the functional $f : \mathscr{X} \times \mathscr{X} \to \Gamma$, defined by the formula $f(x, y) = < Ux, y >$, is continuous and Hermitian. If we fix $y \in \mathscr{X}$, then the function $f(\cdot, y)$ is continuous and linear. Consequently, in accordance to the Riesz' theorem, to each vector y there corresponds a (unique) vector $y^* \in \mathscr{X}$, such that

$$f(x, y) = < x, y^* >$$

holds at all $x \in \mathscr{X}$. It remains to note $y^* = Uy$, and to repeat the reason in the proof of proposition 4.20. ◇

4.28. Definition. The operator U^*, introduced by theorem 4.27 from above, is called *adjoint* of U. If $U^* = U$, i.e. the equality $< Ux, y > = < x, Uy >$ holds at all $x, y \in \mathscr{X}$, then U is called *self-adjoint* operator.

The self-adjoint operators have remarkable spectral properties:

4.29. Proposition. Let $(\mathscr{X}, <\cdot,\cdot>)$ be a Hilbert space. If $U : \mathscr{X} \to \mathscr{X}$ is a self-adjoint operator, then:

a) All its proper values are real (but not necessarily simple);
b) The proper vectors, which correspond to different proper values, are orthogonal each other;
c) If $\mathscr{X} = \Gamma^n$, then there exists a base consisting of proper vectors;
d) A matrix $A \in \mathcal{M}_n(\Gamma)$ represents U iff it is Hermitian, i.e. $A = \overline{A}^T$.

Proof. a) Replacing $Ux = \lambda x$ in the very definition $< Ux, y > = < x, Uy >$, we obtain $\lambda = \overline{\lambda}$, i.e. $\lambda \in \mathbb{R}$.

b) If $Ux = \lambda x$, and $Uy = \mu y$, where $\lambda, \mu \in \mathbb{R}$, $\lambda \neq \mu$, then the fact that U is self-adjoint leads to $(\lambda - \mu) < x, y > = 0$. Consequently, we have $x \perp y$.

c) If x_1 is a proper vector of U, then $x_1^{\perp}$ is an invariant subspace of U, and the restriction U_1 of U to this subspace is a self-adjoint operator too. Let x_2 be a proper vector of U_1, and let U_2 be the restriction of U to $\{x_1, x_2\}^{\perp}$. Here we find another proper vector, say x_3, etc. Because the dimension of $\mathscr{X} = \Gamma^n$ is finite, namely n, this process stops at the proper vector x_n.

d) If A represents U, the relation $< Ux, y > = < x, Uy >$ becomes

$$X^T A^T \overline{Y} = X^T \overline{A}\, \overline{Y}.$$

This means that the equalities $a_{ij} = \overline{a_{ji}}$ hold for all $i, j = \overline{1,n}$. ◇

Now we can discuss the fundamental theorem concerning the canonical form of a quadratic function on $\mathbb{R}^n$.

4.30. <u>Theorem</u>. For each quadratic from there exists an orthogonal base of $\mathbb{R}^n$, in which it reduces to a sum of $\pm$ squares.

<u>Proof</u>. Let the quadratic form φ be generated by the symmetric bilinear form f. The operator U, for which we have $f(x, y) = < x, Uy >$, obviously is self-adjoint. According to the above property 4.29. c), it has an orthogonal system of proper vectors. Because $a_{ij} = f(e_i, e_j) = < e_i, Ue_j >$, it follows that $a_{ij} = 0$ whenever $i \neq j$. Consequently, the system of proper vectors forms the sought base. ◇

4.31. <u>Remark</u>. The matrix A, which represents a quadratic form φ in its canonical form, has a diagonal shape. If this matrix contains some zeros in the diagonal (hence $\det A = 0$), we say that the form φ is *degenerate*. The other elements of the diagonal are either +1 or –1 (in the complex case the sign doesn't matter). Relative to the number p of positive, q of negative, and r of null coefficients of the squares in different canonical forms, the following theorem is very important:

4.32. <u>Theorem</u>. (Sylvester's *inertia* law) The numbers p, q, and r, of positive, negative, respectively null coefficients, are the same for all the canonical forms of a given (real) quadratic form.

The proof is purely algebraic, and therefore it is omitted here (see some algebra treatises). According to this "*law of inertia*", the triplet (p, q, r) represents an intrinsic property of each quadratic form, in the sense that it is the same in all the bases mentioned in theorem 4.30. This triplet is frequently called *signature*. Its usefulness is primarily seen in the process of classifying the quadratic forms, which, obviously, should be based on some intrinsic properties of these forms.

4.33. <u>Classification</u>. Let $\varphi : \mathscr{X} \to \mathbb{R}$ be a quadratic function on the real linear space $\mathscr{X}$. We distinguish the following situations:

- There exist $x, y \in \mathscr{X}$ such that $\varphi(x) > 0$ and $\varphi(y) < 0$, when we say that φ is *indefinite*; in the contrary case we say that φ is *semi-definite.*
- The semi-definite quadratic function φ vanishes only at the origin, i.e. $x \neq 0 \Rightarrow \varphi(x) \neq 0$. In this case we say that φ is *definite*.
- The semi-definite (possibly definite) quadratic function φ takes only positive values, i.e. $\varphi(x) \geq 0$ holds at all $x \in \mathscr{X}$, when we say that φ is *positive*. In the remaining case, when $\varphi(x) \leq 0$ holds at all $x \in \mathscr{X}$, we say that φ is *negative*.

We may easily reformulate this classification in terms of signature:

4.34. Proposition. If $\varphi:\mathbb{R}^n\to\mathbb{R}$ is a quadratic form of signature (p, q, r), then the following characterizations are valid:

a) φ is indefinite iff $p\neq 0\neq q$;

b) φ is positively (negatively) semi-definite iff $q=0$ ($p=0$);

c) φ is positively (negatively) definite iff $q=r=0$ ($p=r=0$).

4.35. Remark. Establishing the type of a quadratic form is useful in the study of the local extrema. Therefore according to the above proposition, our interest is to know more methods for finding the signature. The first one is directly based on the canonical form, which can be obtained by following the proof of the proposition 4.29.c. More exactly:

4.36. Theorem. A (real) quadratic form is positively (negatively) definite iff the proper values of the associated matrix are all positive (respectively negative). If this matrix has both positive and negative proper values, then the quadratic form is indefinite.

Now we mention without proof another test of definiteness, which is very useful in practice because its hypothesis asks to evaluate determinants:

4.37. Theorem. (Sylvester) Let φ be a quadratic form, and let

$$\Delta_1 = a_{11}, \Delta_2 = a_{11}\, a_{22} - a_{12}^2, \ldots, \Delta_n = \det A$$

be the principal minors of the associated matrix A. We have:

a) $\Delta_1 > 0, \Delta_2 > 0, \ldots, \Delta_n > 0$ iff φ is positively definite;

b) $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \ldots,(-1)^n \Delta_n > 0$ iff φ is negatively definite.

We conclude this section with a significant property of the definite forms, which will be used to obtain sufficient conditions for the existence of a local extremum.

4.38. Theorem. If $\varphi:\mathbb{R}^n\to\mathbb{R}$ is a positively definite quadratic form, then there exists $k > 0$ such that the inequality

$$\varphi(x) \geq k\,\|x\|^2$$

holds at all $x\in\mathbb{R}^n$.

Proof. Let $S = \{x\in\mathbb{R}^n : \|x\| = 1\}$ be the unit sphere in $\mathbb{R}^n$. Because S is bounded and closed, hence compact, and the function φ is continuous, there exists $k = \inf\{\varphi(x) : x\in S\} = \varphi(x_0) \geq 0$, where $x_0\in S$. More precisely, we have $x_0 \neq 0$, so that $k > 0$. Now, at each $x\in\mathbb{R}^n \setminus \{0\}$, we evaluate:

$$\frac{\varphi(x)}{\|x\|^2} = \frac{f(x,x)}{\|x\|^2} = f\left(\frac{1}{\|x\|}x, \frac{1}{\|x\|}x\right) = \varphi\left(\frac{1}{\|x\|}x\right) \geq k,$$

where f is the symmetric bilinear form that generates φ. This leads to the claimed relation at all $x\neq 0$. It remains to remark that it is obviously verified at the origin. ◇

PROBLEMS §III.4.

1. Show that every linear operator, which acts between spaces of finite dimension, is continuous. In particular, consider the derivation on the space of all polynomial functions defined on $[a, b] \subset \mathbb{R}$, which have the degree smaller than or equal to a fixed $n \in \mathbb{N}$.

Hint. Each linear operator is represented by a matrix $A = (a_{ik})$, i.e.

$$y_i = \sum_{k=1}^{n} a_{ik}\, x_k = \langle a_i , \bar{x} \rangle$$

holds for all $i = \overline{1, m}$. Using the fundamental inequality in Γ^n , we deduce that $|y_i| \leq \|a_i\|\, \|x\|$, hence $\|y\| \leq \mu\, \|x\|$, where

$$\mu = \sqrt{\sum_{i=1}^{m}\sum_{k=1}^{n} |a_{ik}|^2} .$$

For polynomial functions, if we refer to the base $\{1, t, t^2, \ldots, t^n\}$, then we may describe the derivation as a change of coefficients, namely

$$(c_0, c_1, \ldots, c_n) \mapsto (c_1, 2c_2, \ldots, nc_n).$$

In other terms, we represent the operator of derivation by the matrix

$$A = \begin{pmatrix} 0\,1\,0 \ldots 0 \\ 0\,0\,2 \ldots 0 \\ \ldots\ldots\ldots \\ 0\,0\,0 \ldots n \\ 0\,0\,0 \ldots 0 \end{pmatrix} .$$

2. Show that the operator of integration on a compact $K = [a, b] \subset \mathbb{R}$, with a continuous *nucleus* $A : K \times K \to \mathbb{R}$, defined by the formula

$$y(s) = \int_a^b A(t,s)\, x(t)\, dt ,$$

is continuous relative to the $\sup$ norm of the space $\mathbf{C}_{\mathbb{R}}(K)$.

Hint. We cannot represent this operator by a matrix, because it acts on a space of infinite dimension. However, if we note

$$\mu = (b - a) \sup \{ | A(t, s)| : t, s \in K\},$$

we may evaluate

$$| y(s)| \leq \int_a^b |A(t, s)|\, |x(t)|\, \mathrm{d}t \leq \int_a^b |A(t, s)| \mathrm{d}t \leq \mu\, \|x\|.$$

The asked continuity follows from the inequality $\|y\| \leq \mu\, \|x\|$.

3. Show that the operator of derivation, $\mathbf{D} : \mathbf{C}_{\mathbb{R}}^{1}(I) \to \mathbf{C}_{\mathbb{R}}(I)$, where $I \subseteq \mathbb{R}$, is not continuous relative to the sup norms of $\mathbf{C}_{\mathbb{R}}^{1}(I)$ and $\mathbf{C}_{\mathbb{R}}(I)$.
Hint. **D** is discontinuous on $\mathbf{C}_{\mathbb{R}}^{1}(I)$, since it is discontinuous at the origin. In fact, let the functions $x_n : I \to \mathbb{R}$, of the sequence (x_n), be defined by

$$x_n(t) = \frac{\sin nt}{\sqrt{n}}.$$

This sequence tends to zero when $n \to \infty$, since $\|x_n\| \le n^{-1/2}$. However, the sequence $(x_n^{/})$, of derivatives, is divergent, since $x_n^{/}(t) = \sqrt{n}\cos nt$.

4. Using a geometric interpretation of the equation f = const. , evaluate the norm of the function $f: \mathbb{R}^3 \to \mathbb{R}$, of values $f(x, y, z) = x + y + z$.
Hint. Find $\alpha \in \mathbb{R}$ such that the plane π, of equation $f(x, y, z) = \alpha$, is tangent to the sphere of equation $x^2 + y^2 + z^2 = 1$. Because $\pi \perp (1, 1, 1)$, it follows that the point of tangency belongs to the straight line $\{\lambda(1, 1, 1) : \lambda \in \mathbb{R}\}$.

5. On the space $\mathbf{C}_{\mathbb{R}}([-1, +1])$, endowed with the sup norm, we define the function $f : \mathbf{C}_{\mathbb{R}}([-1, +1]) \to \mathbb{R}$, of values $f(x) = x(0)$. Show that f is linear and continuous, and find its norm.
Hint. The linearity is immediate. If $\|x\| = \sup \{|x(t)| : t \in [-1, +1]\} \le 1$, then obviously $|f(x)| \le 1$, hence $\|f\| \le 1$. Functions like $x(t) = \cos t$, which have the norm $\|x\| = |x(0)| = 1$, show that the sup value is attained, i.e. $\|f\| = 1$.

6. Show that every full sphere, in a normed linear space $(\mathscr{X}, \|.\|)$, is convex, but no straight line is entirely contained in such a sphere. Identify the linear operator $U:\mathscr{X} \to \mathscr{Y}$, where $(\mathscr{Y}, \|.\|)$ is another normed space, for which there exists $M > 0$ such that $\|U(x)\| < M$ holds at all $x \in \mathscr{X}$.
Hint. We may take the center of the sphere at the origin. If $x + \lambda y \in S(0, r)$ is allowed for arbitrary $\lambda \in \Gamma$, including $\lambda \to \infty$, then from the inequality

$$|\lambda|\,\|y\| = \|\lambda y\| \le \|x\| + \|x + \lambda y\| \le \|x\| + r$$

it follows that $y = 0$. Because U carries a straight line into another straight line, the single "*bounded* on $\mathscr{X}$ " linear operator is the null one.

7. Let U be a self-adjoint operator on the Euclidean space $\mathscr{X} = \Gamma^n$ (of finite dimension). Show that:
a) There exists $x_0 \in \mathscr{X}$ such that $\|x_0\| = 1$ and $\|U\| = \|U(x_0)\|$;
b) This x_0 is a proper vector of U^2 ;
c) The proper value of U^2, corresponding to x_0 , is $\lambda = \|U\|^2$;
d) Either + ||U|| or – ||U|| is a proper value of U.
Hint. a) U is continuous, hence the supremum in the definition of $\|U\|$ is attained on the unit sphere, which is a compact set.

b) If we note $U(x_0) = y_0$, then we successively obtain

$$\|U\|^2 = < U(x_0), U(x_0) > = < y_0 , U(x_0) > =$$
$$= < U(y_0), x_0 > \leq \|U(y_0)\| \, \|x_0\| = \|U(y_0)\| \leq \|U\| \, \|y_0\| = \|U\| \, \|U(x_0)\| = \|U\|^2 .$$

The resulting equality $< U(y_0), x_0 > = \|U(y_0)\| \, \|x_0\|$ is possible in a single case, namely when $U(y_0) = \lambda \, x_0$.

c) From $U^2(x_0) = \lambda \, x_0$ we deduce $\|U\|^2 = < U(y_0), x_0 > = \lambda$.

d) We may write the relation $U^2(x_0) = \|U\|^2 x_0$ in the form

$$(U - \|U\| \, I)(U + \|U\| I)(x_0) = 0.$$

If $z_0 = (U + \|U\| \, I)(x_0) \neq 0$, then $(U - \|U\| \, I)(z_0) = 0$, hence z_0 is a proper vector corresponding to the proper value $\|U\|$. If $z_0 = 0$, then the expression of z_0 directly gives $U(x_0) = - \|U\| \, I \, x_0$.

8. Find the symmetric bilinear forms that generate the following quadratic forms, and bring them into a canonical form:

a) $\varphi = x^2 + 2y^2 + 3z^2 - 4xy - 4yz$;

b) $\psi = 2xy + 2xz - 2xt - 2yz + 2yt + 2 zt$;

c) $\chi = 2\sum_{i<j}^{n} x_i \, x_j$.

<u>Hint</u>. a) The matrix A, associated to φ , has three distinct proper values, namely $\lambda_1 = -1$, $\lambda_2 = 2$, and $\lambda_3 = 5$, hence a canonical form of φ is

$$- u^2 + 2v^2 + 5w^2.$$

b) The matrix of the corresponding bilinear form is

$$A = \begin{pmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & -1 & 1 \\ 1 & -1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix}.$$

It has a simple proper value $\lambda_1 = -3$ and a triple one $\lambda_2 = 1$. The new base may consist of (1, –1, –1, 1), which is a proper vector corresponding to λ_1, and three orthogonal solutions of the equation $AX = X$, which furnishes the proper vectors of $\lambda_2 = 1$. Because this equation reduces to $x - y - z + t = 0$, we may chose the vectors (1, 1, 0, 0), (1, 0, 1, 0), (0, 0, 1, 1). The resulting canonical form is $u^2 + v^2 + w^2 - 3s^2$.

c) The attached symmetric matrix has two proper values, namely $\lambda_1 = n-1$, and $\lambda_2 = -1$, of multiplicity $(n-1)$. A solution of the equation $Ax = (n-1)x$ is (1, 1, …, 1), while equation $Ax = - x$ reduces to $x_1 + \ldots + x_n = 0$.

A canonical form is $\chi = x_1'^2 - \sum_{i=1}^{n} x_i'^2$.

9. Let $\mathscr{X}$ be the linear space of real polynomial functions with degree not exceeding 2, defined on [0, 1]. Show that the function $f: \mathscr{X} \times \mathscr{X} \to \mathbb{R}$,

$$f(x, y) = \int_0^1 x(t)y(t)\, dt,$$

is a bilinear symmetric form on $\mathscr{X}$, and find its matrix in the base $\{1, t, t^2\}$. Verify directly, and using a canonical form, that f generates a positively definite quadratic form.

Hint. According to the formula $a_{ij} = f(e_i, e_j)$, we obtain

$$A = \begin{pmatrix} f(1,1) & f(1,t) & f(1,t^2) \\ f(t,1) & f(t,t) & f(t,t^2) \\ f(t^2,1) & f(t^2,t) & f(t^2,t^2) \end{pmatrix} = \begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{pmatrix}.$$

To prove the positiveness of the generated quadratic form, we may use the Sylvester's test 4.37.

10. Discuss the signature of the following quadratic forms upon the parameters $a, b, c \in \mathbb{R}$:

1) $\varphi = ax^2 + 2bxy + cy^2$;
2) $\psi = 2x^2 + 2xy + 2axz + y^2 + 2yz + az^2$;
3) $\chi = x^2 + 2xy + 2axz + y^2 - 2yz + az^2$.

Hint. 1) The cases when some parameters vanish are immediate. If $a \neq 0$, we may isolate a square, and write φ in the form

$$\frac{1}{a}[a^2(x + \frac{b}{a}y)^2 + (ac - b^2)\, y^2].$$

2) Use the Sylvester's test; ψ is positive at $a \in (1, 2)$.
3) In the attached matrix, the second minor vanishes; hence the Sylvester's test doesn't work any more. The form χ is degenerated at the value $a = -1$, and indefinite at any $a \in \mathbb{R}$.

11. Find the extreme values (if there are some) of the following polynomial functions of the second degree:

$f(x, y, z) = x^2 + y^2 + z^2 - xy + 2z - 3x;$
$g(x, y, z) = x^2 + y^2 - 2z^2 + 2x - 2y + 3.$

Hint. Realize translations of the origin in $\mathbb{R}^3$, and make evident some quadratic forms. In particular, replacing $x = u - 1$, $y = v + 1$, and $z = w$, brings g into the form $u^2 + v^2 - w^2 + 1$. Because the involved quadratic form is indefinite, $g(-1, 1, 0)$ isn't extreme value, i.e. in any neighborhood of, $(-1, 1, 0)$ there are points where g takes both greater and smaller values than $1 = g(-1, 1, 0)$.

CHAPTER IV. DIFFERENTIABILITY

§ IV.1. REAL FUNCTIONS OF A REAL VARIABLE

1.1. <u>Linear approximations</u>. In order to define the notion of *differential* we will analyze the process of approximating a real function of a real variable by a linear function. Let us consider that the function $f: (a, b) \to \mathbb{R}$ is derivable at the point $x_0 \in (a, b) \subseteq \mathbb{R}$ (see fig. IV.1.1), i.e. then exists the tangent to the graph of f at M_0 .

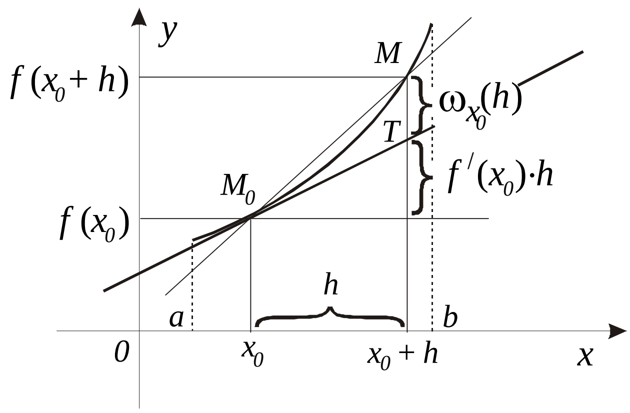


Fig.IV.1.1.

Taking into account the signs of the increments, we obtain the equality

$$f(x_0 + h) - f(x_0) = f^{/}(x_0) \cdot h + \omega_{x_0}(h) \ ,$$

which indicates a possibility of approximating $f(x_0 + h)$ by $f(x_0) + f^{/}(x_0)\, h$.

The error of this approximation is $\omega_{x_0}(h)$. In addition, we may consider that this method provides a "good approximation", in the sense that the existence of $f^{/}(x_0)$ leads to

$$\lim_{h\to 0}\left|\frac{f(x_0+h)-f(x_0)}{h} - f^{/}(x_0)\right| = \lim_{h\to 0}\frac{\omega_{x_0}}{h} = 0\,,$$

which means that $\omega_{x_0}(h)$ tends to zero faster than h does. Geometrically, using the graph of f, this property of the error shows that the secant $M_0 M$ tends to the tangent $M_0 T$ whenever $h \to 0$. Such an approximation is said to be <u>*linear*</u> because the function $L_{x_0} : \mathbb{R} \to \mathbb{R}$, defined by $L_{x_0}(h) = f^{/}(x_0) \cdot h$, is linear (note that $f^{/}(x_0)$ is a fixed number here, and h is the argument).

As a conclusion, the above linear approximation is possible because there exists the linear functional L_{x_0} such that

$$\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)-L_{x_0}(h)}{h} = 0\,. \qquad (*)$$

The linear functions like L_{x_0} will be the object of the present chapter. As usually, $\mathscr{L}(\mathscr{X}, \mathscr{Y})$ will denote the set of all linear functions $L: \mathscr{X} \to \mathscr{Y}$.

1.2. <u>Definition</u>. Let $A \subseteq \mathbb{R}$ be an open set, $f: A \to \mathbb{R}$ be an arbitrary function, and $x_0 \in A$ be a fixed point. We say that f is *differentiable* at x_0 iff there exists $L_{x_0} \in \mathscr{L}(\mathbb{R}, \mathbb{R})$ such that (*) holds. The linear (and also continuous) function L_{x_0} is called *differential* of f at x_0, and by tradition it is noted df_{x_0}.

1.3. <u>Theorem</u>. A real function of one real variable, say $f: A \to \mathbb{R}$, $A \subseteq \mathbb{R}$, is differentiable at $x_0 \in A$ iff it is derivable at this point. In addition, the values of the differential are

$$df_{x_0}(h) = f'(x_0)\, h$$

at each $h \in \mathbb{R}$,

<u>Proof</u>. If f is differentiable at x_0, then condition (*) holds. Now, let us remark that the linear functions on $\mathbb{R}$ have the form $L_{x_0}(h) = c \cdot h$, for some $c \in \mathbb{R}$. In fact, because linearity means additivity and homogeneity, if we put $k=1$ in $L_{x_0}(kh) = h L_{x_0}(k)$, we obtain $L_{x_0}(h) = h L_{x_0}(1)$, where h is arbitrary in R. Consequently, we have $L_{x_0}(h) = c \cdot h$, with $c = L_{x_0}(1)$. If we replace this expression in (*), we obtain

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = c,$$

which shows that f is derivable at x_0, and $f'(x_0) = c$. In addition,

$$L_{x_0}(h) \overset{not.}{=} df_{x_0}(h) = c \cdot h = f'(x_0)\, h.$$

Conversely, the existence of the derivative $f'(x_0)$ may be written as

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0) - f'(x_0)h}{h} = 0,$$

where $f'(x_0)\, h = L_{x_0}(h)$ is the searched linear function. ◇

1.4. <u>Remark</u>. The concrete calculation of the differential df_{x_0} of a function $f: A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, at a point $x_0 \in A$, simply reduces to the calculation of $f'(x_0)$. Both $f'(x_0)$ and df_{x_0} involve a limiting process, locally at x_0.

The first part of the above proof shows that the linear spaces $\mathscr{L}(\mathbb{R}, \mathbb{R})$ and $\mathbb{R}$ are isomorphic. However, the main point of theorem 1.3 is the strong connection between differential and derivative, which explains why some authors identify the terms "*differentiation*" and "*derivation*". To be more specific, we will use the word "*differentiation*" in the sense of "*calculating a differential*".

If we calculate the differential df_{x_0} at each $x_0 \in A$, we obtain a function on A, with values in $\mathscr{L}(\mathbb{R}, \mathbb{R})$, which is a *differential on a set*. More exactly:

1.5. Definition. A function $f: A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, is *differentiable on the set A* iff it is differentiable at each point $x \in A$. The *differential* of f on A, noted $df: A \to \mathscr{L}(\mathbb{R}, \mathbb{R})$, is defined by the formula

$$df(x) = df_x \ .$$

1.6. Remarks. According to theorem 1.3, we have

$$df(x)(h) = df_x(h) = f^{/}(x) \cdot h \ .$$

If we apply this formula to the identity of A, noted $\imath : A \to A$, and defined by $\imath(x) = x$, then we obtain the derivative $\imath^{/}(x) = 1$, and the differential

$$d\imath_x(h) = h \ .$$

Consequently, for an arbitrary function f, we have

$$df(x)(h) = f^{/}(x)\, d\imath_x(h) = f^{/}(x)\, d\imath(x)(h) \ ,$$

or, renouncing to mention the variable h, which is arbitrary in $\mathbb{R}$,

$$df(x) = f^{/}(x)\, d\imath(x).$$

Since x is arbitrary in A, we may omit it, and write the relation between the functions, that is $df = f^{/}(x)\, d\imath$. Because of the tradition to note $d\imath = dx$, we finally obtain a symbolic form of the differential, namely

$$df = f^{/} dx. \qquad (**)$$

This is the simplest, but formalistic way to correlate the differential and the derivative. It is very useful in formulating the general properties of the differential, but its exact meaning has the chief importance in practice (see the exercises at the end of the section).

1.7. Theorem. Let $A \subseteq \mathbb{R}$ be an open set. The differential has the following properties:

a) Every differentiable function on A is continuous on A.

b) If the function $f: A \to B \subseteq \mathbb{R}$ is differentiable on A, and $g: B \to \mathbb{R}$ is differentiable on B, then $g \circ f: A \to \mathbb{R}$ is differentiable on A, and

$$d(g \circ f) = (g^{/} \circ f)\, df.$$

c) If the functions $f, g: A \to \mathbb{R}$ are differentiable on A, then $f + g$, λf, fg, and f/g (where defined) are differentiable on A, and:

$$d(f+g) = df + dg$$
$$d(\lambda f) = \lambda df$$
$$d(fg) = f\, dg + g\, df$$
$$d(f/g) = \frac{gdf - fdg}{g^2}.$$

Proof. a) According to theorem 1.3, the properties of differentiability and derivability are equivalent, but derivable functions are continuous.

b) By a direct calculation of $d(g \circ f)_x(h)$, we are lead to the formula

$$d(g \circ f) = (g \circ f)^{/}\, dx = (g^{/} \circ f)\, f^{/} dx = (g^{/} \circ f)\, df \ .$$

c) The problem reduces to the derivation of a quotient. Following the definition of $d(f/g)_x(h)$, and erasing the variables h and x, we obtain:

$$d\left(\frac{f}{g}\right)=\left(\frac{f}{g}\right)^{/}dx=\frac{f^{/}g-fg^{/}}{g^2}dx=\frac{g(f^{/}dx)-f(g^{/}dx)}{g^2}=\frac{g\,df-f\,dg}{g^2}\ .$$

Alternatively, (**) allows formal proofs of b) and c). ◇

1.8. <u>Remark</u>. To solve practical problems, especially approximations, we may use the notion of differential in a simpler sense. For example (see [DB], etc.), the differential of a function $y = f(x)$ is the *principal part* of its increment $\Delta y = f(x + \Delta x) - f(x)$, which is linear relative to the increment $\Delta x = h$. Formally, the differential of f is defined by $dy = f^{/}dx$, which is considered equivalent to the derivative $f^{/} = \dfrac{dy}{dx}$.

The problem of approximation reduces to the replacement of Δy by dy. For example, for $y = 3x^2 - x$, $x=1$, and $\Delta x = 0.01$, we obtain $\Delta y = 0.0503$ and $dy = 0.0500$.

Finally, we mention that in the process of evaluating the error of such a linear approximation, we need some formulas for the remainder of the Taylor series.

1.9. <u>Example</u>. Let us consider that we have to evaluate $\sqrt[3]{8.1}$. Of course, we know that $\sqrt[3]{8}=2$, and for $\sqrt[3]{8.1}$ we can specify a finite number of exact digits only. The number of exact digits is determined by practical reasons, say four in this case. According to the formula of linear approximation

$$f(x+h)-f(x)\cong f^{/}(x)\cdot h$$

in the case $f(x)=\sqrt[3]{x}$, $x = 8$, and $h = 0.1$, we obtain $\sqrt[3]{8.1}\approx 2.0083$.

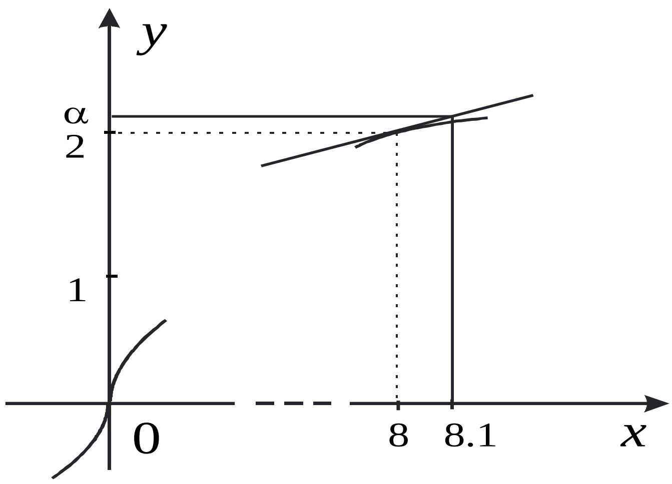


Fig. IV.1.2.

We remark that $\sqrt[3]{0.1}$ is not available on this way using $f^{/}(0)$!

PROBLEMS §IV.1.

1. Using the linear approximation, calculate $\sqrt[3]{8.5}$, *arcsin* 0.51, and the area of a circle of radius $r = 3.02$ *cm* .

Hint. $8.5 = 2^3 + 0.5$, hence for $f(x) = \sqrt[3]{x}$ we take $x_0 = 2$ and $h = 0.5$. Similarly, for $f(x) = arcsin\ x$, we consider $x_0 = \frac{1}{2}$ and $h = 0.01$. The area of the circle approximately equals

$$\mathcal{A} \cong \pi \cdot 3^2 + 2\,\pi \cdot 3 \cdot 0.02 = 9\,\pi + 0.12\,\pi \cong 28.66\ .$$

2. Using the linear approximation, find the solution of the equation

$$15\cos x - 13\sin x = 0$$

in the interval $(0, \pi/2)$.

Hint. $x = arctg\,(1 + \frac{2}{13})$ may be approximated taking $x_0 = 1$ and $h = \frac{2}{13}$.

3. Find the differentials of the following functions on the indicated sets of definition (intervals):

a) $f: \mathbb{R} \to \mathbb{R}$, f linear; show that $df_x = f$;

b) $f: (-1, +\infty) \to \mathbb{R}$, $f(x) = ln(1 + x)$;

c) $f: \mathbb{R} \to \mathbb{R}$, $f(x) = \begin{cases} \sin^2 x & \text{if } x > 0 \\ x^2 & \text{if } x \le 0\ . \end{cases}$

4.Study the differentiability of the function $f: \mathbb{R} \to \mathbb{R}$, of values

$$f(x) = \begin{cases} -x & \text{if } x \le 0 \\ x^2 e^{-x} & \text{if } x > 0\ . \end{cases}$$

Hint. f is not derivable at the origin.

5.Using the *Ohm's Law* $I = E/R$, show that a small change in the current, due to a small change in the resistance, may be approximately calculated by the formula

$$\Delta I = -\ \frac{I}{R}\Delta R.$$

6. Let us imagine a thin thread along the equator of the earth, and the length of this thread increases by 1*m*. If this thread is arranged in a concentric to the equator circle, can a cat pass through the resulting space?

Hint. From the formula $L = 2\pi r$ we deduce that $dr = \frac{1}{2\pi} dL$. The increment $dL = 100\ cm$ of the circumference corresponds to an increment $dr > 15\ cm$ of the radius, so the answer is "Yes".

§ IV.2. FUNCTIONS BETWEEN NORMED SPACES

As we already have seen in the previous section, the differential of a function $f: A\to\mathbb{R}$, where $A\subseteq\mathbb{R}$, is a function $df: A\to\mathscr{L}(\mathbb{R}, \mathbb{R})$, hence it ranges in a linear normed space. The higher order differentials should be defined on normed linear spaces. Similarly, the vector functions of one or more variables, as well as their differentials, act between normed spaces. This is the reason why we have to extend the differential of a real function of one real variable to that of a function between normed spaces.

2.1. <u>Definition</u>. Let us consider two normed spaces $(\mathscr{X}, \|\cdot\|_1)$, $(\mathscr{Y}, \|\cdot\|_2)$, an open set $A\subseteq\mathscr{X}$, $x_0\in A$, and a function $f: A\to\mathscr{Y}$. As usually, we note by $\mathscr{B}(\mathscr{X}, \mathscr{Y})$ the set of all linear and continuous functions from $\mathscr{X}$ to $\mathscr{Y}$. We say that f is *differentiable* (in the Fréchet's sense) at the point x_0 iff there exists $L_{x_0}\in\mathscr{B}(\mathscr{X}, \mathscr{Y})$ such that

$$\lim_{\|h\|_1\to 0}\frac{\left\| f(x_0+h)-f(x_0)-L_{x_0}(h)\right\|_2}{\|h\|_1}=0.$$

The linear and continuous function L_{x_0} is called *differential* of f at x_0 (in the Fréchet's sense). By tradition, L_{x_0} is frequently noted df_{x_0}.

2.2. <u>Remark</u>. If we note $U=\{h\in\mathscr{X}: x_0+h\in A\}$ and $\omega_{x_0}: U\to\mathscr{Y}$, where

$$\omega_{x_0}(h)=f(x_0+h)-f(x_0)-L_{x_0}(h),$$

then the condition of differentiability reduces to

$$\lim_{\|h\|_1\to 0}\frac{\left\|\omega_{x_0}(h)\right\|_2}{\|h\|_1}=0.$$

As for real functions, we may consider that L_{x_0} is a linear approximation of $f(x_0+h)-f(x_0)$ in a neighborhood of x_0. Establishing the differentiability of a function is possible only if we have a good knowledge of $\mathscr{B}(\mathscr{X}, \mathscr{Y})$. More than this, even if the differentiability is assured, there remains the concrete problem of writing the differential df_{x_0}.

The following theorem is useful in this respect:

2.3. <u>Theorem</u>. If $f: A\to\mathscr{Y}$, $A\subseteq\mathscr{X}$, is differentiable at $x_0\in A$, then:

a) The value of df_{x_0} at any $h\in\mathscr{X}$ is

$$df_{x_0}(h)=\lim_{t\to 0}\frac{f(x_0+th)-f(x_0)}{t},$$

where $t\in\mathbb{R}$ (this limit is usually called *weak*, or *Gateaux*' *differential*);

b) df_{x_0} is uniquely determined by f and x_0.

Proof. a) The case $h = 0$ is obvious. For $h \neq 0$, we may remark that

$$\left\| \frac{f(x_0+th)-f(x_0)}{t} - df_{x_0}(h) \right\|_2 = \|h\|_1 \frac{\left\| f(x_0+th)-f(x_0)-df_{x_0}(th) \right\|_2}{\|th\|_1}$$

and $t \to 0$ implies $\|th\|_1 \to 0$ for each fixed $h \in \mathscr{X}$. Consequently, because f is supposed differentiable, the claimed formula of df_{x_0} follows from

$$\lim_{t\to 0} \left\| \frac{f(x_0+th)-f(x_0)}{t} - df_{x_0}(h) \right\|_2 = 0 .$$

b) Accordingly to a), $df_{x_0}(h)$ is obtained as a limit, but in normed spaces the limit is unique. Since h is arbitrary, df_{x_0} is unique too. ◇

2.4. Remark. The hypothesis that function f is differentiable is essential in the above theorem, i.e. it assures the existence of the Gateaux' differential

$$\lim_{t\to 0} \frac{f(x_0+th)-f(x_0)}{t} \overset{not.}{=} \gamma_{x_0}(h) .$$

Conversely, the existence of γ_{x_0} at arbitrary $h \in \mathscr{X}$ does not mean that f is differentiable at x_0 . More exactly, it may happen that either $\gamma_{x_0}(h)$ is not linear, or the quotient $\dfrac{\|\omega_{x_0}(h)\|_2}{\|h\|_1}$ has no limit (even if γ_{x_0} is linear). Simple examples of these possibilities are the functions $f: \mathbb{R}^2 \to \mathbb{R}$, defined by

$$f(x,y) = \begin{cases} \dfrac{x|y|}{\sqrt{x^2+y^2}} & if \quad (x,y) \neq (0,0) \\ 0 & if \quad x = y = 0 \end{cases} ,$$

respectively $g: \mathbb{R}^2 \to \mathbb{R}$, defined by

$$g(x,y) = \begin{cases} \dfrac{x^3 y}{x^6+y^2} & if\ (x,y) \neq (0,0) \\ 0 & if \quad (x,y) = (0,0) . \end{cases}$$

The differentiability at a point is naturally extended to a set:

2.5. Definition. In the conditions of Definition 2.1, we say that function f is *differentiable on the set A* iff it is differentiable at each point $x_0 \in A$. In this case, the function $df: A \to \mathscr{B}(\mathscr{X}, \mathscr{Y})$, defined by $df(x) = df_x$, is called the *differential* of f on A . If df is continuous, then we say that f is of class $\mathbf{C}^1$ on A, and we note $f \in \mathbf{C}^1_{\mathscr{Y}}(A)$.

The general properties of the differential of a real function depending on a single real variable, expressed in theorem 1.7 of the previous section, remain valid for functions between normed linear spaces:

2.6. <u>Theorem</u>. a) Each differentiable function is continuous;
b) If $f: A\to B$, where $A\subseteq\mathscr{X}$ and $B\subseteq\mathscr{Y}$ are open sets, is differentiable at the point $x_0 \in A$, and $g: B\to\mathscr{Z}$ is differentiable at $f(x_0)$, where $(\mathscr{Z}, \|\cdot\|_3)$ is another normed space, then $g\circ f: A\to\mathscr{Z}$ is differentiable at x_0, and

$$d(g\circ f)_{x_0} = dg_{f(x_0)}\circ df_{x_0};$$

c) If $f, g : A\to\mathscr{Y}$ are differentiable functions on A and $\lambda\in\mathbb{R}$, then $f+g$ and λf are also differentiable on A, and $d(f+g) = df + dg$, $d(\lambda f) = \lambda df$.
<u>Proof</u>. a) Differently from theorem IV.1.7, the proof shall be directly based on the definitions. Let $f: A\to\mathscr{Y}$ be a differentiable function at an arbitrary point $x_0 \in A$. According to the corollary III.4.6, the continuity of the differential shows that for every $h\in\mathscr{X}$ we have $\|df_{x_0}(h)\|_2 \le \|df_{x_0}\|\cdot\|h\|_1$. In the terms of the remark 2.2, the condition of differentiability takes the form

$$\lim_{\|h\|_1\to 0}\frac{\|\omega_{x_0}(h)\|_2}{\|h\|_1} = 0 .$$

In particular, for $\varepsilon = 1$, there exists $\delta_1 > 0$, such that $\frac{\|\omega_{x_0}(h)\|_2}{\|h\|_1}<1$, that is $\|\omega_{x_0}(h)\|_2 < \|h\|_1$, holds at each $h\in U$ for which $\|h\|_1 < \delta_1$. Consequently, at all such h, the following relations hold:

$$\|f(x_0+h) - f(x_0)\|_2 = \|df_{x_0}(h) + \omega_{x_0}(h)\|_2 \le \|df_{x_0}(h)\|_2 + \|\omega_{x_0}(h)\|_2 <$$
$$< \|df_{x_0}\|\cdot\|h\|_1 + \|h\|_1 = \left(\|df_{x_0}\| + 1\right)\cdot\|h\|_1 .$$

Now, for arbitrary $\varepsilon > 0$, we consider $\delta = \min\left\{\delta_1, \varepsilon/\left(\|df_{x_0}\|+1\right)\right\} > 0$. It is easy to see that for all $h\in U$, for which $\|h\|_1 < \delta$, we have

$$\|f(x_0+h) - f(x_0)\|_2 < \left(\|df_{x_0}\|+1\right)\cdot\|h\|_1 < \varepsilon,$$

which proves the continuity of f at x_0. Since x_0 was arbitrary in A, we conclude that f is continuous on A.

b) Besides U, involved in the differentiability of f, let us consider

$$V = \{s\in\mathscr{Y}: f(x_0) + s\in B\}.$$

Since A and B are open, we have $U\ne\varnothing$, $V\ne\varnothing$. The differentiability of f at x_0 means that there exist $df_{x_0}\in\mathscr{B}(\mathscr{X}, \mathscr{Y})$ and $\omega^f_{x_0}: U\to\mathscr{Y}$, such that $f(x_0+h) - f(x_0) = df_{x_0}(h) + \omega^f_{x_0}(h)$ holds at each $h\in U$, and

$$\lim_{\|h\|_1\to 0}\frac{\|\omega^f_{x_0}(h)\|_2}{\|h\|_1} = 0 .$$

Similarly, the differentiability of g at $f(x_0)$ assumes the existence of the functions $dg_{f(x_0)} \in \mathscr{B}(\mathscr{Y}, \mathscr{Z})$ and $\omega^g_{f(x_0)} : V \to \mathscr{Z}$, such that

$$g(f(x_0)+s) - g(f(x_0)) = dg_{f(x_0)}(s) + \omega^g_{f(x_0)}(s)$$

holds at all $s \in V$, and $\lim\limits_{\|s\|_2 \to 0} \dfrac{\left\|\omega^g_{f(x_0)}(s)\right\|_3}{\|s\|_2} = 0$.

Since $s = f(x_0 + h) - f(x_0) \in V$ whenever $h \in U$, we obtain:

$$g(f(x_0+h)) - g(f(x_0)) = dg_{f(x_0)}(f(x_0+h) - f(x_0)) + \omega^g_{f(x_0)}(s).$$

In terms of composed functions, this relation shows that

$$(g \circ f)(x_0+h) - (g \circ f)(x_0) = dg_{f(x_0)}\left(df_{x_0}(h) + \omega^f_{x_0}(h)\right) + \omega^g_{f(x_0)}(s).$$

Because $dg_{f(x_0)}$ is a linear function, it follows that

$$\begin{aligned}(g \circ f)(x_0+h) &- (g \circ f)(x_0) = \\ &= dg_{f(x_0)}\left(df_{x_0}(h)\right) + dg_{f(x_0)}\left(\omega^f_{x_0}(h)\right) + \omega^g_{f(x_0)}(s) = \\ &= \left(dg_{f(x_0)} \circ df_{x_0}\right)(h) + dg_{f(x_0)}\left(\omega^f_{x_0}(h)\right) + \omega^g_{f(x_0)}\left(f(x_0+h) - f(x_0)\right).\end{aligned}$$

The membership $dg_{f(x_0)} \circ df(x_0) \in \mathscr{B}(\mathscr{X}, \mathscr{Z})$ immediately follows from the previous hypotheses $df_{x_0} \in \mathscr{B}(\mathscr{X}, \mathscr{Y})$ and $dg_{f(x_0)} \in \mathscr{B}(\mathscr{Y}, \mathscr{Z})$, so it remains to define ω_{x_0} for the final proof of the differentiability of $f \circ g$. In this respect we define the function $\omega_{x_0} : U \to \mathscr{Z}$, which takes the values

$$\omega_{x_0}(h) = dg_{f(x_0)}\left(\omega^f_{x_0}(h)\right) + \omega^g_{f(x_0)}\left(f(x_0+h) - f(x_0)\right),$$

and we show that it satisfies the condition

$$\lim_{\|h\|_1 \to 0} \frac{\left\|\omega_{x_0}(h)\right\|_3}{\|h\|_1} = 0.$$

In fact, if $\varepsilon > 0$ is arbitrarily given, let us consider

$$\varepsilon' = \frac{\varepsilon}{1 + \left\|df_{x_0}\right\| + \left\|dg_{f(x_0)}\right\|} > 0.$$

According to Corollary III.4.6, the continuity of $dg_{f(x_0)}$ implies that

$$\left\|dg_{f(x_0)}(\omega^f_{x_0}(h))\right\|_3 \le \left\|dg_{f(x_0)}\right\| \cdot \left\|\omega^f_{x_0}(h)\right\|_2$$

holds at each $h \in U$. Relative to the differentiability of f, we recall that

$$\lim_{\|h\|_1 \to 0} \frac{\left\|\omega^f_{x_0}(h)\right\|_2}{\|h\|_1} = 0,$$

i.e. for arbitrary $\varepsilon' > 0$, there exists $\delta_1 > 0$, such that the inequality

$$\left\| \omega_{x_0}^{f}(h) \right\|_2 < \varepsilon' \cdot \|h\|_1$$

holds at each $h \in U$, for which $\|h\|_1 < \delta_1$. As a partial result, we retain that, under the mentioned conditions, we have the inequality

$$\left\| dg_{f(x_0)}(\omega_{x_0}^{f}(h)) \right\|_3 < \varepsilon' \cdot \left\| dg_{f(x_0)} \right\| \cdot \|h\|_1 . \qquad (*)$$

Similarly, the differentiability of g assures the relation

$$\lim_{\|s\|_2 \to 0} \frac{\left\| \omega_{f(x_0)}^{g}(s) \right\|_3}{\|s\|_2} = 0 .$$

In more details, this means that for arbitrary $\varepsilon' > 0$, there exists $\eta > 0$, such that the inequality $\left\| \omega_{f(x_0)}^{g}(s) \right\|_3 < \varepsilon' \cdot \|s\|_2$ is valid at each $s \in V$, for which $\|s\|_2 < \eta$. Because f is continuous at x_0, as already proved at part a), to this η there corresponds some $\delta_2 > 0$, such that $h \in U$, and $\|h\|_1 < \delta_2$ are sufficient conditions for $\| f(x_0 + h) - f(x_0) \|_2 < \eta$. Consequently, if $h \in U$, and $\|h\|_1 < \delta_2$, then we may replace s in the above inequality, and we obtain

$$\left\| \omega_{f(x_0)}^{g}\big(f(x_0 + h) - f(x_0) \big) \right\|_3 < \varepsilon' \cdot \| f(x_0 + h) - f(x_0) \|_2 .$$

On the other hand, when proving part a) of the present theorem, we have established that there exists $\delta_3 > 0$, such that $h \in U$, and $\|h\|_1 < \delta_3$ imply

$$\| f(x_0 + h) - f(x_0) \|_2 < \left(\left\| df_{x_0} \right\| + 1 \right) \cdot \|h\|_1 .$$

Consequently, if $h \in U$, such that $\|h\|_1 < \min\{\delta_2, \delta_3\}$, then

$$\left\| \omega_{f(x_0)}^{g}\big(f(x_0 + h) - f(x_0) \big) \right\|_3 < \varepsilon' \cdot \left(1 + \left\| df_{x_0} \right\| \right) \cdot \|h\|_1 . \qquad (**)$$

Finally, let us define $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Because (*) and (**) are valid at all $h \in U$, for which $\|h\|_1 < \delta$, it follows that at these points we have

$$\left\| \omega_{x_0}(h) \right\|_3 < \varepsilon' \cdot \left(1 + \left\| df_{x_0} \right\| + \left\| dg_{f(x_0)} \right\| \right) \cdot \|h\|_1 .$$

This inequality shows that the inequality $\dfrac{\left\| \omega_{x_0}(h) \right\|_3}{\|h\|_1} < \varepsilon$ holds at all $h \in U$, for which $0 < \|h\|_1 < \delta$. In other words, this means $\lim\limits_{\|h\|_1 \to 0} \dfrac{\left\| \omega_{x_0}(h) \right\|_3}{\|h\|_1} = 0$, which shows that $g \circ f$ is differentiable at x_0, and its differential is

$$d(g \circ f)_{x_0} = dg_{f(x_0)} \circ df_{x_0} .$$

To accomplish the proof, we recall the uniqueness of a differential.

c) Let $f, g : A \to \mathscr{Y}$ be differentiable functions at $x_0 \in A$, and let $\lambda \in \mathbb{R}$ be arbitrary. Because df_{x_0}, $dg_{x_0} \in \mathscr{B}(\mathscr{X}, \mathscr{Y})$, and $\mathscr{B}(\mathscr{X}, \mathscr{Y})$ is a linear space, it follows that $df_{x_0} + dg_{x_0}$, $\lambda\, df_{x_0} \in \mathscr{B}(\mathscr{X}, \mathscr{Y})$ too. It is easy to see that at each non null $h \in U$, we have

$$\frac{\left\| (f+g)(x_0+h) - (f+g)(x_0) - (df_{x_0} + dg_{x_0})(h) \right\|_2}{\|h\|_1} \le$$

$$\le \frac{\left\| f(x_0+h) - f(x_0) - df_{x_0}(h) \right\|_2}{\|h\|_1} + \frac{\left\| g(x_0+h) - g(x_0) - dg_{x_0}(h) \right\|_2}{\|h\|_1} .$$

This inequality shows that $f + g$ is differentiable at x_0, and its differential has the (unique) value $d(f+g)_{x_0} = df_{x_0} + dg_{x_0}$. Since x_0 is arbitrary in A, we obtain the claimed relation, $d(f + g) = df + dg$.

By analogy, at each $h \in U \setminus \{0_{\mathscr{X}}\}$, we may write the relation

$$\frac{\left\| (\lambda f)(x_0+h) - (\lambda f)(x_0) - (\lambda \cdot df_{x_0})(h) \right\|_2}{\|h\|_1} =$$

$$= |\lambda| \cdot \frac{\left\| f(x_0+h) - f(x_0) - df_{x_0}(h) \right\|_2}{\|h\|_1} .$$

Consequently, the differentiability of f at x_0 implies the differentiability of (λf) at this point. In addition, $d(\lambda f)_{x_0} = \lambda \cdot df_{x_0}$ is the (unique) value of this differential at x_0. Taking into account that x_0 was arbitrary in A, we see that λf is differentiable on A, and $d(\lambda f) = \lambda \cdot df$. ◇

Now we can define the higher order differentials as follows:

2.7. <u>Definition</u>. Let f be a function in the conditions of the definition 2.1. We say that f is *two times differentiable* at x_0 iff:

a) f is differentiable on an open neighborhood V of x_0;

b) The function $df : V \to \mathscr{B}(\mathscr{X}, \mathscr{Y})$ is differentiable at x_0.

In this case, the differential of df at x_0 is called *second order differential* of f at x_0, and it is noted $d^2 f_{x_0}$. Briefly,

$$d^2 f_{x_0} = d(df)_{x_0} \in \mathscr{B}(\mathscr{X}, \mathscr{B}(\mathscr{X}, \mathscr{Y})).$$

If f is two times differentiable at each $x \in A$, we say that f is *two times differentiable on A*. The function $d^2 f : A \to \mathscr{B}(\mathscr{X}, \mathscr{B}(\mathscr{X}, \mathscr{Y}))$, defined by

$$(d^2 f)(x) = d^2 f_x = d(df)_x$$

is called *second order differential* of f on A.

2.8. <u>Remark</u>. In the terms of § III.4, and particularly according to theorem III.4.18, the space $\mathscr{B}(\mathscr{X}, \mathscr{B}(\mathscr{X}, \mathscr{Y}))$ is isometrically isomorphic with the space $\mathscr{B}_2(\mathscr{X}, \mathscr{X}; \mathscr{Y})$ of all bilinear functions on $\mathscr{X} \times \mathscr{X}$, which range in $\mathscr{Y}$. Consequently, we may consider that $d^2 f_{x_0} \in \mathscr{B}_2(\mathscr{X}, \mathscr{X}; \mathscr{Y})$.

2.9. <u>Proposition</u>. $d^2x = 0$, where x denotes the independent (real) variable.

<u>Proof</u>. By definition 2.7, $d^2x = d(dx)$. As already discussed in the previous section, dx stands for $d\iota$, where $\iota : \mathbb{R} \to \mathbb{R}$ is the identity of $\mathbb{R}$ (defined by $\iota(x) = x$ at all $x \in \mathbb{R}$). Because $\iota^{/}(x) = 1$, the differential of the identity takes the values $dx(h) = h$ at each $x \in \mathbb{R}$. Consequently, $d\iota = \iota$, i.e. $d\iota$ is constant. The claimed relation $d(dx) = 0$ holds because the differential of a constant function always vanishes. ◇

2.10. <u>Proposition</u>. If $f \in \mathscr{B}(\mathscr{X}, \mathscr{Y})$, i.e. f is linear and continuous, then:

a) f is differentiable on $\mathscr{X}$,

b) $df_{x_0} = f$ at each $x_0 \in \mathscr{X}$, and

c) $d^2 f = 0$.

<u>Proof</u>. For a) and b) we check $\| f(x_0 + h) - f(x_0) - f(h) \|_2 = 0$.

To prove c) we interpret b) as showing that df is constant. ◇

In the last part of this section we will deduce properties of the differential in the case when $\mathscr{X}$ and / or $\mathscr{Y}$ reduce to $\mathbb{R}^n$, for some $n \geq 1$.

2.11. <u>Proposition</u>. Let function $f : A \to \mathbb{R}$ be defined on the open set $A \subseteq \mathbb{R}$. If the function f is two times derivable, then it is two times differentiable, and $d^2 f_{x_0}(h, k) = f^{//}(x_0)\, h \cdot k$ holds at each $x_0 \in A$.

Conversely, if f is two times differentiable, then it is two times derivable, and we have $f^{//}(x_0) = d^2 f_{x_0}(1, 1)$ at each $x_0 \in A$.

<u>Proof</u>. Let f be two times derivable. Because the map $(h, k) \mapsto f^{//}(x_0)\, h \cdot k$ is bilinear, the function $L_{x_0}(h) : \mathbb{R} \to \mathbb{R}$, of values $L_{x_0}(h)(k) = f^{//}(x_0)\, h \cdot k$, is linear. In the very definition of the second order differential we have

$$\frac{1}{|h|} \left\| (df)(x_0 + h) - (df)(x_0) - L_{x_0}(h) \right\| =$$

$$= \frac{1}{|h|} \sup_{k \neq 0} \left(\frac{1}{|k|} \left| (df)(x_0 + h)(k) - df(x_0)(k) - f^{//}(x_0) h \cdot k \right| \right) =$$

$$= \frac{1}{|h|} \sup_{k \neq 0} \left(\frac{1}{|k|} \left| f^{/}(x_0 + h) \cdot k - f^{/}(x_0) \cdot k - f^{//}(x_0) \cdot h \cdot k \right| \right) =$$

$$= \frac{1}{|h|} \left| f^{/}(x_0 + h) - f^{/}(x_0) - f^{//}(x_0) \cdot h \right| .$$

Because $f^{/}$ is derivable at x_0, this expression has the limit zero when h tends to zero, hence df is differentiable at x_0, and $d^2 f_{x_0}(h, k) = f^{//}(x_0)\, h{\cdot}k$.

Conversely, if f is two times differentiable, then df is supposed to exist in a neighborhood of x_0. According to theorem IV.1.3, $f^{/}(x)$ exists at all x in this neighborhood, and $df_x(k) = f^{/}(x)\, k$. But f is two tomes differentiable, hence there exists $d^2 f_{x_0}$, such that $d^2 f_{x_0}(h,k) = d^2 f_{x_0}(1,1)\cdot h \cdot k$, and

$$\lim_{h\to 0}\left(\frac{1}{|h|}\sup_{k\neq 0}\frac{\left|f^{/}(x_0+h)\cdot k - f^{/}(x_0)\cdot k - d^2 f_{x_0}(1,1)\cdot h\cdot k\right|}{|k|}\right) = 0 .$$

In other terms, there exists the limit

$$d^2 f_{x_0}(1,1) = \lim_{h\to 0}\frac{f^{/}(x_0+h) - f^{/}(x_0)}{h} \overset{not.}{=} f^{//}(x_0) .$$

Consequently, $f^{//}(x_0) = d^2 f_{x_0}(1, 1)$. ◇

2.12. <u>Proposition</u>. Let A be an open subset of $\mathbb{R}$, and let $f: A\to\mathbb{R}^m$, $m > 0$, be a vector function of components $f^{\,i}: A\to\mathbb{R}$, $i = 1, 2, \dots, m$. Function f is differentiable at a point $x_0 \in A$ iff all of its components are differentiable, and, in this case, its differential has the form

$$df_{x_0} = (df^1_{x_0}, df^2_{x_0}, ..., df^m_{x_0})$$

<u>Proof</u>. The differentiability of f involves a limiting process, which reduces to limits on each component. On the other hand, the general form of a linear function $L\in \mathscr{L}(\mathbb{R}, \mathbb{R}^m) = \mathscr{B}(\mathbb{R}, \mathbb{R}^m)$ is

$$L(h) = (c_1 h, c_2 h, \dots, c_m h) .$$

In particular, the differential of f at x_0 has the same form,

$$df_{x_0}(h) = (c_1 h, c_2 h, \dots, c_m h),$$

where $c_i = f^{\,i/}(x_0)$ for all $i = 1, 2, \dots, m.$ ◇

<u>2.13. Remark</u>. We may use the above proposition with the aim of writing the differential of a vector function as the differential of a real function of one real variable (if necessary, see the previous section). To obtain that form, we first introduce the *derivative* of the vector function f at x_0 by

$$f^{/}(x_0) = \left(f^{1/}(x_0), f^{2/}(x_0), ..., f^{m/}(x_0)\right) .$$

Using this notion, the expression of the differential of f at x_0 , established in proposition 2.12, becomes

$$df_{x_0}(h) = f^{/}(x_0)\cdot h .$$

The only difference between the two cases is that "·" stands here for the product of a *vector*, namely $f^{/}(x_0)$, by a scalar $h\in\mathbb{R}$.

For studying the differential of a function of more than one variables, we need the following specific notions:

2.14. Definition. Let us consider $f: A\to\mathbb{R}$, where $A\subseteq\mathbb{R}^p$ is open, and $p > 1$. We say that f is *partially derivable* at the point $x_0 = (x_1^0, \ldots, x_p^0) \in A$, relative to the i'th variable, iff there exists the finite limit (in $\mathbb{R}$),

$$\lim_{h_i\to 0}\frac{f(x_1^0,\ldots,x_{i-1}^0,x_i^0+h_i,x_{i+1}^0,\ldots,x_p^0)-f(x_0)}{h_i}\overset{not.}{=}\frac{\partial f}{\partial x_i}(x_0).$$

If so, this limit is called *partial derivative* of f relative to x_i. In other words, to obtain the i'th partial derivative, we fix the variables different from x_i, and we derive as usually relative to this variable.

If f is partially derivable relative to all of its variables, at the point x_0, then the vector $\left(\frac{\partial f}{\partial x_1}(x_0),\ldots,\frac{\partial f}{\partial x_p}(x_0)\right)$ is said to be the *gradient* of f at x_0, and we note it $(grad\, f)(x_0)$, or simply $grad\, f(x_0)$.

2.15. Proposition. Let us consider a function $f: A\to\mathbb{R}$, where $A\subseteq\mathbb{R}^p$ is open, and $p > 1$. If f is differentiable at $x_0 = (x_1^0, \ldots, x_p^0) \in A$, then it has all the partial derivatives at x_0, and its differential has the form

$$df_{x_0}(h) = \langle grad\, f(x_0), h\rangle,$$

where $h = (h_1, \ldots, h_p) \in \mathbb{R}^p$, and $\langle\, .\, ,\, .\, \rangle$ denotes the scalar product of the Euclidean space $\mathbb{R}^p$. In other terms, we *represent* the differential $df_{x_0}(h)$ by the gradient of f at x_0, according to the formula

$$df_{x_0}(h) = \frac{\partial f}{\partial x_1}(x_0)\cdot h_1+\ldots+\frac{\partial f}{\partial x_p}(x_0)\cdot h_p.$$

Proof. By hypothesis, there exists $L_{x_0} \in \mathscr{L}(\mathbb{R}^p, \mathbb{R}) = \mathscr{B}(\mathbb{R}^p, \mathbb{R})$, such that

$$\lim_{\|h\|\to 0}\frac{\left|f(x_0+h)-f(x_0)-L_{x_0}(h)\right|}{\|h\|}=0.$$

A linear function from $\mathbb{R}^p$ to $\mathbb{R}$ is specified by $c_1, \ldots, c_p \in\mathbb{R}$, in the formula

$$L_{x_0}(h) = c_1 h_1 + \ldots + c_p h_p.$$

In our concrete case, we have to express the constants $c_1, \ldots, c_p$ by f. With this purpose, for each $i\in\{1, \ldots, p\}$, we consider increments of the form

$$h = (0, \ldots, 0, h_i, 0, \ldots, 0),$$

such that $L_{x_0}(h) = c_i h_i$. The differentiability of f leads to

$$\lim_{h_i\to 0}\left[\frac{f(x_1^0,\ldots,x_i^0+h_i,\ldots,x_p^0)-f(x_0)}{h_i}-c_i\right]=0,$$

which shows that $c_i = \frac{\partial f}{\partial x_i}(x_0)$ for each $i = 1, 2, \ldots, p$. In conclusion, the differential takes the form

$$df_{x_0}(h) = \frac{\partial f}{\partial x_1}(x_0)\cdot h_1 + \ldots + \frac{\partial f}{\partial x_p}(x_0)\cdot h_p = < grad\, f(x_0), h > ,$$

which gives a representation of df_{x_0} by $grad\, f(x_0)$. ◇

2.16. <u>Remarks</u>. 1) The simple existence of $(grad\, f)(x_0)$ does not assure the differentiability of f at x_0. For example, the function $f:\mathbb{R}^2 \to \mathbb{R}$, of values

$$f(x, y) = \begin{cases} 1 & if\ x \neq 0\ and\ y \neq 0 \\ 0 & if\ either\ x = 0\ or\ y = 0 \end{cases},$$

is partially derivable at the origin, and $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0$. However, it is not continuous at (0, 0), hence it is not differentiable at this point. In the next section we will see that the continuity of the partial derivatives is a sufficient condition for its differentiability.

2) The most frequent form of the differential of a function depending on several real variables is

$$df = \frac{\partial f}{\partial x_1}dx_1 + \frac{\partial f}{\partial x_2}dx_2 + \ldots + \frac{\partial f}{\partial x_p}dx_p.$$

To obtain the precise meaning of this formula, we have to consider the *projections* $P_i : \mathbb{R}^p \to \mathbb{R}$, expressed by $P_i(x_1, \ldots, x_i, \ldots, x_p) = x_i$. . It is easy to see that each projection P_i, $i = 1, 2, \ldots, p$, is a linear and continuous function, i.e. $P_i \in \mathscr{L}(\mathbb{R}^p, \mathbb{R}) = \mathscr{B}(\mathbb{R}^p, \mathbb{R})$. In addition, each P_i is differentiable at any $x_0 \in \mathbb{R}^p$, and $dP_i(x_0) = P_i$. By convention (and tradition) we note $dP_i = dx_i$, so that $(dx_i)_{x_0}(h_1, \ldots, h_p) = P_i(h_1, \ldots, h_p) = h_i$ for all indices $i = 1, 2, \ldots, p$. Replacing these expressions of h_i in the differential of f , we obtain the formula

$$df_{x_0}(h) = \frac{\partial f}{\partial x_1}(x_0)(dx_1)_{x_0}(h) + \ldots + \frac{\partial f}{\partial x_p}(x_0)(dx_p)_{x_0}(h).$$

Because h is arbitrary in $\mathbb{R}^p$, we obtain the relation

$$df_{x_0} = \left(\frac{\partial f}{\partial x_1}dx_1\right)_{x_0} + \ldots + \left(\frac{\partial f}{\partial x_p}dx_p\right)_{x_0}.$$

It remains to take into consideration the fact that x_0 is arbitrary in the set $A \subseteq \mathbb{R}^p$, on which f is differentiable. In applications we sometimes meet the formula $df = < grad\, f, dx >$, where $dx = (dx_1, dx_2, \ldots, dx_p)$.

The following notions and notations will be useful in introducing the differential of a vector function of several real variables.

2.17. <u>Definition</u>. Let the spaces $\mathscr{X} = \mathbb{R}^p$ and $\mathscr{Y} = \mathbb{R}^m$ be endowed with their Euclidean norms. We note by A an open set of $\mathscr{X}$, and we define the vector function $f : A \to \mathscr{Y}$, of components $f_i : A \to \mathbb{R}$, $I = 1,2, \ldots, m$, which are considered partially derivable. Then the matrix

$$\mathrm{J}f(x_0) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) \ldots \frac{\partial f_1}{\partial x_p}(x_0) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) \ldots \frac{\partial f_m}{\partial x_p}(x_0) \end{pmatrix}$$

is called *Jacobi's matrix* of f at $x_0 \in A$. Obviously, at each x_0 fixed in A, we have $\mathrm{J}f(x_0) \in \mathcal{M}_{m,p}(\mathbb{R})$. In the case $m = p$, the determinant

$$\mathrm{Det}\,(\mathrm{J}f(x_0)) \overset{not.}{=} \frac{D(f_1,\ldots,f_m)}{D(x_1,\ldots,x_m)}(x_0)$$

is called *Jacobian* of f at x_0 .

2.18. <u>Proposition</u>. If the function f in the above definition is differentiable at $x_0 \in A$, then its differential has the value

$$df_{x_0}(h) = \mathrm{J}f(x_0)\cdot h = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) \ldots \frac{\partial f_1}{\partial x_p}(x_0) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \frac{\partial f_m}{\partial x_1}(x_0) & \frac{\partial f_m}{\partial x_2}(x_0) \ldots \frac{\partial f_m}{\partial x_p}(x_0) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}.$$

<u>Proof.</u> Because $df_{x_0} \in \mathscr{L}(\mathbb{R}^p, \mathbb{R}^m)$, i.e. it is a linear operator acting between finite dimensional linear spaces, it follows that its value at an arbitrary point $h = (h_1, \ldots, h_p)^T \in \mathbb{R}^p$ is given by the formula $df_{x_0}(h) = (c_{ij})\, h$. What remains is to see how the constants c_{ij} depend on f . If $f_1, \ldots, f_m$ are the components of f, then, similarly to proposition 2.12, it follows that they are differentiable at x_0, and the general form of their differentials is

$$(df_i)_{x_0}(h) = c_{i1} h_1 + \ldots + c_{ip} h_p\ ,$$

where $i = 1, \ldots, m$. More than this, according to proposition 2.15, we know that the values of these differentials are given by the formula

$$(df_i)_{x_0}(h) = \frac{\partial f_i}{\partial x_1}(x_0)h_1 + \ldots + \frac{\partial f_i}{\partial x_p}(x_0)h_p$$

for each $i = 1, \ldots, m$. Consequently, $c_{ij} = \frac{\partial f_i}{\partial x_j}(x_0)$. ◇

2.19. Remark. a) As before (see Remark 2.16), the existence of the partial derivatives $\frac{\partial f_i}{\partial x_j}(x_0)$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, p$, which represent the elements of J$f(x_0)$, does not assure the differentiability of f at x_0 .
b) We may write the differential of a vector function of several variables in a symbolic form too, by using the projections $P_j : \mathbb{R}^p \to \mathbb{R}$, $j = 1, \ldots, p$. In fact, if we recall the traditional notation $dP_i = dx_i$, then we have

$$df_{x_0} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0)(dx_1)_{x_0} + \ldots + \frac{\partial f_1}{\partial x_p}(x_0)(dx_p)_{x_0} \\ \cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \\ \frac{\partial f_m}{\partial x_1}(x_0)(dx_1)_{x_0} + \ldots + \frac{\partial f_m}{\partial x_p}(x_0)(dx_p)_{x_0} \end{pmatrix}.$$

If f is differentiable at any $x_0 \in A \subseteq \mathbb{R}^p$, then we may omit to mention x_0 , and so we obtain another symbolic form of the differential, namely

$$df = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}dx_1 + \ldots + \frac{\partial f_1}{\partial x_p}dx_p \\ \cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \;\cdot \\ \frac{\partial f_m}{\partial x_1}dx_1 + \ldots + \frac{\partial f_m}{\partial x_p}dx_p \end{pmatrix}.$$

This one leads to the shortest form of the differential, which is

$$df = (\mathrm{J}f) \cdot \left(dx_1 \ldots dx_p\right)^T = \mathrm{J}f\, dx.$$

If we compare the formula $df = f' dx$, which represents the differential of a real function depending on a single real variable, to the similar formulas $df = < grad\, f, dx >$ and $df = \mathrm{J}f\, dx$ for functions of several (real) variables, then we see that *grad f* and Jf stand for f' . In this sense, we can interpret *grad f* and Jf as representing the *derivative* of a real, respectively vectorial function of several variables.

2.20. Approximating functions of several variables. The differential of a real function of several variables (sometimes called total differential) may be interpreted like a linear approximation of the (total) increment of the function, corresponding to the increments $\Delta x = dx$, $\Delta y = dy$, etc. of the variables. For example, if $f(x, y) = x^2 + xy - y^2$, then

$$\Delta f(x, y) = f(x + \Delta x, y + \Delta y) - f(x, y) =$$
$$= [(2x + y)\, \Delta x + (x - 2y)\, \Delta y] + [\Delta x^2 + \Delta x \Delta y - \Delta y^2] .$$

The *principal* part of this increment is the differential

$$df_{(x, y)} = (2x + y)\, dx + (x - 2y)\, dy ,$$

which expresses a linear approximation of the forthcoming error.

PROBLEMS § IV.2.

1. Show that every continuous bilinear function $f \in \mathcal{B}_2(\mathcal{X}_1, \mathcal{X}_2; \mathcal{Y})$ is differentiable at each $(x_1^0, x_2^0) \in \mathcal{X}_1 \times \mathcal{X}_2 = \mathcal{X}$, and

$$df_{(x_1^0, x_2^0)}(h,k) = f(x_1^0, k) + f(h, x_2^0).$$

Hint. Evaluate

$$\frac{\left\|\omega_{(x_1^0,x_2^0)}(h,k)\right\|_2}{\|(h,k)\|_1} = \frac{\|f(h,k)\|_2}{\sqrt{\|h\|^2+\|k\|^2}} \le \frac{\|f\|\cdot\|h\|\cdot\|k\|}{\sqrt{\|h\|^2+\|k\|^2}} \le \|f\|\cdot\|h\|.$$

2. Show that for real functions f and g, depending on several real variables, the ordinary rules of differentiation remain valid, i.e.

$$d(f+g) = df + dg\ ;\ d(fg) = f\,dg + g\,df\ ;\ d\left(\frac{f}{g}\right) = \frac{gdf - fdg}{g^2}\ .$$

3. Show that the relative error of a product of real functions of several variables approximately equals the sum of the relative errors of the factors.

4. Evaluate the function $f(x, y) = x^2 \sin y$ at $x=1.1$, $y=33°$, approximately.
Hint. Use the formula of the linear approximation

$$f(x_0+h, y_0+k) \cong f(x_0,y_0) + \frac{\partial f}{\partial x}(x_0,y_0)\cdot h + \frac{\partial f}{\partial y}(x_0,y_0)\cdot k\ .$$

5. Compute $1.02^{3.01}$ approximately.
Hint. Take $f(x, y) = x^y$, $x_0 = 1$, $y_0 = 3$, $\Delta x = 0.02$, $\Delta y = 0.01$, and compute $1.02^{3.01} \cong 1 + df_{(1,3)}(0.02, 0.01)$.

6. The measurements of a triangle ABC yield the following approximate values: side $a = 100m \pm 2m$, side $b = 200m \pm 3m$ and angle $C = 60° \pm 1°$. To what degree of accuracy can be computed side c?
Hint. Express c by the generalized Pythagorean formula and compute dc.

7. Show that the function $f:\mathbb{R}^2 \to \mathbb{R}$, which takes the values

$$f(x,y) = \begin{cases} \dfrac{xy}{\sqrt{x^2+y^2}} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0)\ , \end{cases}$$

is continuous and has partial derivatives on the entire $\mathbb{R}^2$, but it is not differentiable at (0, 0).

§ IV.3. FUNCTIONS OF SEVERAL REAL VARIABLES

We will use the notions introduced in the previous section to develop in more details the differential calculus of real functions depending on several real variables. We begin by the *directional derivative* of a *scalar field*:

3.1. <u>Definition</u>. The real functions depending on several real variables are named *scalar fields*. If $f: A \to \mathbb{R}$ is a scalar field, i.e. A is an open subset of $\mathbb{R}^p$, $p \geq 1$, then the sets of the form

$$\{x \in A : f(x) = c\} = S_c \ ,$$

where $c \in f(A)$, are called *level surfaces*. In particular, each point $x_0 \in A$ belongs to a level surface, namely that one for which $c = f(x_0)$. We say that f is *smooth* iff it has continuous (partial) derivatives at each $x_0 \in A$.

Let us choose a point $x_0 \in A$ and a unit vector $\ell = (\ell_1, \ell_2, ..., \ell_p) \in \mathbb{R}^p$ (some authors call ℓ *direction*, and specify the condition $\|\ell\| = 1$). We say that the field f is *derivable in the direction* ℓ *at the point* x_0 iff there exists the limit

$$\lim_{t \to 0} \frac{f(x_0 + t\ell) - f(x_0)}{t} \overset{not.}{=} \frac{\partial f}{\partial \ell}(x_0).$$

If so, then we call it *derivative* of f *in the direction* ℓ at the point x_0 .

3.2. <u>Remarks</u>. *a*) The derivative in a direction may be interpreted in terms of level surfaces. In particular, in $\mathbb{R}^3$, we may illustrate it as in Fig.IV.3.1.

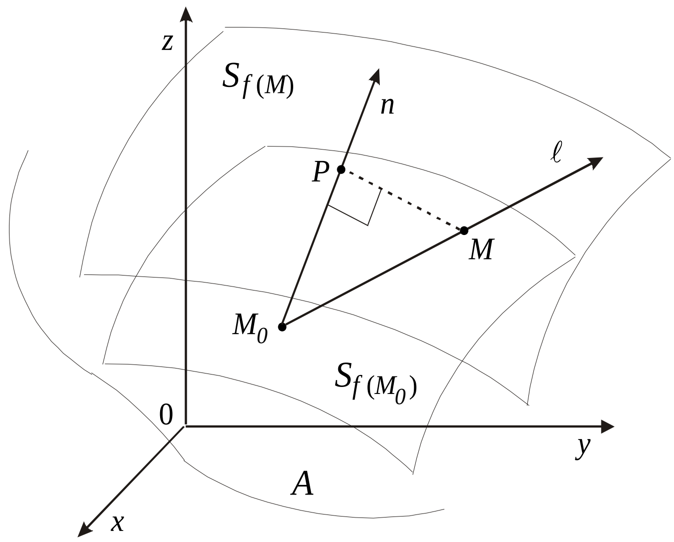


Fig. IV.3.1.

Because $\|\ell\| = 1$, we have $M_0 M = \|t\ell\| = t$, hence the directional derivative of f at M_0 allows the more geometrical construction

$$\frac{\partial f}{\partial \ell}(M_0) = \lim_{M \to M_0} \frac{f(M) - f(M_0)}{M_0 M} \ .$$

In particular, we can see that the most rapid variation of the field f at the point M_0 is obtained in the direction of the normal to the level surface (which exists for smooth fields), when $M_0 M$ is minimal.

b) The derivative of f in a direction ℓ at the point $x_0 \in A$ reduces to the usual derivative of $g:(-\delta, \delta) \to \mathbb{R}$ at $0 \in \mathbb{R}$, where $g(t) = f(x_0 + tl)$. The domain of g is determined by the number $\delta > 0$, which is chosen to obey the condition that $x_0 + tl \in A$ for all $t \in (-\delta, \delta)$ (possibly because A is open!).

c) When we speak of a *direction*, we tacitly include the *orientation* of the vector ℓ. The derivative in a direction depends on this orientation, in the sense that reversing the orientation changes the sign. More exactly,

$$\frac{\partial f}{\partial(-\ell)}(x_0) = -\frac{\partial f}{\partial \ell}(x_0).$$

d) Each partial derivative is a derivative in the direction of one of the coordinate axes. More exactly, if $\ell_k = (0,...,0,1,0,...,0)$, where 1 stands on the k^{th} position, denotes the direction of the k^{th} axis, $k \in \{1, \dots, p\}$, then

$$\frac{\partial f}{\partial x_k}(x_0) = \frac{\partial f}{\partial \ell_k}(x_0) \; .$$

Now we can extend the previous result concerning the partial derivability of a differentiable function, according to the following:

3.3. Theorem. If $f: A \to \mathbb{R}$, $A \subseteq \mathbb{R}^p$, is a differentiable field at $x_0 \in A$, then it is derivable in any direction $\ell \in \mathbb{R}^p$, and

$$\frac{\partial f}{\partial \ell}(x_0) = d\, f_{x_0}(\ell) \; .$$

Proof. According to theorem IV.2.3, the differential at ℓ equals

$$df_{x_0}(\ell) = \lim_{t \to 0} \frac{f(x_0 + t\ell) - f(x_0)}{t} \; .$$

In the terms of the above definition, this limit means $\frac{\partial f}{\partial \ell}(x_0)$. ◇

3.4. Consequences. a) The derivative of the scalar field f in the direction ℓ, at the point x_0, equals the projection of *grad* $f(x_0)$ on the unit vector ℓ. In fact, this is a consequence of the relations $\|\ell\| = 1$, and

$$\frac{\partial f}{\partial \ell}(x_0) = d\, f_{x_0}(\ell) = \langle grad\, f(x_0), \ell \rangle = \|grad\; f(x_0)\| \cdot \cos(\angle MM_0P).$$

b) The greatest value of such a projection is obtained when the vectors ℓ and *grad* $f(x_0)$ are parallel. This means $\ell \| n$, where n denotes the normal to the level surface at x_0. In this case, since $\|n\| = 1$ too, we have

$$\frac{\partial f}{\partial n}(x_0) = \langle grad\; f(x_0), n \rangle = \|grad\; f(x_0)\| \; .$$

In $\mathbb{R}^3$, after calculating $(grad\ f)(x_0)$, we can find $\frac{\partial f}{\partial \ell}(x_0)$ using the sphere of diameter $grad\, f(x_0)$, as in Fig.IV.3.2., where it equals $M_0 Q$.

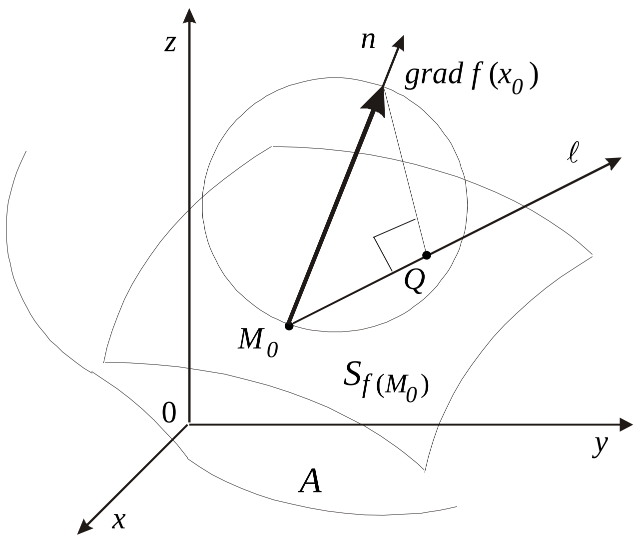


Fig. IV.3.2.

The following sufficient condition of differentiability is very useful in many practical problems:

3.5. Theorem. If f is partially derivable on the neighborhood V of x_0 and all its partial derivatives $\frac{\partial f}{\partial x_k}: V \to \mathbb{R}$, $k = \overline{1, p}$, are continuous at x_0, then this function is differentiable at x_0. By extension to A, $f \in \mathbf{C}^1_{\mathbb{R}}(A)$ implies the differentiability of f on A.

Proof. Let us note $x_0 = (x_1^0, x_2^0, ..., x_p^0) \in A$, and $U = \{h \in \mathbb{R}^p : x_0 + h \in A\}$. If $h = (h_1, h_2, ..., h_p) \in U$, including the case of some $h_k = 0$, then we may decompose the increment $f(x_0 + h) - f(x_0)$ in the following sum:

$$\left[f(x_1^0 + h_1, x_2^0 + h_2, ..., x_p^0 + h_p) - f(x_1^0, x_2^0 + h_2, ..., x_p^0 + h_p) \right] +$$
$$+ \left[f(x_1^0, x_2^0 + h_2, ..., x_p^0 + h_p) - f(x_1^0, x_2^0, x_3^0 + h_3, ..., x_p^0 + h_p) \right] + ... +$$
$$+ \left[f(x_1^0, x_2^0, ..., x_{p-1}^0, x_p^0 + h_p) - f(x_1^0, x_2^0, ..., x_{p-1}^0, x_p^0) \right].$$

Using the Lagrange's theorem on finite increments, successively applied to each of the square brackets from above, we obtain

$$f(x_0 + h) - f(x_0) = h_1 \cdot \frac{\partial f}{\partial x_1}(\xi_1, x_2^0 + h_2, ..., x_p^0 + h_p) +$$
$$h_2 \cdot \frac{\partial f}{\partial x_2}(x_1^0, \xi_2, x_3^0 + h_3, ..., x_p^0 + h_p) + \ldots + h_p \cdot \frac{\partial f}{\partial x_p}(x_1^0, ..., x_{p-1}^0, \xi_p),$$

where ξ_k is between x_k^0 and $x_k^0 + h_k$ for all $k = \overline{1, p}$.

Now, let us consider the linear function $L_{x_0} : \mathbb{R}^p \to \mathbb{R}$, of values

$$L_{x_0}(h_1, \dots, h_p) = \frac{\partial f}{\partial x_1}(x_0) \cdot h_1 + \frac{\partial f}{\partial x_2}(x_0) \cdot h_2 + \dots + \frac{\partial f}{\partial x_p}(x_0) \cdot h_p \ .$$

To show that L_{x_0} is the differential of *f* at x_0, we first evaluate

$$f(x_0 + h) - f(x_0) - L_{x_0}(h) =$$

$$= \left[\frac{\partial f}{\partial x_1}(\xi_1, x_2^0 + h_2, \dots, x_p^0 + h_p) - \frac{\partial f}{\partial x_1}(x_0) \right] \cdot h_1 +$$

$$+ \left[\frac{\partial f}{\partial x_2}(x_1^0, \xi_2, x_3^0 + h_3, \dots, x_p^0 + h_p) - \frac{\partial f}{\partial x_2}(x_0) \right] \cdot h_2 + \dots +$$

$$+ \left[\frac{\partial f}{\partial x_p}(x_1^0, \dots, x_{p-1}^0, \xi_p) - \frac{\partial f}{\partial x_p}(x_0) \right] \cdot h_p \ .$$

Then, the differentiability of *f* at x_0 follows from the inequality

$$\frac{\left| f(x_0 + h) - f(x_0) - L_{x_0}(h) \right|}{\|h\|} \le$$

$$\le \left| \frac{\partial f}{\partial x_1}(\xi_1, x_2^0 + h_2, \dots, x_p^0 + h_p) - \frac{\partial f}{\partial x_1}(x_0) \right| \cdot \frac{|h_1|}{\|h\|} +$$

$$+ \left| \frac{\partial f}{\partial x_2}(x_1^0, \xi_2, x_3^0 + h_3, \dots, x_p^0 + h_p) - \frac{\partial f}{\partial x_2}(x_0) \right| \cdot \frac{|h_2|}{\|h\|} + \dots +$$

$$+ \left| \frac{\partial f}{\partial x_p}(x_1^0, \dots, x_{p-1}^0, \xi_p) - \frac{\partial f}{\partial x_p}(x_0) \right| \cdot \frac{|h_p|}{\|h_p\|} \ .$$

In fact, since $\frac{|h_k|}{\|h\|} \le 1$ for all $k = \overline{1, p}$, and the partial derivatives of *f* are continuous at x_0, it follows that

$$\lim_{\|h\| \to 0} \frac{\left| f(x_0 + h) - f(x_0) - L_{x_0}(h) \right|}{\|h\|} = 0 \ .$$

The last assertion of the theorem follows by applying the former result to arbitrary $x_0 \in A$. ◇

We mention that the continuity of the partial derivatives is still not necessary for differentiability (see problem 2 at the end of the section).

The rule of differentiating *composed functions*, established in theorem IV.2.6-b, has important consequences concerning the *partial derivatives* of such functions:

3.6. __Theorem__. Let us consider the functions $f: A \to B$, $g : B \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^p$ and $B \subseteq \mathbb{R}^n$ are open sets. If f is differentiable at $x_0 \in A$ and g is differentiable at $f(x_0) \in B$, then $g \circ f : A \to \mathbb{R}^m$ is differentiable at x_0 and for the corresponding Jacobi matrixes we have

$$\mathrm{J}(g \circ f)(x_0) = \mathrm{J}g(f(x_0)) \cdot \mathrm{J}f(x_0). \qquad (*)$$

Proof. The differentiability of $g \circ f$ was established in theorem IV.2.6-b, as well as the formula $d(g \circ f)_{x_0} = dg_{f(x_0)} \circ df_{x_0}$. According to proposition IV.2.18, we represent these differentials by Jacobi matrixes. Using a well-known theorem from linear algebra (see [AE], [KA], etc.), the matrix of the composed linear operators, in our case $dg_{f(x_0)}$ and df_{x_0}, equals the product of the corresponding matrixes. ◇

3.7. __Remark__. The formula (*) from above concentrates all the rules we need to write the partial derivatives of composed functions, as for example:

a) The case $p = m = 1$, $n = 2$. If t is the variable of f, and u, v are the variables of g, then formula (*) becomes:

$$(g \circ f)'(t) = \begin{pmatrix} \frac{\partial g}{\partial u}(f_1(t), f_2(t)) & \frac{\partial g}{\partial v}(f_1(t), f_2(t)) \end{pmatrix} \cdot \begin{pmatrix} f_1'(t) \\ f_2'(t) \end{pmatrix} =$$

$$= \frac{\partial g}{\partial u}\big(f_1(t), f_2(t)\big) \cdot f_1'(t) + \frac{\partial g}{\partial v}\big(f_1(t), f_2(t)\big) \cdot f_2'(t) .$$

b) The case $p = n = 2$, $m = 1$. Let x, y be the variables of f and u, v be the variables of g. From (*) we obtain:

$$\begin{pmatrix} \frac{\partial (g \circ f)}{\partial x}(x,y) & \frac{\partial (g \circ f)}{\partial y}(x,y) \end{pmatrix} =$$

$$= \begin{pmatrix} \frac{\partial g}{\partial u}(f_1(x,y), f_2(x,y)) & \frac{\partial g}{\partial v}(f_1(x,y), f_2(x,y)) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f_1}{\partial x}(x,y) & \frac{\partial f_1}{\partial y}(x,y) \\ \frac{\partial f_2}{\partial x}(x,y) & \frac{\partial f_2}{\partial y}(x,y) \end{pmatrix} .$$

Consequently, the partial derivatives of $h = g \circ f$ are:

$$\frac{\partial h}{\partial x}(x,y) = \frac{\partial g}{\partial u}(f_1(x,y), f_2(x,y)) \frac{\partial f_1}{\partial x}(x,y) + \frac{\partial g}{\partial v}(f_1(x,y), f_2(x,y)) \frac{\partial f_2}{\partial x}(x,y)$$

$$\frac{\partial h}{\partial y}(x,y) = \frac{\partial g}{\partial u}(f_1(x,y), f_2(x,y)) \frac{\partial f_1}{\partial y}(x,y) + \frac{\partial g}{\partial v}(f_1(x,y), f_2(x,y)) \frac{\partial f_2}{\partial y}(x,y)$$

Generalizing these formulas, we may retain that each partial derivative of a composed function equals the sum of specific products taken over all the components, which contain the chosen variable.

c) The case $p = m = n$. In this case, it is recommended to interpret f and g as transformations of $\mathbb{R}^n$. Because the determinant of a product of matrixes equals the product of the corresponding determinants, we obtain:

$$\frac{D((g \circ f)_1,...,(g \circ f)_n)}{D(x_1,...,x_n)} = \frac{D(g_1,...,g_n)}{D(u_1,...,u_n)} \cdot \frac{D(f_1,...,f_n)}{D(x_1,...,x_n)}.$$

Another important topic of the differential calculus refers to the higher order partial derivatives, which are introduced by the following

3.8. Definition. Let $f: A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^p$ is open, be partially derivable relative to x_k , $1 \leq k \leq p$, in a neighborhood $V \subseteq A$ of $x_0 \in A$. If $\frac{\partial f}{\partial x_k}: V \to \mathbb{R}$ is also derivable relative to x_j , $1 \leq j \leq p$, at x_0, then we say that f is *two times derivable* at x_0 relative to x_k and x_j . The second order derivative is noted

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_k}\right)(x_0) = \frac{\partial^2 f}{\partial x_j \partial x_k}(x_0).$$

If all second order derivatives there exist and are continuous on A, then we say that f is of class $\mathbf{C}^2$, and we write $f \in \mathbf{C}^2_{\mathbb{R}}(A)$.

By induction, we define the *higher order derivatives*, k_1 times relative to x_1 , and so on, k_p times relative to x_p , where $k_1 + \ldots + k_p = n$, noted

$$\frac{\partial^n f}{\partial x_1^{k_1} \ldots \partial x_p^{k_p}}.$$

Similarly we define classes $\mathbf{C}^n$ and $\mathbf{C}^\infty$.

3.9. Remark. The higher order derivatives depend on the order in which we realize each derivation. In particular, $\frac{\partial^2 f}{\partial x_j \partial x_k}$ means that we have first derived relative to x_k , and after that relative to x_j . The result is generally different from $\frac{\partial^2 f}{\partial x_k \partial x_j}$, where we have derived in the inverse order. For example, let us consider the function $f: \mathbb{R}^2 \to \mathbb{R}$, of values

$$f(x,y) = \begin{cases} xy\dfrac{x^2 - y^2}{x^2 + y^2} & if\ (x,y) \neq (0,0) \\ 0 & if\ (x,y) = (0,0) \end{cases}.$$

A direct evaluation shows that $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1 \neq \frac{\partial^2 f}{\partial y \partial x}(0,0) = -1$.

Such situations justify our interest in knowing sufficient conditions for the equality of the mixed partial derivatives realized in different orders:

3.10 <u>Theorem</u>. (due to *H. A. Schwarz*) If $f \in \mathbf{C}^2_{\mathbb{R}}(A)$, where $A \subseteq \mathbb{R}^p$ is an open set, then the equality

$$\frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j}$$

holds on A, for all $j, k = 1, 2, \dots, p$.

<u>Proof</u>. It is sufficient to discuss the case $p = 2$. Let us consider the auxiliary functions $\varphi(x, y) = f(x, y) - f(x_0, y)$ and $\psi(x, y) = f(x, y) - f(x, y_0)$, so that

$$\varphi(x, y) - \varphi(x, y_0) = \psi(x, y) - \psi(x_0, y).$$

By applying the Lagrange's theorem to each side of this equality we obtain:

$$\frac{\partial \varphi}{\partial y}(x, \eta_1) \cdot (y - y_0) = \frac{\partial \psi}{\partial x}(\xi_1, y) \cdot (x - x_0).$$

It remains to replace φ and ψ, namely to apply the finite increments' theorem, and finally to use the continuity of the mixed derivatives. ◇

3.11. <u>Remark</u>. The continuity of the mixed derivatives is <u>not necessary</u> for their equality. This is visible in the case of $f : \mathbb{R}^2 \to \mathbb{R}$, of values

$$f(x, y) = \begin{cases} y^2 \ln\left(1 + \dfrac{x^2}{y^2}\right) & if \quad y \neq 0 \\ 0 & if \quad y = 0 \ . \end{cases}$$

In fact, a direct calculation leads to

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = \begin{cases} \dfrac{4x^3 y}{(x^2 + y^2)^2} & if \quad y \neq 0 \\ 0 & if \quad y = 0 \ . \end{cases}$$

This function is not continuous at the origin. However,

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{x \to 0} \frac{1}{x} \cdot \left[\frac{\partial f}{\partial y}(x,0) - \frac{\partial f}{\partial y}(0,0)\right] = 0,$$

hence the equality $\dfrac{\partial^2 f}{\partial y \partial x}(0,0) = 0 = \dfrac{\partial^2 f}{\partial x \partial y}(0,0)$ is valid.

Now, let us study the relations between the second order differentials and the second order derivatives of a real function of several variables. This study is based on a specific notion involving the partial derivatives:

3.12. <u>Definition</u>. Let us consider that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^p$ is an open set, has all second order partial derivatives at $x_0 \in A$. The matrix

$$\mathbf{H} f(x_0) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)\right)_{i, j = 1, \dots, p}$$

is called *Hesse matrix* (or *Hessean*) of f at x_0.

Obviously, the Hessean matrix of $f \in \mathbf{C}^2_{\mathbb{R}}(A)$ is symmetric by virtue of Schwarz' theorem 3.10. To see how the Hessean matrix represents a second order differential, we recall that $d^2 f_{x_0}$ is a bilinear function, and matrixes represent bilinear functions on finite dimensional spaces. More exactly:

3.13. __Lemma__. Let $\mathcal{B}$ denote a fixed orthonormal basis in $\mathbb{R}^p$ (e.g. the canonical one, as in I.3.8b). To each bilinear function $\mathbf{B} : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ there corresponds a square matrix $B = (b_{ij})_{i,j=1,\dots,p}$ such that $\mathbf{B}(x, y) = x^T B\, y$, i.e.

$$\mathbf{B}(x, y) = \begin{pmatrix} x_1 & \cdots & x_p \end{pmatrix} \cdot \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & & \vdots \\ b_{p1} & \cdots & b_{pp} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} .$$

__Proof__. Referred to the (canonical) base $\mathcal{B} = \{e_1, \dots, e_p\}$ of $\mathbb{R}^p$, we have

$$x = \sum_{i=1}^{p} x_i e_i \text{ and } y = \sum_{j=1}^{p} y_j e_j .$$

If we note $b_{ij} = \mathbf{B}(e_i, e_j)$, then $\mathbf{B}(x, y) = \sum_{i,j=1}^{p} b_{ij} x_i y_j$. ◇

In our case, we have to express the components b_{ij} by the corresponding values of $\mathbf{B} = d^2 f_{x_0}$, using the second order partial derivatives of f.

3.14. __Theorem__. Let $\mathcal{B} = \{e_1, \dots, e_p\}$ be the canonical basis of $\mathbb{R}^p$. If the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^p$ is an open set, is two times differentiable at $x_0 \in A$, then:

a) f is twice partially derivable relative to all its variables;

b) $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) = d^2 f_{x_0}(e_i, e_j)$ holds for all $i, j = 1, \dots, p$;

c) $d^2 f_{x_0}$ is represented by the Hessean matrix $\mathbf{H} f(x_0)$, i.e.

$$d^2 f_{x_0}(h, k) = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial y_j}(x_0) h_i k_j .$$

__Proof.__ For the assertions a) and b), we start with the hypothesis that there exists $d^2 f_{x_0} \in \mathscr{B}(\mathbb{R}^p, \mathscr{B}(\mathbb{R}^p, \mathbb{R}))$ such that

$$\lim_{\|h\| \to 0} \frac{\left\| df_{x_0+h} - df_{x_0} - d^2 f_{x_0}(h, .) \right\|}{\|h\|} = 0 .$$

Using the expression of the norm of a linear function, we obtain

$$\lim_{\|h\|\to 0} \frac{\left|df_{x_0+h}(k) - df_{x_0}(k) - d^2 f_{x_0}(h,k)\right|}{\|h\|\|k\|} = 0 .$$

Let us remark that $df_{x_0}(e_j) = \frac{\partial f}{\partial x_j}(x_0)$, and $d^2 f_{x_0}(h, k) = u \cdot d^2 f_{x_0}(e_i , e_j)$ whenever $k = e_j$ and $h = u\, e_i$ for some $u \in \mathbb{R}$. If $u \to 0$, then $h \to 0$, hence

$$\lim_{u\to 0} \left| \frac{\frac{\partial f}{\partial x_j}(x_0 + ue_i) - \frac{\partial f}{\partial x_j}(x_0)}{u} - d^2 f_{x_0}(e_i, e_j) \right| = 0 .$$

c) We apply the above lemma 3.13 to $\mathbf{B} = d^2 f_{x_0}$. ◇

3.15. **<u>Remark</u>.** The existence of the second order partial derivatives <u>does not generally assure</u> the twice differentiability of a function. To exemplify this fact we may use the same function $f: \mathbb{R}^2 \to \mathbb{R}$ as in remark 3.11, namely

$$f(x, y) = \begin{cases} y^2 \ln\left(1 + \frac{x^2}{y^2}\right) & if \quad y \neq 0 \\ 0 & if \quad y = 0 \ . \end{cases}$$

In fact, it has null partial derivatives of the first and second order at (0, 0), so if we suppose the double differentiability, then we should have

$$df_{(0,0)} = \theta_{\mathcal{B}(\mathbb{R}^2,\mathbb{R})}, \text{ and } d^2 f_{(0,0)} = \theta_{\mathcal{B}(\mathbb{R}^2,\mathcal{B}(\mathbb{R}^2,\mathbb{R}))} .$$

Consequently, from

$$\lim_{\|h\|\to 0} \frac{\left\|df_h - df_{(0,0)} - d^2 f_{(0,0)}(h)\right\|}{\|h\|} = 0$$

we deduce

$$\lim_{\|h\|\to 0} \frac{\|df_h\|}{\|h\|} = 0 = \lim_{\|h\|\to 0} \frac{|df_h(k)|}{\|h\|\|k\|} = \lim_{\|h\|\to 0} \frac{\left|\frac{\partial f}{\partial x}(h_1,h_2)k_1 + \frac{\partial f}{\partial y}(h_1,h_2)k_2\right|}{\|h\|\|k\|} .$$

On the other hand, evaluating $\frac{\partial f}{\partial x}(x, y)$ and $\frac{\partial f}{\partial y}(x, y)$ at $(x, y) \neq (0, 0)$, and considering $k = (0,1)$, $h = \left(u, \frac{u}{\sqrt{e-1}}\right)$, we obtain $\frac{\|df_h\|}{\|h\|} = \frac{2}{e\sqrt{e}}$, which does not tend to *0*. The contradiction shows that $d^2 f$ does not exist at (0, 0).

As in the case of the first differential (compare to theorem 3.5 from above), we can show that the continuity of the second order derivatives is a sufficient condition for the double differentiability:

3.16. Theorem. Let $A \subseteq \mathbb{R}^p$ be an open set, let $x_0 \in A$ be fixed, and let V be a neighborhood of x_0, such that $V \subseteq A$. If the function $f : A \to \mathbb{R}$ has partial derivatives of the second order on V, which are continuous at x_0, then this function is twice differentiable at x_0. In addition, if $f \in \mathbf{C}^2(A)$, then f is two times differentiable on A.

Proof. If we note $h = (h_1, h_2, ..., h_p)$, and $k = (k_1, k_2, ..., k_p)$, then the map

$$(h,k) \to \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \cdot h_i \cdot k_j \ ,$$

is bilinear from $\mathbb{R}^p \times \mathbb{R}^p$ to $\mathbb{R}$. To each $h \in \mathbb{R}^p \setminus \{\theta_{\mathbb{R}^p}\}$ we attach a linear function $L_{x_0}(h)$, which carries $k \in \mathbb{R}^p$ to $\sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \cdot h_i \cdot k_j$. To show that df is differentiable at x_0, we first evaluate:

$$\frac{\left\| (df)(x_0+h) - (df)(x_0) - L_{x_0}(h) \right\|}{\|h\|} =$$

$$= \frac{1}{\|h\|} \cdot \sup_{k \neq \theta_{\mathbb{R}^p}} \frac{\left| df_{x_0+h}(k) - df_{x_0}(k) - L_{x_0}(h)(k) \right|}{\|k\|} =$$

$$\frac{1}{\|h\|} \cdot \sup_{k \neq \theta_{\mathbb{R}^p}} \frac{\left| \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}(x_0+h) \cdot k_j - \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}(x_0) \cdot k_j - \sum_{j=1}^{p} \sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i k_j \right|}{\|k\|} =$$

$$= \frac{1}{\|h\|} \cdot \sup_{k \neq \theta_{\mathbb{R}^p}} \frac{\left| \sum_{j=1}^{p} \left[\frac{\partial f}{\partial x_j}(x_0+h) - \frac{\partial f}{\partial x_j}(x_0) - \sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i \right] \cdot k_j \right|}{\|k\|} \leq$$

$$\leq \frac{1}{\|h\|} \cdot \sup_{k \neq \theta_{\mathbb{R}^p}} \frac{\sum_{j=1}^{p} \left| \frac{\partial f}{\partial x_j}(x_0+h) - \frac{\partial f}{\partial x_j}(x_0) - \sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i \right| \cdot \left| k_j \right|}{\|k\|} \leq$$

$$\leq \frac{1}{\|h\|} \sum_{j=1}^{p} \left| \frac{\partial f}{\partial x_j}(x_0+h) - \frac{\partial f}{\partial x_j}(x_0) - \sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i \right| .$$

According to theorem 3.5, the continuity of the second order derivatives is sufficient for the differentiability of the first order derivatives, hence

$$\sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \cdot h_i = d\left(\frac{\partial f}{\partial x_j}\right)_{x_0}(h),$$

and

$$\lim_{\|h\| \to 0} \frac{\left| \frac{\partial f}{\partial x_j}(x_0 + h) - \frac{\partial f}{\partial x_j}(x_0) - d\left(\frac{\partial f}{\partial x_j}\right)_{x_0}(h) \right|}{\|h\|} = 0.$$

Because this is valid for each $j = 1, ..., p$, we may conclude that

$$\frac{\left\| (df)(x_0 + h) - (df)(x_0) - L_{x_0}(h) \right\|}{\|h\|} \leq$$

$$\leq \sum_{j=1}^{p} \frac{\left| \frac{\partial f}{\partial x_j}(x_0 + h) - \frac{\partial f}{\partial x_j}(x_0) - \sum_{i=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) h_i \right|}{\|h\|} \underset{\|h\| \to 0}{\to} 0.$$

Consequently, f is two times differentiable at x_0 and

$$d^2 f_{x_0}(h,k) = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0) \cdot h_i \cdot k_j \ .$$

The last assertion of the theorem follows from the fact that if $f \in \mathbf{C}^2(A)$, then the initial hypotheses hold at each $x_0 \in A$. ◇

3.17. <u>Remarks</u>. 1) The condition of continuity of the second order partial derivatives is generally <u>not necessary</u>, for the second order differentiability. For example, we may consider the function $f: \mathbb{R}^2 \to \mathbb{R}$, defined by

$$f(x,y) = \begin{cases} (x^2 + y^2)^2 \sin \frac{1}{\sqrt{x^2 + y^2}} & if \quad (x,y) \neq (0,0) \\ 0 & if \quad (x,y) = (0,0) \ . \end{cases}$$

This function is two times differentiable at the origin, and $d^2 f_{(0,0)}(h)(k) = 0$ at all $h, k \in \mathbb{R}^2$ (see problem 3 at the end of the section), but its partial derivatives are not continuous at (0, 0).

2) Similarly to the first order differential (see remark IV.2.16.b), we may write the second order differential in a symbolic form too. We obtain it if we take into consideration the projections $P_i : \mathbb{R}^p \to \mathbb{R}$, defined by $P_i(x) = x_i$ for all $i = 1, \ldots, n$. Because of the tradition to note $dP_i = dx_i$, we have

$$(dx_i)_{x_0}(h) = P_i(h_1, \ldots, h_p) = h_i$$

for all $i = 1, 2, \ldots, p$. As a consequence of the above theorem 3.14.c, we represent the second order differential by the formula

$$d^2 f_{x_0}(h, k) = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial y_j}(x_0) h_i k_j =$$

$$= \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial y_j}(x_0) \cdot \left[(dx_i)_{x_0}(h)\right] \cdot \left[(dx_j)_{x_0}(k)\right].$$

Since h and k are arbitrary, a natural temptation of raising this equality to the function level is justified. However, we cannot do it directly, since

$$\left[(dx_i)_{x_0}(h)\right] \cdot \left[(dx_j)_{x_0}(k)\right] \neq \left(\left[(dx_i)_{x_0}\right] \cdot \left[(dx_j)_{x_0}\right]\right)(h,k).$$

To overpass this difficulty, we replace the usual product of functions

$$(\varphi \cdot \psi)(h,k) = [\varphi(h,k)] \cdot [\psi(h,k)]$$

by the so called *tensor product* of functions, which takes the values

$$(\varphi \otimes \psi)(h,k) = [\varphi(h)] \cdot [\psi(k)].$$

Consequently, the second order differential at x_0 takes the form

$$d^2 f_{x_0} = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial y_j}(x_0) \cdot \left[(dx_i)_{x_0}\right] \otimes \left[(dx_j)_{x_0}\right].$$

Again, because x_0 is arbitrary, we may briefly write

$$d^2 f = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i \otimes dx_j .$$

Beside this rigorous way of writing the second order differential, used in [CI], [PM$_1$], etc., there are books where we still find the "simpler" forms

$$d^2 f_{x_0} = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial y_j}(x_0) \cdot \left[(dx_i)_{x_0}\right] \cdot \left[(dx_j)_{x_0}\right],$$

respectively

$$d^2 f = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j} dx_i dx_j .$$

3) Based on multi-linear forms we obtain similar results for higher order differentials. More exactly, if f is a function of class $\mathbf{C}^n$, then its differential of order n takes the symbolic form

$$d^n f = \sum_{i_1, i_2, \ldots, i_n = 1}^{p} \frac{\partial^n f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} dx_{i_1} \otimes dx_{i_2} \otimes \ldots \otimes dx_{i_n} .$$

For the sake of simplicity, we may accept to write

$$d^n f = \sum_{i_1, i_2, \ldots, i_n = 1}^{p} \frac{\partial^n f}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} dx_{i_1} dx_{i_2} \ldots dx_{i_n} ,$$

but the exact meaning of this formula needs a special explanation.

Another important topic of the analysis of a real function of several variables concerns the Taylor's formula, which is of great use in the study of extrema. We will deduce this formula from the similar one, expressed for real functions of a single real variable, previously presented at II.3.24. First, we reformulate this result in the following convenient form:

3.18. Theorem. Let g: $I\to\mathbb{R}$ be a function of class $\mathbf{C}^{n+1}$ on the interval $I\subseteq\mathbb{R}$, and let us fix $x_0\in I$. For every $x\in I$ there exists θ (depending on x) between x_0 and x, such that

$$g(x) = g(x_0) + \frac{1}{1!}dg_{x_0}(x-x_0) + \frac{1}{2!}d^2g_{x_0}(x-x_0,x-x_0) + \ldots +$$
$$+\frac{1}{n!}d^n g_{x_0}(\underbrace{x-x_0,\ldots,x-x_0}_{n\ \text{times}}) + \frac{1}{(n+1)!}d^{n+1}g_\theta(\underbrace{x-x_0,\ldots,x-x_0}_{(n+1)\ \text{times}}).$$

We recall that, for shortness, we frequently use the notation $x-x_0 = h$, and

$$d^n f_{x_0}(\underbrace{h,\ldots,h}_{n\,\text{times}}) = d^n f_{x_0}(h^n) = d^n f_{x_0}(x-x_0)^n.$$

Before formulating the similar result in $\mathbb{R}^p$, we recall that by *interval* of end-points x_1 and x_2 in this space we understand the set

$$[x_1, x_2] = \{x = (1-t)x_1 + t\,x_2 \in \mathbb{R}^p : t\in[0,1]\}.$$

Now we can show that the Taylor's formula keeps the same form for real functions of several real variables:

3.19. Theorem. If $f: A\to\mathbb{R}$ is a function of class $\mathbf{C}^{n+1}$ on the open and *convex* set $A\subseteq\mathbb{R}^p$, and $x_0\in A$ is fixed, then to each $x\in A$ there corresponds some $\xi\in[x_0, x]$ such that

$$f(x) = f(x_0) + \frac{1}{1!}df_{x_0}(x-x_0) + \frac{1}{2!}d^2 f_{x_0}(x-x_0)^2 + \ldots +$$
$$+\frac{1}{n!}d^n f_{x_0}(x-x_0)^n + \frac{1}{(n+1)!}d^{n+1}f_\xi(x-x_0)^{n+1}.$$

As before, this is called Taylor's formula of the function f around x_0, with the *remainder* (i.e. the last term) *in Lagrange's form*.

Proof. Let us introduce the auxiliary function g: $[0, 1]\to\mathbb{R}$, defined by

$$g(t) = f((1-t)\,x_0 + t\,x).$$

This function is well defined because A is convex, hence $[x_0, x]\subseteq A$ holds for every $x\in A$. From the hypothesis it follows that g is $n+1$ times derivable on [0, 1], and according to the above theorem 3.18, the formula

$$g(1) = g(0) + \sum_{j=1}^{n}\frac{1}{j\,!}g^{(j)}(0) + \frac{1}{(n+1)!}g^{(n+1)}(\theta)$$

holds for some $\theta\in(0, 1)$. If we note $x_0 = (x_1^0, \ldots, x_p^0)$, $x = (x_1, \ldots, x_p)$, and we calculate the derivatives as for composed functions, then we obtain

$$g(0) = f(x_0)$$

$$g^{/}(0) = \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}(x_0)(x_j - x_j^0) = df_{x_0}(x - x_0)$$

$$g^{//}(0) = \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)(x_i - x_i^0)(x_j - x_j^0) = d^2 f_{x_0}(x - x_0)^2$$

....

$$g^{(n)}(0) = \sum_{i_1,..,i_n=1}^{p} \frac{\partial^n f}{\partial x_{i_1} ... \partial x_{i_n}}(x_0)(x_{i_1} - x_{i_1}^0)...(x_{i_n} - x_{i_n}^0) = d^n f_{x_0}(x - x_0)^n .$$

Similarly, for $\xi = (1 - \theta)\, x_0 + \theta\, x \in [x_0 , x]$, it follows

$$g^{(n+1)}(\theta) = d^{n+1} f_{\xi}(x - x_0)^{n+1} .$$

To accomplish the proof it is sufficient to see that $g(1) = f(x)$. ◇

In order for us to study the extreme values of a real function of several variables we shall precise some terms:

3.20. Definition. Let $f : A \to \mathbb{R}$ be an arbitrary function defined on the open set $A \subseteq \mathbb{R}^p$. The point $x_0 \in A$ is called a *local maximum* of f iff there exists a neighborhood V of x_0, $V \subseteq A$, such that for all $x \in V$ we have

$$f(x) - f(x_0) \le 0 .$$

Dually, the *local minimum* is defined by the converse inequality

$$f(x) - f(x_0) \ge 0 .$$

The local maximum and the local minimum points of f are called *extremum* (extreme) *points* of f.

If all the partial derivatives of f are void at x_0, i.e.

$$\frac{\partial f}{\partial x_j}(x_0) = 0 \ , j = 1, \ldots, p,$$

then we say that x_0 is a *stationary* point of f.

Now, we can formulate necessary conditions for extremes:

3.21. Theorem. (Fermat) Let us consider that $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^p$ is an open set, has all the partial derivatives at $x_0 \in A$. If x_0 is an extreme of f then x_0 is a stationary point of f.

Proof. Let $r > 0$ be such that $S(x_0, r) \subseteq A$ and for all $x \in S(x_0 , r)$ the difference $f(x) - f(x_0)$ has a constant sign. Let $\{e_1, \ldots, e_p\}$ denote the canonical base of the linear space $\mathbb{R}^p$. For each $j = 1, \ldots, p$ we define the function $g_j : (-r, +r) \to \mathbb{R}$ by

$$g_j(t) = f(x_0 + t\, e_j) .$$

Obviously, $t_0 = 0$ is an extreme point of g_j for each $j = 1, \ldots, p$. In addition, g_j is derivable at $t_0 = 0$, and its derivative equals

$$\frac{dg_j}{dt}(0) = \frac{\partial f}{\partial x_j}(x_0) .$$

Applying the Fermat's theorem, for each g_j, where $j = 1, \ldots, p$, we obtain $g'_j(0) = 0$, hence x_0 is a stationary point of f. ◇

3.22. Remarks. a) In order to find the local extreme points of a derivable function we primarily determine the stationary points by solving the system:

$$\begin{cases} \dfrac{\partial f}{\partial x_1} = 0 \\ \vdots \\ \dfrac{\partial f}{\partial x_p} = 0 \end{cases} .$$

However, we shall carefully continue the investigation because not all stationary points are extremes. For example, the function $f : \mathbb{R}^2 \to \mathbb{R}$, of values $f(x, y) = x^3 - y^3$, is stationary at the origin (0, 0), i.e.

$$\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 0 ,$$

but the difference $f(x, y) - f(0, 0)$ changes its sign on any neighborhood of this point. In other words, we need sufficient conditions to establish which stationary point is extreme and which is not.

b) The sufficient conditions for extreme points will be based on the study of the second order differential. In fact, according to the Fermat's theorem, since $df_{x_0} = 0$, the increment of f takes the form

$$f(x) - f(x_0) = d^2 f_\xi(x - x_0, x - x_0) .$$

More than this, the second order differential is calculated in a particular case $h = x - x_0 = k$, when it reduces to a *quadratic form* (see § II.4). We recall that, generally speaking, $\varphi : \mathbb{R}^p \to \mathbb{R}$ is named *quadratic form* iff there exists a bilinear symmetric (and continuous) function $\Phi : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, such that the equality

$$\varphi(x) = \Phi(x, x)$$

holds at each $x \in \mathbb{R}^p$.

3.23. Theorem. Let $f : A \to \mathbb{R}$ be a function of class $\mathbf{C}^2_{\mathbb{R}}(A)$, where $A \subseteq \mathbb{R}^p$ is an open and connected set, and let $x_0 \in A$ be a stationary point of f. If the quadratic form $\varphi : \mathbb{R}^p \to \mathbb{R}$, defined by

$$\varphi(h) = d^2 f_{x_0}(h, h)$$

is positively (negatively) defined, then f has a local minimum (respectively maximum) at the point x_0 .

Proof. Let us consider that φ is positively defined. Then, there exists $m>0$ such that $\varphi(x - x_0) \geq m \| x - x_0 \|^2$ for all $x \in \mathbb{R}^p$. On the other hand, from the Taylor's formula for $n = 1$, at the stationary point x_0, we deduce

$$f(x) - f(x_0) =$$

$$= \frac{1}{2} d^2 f_\xi (x - x_0, x - x_0) = \frac{1}{2} \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j} (\xi)(x_i - x_i^0)(x_j - x_j^0) =$$

$$= \frac{1}{2} d^2 f_{x_0} (x - x_0, x - x_0) +$$

$$+ \frac{1}{2} \sum_{i,j=1}^{p} \left[\frac{\partial^2 f}{\partial x_i \partial x_j} (\xi) - \frac{\partial^2 f}{\partial x_i \partial x_j} (x_0) \right] (x_i - x_i^0)(x_j - x_j^0).$$

If we note the last sum by $\alpha(x)$, then the above equality becomes

$$f(x) - f(x_0) = \frac{1}{2} \varphi(x - x_0) + \alpha(x) .$$

The continuity of the second order derivatives of f at x_0 leads to

$$\lim_{x \to x_0} \frac{\alpha(x)}{\| x - x_0 \|^2} = 0 .$$

Now, let $r > 0$ be chosen such that at each $x \in S(x_0, r)$ we have

$$\frac{\alpha(x)}{\| x - x_0 \|^2} + \frac{m}{2} > 0 .$$

Consequently, at each x in this neighborhood, the inequality

$$f(x) - f(x_0) \geq \left[\frac{m}{2} + \frac{\alpha(x)}{\| x - x_0 \|^2} \right] \cdot \| x - x_0 \|^2 \geq 0 ,$$

i.e. x_0 is a minimum point of f.

Similarly, we discuss the case of a maximum. ◇

3.24. Corollary. In the case $p = 2$, let $f : A \to \mathbb{R}$ be of class $\mathbf{C}_{\mathbb{R}}^2(A)$, and let (x_0, y_0) be a stationary point of f. If we note

$$a = \frac{\partial^2 f}{\partial x^2}(x_0, y_0) \text{ and } \delta = \frac{\partial^2 f}{\partial x^2}(x_0, y_0) \frac{\partial^2 f}{\partial y^2}(x_0, y_0) - \left[\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) \right]^2 ,$$

then, according to the above tests of positivity (negativity), we distinguish the following possibilities:

1. if $a > 0$ and $\delta > 0$, then f has a minimum at x_0
2. if $a < 0$ and $\delta > 0$, then f has a maximum at x_0
3. if $\delta < 0$, then x_0 is no extreme point of f.

3.25. Remark. There are situations when we are not able to establish the nature of a stationary point by using the results from above. For functions of two variables, we mention the following such cases:

a) $\delta = 0$ in the previous corollary. We must directly study the sign of the difference $f(x) - f(x_0)$.

b) $d^2 f_{x_0} = \theta_{\mathscr{B}(\mathbb{R}^p, \mathbb{R})}$. The investigation of the higher order differentials is necessary to obtain information about $f(x) - f(x_0)$.

c) $f : K \to \mathbb{R}$, where $K \subseteq \mathbb{R}^p$ is a compact set. If f is continuous, then f is bounded and the extreme values are effectively reached at some points, but it is possible these points to be in $K \setminus \overset{\circ}{K} = \operatorname{Fr} K$.

Practically, in each case may occur more situations, as in the following:

3.26. <u>Example</u>. Let us study the extreme values of the function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$f(x, y) = x^3 y^2 (6 - x - y) .$$

We obtain the stationary points by solving the system

$$\begin{cases} \dfrac{\partial f}{\partial x}(x, y) \equiv 18x^2y^2 - 4x^3y^2 - 3x^2y^3 = 0 \\ \dfrac{\partial f}{\partial y}(x, y) \equiv 12x^3y - 2x^4y - 3x^3y^2 = 0 \end{cases} .$$

The stationary points have one of the forms:

(a) $(x_1, y_1) = (3, 2)$; (b) $(x_0, 0)$, $(\forall)\ x_0 \in \mathbb{R}$; or (c) $(0, y_0)$, $(\forall)\ y_0 \in \mathbb{R}$.

By evaluating the second order derivatives in these cases, we obtain the following Hessean matrices:

(a) $\mathbf{H} f(3, 2) = \begin{pmatrix} -144 & -108 \\ -108 & -162 \end{pmatrix}$

(b) $\mathbf{H} f(x_0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 2x_0^3(6 - x_0) \end{pmatrix}$

(c) $\mathbf{H} f(0, y_0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

In case (a), following the Sylvester's test, we evaluate $a = -144 < 0$, and $\delta = 2^4 \cdot 3^6 > 0$, hence $(x_1, y_1) = (3, 2)$ is a *local maximum*.

The case (b) contains the following sub-cases:

(b_1) $x_0 = 0$,
(b_2) $x_0 = 6$,
(b_3) $x_0 \in (0, 6)$, and
(b_4) $x_0 \in (-\infty, 0) \cup (6, \infty)$.

In the sub-case (b_1) the above techniques are useless because all partial derivatives are null at (0, 0) up to the 5th order. In this situation, we have to find other ways to study the sign of the difference $f(x, y) - f(0, 0)$. Because in "small" neighborhoods of the origin we have $6 - x - y > 0$, we deduce

$$\operatorname{sign}\, [f(x, y) - f(0, 0)] = \operatorname{sign}\, [x^3 y^2 (6 - x - y)] = \operatorname{sign} x .$$

This shows that $f(x, y) - f(0, 0)$ changes its sign in any neighborhood of (0, 0), since x does. Consequently, (0, 0) is not an extreme point of f.

We may analyze the sub-case (b_2) by using the third differential, which is

$$d^3 f_{(6,0)}(x-6, y)^3 = -36^2(x-6)\,[y^2 + (x-6)^2]\,.$$

Obviously, $d^3 f_{(6,0)}$ changes its sign in any neighborhood of (6, 0). Because

$$\operatorname{sign}[f(x, y) - f(6, 0)] = \operatorname{sign} d^3 f_{(6,0)}(x-6, y)^3\,,$$

it follows that (6, 0) is not an extreme point of f.

We may reduce the analysis of the sub-cases (b_3) and (b_4) to the study of the second order differential, since

$$f(x, y) - f(x_0, 0) = \frac{1}{2} d^2 f_{(\xi,\eta)}(x - x_0)^2 = \xi^3 (6 - \xi) \cdot y^2\,,$$

where ξ is laying between x_0 and x. Consequently, if $x_0 \in (0, 6)$, then there exists a neighborhood of $(x_0, 0)$ where we have $f(x, y) - f(x_0, 0) \geq 0$, with equality at that point only, hence each point of the form $(x_0, 0)$ is a local minimum of f. Similarly, if $x_0 \in (-\infty, 0) \cup (6, \infty)$, then each point $(x_0, 0)$ is a local maximum of f.

In case (c), the increment of f takes the form:

$$f(x, y) - f(0, y_0) = \frac{1}{3!} d^3 f_{(0,y_0)}(x, y - y_0)^3 + \ldots$$

Because $d^3 f_{(0,y_0)}(x, y - y_0)^3 = 6y_0^2(6 - y_0) \cdot x^3$, we have to distinguish the following sub-cases:

(c_1) $y_0 = 0$,

(c_2) $y_0 = 6$, and

(c_3) $y_0 \in \mathbb{R} \setminus \{0, 6\}$.

Sub-case (c_1) coincides with (b_1). The sub-cases (c_2) and (b_2) are similar. In fact, from the equalities

$$\operatorname{sign}[f(x, y) - f(0, 6)] = \operatorname{sign}[x^3 y^2 (6 - x - y)] = \operatorname{sign}[x(6 - x - y)],$$

it follows that $f(x, y) - f(0, 6)$ changes its sign on any neighborhood of the point (0, 6). Consequently, (0, 6) is not an extreme point of f.

In the sub-case (c_3), we have $d^3 f_{(0,y_0)} \neq 0$, but because of x^3, it does not keep a constant sign on a neighborhood of $(0, y_0)$. So we conclude that these points are not extremes of f.

PROBLEMS § IV.3.

1. Calculate the Jacobians of the functions f, $g : A \to \mathbb{R}^3$, where A is an open set in $[0, \infty) \times \mathbb{R}^2$, and the values of these functions are

$$f(\rho, \theta, \varphi) = (\rho \sin\theta \cos\varphi, \rho \sin\theta \sin\varphi, \rho \cos\theta)$$
$$g(\rho, \varphi, z) = (\rho \cos\varphi, \rho \sin\varphi, z).$$

Hint. Write the Jacobi's matrixes, then evaluate the determinants (as in definition 2.17). $\text{Det}(\mathbf{J}f(\rho, \theta, \varphi)) = \rho^2 \sin\theta$; $\text{Det}(\mathbf{J}g(\rho, \varphi, z)) = \rho$.

2. Show that the partial derivatives of the function $f: \mathbb{R}^2 \to \mathbb{R}$, of values

$$f(x, y) = \begin{cases} (x^2 + y^2)\sin(x^2 + y^2)^{-1/2} & \text{if } (x, y) \neq (0,0) \\ 0 & \text{if } (x, y) = (0,0) \end{cases}$$

are discontinuous at (0, 0), but f is differentiable at this point.

Hint. At (0, 0) we have

$$\frac{\partial f}{\partial x}(0,0) = \lim_{x\to 0} \frac{x^2 \sin\frac{1}{|x|} - 0}{x} = 0,$$

and otherwise

$$\frac{\partial f}{\partial x}(x,y) = 2x\sin\frac{1}{\sqrt{x^2+y^2}} - \frac{2x}{\sqrt{x^2+y^2}}\cos\frac{1}{\sqrt{x^2+y^2}} .$$

To see the discontinuity of this derivative at (0, 0), consider particular sequences, or analyze $\lim\limits_{x\to 0}\frac{\partial f}{\partial x}(x,0) = \lim\limits_{x\to 0}\left(2x\sin\frac{1}{|x|} - \frac{2x}{|x|}\cos\frac{1}{|x|}\right)$, etc .

Similar results concern $\frac{\partial f}{\partial y}$. However, f is differentiable at (0, 0), and $df_{(0,0)} = \theta_{\mathcal{B}(\mathbb{R}^2,\mathbb{R})}$, since $\lim\limits_{\|h\|\to 0} \frac{|f(h_1,h_2) - f(0,0) - 0|}{\|h\|} = 0$.

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

$$f(x, y) = \begin{cases} (x^2 + y^2)^2\sin(x^2 + y^2)^{-1/2} & \text{if } (x, y) \neq (0,0) \\ 0 & \text{if } (x, y) = (0,0) . \end{cases}$$

Show that:

a) This function is two times differentiable at (0,0), but its second order partial derivatives are not continuous at this point, and

b) The equality $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ holds on the entire $\mathbb{R}^2$ (without using the Schwarz' theorem 3.10).

Hint. If we note $\rho = \sqrt{x^2 + y^2}$, then the first order partial derivatives are

$$\frac{\partial f}{\partial x}(x,y) = \begin{cases} 4x\rho^2 \sin\frac{1}{\rho} - x\rho\cos\frac{1}{\rho} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases}$$

$$\frac{\partial f}{\partial y}(x,y) = \begin{cases} 4y\rho^2 \sin\frac{1}{\rho} - y\rho\cos\frac{1}{\rho} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{if } (x,y) = (0,0) \end{cases} .$$

These derivatives are everywhere continuous, hence f is differentiable. In particular, $df_{(0,0)} = \theta_{\mathscr{B}(\mathbb{R}^2,\mathbb{R})}$. The second order partial derivatives are not continuous at (0,0), but

$$\frac{\left\| df_h - df_0 - \theta_{\mathscr{L}(\mathbb{R}^2,\mathbb{R})} \right\|}{\|h\|} = \frac{\|df_h\|}{\|h\|} = \frac{1}{\|h\|} \cdot \sup_{k \neq \theta_{\mathbb{R}^2}} \frac{|df_h(k)|}{\|k\|} =$$

$$= \frac{1}{\|h\|} \cdot \sup_{k \neq \theta_{\mathbb{R}^2}} \frac{\left| \frac{\partial f}{\partial x}(h_1,h_2) \cdot k_1 + \frac{\partial f}{\partial y}(h_1,h_2) \cdot k_2 \right|}{\|k\|} \leq$$

$$\leq \frac{1}{\|h\|} \cdot \left[\left| \frac{\partial f}{\partial x}(h_1,h_2) \right| + \left| \frac{\partial f}{\partial y}(h_1,h_2) \right| \right] =$$

$$= \frac{1}{\|h\|} \cdot \left[\left| 4h_1\|h\|^2 \sin\frac{1}{\|h\|} - h_1\|h\|\cos\frac{1}{\|h\|} \right| + \left| 4h_2\|h\|^2 \sin\frac{1}{\|h\|} - h_2\|h\|\cos\frac{1}{\|h\|} \right| \right] =$$

$$= (|h_1| + |h_2| \cdot \left| 4\|h\|\sin\frac{1}{\|h\|} - \cos\frac{1}{\|h\|} \right| \underset{\|h\| \to 0}{\longrightarrow} 0 .$$

Consequently, f is two times differentiable at (0, 0), and at each $h \in \mathbb{R}^2$ we have $d^2 f_{(0,0)}(h) = \theta_{\mathscr{B}(\mathbb{R}^2,\mathbb{R})}$, i.e. $d^2 f_{(0,0)}(h)(k) = 0$ at all $h, k \in \mathbb{R}^2$.

4. Using the fact that $grad\, f(x_0) \parallel n$, where n is the normal to the level surface $S_{f(x_0)}$, at x_0 , determine the components of n for a surface of *explicit* equation $z = \varphi(x, y)$ in $\mathbb{R}^3$.

Hint. Consider $f(x, y, z) = \varphi(x, y) - z$, so that

$$grad\, f(x_0, y_0, z_0) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)_{(x_0,y_0,z_0)} = \left(\frac{\partial \varphi}{\partial x}(x_0,y_0), \frac{\partial \varphi}{\partial y}(x_0,y_0), -1 \right).$$

Using the Monge's notation $p = \frac{\partial \varphi}{\partial x}$, $q = \frac{\partial \varphi}{\partial y}$ it follows that $n \parallel (p, q, -1)$.

5. Calculate the derivative of the function $f:\mathbb{R}^3\to\mathbb{R}$, of values

$$f(x, y, z) = x^2 + 2y^2 + 3z^2$$

at the point (2, 1, 1), in the direction of the unit vector $\vec{e} = \dfrac{\vec{i}-\vec{k}}{\sqrt{2}}$.

<u>Hint</u>. $\dfrac{\partial f}{\partial e}(x_0, y_0, z_0) = df_{(x_0,y_0,z_0)}(\vec{e}) = <(grad\ f)(x_0, y_0, z_0), \vec{e}>$.

6. In what directions $\vec{e}$, function $f(x, y) = \sqrt[4]{xy^2}$ has derivatives at (0, 0)?
<u>Hint</u>. Function f is not differentiable at (0,0), so we have to apply the very definition of the derivative in a direction. If $\vec{e} = \vec{i}\cos\alpha + \vec{j}\sin\alpha$, then

$$\frac{\partial f}{\partial e}(0,0) = \lim_{t\to 0}\frac{f(t\cos\alpha, t\sin\alpha)}{t} = \lim_{t\to 0}\frac{\sqrt[4]{t^3\cos\alpha\sin^2\alpha}}{t}.$$

This limit exists only for $\alpha = k\dfrac{\pi}{2}$, k = 0, 1, 2, 3.

7. Find the unit vector $\vec{e}$, which is tangent to the plane curve of implicit equation $x^2 + y^2 - 2x = 0$, at the point $\left(\dfrac{1}{2}, \dfrac{\sqrt{3}}{2}\right)$. Evaluate the derivative of the function $f(x, y) = arctg\ \dfrac{y}{x}$ at that point, in the direction of $\vec{e}$.
<u>Hint</u>. Use geometric interpretations, or derive in the explicit equation of the curve, $y = +\sqrt{2x - x^2}$, to find $\vec{e}$. Differentiate f at that point.

8. Let us consider the function $f(x, y) = x^3 + xy^2$ and the point a = (1, 2). Calculate the partial derivatives of first and second order, and write the first and second order differentials of f at a. Find the derivative of f at a in the direction $\vec{e} = (\cos\alpha, \sin\alpha)$. In what sense could we speak of a second order derivative of f at a in the fixed direction $\vec{e}$?
<u>Hint</u>. The differentials are $df_{(1,2)}(h,k) = \dfrac{\partial f}{\partial x}(1,2)\cdot h + \dfrac{\partial f}{\partial y}(1,2)\cdot k = 7h + 4k$, and $d^2 f_{(1,2)}(h,k) = \dfrac{\partial^2 f}{\partial x^2}(1,2)\cdot h^2 + 2\dfrac{\partial^2 f}{\partial x\partial y}(1,2)\cdot hk + \dfrac{\partial^2 f}{\partial y^2}(1,2)\cdot k^2$, hence $d^2 f_{(1,2)}(h,k) = 6h^2 + 8hk + 2k^2$. If we fix $\vec{e}$, then $\dfrac{\partial f}{\partial \vec{e}}(x, y)$ is defined on a neighborhood of a, where it is differentiable. Its derivative in the direction e plays the role of the second order derivative of f in this direction.

9. Show that the functions of values $u(x, y) = arctg\ \frac{y}{x}$ and $v(x, y) = ln\ \frac{1}{r}$, where $r = \sqrt{(x-a)^2 + (y-b)^2}$, are harmonic where defined.

Hint. u and v satisfy the Laplace's equation $\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0$.

10. We note $u(t, x) = A\ sin(c\lambda t + \mu)\ sin\ \lambda x$ and $v = \varphi(x - ct) + \psi(x + ct)$, where A, λ, μ, c are constants, and φ, ψ are arbitrary functions of class $\mathbf{C}^2$. Show that u and v satisfy the D'Alembert equation of the oscillating string.

Hint. The D'Alembert's equation is $\frac{1}{c^2}\frac{\partial^2 f}{\partial t^2} - \frac{\partial^2 f}{\partial x^2} = 0$.

11. Write the Taylor polynomial of the n^th^ degree for $f(x, y) = e^{x+y}$ at the point $(x_0, y_0) = (1, -1)$. What happens when $n \to \infty$? What should mean

$$d^n z = \left(dx\frac{\partial}{\partial x} + dy\frac{\partial}{\partial y} \right)^n z\ ?$$

Hint. Note $x + y = t$ and observe that the Taylor's series of e^t is absolutely and almost uniformly convergent to this function. The symbolic formula describes the higher order differential as a formally expanded binomial.

12. Let us note $r = \sqrt{x^2 + y^2}$ and let n be the normal to the circle of center (0, 0) and radius r. Show that $\frac{\partial\left(\frac{1}{r}\right)}{\partial n} = \frac{d\left(\frac{1}{r}\right)}{dr} = -\frac{1}{r^2}$, and sketch the vector field $grad\ \frac{1}{r}$ on $\mathbb{R}^2 \setminus \{0\}$.

Hint. The vectors n, r, and $grad\ \frac{1}{r}$ are collinear.

13. Using the Hessean matrix, calculate the second order differential of the function $f(x, y) = e^{xy}$ at a current point $(x_0, y_0) \in \mathbb{R}^2$.

14. Calculate the second order derivatives of $u(x, y) = f(x^2 + y^2, x^2 - y^2, xy)$, where f is a function of class $\mathbf{C}^2$ on $\mathbb{R}^2$. Express $d^2 f$ as a differential of df in the case of $f(u, v, w) = u - 2v^2 + 3vw$.
Hint. $d^2x=0$ only if x is an independent variable!

15. Using Taylor's formulas up to the second order terms, approximate $\sqrt[3]{0.98}$, $(0.95)^{2.01}$, $\cos 1^\circ$, $e^{0.1} \sin 1^\circ$. Give geometrical interpretation to the results, as quadratic approximations.
Hint. Use Taylor's formulas for one and two variables.

16. Test the following function for an extremum: $z = x^3 + 3xy^2 - 15x - 12y$.
Solution. (2, 1) is a local minimum, and (–2, –1) is a local maximum, but at the stationary points (1, 2) and (–1, –2) there is no extremum.

17. Break up a positive number a into three nonnegative numbers so that their product be the greatest possible.
Hint. If we denote the three numbers by x, y and $a - x - y$, then we are led to find the maximum of the function $f(x, y) = xy(a - x - y)$ in the triangle $x \geq 0$, $y \geq 0$ and $x + y \leq a$. The unique stationary point is $(a/3, a/3)$. The second order differential shows that it is indeed a maximum point.

18. Test the following functions for points of maximum and minimum:
(a) $f(x, y) = x^2 + xy + y^2 - 2x - y$
(b) $g(x, y) = (x^2 + y^2)\, e^{-(x^2+y^2)}$
(c) $h(x, y) = \dfrac{1 + x - y}{\sqrt{1 + x^2 + y^2}}$.

Solutions. (a) f takes the minimum value –1 at the point (1,0).
(b) $g_{min} = 0$ at (0, 0) and $g_{max} = \dfrac{1}{e}$ at the points of the circle $x^2 + y^2 = 1$;
(c) $h_{max} = \sqrt{3}$ at (1, –1).

19. Show that the function $f(x, y) = (1 + e^x) \cos y - x e^x$ has infinite many points of maximum but no minimum.
Hint. Find the stationary points, and study the higher order differentials at these points.

20. Find the increment of the function $f(x, y) = x^3 - 2y^3 + 3xy$ when passing from (1, 2) to $(1 + h, 2 + k)$. Determine all functions of class $\mathbf{C}^\infty$ on $\mathbb{R}^2$ for which this increment is a polynomial in h and k.
Hint. We have $d^4 f_{(x_0, y_0)} = 0$. From $d^n f = 0$ on A, we deduce $d^{n-1} f = const.$, hence f must be a polynomial in x, y. Generally speaking, the polynomial functions are characterized by null differentials of higher orders.

§ IV.4. IMPLICIT FUNCTIONS

Until now we have studied only explicit functions, for which it is explicitly indicated what operations on the variables lead to the value of the function. The typical notation was $y = f(x_1, \dots, x_n)$. However, we may express the dependence of y on the variables x_1 up to x_n by a condition of the form $F(x_1, \dots, x_n; y) = 0$, where making $y = f(x_1, \dots, x_n)$ explicit is either impossible, or non-convenient (e.g. non-unique, too complicated, useless, etc.). The aim of this section is to clarify how to obtain the explicit functions (Theorems 4.3 and 4.6 below), and to deduce several theoretical and practical consequences of this result, concerning the local inversion, smooth transformations, and conditional extrema.

4.1. Example. The equation of the unit circle in the plane, $x^2 + y^2 = 1$, establishes a dependence of y on x, but this curve cannot be the graph of a function $y = f(x)$, because to each $x \in (-1, +1)$ there correspond two values of y. However, excepting the points $(-1, 0)$ and $(1, 0)$, for each (x_0, y_0) belonging to the circle there exists a neighborhood $V \in \mathscr{V}(x_0, y_0)$, such that the arc of the circle, which is contained in V, actually is the graph of some explicit function, namely $y = \sqrt{1 - x^2}$, or $y = -\sqrt{1 - x^2}$.

Our purpose is to generalize this case, but primarily we have to precise some notions we deal with.

4.2. Definition. Let $D \subseteq \mathbb{R}^2$ be an open set, and let $F : D \to \mathbb{R}$ be a function. We note by $D_x = P_x(D)$ the x-projection of D and we choose $A \subseteq D_x$. Each function $f : A \to \mathbb{R}$, which verifies the equation $F(x, y) = 0$, when we replace $y = f(x)$, i.e. $F(x, f(x)) \equiv 0$ on A, is called *solution* of this equation. If this solution is unique, we say that f is *implicitly defined* by the equation $F(x, y) = 0$, or, in short, f is an *implicit function* (of one variable).

4.3. Theorem. Let us consider an open set $D \subseteq \mathbb{R}^2$, a point $(x_0, y_0) \in D$, and a function $F : D \to \mathbb{R}$. If the following conditions hold

1) $F(x_0, y_0) = 0$,
2) F is of class $\mathbf{C}^1$ on a neighborhood W of (x_0, y_0), and
3) $\dfrac{\partial F}{\partial y}(x_0, y_0) \neq 0$, then

(a) There exist $U \in \mathscr{V}(x_0)$, $V \in \mathscr{V}(y_0)$, and $f : U \to V$, which is the unique solution of the equation $F(x, y) = 0$, such that $f(x_0) = y_0$;

(b) $f \in C_R^1(U)$, and for every $x \in U$ we have

$$f'(x) = -\left[\frac{\partial F}{\partial x}(x, f(x))\right] \Big/ \left[\frac{\partial F}{\partial y}(x, f(x))\right] \qquad (*)$$

(c) If $F \in \mathbf{C}^k_{\mathbb{R}}(W)$, then $f \in \mathbf{C}^k_{\mathbb{R}}(U)$, for all $k \in \mathbb{N}^*$.

Proof. (a) To make a choice, let us suppose that $\frac{\partial F}{\partial y}(x_0, y_0) > 0$. Because $\frac{\partial F}{\partial y}$ is continuous on W, there exist $a > 0$ and $b > 0$ such that $\frac{\partial F}{\partial y}(x, y) > 0$ for all (x, y) which satisfy the inequalities $|x - x_0| < a$ and $|y - y_0| < b$. Consequently, the function $y \mapsto F(x_0, y)$, defined on $(y_0 - b, y_0 + b)$, is strictly increasing. In particular, because the inequalities

$$y_0 - b < y_0 - \varepsilon < y_0 < y_0 + \varepsilon < y_0 + b$$

hold for any $\varepsilon \in (0, b)$, it follows that

$$F(x_0, y_0 - \varepsilon) < F(x_0, y_0) = 0 < F(x_0, y_0 + \varepsilon).$$

From the second hypothesis we deduce that the functions

$$x \mapsto F(x, y_0 - \varepsilon)$$
$$x \mapsto F(x, y_0 + \varepsilon)$$

are continuous on $(x_0 - a, x_0 + a)$, hence there exists $\delta \in (0, a)$ such that

$$F(x, y_0 - \varepsilon) < 0 < F(x, y_0 + \varepsilon)$$

holds for all $x \in (x_0 - \delta, x_0 + \delta)$. Using the third hypothesis in the sense that $\frac{\partial F}{\partial y}(x, y) > 0$ holds whenever (x, y) satisfy $|x - x_0| < a$ and $|y - y_0| < b$, it follows that the function $y \mapsto F(x, y)$ is strictly increasing on $[y_0 - \varepsilon, y_0 + \varepsilon]$ for each fixed $x \in (x_0 - \delta, x_0 + \delta)$. Being continuous on this interval, it has the Darboux property, hence there exists a *unique* $y \in (y_0 - \varepsilon, y_0 + \varepsilon)$ such that $F(x, y) = 0$. In brief, we have constructed the function

$$f : U = (x_0 - \delta, x_0 + \delta) \to V = (y_0 - \varepsilon, y_0 + \varepsilon),$$

such that $F(x, y = f(x)) = 0$ holds at each $x \in V$, i.e. $F(x, f(x)) \equiv 0$ is valid on the set U. Using the uniqueness of the function $y = f(x)$, and the fact that $F(x_0, y_0) = 0$, it follows that $f(x_0) = y_0$.

(b) Primarily we show that f is continuous on U. In fact, in the above construction, δ depends on ε, and $f(x) \in V$ means $|f(x) - y_0| < \varepsilon$. If we repeat this construction for another $\varepsilon' > 0$, then we find $\delta' > 0$ and function $f_1 : (x_0 - \delta', x_0 + \delta') \to (y_0 - \varepsilon', y_0 + \varepsilon')$, such that $|f_1(x) - y_0| < \varepsilon'$ holds whenever $|x - x_0| < \delta'$. The uniqueness of f, and the equality of f and f_1 at x_0, i.e. $f(x_0) = f_1(x_0) = y_0$, lead to $f(x) = f_1(x)$ at all $x \in (x_0 - \delta', x_0 + \delta')$.

Now, let us analyze the derivability of f at an arbitrary point $x^* \in U$, where $f(x^*) = y^* \in V$. If we write the Taylor formula at (x^*, y^*) for $n = 0$, then

$$F(x, y) - F(x^*, y^*) = \frac{\partial F}{\partial x}(\xi, \eta)(x - x^*) + \frac{\partial F}{\partial y}(\xi, \eta)(y - y^*)$$

where (ξ, η) is lying between (x^*, y^*) and (x, y). In particular, if we replace y by $f(x)$, then at each $x \in U$, $x \neq x^*$, we obtain

$$\frac{\partial F}{\partial x}(\xi, \eta)(x - x^*) + \frac{\partial F}{\partial y}(\xi, \eta)\left[f(x) - f(x^*)\right] = 0.$$

Equivalently, this relation takes the form

$$\frac{f(x)-f(x^*)}{x-x^*}=-\frac{\frac{\partial F}{\partial x}(\xi,\eta)}{\frac{\partial F}{\partial y}(\xi,\eta)}.$$

Because f is continuous at x^*, if x tends to x^* it follows that $f(x)\to f(x^*)$, $\xi\to x^*$, and $\eta\to y^*=f(x^*)$. Consequently, the derivability of f follows from the continuity of $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ at (x^*, y^*), and the value of this derivative is

$$f^{/}(x^*)=-\frac{\frac{\partial F}{\partial x}(x^*,f(x^*))}{\frac{\partial F}{\partial y}(x^*,f(x^*))}.$$

In particular, this shows that $f^{/}$ is continuous at all $x^*\in U$.
(c) Since $k\in\mathbb{N}^*$, we will use mathematical induction. Case $k = 1$ is just (b) from above. Let us suppose that the property is valid for $k = n$. In order to prove it for $k = n + 1$, it is sufficient to remark that from $F\in\mathbf{C}^{n+1}_{\mathbb{R}}(W)$, it follows $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}\in\mathbf{C}^{n}_{\mathbb{R}}(W)$. According to (*), we have $f^{/}\in\mathbf{C}^{n}_{\mathbb{R}}(U)$, and finally $f\in\mathbf{C}^{n+1}_{\mathbb{R}}(U)$. ◇

4.4. <u>Remarks</u>. (a) The above theorem remains valid if instead of $x\in\mathbb{R}$ we take $x=(x_1,\ldots,x_p)\in\mathbb{R}^p$ for some $p>1$, and the proof is similar. In this case function f depends on p real variables, and for all $j = 1, \ldots, p$ we have

$$\frac{\partial f}{\partial x_j}(x)=-\frac{\frac{\partial F}{\partial x_j}(x,f(x))}{\frac{\partial F}{\partial y}(x,f(x))}. \qquad (**)$$

(b) Another extension of theorem 4.3 refers to the number of conditions. For example, the system

$$\begin{cases}F(x,y,z)=0\\ G(x,y,z)=0\end{cases}$$

defines a vector implicit function of components

$$\begin{cases}y=f(x)\\ z=g(x)\end{cases}$$

In fact, if $z = h(x, y)$ is an explicit function defined by $F = 0$, i.e. this equation becomes $F(x, y, h(x, y)) \equiv 0$, then the second equation, which is $G(x, y, h(x, y)) \equiv 0$, yields $y = f(x)$. Finally, $g(x) = h(x, f(x))$.

(c) Similarly to (a) and (b) from above, the system

$$\begin{cases} F(x,y;u,v)=0 \\ G(x,y;u,v)=0 \end{cases}$$

implicitly defines the functions

$$\begin{cases} u = f(x,y) \\ v = g(x,y) \end{cases}$$

which can be considered a *vector* implicit function. To be more specific, we introduce the following:

4.5. <u>Definition</u>. Let us consider a system of equations

$$\begin{cases} F_1(x_1,...,x_p;y_1,...,y_m)=0 \\ \quad \dots \\ F_m(x_1,...,x_p;y_1,...,y_m)=0 \end{cases}$$

where the functions $F_i : D\to\mathbb{R}$ are defined on the same open set $D \subseteq \mathbb{R}^{p+m}$ for all $i = 1, .., m$. Let also $A\subseteq\mathbb{R}^p$ be a set consisting of those $x = (x_1 ,\dots, x_p)$, for which there exists $y = (y_1 ,\dots, y_m)\in\mathbb{R}^m$ such that $(x, y)\in D$. A set of functions $\{f_k : A\to\mathbb{R};\ k = 1, \dots, m\}$ is called *solution* of the given system of equations on A iff for all $x\in A$ and $i = 1, \dots, n$, we have

$$F_i (x_1, \dots, x_p ; f_1(x), \dots, f_m(x)) = 0 .$$

If the set of solutions is unique, we say $f_1 , \dots, f_m$ are *implicit functions* defined by the given system relative to the variables $(y_1 , ..., y_m)$.

For brevity, instead of several functions $F_1 ,\dots, F_m$ we may speak of a single vector function F, of components $F_1 ,\dots, F_m$. Similarly, the functions $f_1, ..., f_m$ define a vector function f. Using these notations, we may extend the implicit function theorem 4.3 to vector functions, namely:

4.6. <u>Theorem</u>. If at $(x_0 , y_0) \in D$ (in the above terminology) we have:

1) $F(x_0 , y_0) = 0$ (i.e. $F_i (x_0 , y_0) = 0$ for all $i = 1, \dots, m$),
2) F is of class $\mathbf{C}^1$ on a neighborhood W of (x_0 , y_0), and
3) $\Delta = \dfrac{D(F_1,...,F_m)}{D(y_1,...,y_m)}(x_0,y_0)\neq 0$, then:

(a) There exists $U \in \mathscr{V}(x_0)$ and a unique solution $f : U \to V$ of the equation $F = 0$, such that $f(x_0) = y_0$,

(b) $f\in \mathbf{C}^1_{\mathbb{R}^m}(U)$, and for all $i = 1, \dots, m$ and $j = 1, \dots, p$, we have

$$\frac{\partial f_i}{\partial x_j} = -\frac{\dfrac{D(F_1,...,F_m)}{D(y_1,...,y_{i-1},x_j,y_{i+1},...,y_m)}}{\dfrac{D(F_1,...,F_m)}{D(y_1,...,y_{i-1},y_i,y_{i+1},....,y_m)}} \qquad (***)$$

(c) If $F\in \mathbf{C}^k_{\mathbb{R}^m}(W)$ for some $k \in\mathbb{N}^*$, then $f\in \mathbf{C}^k_{\mathbb{R}^m}(U)$ too.

Proof. (a) We reason by mathematical induction over m. The verification step is contained in Theorem 4.3, where $m = 1$. In the second step we have to show that for every $m \in \mathbb{N}$, from the hypothesis of validity up to $m - 1$ it follows the validity for m.

Because $\Delta \neq 0$, there exists at least one non-null minor of order $m - 1$. For simplicity, let it be

$$\frac{D(F_1,...,F_{m-1})}{D(y_1,...,y_{m-1})}(x_0, y_0) \neq 0\,.$$

Since the theorem is supposed to be true for $m - 1$, the system

$$\begin{cases} F_1(x_1,...,x_p; y_1,...,y_{m-1}, y_m) = 0 \\ \ \ \cdots \\ F_{m-1}(x_1,...,x_p; y_1,...,y_{m-1}, y_m) = 0 \end{cases}$$

defines $m - 1$ implicit functions $y_1, \ldots, y_{m-1}$ in a neighborhood of (x_0, y_0). More exactly, if we note $x_0 = (x_1^0,...,x_p^0)$, and $y_0 = (y_1^0,...,y_m^0)$, then there exist $U' \in \mathscr{V}(x_0)$, $V' = V_1' \times V_2' \times ... \times V_m' \in \mathscr{V}(y_0)$, and $m - 1$ functions

$$y_1 = h_1(x, y_m) : U' \times V_m' \to V_1'$$
$$y_2 = h_2(x, y_m) : U' \times V_m' \to V_2'$$
$$\cdots$$
$$y_{m-1} = h_{m-1}(x, y_m) : U' \times V_m' \to V_{m-1}'$$

such that $h_k(x_0, y_m^0) = y_k^0$ for all $k = \overline{1, m-1}$. In addition, we have

$$F_i(x, h_1(x, y_m),..., h_{m-1}(x, y_m), y_m) = 0\,,\ \forall\, i = \overline{1, m-1}\,,$$

at each $(x, y_m) \in U' \times V_m'$. Because $h_1, \ldots, h_{m-1}$ are functions of class $\mathbf{C}^1$ on the neighborhood $U' \times V_m'$, the initial system

$$F_i(x; y_1,..., y_{m-1}, y_m) = 0\,,\ i = \overline{1, m}\ ,$$

is equivalent (at least on $U' \times V_1' \times V_2' \times ... \times V_m'$) to the system

$$\begin{cases} y_1 = h_1(x, y_m) \\ y_2 = h_2(x, y_m) \\ \cdots \\ y_{m-1} = h_{m-1}(x, y_m) \\ F_m(x, h_1(x, y_m),..., h_{m-1}(x, y_m), y_m) = 0\ . \end{cases}$$

Now, let us consider a helping function $\varphi : U' \times V_m' \to \mathbb{R}$, of values

$$\varphi(x, y_m) = F_m(x, h_1(x, y_m),..., h_{m-1}(x, y_m), y_m)\,.$$

It is easy to see that φ satisfies conditions 1) and 2) of theorem 4.3, so that $\varphi(x, y_m) = 0$ implicitly defines y_m in a neighborhood of (x_0, y_m^0).

Hypothesis 3) of theorem 4.3 is also fulfilled, i.e. $\frac{\partial \varphi}{\partial y_m}(x_0, y_m^0) \neq 0$. In fact, deriving in respect to y_m, we obtain the following system of conditions

$$\begin{aligned} &\frac{\partial F_1}{\partial y_1} \cdot \frac{\partial h_1}{\partial y_m}(x_0, y_m^0) + \ldots + \frac{\partial F_1}{\partial y_{m-1}} \cdot \frac{\partial h_{m-1}}{\partial y_m}(x_0, y_m^0) + \frac{\partial F_1}{\partial y_m} = 0 \\ &\vdots \\ &\frac{\partial F_{m-1}}{\partial y_1} \cdot \frac{\partial h_1}{\partial y_m}(x_0, y_m^0) + \ldots + \frac{\partial F_{m-1}}{\partial y_{m-1}} \cdot \frac{\partial h_{m-1}}{\partial y_m}(x_0, y_m^0) + \frac{\partial F_{m-1}}{\partial y_m} = 0 \\ &\frac{\partial \varphi}{\partial y_m}(x_0, y_m^0) = \frac{\partial F_m}{\partial y_1} \cdot \frac{\partial h_1}{\partial y_m}(x_0, y_m^0) + \ldots + \frac{\partial F_m}{\partial y_{m-1}} \cdot \frac{\partial h_{m-1}}{\partial y_m}(x_0, y_m^0) + \frac{\partial F_m}{\partial y_m}, \end{aligned}$$

where the derivatives of F_1, …, F_m are evaluated at (x_0, y_0). The value of the Jacobian $\frac{D(F_1,\ldots,F_m)}{D(y_1,\ldots,y_m)}(x_0, y_0)$ is preserved if to its last column we add

$$\text{(column 1)} \cdot \frac{\partial h_1}{\partial y_m}(x_0, y_m^0) + \ldots + (\text{column } m-1) \cdot \frac{\partial h_{m-1}}{\partial y_m}(x_0, y_m^0).$$

According to the above formulas, we obtain

$$0 \neq \frac{D(F_1,\ldots,F_m)}{D(y_1,\ldots,y_m)}(x_0, y_0) = \begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_{m-1}} & \frac{\partial F_1}{\partial y_m} \\ \vdots & & \vdots & \vdots \\ \frac{\partial F_{m-1}}{\partial y_1} & \cdots & \frac{\partial F_{m-1}}{\partial y_{m-1}} & \frac{\partial F_{m-1}}{\partial y_m} \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_{m-1}} & \frac{\partial F_m}{\partial y_m} \end{vmatrix}(x_0, y_0) =$$

$$= \begin{vmatrix} \frac{\partial F_1}{\partial y_1}(x_0, y_0) & \cdots & \frac{\partial F_1}{\partial y_{m-1}}(x_0, y_0) & 0 \\ \vdots & & \vdots & \vdots \\ \frac{\partial F_{m-1}}{\partial y_1}(x_0, y_0) & \cdots & \frac{\partial F_{m-1}}{\partial y_{m-1}}(x_0, y_0) & 0 \\ \frac{\partial F_m}{\partial y_1}(x_0, y_0) & \cdots & \frac{\partial F_m}{\partial y_{m-1}}(x_0, y_0) & \frac{\partial \varphi}{\partial y_m}(x_0, y_m^0) \end{vmatrix} =$$

$$= \frac{D(F_1,\ldots,F_{m-1})}{D(y_1,\ldots,y_{m-1})}(x_0, y_0) \cdot \frac{\partial \varphi}{\partial y_m}(x_0, y_m^0).$$

Consequently $\frac{\partial \varphi}{\partial y_m}(x_0, y_m^0) \neq 0$, hence theorem 4.3 is working. This means that there exist a neighborhood $U \times V_m^0 \in \mathscr{V}(x_0, y_m^0)$ and a function

$$f_m : U \to V_m^0$$

such that $\varphi(x, f_m(x)) \equiv 0$ on U, and $f_m(x_0) = y_m^0$. In addition, function f belongs to the class $\mathbf{C}^1_{\mathbb{R}}(U)$, and its derivative is given by (*).

It is easy to see that the initial system ($F_i = 0$, $\forall i = \overline{1,m}$) is equivalent to

$$\begin{cases} y_1 = h_1(x, y_m) \\ y_2 = h_2(x, y_m) \\ \dots \\ y_{m-1} = h_{m-1}(x, y_m) \\ y_m = f_m(x) \end{cases}$$

on U x V, where $V = V_1' \times V_2' \times \ldots \times V_{m-1}' \times V_m^0 \in \mathscr{V}(y_0)$.

If we note $f_1(x) = h_1(x, f_m(x))$, …, $f_{m-1}(x) = h_{m-1}(x, f_m(x))$, then we may conclude that $f = (f_1, \ldots, f_{m-1}, f_m) : U \to V$ is the searched implicit function, i.e. assertion (a) of the theorem is proved.

(b) If we derive relative to x_j in the equations $F_i(x, f_1(x), \ldots, f_m(x)) = 0$, where $i = \overline{1,m}$, then we obtain the system

$$\frac{\partial F_i}{\partial x_j} + \frac{\partial F_i}{\partial y_1} \cdot \frac{\partial f_1}{\partial x_j} + \ldots + \frac{\partial F_i}{\partial y_m} \cdot \frac{\partial f_m}{\partial x_j} = 0, \ \forall i = \overline{1,m}.$$

The Cramer's rule furnishes the entire set of derivatives $\dfrac{\partial f_k}{\partial x_j}$, as in (***).

(c) We reason by induction, like in the proof of theorem 4.3. ◇

4.7. <u>Remarks</u>. (a) The above theorems assure the existence of the implicit functions, but do not offer methods to construct them in practice.
(b) The formulas (*), (**) and (***) are useful in calculating the (partial) derivatives of implicit functions, especially when the explicit expressions are not known. In particular, the formulas (***) follow by Cramer's rule.
(c) The study of the extreme points of implicit functions may be done without getting their explicit form. For example, if $y = f(x)$ is implicitly defined by $F(x, y) = 0$, where $x, y \in \mathbb{R}$, the stationary points (where $y' = 0$) are given by the system:

$$\begin{cases} F(x, y) = 0 \\ \dfrac{\partial F}{\partial x}(x, y) = 0 \\ \dfrac{\partial f}{\partial y}(x, y) \neq 0 \ . \end{cases}$$

Sufficient conditions are expressed by the sign of y'', which may be obtained by deriving in (*) one more time.

Similarly, if the function $z = f(x, y)$ is implicitly defined by $F(x, y, z) = 0$, where $x, y, z \in \mathbb{R}$, then the stationary points are the solutions of the system

$$\begin{cases} F(x,y,z) = 0 \\ \dfrac{\partial F}{\partial x}(x,y,z) = 0 \\ \dfrac{\partial F}{\partial y}(x,y,z) = 0 \\ \dfrac{\partial F}{\partial z}(x,y,z) \neq 0 \end{cases} .$$

The decision about extremes results from the study of *sign* (Δ), where

$$\Delta = \frac{\partial^2 z}{\partial x^2} \cdot \frac{\partial^2 z}{\partial y^2} - \left(\frac{\partial^2 z}{\partial x \partial y} \right)^2 .$$

To obtain the second order partial derivatives of z, which occur in Δ, we do another derivation in formulas (***).

There are three types of problems based on the implicit function theorems like 4.3 and 4.6 from above, namely the *conditional extrema*, the *change of coordinates*, and the *functional dependence*. In the sequel we analyze these problems, in the mentioned order.

We start with a geometric example, which illustrates the strong practical nature of the conditional extrema theory.

4.8. <u>Example</u>. Let us find the point $P(x_0, y_0, z_0) \in \mathbb{R}^3$, which belongs to the plane of equation $x + y + z = 1$, and has the smallest distance to the origin.

To solve the problem, we have to find the minimum of the function

$$f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$$

under the condition $x + y + z = 1$. Of course, we may reduce this problem to a *free* extremum one if we replace $z = 1 - x - y$ in f and we study the forthcoming function of two variables. This method works in the present case because we can make the given restriction explicit. Therefore, we are interested in more general methods, which concerns implicit restrictions.

Generally speaking, the problem may involve more than one restriction. This is the case when we are looking for a point P, which has the smallest distance to the origin, and belongs to the straight line

$$\begin{cases} x + y + z = 1 \\ x - y + 2z = 0 . \end{cases}$$

In this case, again, we can express y and z as functions of x, and reduce the problem to that of a *free* minimum of a function of a single real variable x.

It is useful to remark that the number of conditions equals the number of implicit functions, and it cannot exceed the total number of variables. More exactly, we have to specify the terminology:

4.9. <u>Definition</u>. Let $D \subseteq \mathbb{R}^{p+m}$ be an open set, and let $f : D \to \mathbb{R}$ be a function of class $\mathbf{C}^1$ on D (also called *objective function*). The equations

$$g_i(x_1, \ldots, x_p; y_1, \ldots, y_m) = 0, \quad i = 1, \ldots, m$$

where $g_i : D \to \mathbb{R}$ are functions of class $\mathbf{C}^1$ on D for all $i = 1, \ldots, m$, are called *conditions* (*restrictions* or *coupling equations*). For brevity, we note $x = (x_1, \ldots, x_p)$, $y = (y_1, \ldots, y_m)$ and

$$\boldsymbol{M} = \{(x, y) \in D : g_i(x, y) = 0 \text{ for all } i = 1, \ldots, m\}.$$

The point $(x_0, y_0) \in \boldsymbol{M}$ is called *local extremum* of f under the *conditions* $g_i = 0$ iff there exists a neighborhood V of (x_0, y_0), $V \subseteq D$ such that the increment $f(x, y) - f(x_0, y_0)$ has a constant sign on $V \cap \boldsymbol{M}$.

The following theorem reduces the problem of searching a conditioned extremum to the similar problem without conditions, which is frequently called *unconditional* (or *free*) extremum problem. It is easy to recognize the idea suggested by the above examples, of making the restrictions explicit. The explicit restrictions will work locally, in accordance to the implicit function theorems.

4.10. <u>Theorem</u>. (Lagrange) Let $(x_0, y_0) \in \boldsymbol{M}$ be a conditioned extremum of f as in the above definition. If

$$\frac{D(g_1, \ldots, g_m)}{D(y_1, \ldots, y_m)}(x_0, y_0) \neq 0,$$

then there exists a set of numbers $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, such that the same (x_0, y_0) is stationary point of the function

$$F = f + \lambda_1 g_1 + \ldots + \lambda_m g_m.$$

<u>Proof</u>. According to the implicit function theorem 4.6. from above, the system of conditions $g_i = 0$, $i = 1, \ldots, m$, locally defines m implicit functions $y_i = f_i(x)$, $i = 1, \ldots, m$, of class $\mathbf{C}^1$, such that $f_i(x_0) = y_i^0$ holds for all $i = 1, \ldots, m$. By deriving the relations $g_i(x, f_1(x), \ldots, f_m(x)) = 0$ (on $\boldsymbol{M}$), where $i = 1, \ldots, m$, relative to x_j, $j = 1, \ldots, p$, we obtain:

$$\frac{\partial g_i}{\partial x_j} + \sum_{k=1}^{m} \frac{\partial gi}{\partial y_k} \cdot \frac{\partial f_k}{\partial x_j} = 0, \; i = 1, \ldots, m.$$

On the other hand x_0 is a (free) extremum for $f(x, f_1(x), \ldots, f_m(x))$, hence it is stationary point too, i.e. $\dfrac{\partial f}{\partial x_j}(x_0) = 0$ for all $j = 1, \ldots, p$. In other words, for each fixed $j \in \{1, \ldots, p\}$, the vector

$$\left(1, \frac{\partial f_1}{\partial x_j}(x_0), \ldots, \frac{\partial f_m}{\partial x_j}(x_0)\right) \overset{not.}{=} (u_0, u_1, \ldots, u_m)$$

represents a non-trivial solution of the homogeneous linear system of m+1 equations:

$$\begin{cases} \dfrac{\partial g_i}{\partial x_j}(x_0, y_0)u_0 + \sum_{k=1}^{m} \dfrac{\partial g_i}{\partial y_k}(x_0, y_0)u_k = 0, \quad i = 1,...,m \\ \dfrac{\partial f}{\partial x_j}(x_0, y_0)u_0 + \sum_{k=1}^{m} \dfrac{\partial f}{\partial y_k}(x_0, y_0)u_k = 0 \quad . \end{cases}$$

Consequently, for all j = 1, …, p, we have

$$\frac{D(f, g_1,..., g_m)}{D(x_j, y_1,..., y_m)} = 0 .$$

This fact implies the existence of a linear combination between the lines of this determinant, that is, there exist $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$ such that

$$\begin{cases} \dfrac{\partial f}{\partial x_j}(x_0, y_0) + \sum_{i=1}^{m} \lambda_i \dfrac{\partial g_i}{\partial x_j}(x_0, y_0) = 0 \\ \dfrac{\partial f}{\partial y_k}(x_0, y_0) + \sum_{i=1}^{m} \lambda_i \dfrac{\partial g_i}{\partial y_k}(x_0, y_0) = 0 \qquad k = 1,...,m \ . \end{cases}$$

Because $\lambda_1, \ldots, \lambda_m$ are uniquely determined by the above last m equations, it follows that they make valid the former equation too, for all j = 1, …, p. In other words, this means that (x_0 , y_0) is a stationary (non-conditional) point for $F = f + \lambda_1 g_1 + \ldots + \lambda_m g_m$. ◇

4.11. <u>Remarks</u>. a) In practice, the above theorem is used in the sense that we primarily have to find the solutions $x_1^0, \ldots, x_p^0; y_1^0, \ldots, y_m^0 ; \lambda_1^0, \ldots, \lambda_m^0$ of the system

$$\begin{cases} \dfrac{\partial F}{\partial x_j}(x, y) = 0, \qquad j = 1,...,p \\ \dfrac{\partial F}{\partial y_k}(x, y) = 0, \qquad k = 1,...,m \\ g_i(x, y) = 0, \qquad i = 1,...,m \ . \end{cases}$$

In particular, it gives the stationary points of f under the given conditions.

The selection of the points of real extremum results from the study of the sign of $d^2F_{(x_0, y_0)}$, where we take into account that dx_j , j = 1, …, p, and dy_k, k = 1, …, m , are related by $dg_i(x_0, y_0) = 0$, i = 1, …, m, i.e.

$$\frac{\partial g_i}{\partial x_1}dx_1 + ... + \frac{\partial g_i}{\partial x_p}dx_p + \frac{\partial g_i}{\partial y_1}dy_1 + ... + \frac{\partial g_i}{\partial y_m}dy_m = 0 , \qquad i = 1,...,m .$$

These facts are based on the remark that under the restrictions $g_i(x, y) = 0$, we have $F = f$, $\Delta F = \Delta f$, etc.

b) The above method is useless in the case when the points of extremum belong to the boundary of the domain D, of the objective function f (usually, $\overline{D}$ is compact). In principle, we may treat this case as a problem of conditional extremum, by adding new restrictions, namely the equations of the boundary.

Another important application of the implicit function theorem concerns the invertible functions of several variables. Roughly speaking, to invert the function $f : A \to \mathbb{R}^m$, where $A \subseteq \mathbb{R}^p$, means to solve the vectorial equation $f(x) - y = 0_{\mathbb{R}^m}$, or, more exactly, the system

$$\begin{cases} f_1(x_1,\dots,x_p) - y_1 = 0 \\ \vdots \\ f_m(x_1,\dots,x_p) - y_m = 0 \ . \end{cases}$$

If the differentiability is allowed, and $p = m$, then we naturally try to realize the inversion by the help of theorem 4.6. In this respect, we present the following definition, which introduces the specific terminology:

4.12. <u>Definition</u>. Let A be an open subset of $\mathbb{R}^p$, where $p \geq 1$. Each function $T : A \to \mathbb{R}^p$ is called *transformation* of A. If $T \in \mathbf{C}^1_{\mathbb{R}^p}(A)$, then we call it *smooth transformation* of A. If $T : A \to B \subseteq \mathbb{R}^p$ is a 1:1 (one to one) smooth transformation of A onto B, and T^{-1} is smooth on B, then it is named *diffeomorphism* between A and B.

4.13. <u>Theorem</u>. (Local inversion) Let $T : A \to \mathbb{R}^p$, where $A \subseteq \mathbb{R}^p$ is an open set, be a smooth transformation of A, and let $x_0 \in A$ be fixed. If T, through its components $f_1, \dots, f_p$, satisfies the condition

$$\frac{D(f_1,\dots,f_p)}{D(x_1,\dots,x_p)}(x_0) \neq 0,$$

then there exist some neighborhoods $U \in \mathscr{V}(x_0)$ and $V \in \mathscr{V}(T(x_0))$, such that T is a diffeomorphism between U and V.

<u>Proof</u>. It is sufficient to apply theorem 4.6 to the equation $T(x) - y = 0_{\mathbb{R}^m}$. The resulting implicit function obviously is T^{-1}. Formula (***) shows that T^{-1} is a smooth transformation of V. ◇

4.14. <u>Corollary</u>. In the conditions of the above theorem, if $\varphi_1, \dots, \varphi_p$ are the components of T^{-1}, then

$$\frac{D(\varphi_1,\dots,\varphi_p)}{D(y_1,\dots,y_p)}(T(x_0)) = \left[\frac{D(f_1,\dots,f_p)}{D(x_1,\dots,x_p)}(x_0)\right]^{-1}.$$

Proof. We have $T^{-1}\circ T=\iota$, where ι is the identity on *U*, and $|\mathbf{J}_{\iota}(x_0)|=1$. According to formula (*) in theorem IV.3.6, the Jacobi matrixes are related by the equality

$$\mathbf{J}_{T^{-1}\circ T}(x_0)=\mathbf{J}_{T^{-1}}(T(x_0))\cdot\mathbf{J}_T(x_0)\ .$$

It remains to take the determinants in this formula. ◇

The above inversion theorem has a strong local character, which persists in the case when the Jacobi matrix is different from zero on the entire *A*. It is frequent in practice, as the following examples show:

4.15. Examples a) When we are drawing flat maps from a circle or sphere, we usually realize projections like *T* in Fig.IV.4.1 from below. As a matter of fact, *T* projects points *X* of a half-circle only, namely which correspond to angles $x\in(-\pi/2,+\pi/2)$ at the center *C*. The action of *T* is completely described by function $f:(-\pi/2,+\pi/2)\to\mathbb{R}$, of values $f(x)=\ell\, tg\, x=T(X)$.

Because of the simple form of *f*, we prefer to consider it as *projection* of $\mathscr{C}$ on $\mathbb{R}$, instead of *T* . In addition, we easily obtain the derivative

$$f^{\prime}(x)=\frac{\ell}{\cos^2 x}\ .$$

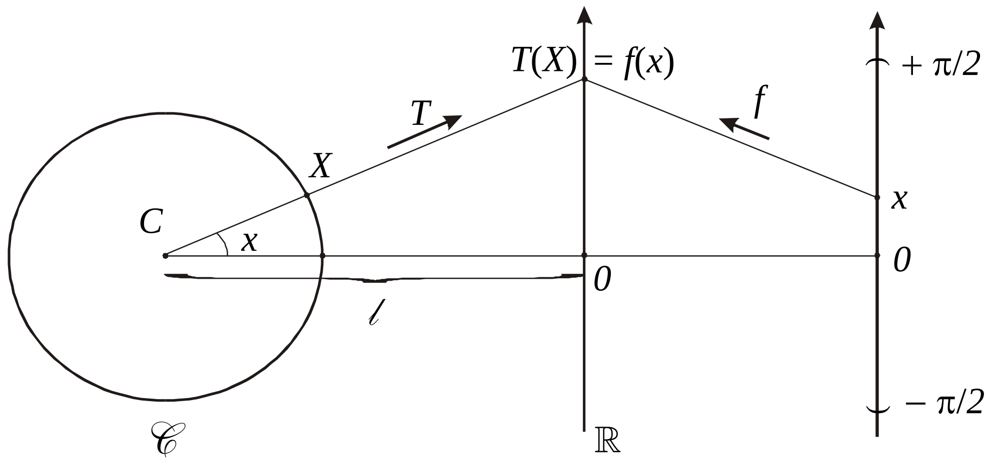


Fig. IV.4.1

To conclude, we may consider that *f* is a local diffeomorphism between $\mathscr{C}$ and $\mathbb{R}$. Anyway there is no global diffeomorphism between them.

b) When we pass from Cartesian to polar coordinates in $\mathbb{R}^2$, then we realize a transformation $T:\mathbb{R}^2\setminus(0,0)\to\mathbb{R}^*_+\times[0,2\pi)$, as in Fig.IV.4.2.

The transformation of a formula from (x,y) to (ρ,θ) reduces to replace *x* and *y* according to the formulas that define T^{-1}, namely

$$\begin{cases}x=\rho\cos\theta\\ y=\rho\sin\theta\ .\end{cases}$$

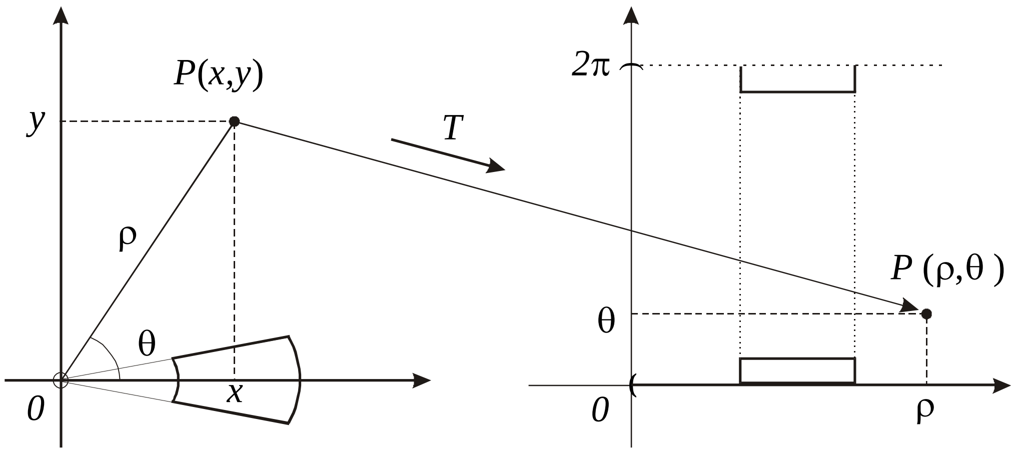


Fig.IV.4.2.

A direct calculus of the Jacobian leads to

$$\left|\mathbf{J}_{T^{-1}}\right| = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & +\rho\cos\theta \end{vmatrix} = \rho\,,$$

hence the local inversion theorem 4.13 is working only if $\rho \neq 0$. More than this, because *sin* and *cos* are periodical functions, the formulas of T^{-1} carry the whole set $\{(\rho,\ \theta + 2k\pi): k \in \mathbb{Z}\}$ to the same point $(x, y) \in \mathbb{R}^2$. In other terms, T can be reversed only from $\mathbb{R}_+ \times [0, 2\pi)$, i.e. rotations around the origin are not allowed in $\mathbb{R}^2$ any more. If we remark that T is discontinuous at points of $\mathbb{R}_+^*$ (e.g. connectedness is not preserved, see Fig.IV.4.2), then the only chance for T to realize a diffeomorphism is obtained by removing the half-line $\mathbb{R}_+^* \times \{0\}$ from its domain of definition. Consequently, T is a diffeomorphism between the sets $A = \mathbb{R}^2 \setminus [\mathbb{R}_+ \times \{0\}]$ and $B = \mathbb{R}_+^* \times (0, 2\pi)$.

Because of its role in this construction, the half-line $\mathbb{R}_+ \times \{0\}$, which has been removed from $\mathbb{R}^2$, is called a *cut* of the plane. Cutting the plane shows another feature of the local character of the inverse function theorem.

To work with local maps means to construct diffeomorphisms similar to T in the example a) from above, combine them in some “atlas”, etc. This technique is specific to the differential geometry of manifolds (see [TK], [UC], etc.), where the local properties furnished by these maps represent the “pieces” of the global properties. However, in analysis we are interested in doing global transformations and changing the coordinates on the entire space, which usually is the *flat* $\mathbb{R}^p$. In this sense, we may place analysis between the geometry on flat spaces, involving continuous transformations, and that of manifolds, where differentiability holds locally.

To be more rigorous, we have to specify some terms:

4.16. <u>Definition</u>. If $A\subseteq\mathbb{R}^p$ is open, and $T : A\to\mathbb{R}^p$ is a diffeomorphism between A and $B = T(A)$, then T is called *change of coordinates* (*variables*). The variables $x_1 ,\dots, x_p$ are called *"old" coordinates* of $x = (x_1 ,\dots, x_p) \in A$, and the components of T at x, namely $f_1(x), \dots, f_p(x)$ are said to be the *"new" coordinates* of x . If, in addition

$$\frac{D(f_1,\dots,f_p)}{D(x_1,\dots,x_p)}(x)\neq 0$$

at each $x\in A$, then we say that T is a *regular*, or *non-degenerate* change of coordinates on A.

The differential calculus, where only regular changes of coordinates are applied, is based on the following simple consequence of theorem 4.6:

4.17. <u>Corollary</u>. Let T be a regular change of coordinates of A, for which we note $T(x) = y = (y_1, \dots, y_p)$. The inverse $T^{-1} = (\varphi_1, \dots, \varphi_p)$ coincides with any local inverse of T, and at each $x\in A$ we have

$$\frac{D(\varphi_1,\dots,\varphi_p)}{D(y_1,\dots,y_p)}(T(x))=\left[\frac{D(f_1,\dots,f_p)}{D(x_1,\dots,x_p)}(x)\right]^{-1}.$$

<u>Proof</u>. The first assertion is a consequence of the uniqueness of the implicit function (theorem 4.6). The second property is an immediate consequence of corollary 4.14 applied at a current point $x\in A$. ◇

In practice, we often have to derive composite functions, which involve transformations of coordinates.

4.18. <u>Example</u>. Write the Laplace equation in polar coordinates in $\mathbb{R}^2$.

In Cartesian coordinates, the Laplace equation is $\frac{\partial^2 u}{\partial x^2}+\frac{\partial\ u}{\partial y^2}=0$. To change the variables means to replace $u(x, y) = v(\rho, \theta)$, where

$$\begin{cases} x=\rho\cos\theta \\ y=\rho\sin\theta \end{cases},\quad \rho>0 \text{ and } \theta\in(0, 2\pi).$$

We calculate the partial derivatives of the composed function to obtain

$$\begin{cases} \dfrac{\partial v}{\partial\rho}=\dfrac{\partial u}{\partial x}\cdot\dfrac{\partial x}{\partial\rho}+\dfrac{\partial u}{\partial y}\cdot\dfrac{\partial y}{\partial\rho} \\ \dfrac{\partial v}{\partial\theta}=\dfrac{\partial u}{\partial x}\cdot\dfrac{\partial x}{\partial\theta}+\dfrac{\partial u}{\partial y}\cdot\dfrac{\partial y}{\partial\theta} \end{cases}$$

The system gives the first order partial derivatives of u. By another derivation of the resulting formulas we obtain the second order partial derivatives of u, such that the given equation becomes

$$\frac{\partial}{\partial\rho}\left(\rho\frac{\partial v}{\partial\rho}\right)+\frac{1}{\rho}\frac{\partial^2 v}{\partial\theta^2}=0.$$

The last application of the implicit function theorem, which we analyze in this section, concerns the functional dependence. Before giving the exact definition and the theoretical results we consider some examples:

4.19. Examples. (a) The *linear dependence* of the functions $f_k : A\to\mathbb{R}$, where $k = 1, \dots, m$ and $A\subseteq\mathbb{R}^p$ is an open set, means that

$$f_m = \lambda_1 f_1 + \dots + \lambda_{m-1} f_{m-1}$$

for some $\lambda_1, \dots, \lambda_{m-1} \in\mathbb{R}$, i.e. for all $x\in A$ we have

$$f_m(x) = \lambda_1 f_1(x) + \dots + \lambda_{m-1} f_{m-1}(x) .$$

(b) The three functions $f_1, f_2, f_3 : \mathbb{R}^3\to\mathbb{R}$, defined by $f_1(x, y, z) = x + y + z$, $f_2(x, y, z) = xy + xz + yz$ and $f_3(x, y, z) = x^2 + y^2 + z^2$, are connected by the relation $f_1^2 - 2f_2 - f_3 = 0$. In other words there exists a function $F\in \mathbf{C}_{\mathbb{R}}^1(\mathbb{R}^3)$, namely $F(u, v, w) = u^2 - 2v - w$, such that the equality

$$F(f_1(x, y, z), f_2(x, y, z), f_3(x, y, z)) = 0$$

holds identically on $\mathbb{R}^3$. Briefly, we note $F(f_1, f_2, f_3) = 0$, and we remark that this dependence is non-linear.

(c) We may formulate the above dependence of f_1, f_2, f_3 in *explicit* form, e.g. $f_3 = f_1^2 - 2f_2$. In this case $f_3 = G(f_1, f_2)$, where $G(u, v) = u^2 - 2v$.

4.20. Definition. We say that the functions $f_1, \dots, f_m : A\to\mathbb{R}$, where $A\subseteq\mathbb{R}^p$ is an open set, are *functionally dependent* iff there exists a function F of class $\mathbf{C}^1$ in a domain of $\mathbb{R}^m$ such that $F(f_1(x), \dots, f_m(x)) = 0$ at each $x\in A$. In the contrary case we say that $f_1, \dots, f_m$ are *functionally independent*.

Alternatively, if we can put the dependence of one function, say f_m, in the explicit form $f_m = G(f_1, \dots, f_{m-1})$, where G is a function of class $\mathbf{C}^1$ in a domain of $\mathbb{R}^{m-1}$, then we say that f_m *functionally depends* on $f_1, \dots, f_{m-1}$.

We start with a sufficient condition for independence:

4.21. Proposition. Let $f_1, \dots, f_m\in \mathbf{C}_{\mathbb{R}}^1(A)$, where $A\subseteq\mathrm{R}^p$ is an open set, and let $m\le p$. If at some $x_0\in A$ we have rank $\mathbf{J}_{(f_1,\dots,f_m)}(x_0) = m$, then $f_1, \dots, f_m$ are functionally independent on a neighborhood of x_0.

Proof. Let us suppose the contrary, i.e. for each $V\in \mathscr{V}(x_0)$ there exists a function $F\in \mathbf{C}^1_{\mathbb{R}^{m-1}}(V)$ of variables $u_1, \dots, u_{m-1}$ such that at all $x\in V$ we have

$$f_m(x) = F(f_1(x), \dots, f_{m-1}(x)).$$

Deriving f_m like a composite function, we obtain

$$\frac{\partial f_m}{\partial x_j}(x) = \sum_{k=1}^{m-1} \frac{\partial F}{\partial u_k}(f_1(x),\dots, f_{m-1}(x))\frac{\partial f_k}{\partial x_j}(x),\ j = 1,\dots, p.$$

In particular, at $x = x_0$, where $f_k(x_0) = u_k^0$ for all $k = 1, \dots, m-1$, we have

$$\frac{\partial f_m}{\partial x_j}(x_0) = \sum_{k=1}^{m-1} \frac{\partial F}{\partial u_k}(u_1^0,\dots,u_{m-1}^0)\frac{\partial f_k}{\partial x_j}(x_0),\ j = 1,\dots, p.$$

Because the numbers $\frac{\partial F}{\partial u_k}(u_1^0,...,u_{m-1}^0)$ are independent of the values of x_j for all $j = 1, 2, \ldots, p$, the above relation shows that in the Jacobian matrix of $f_1, \ldots, f_m$ relative to $x_1, \ldots, x_p$, the last line is a linear combination of the other lines, hence the rank of this matrix is less than m. ◇

4.22. Remark. The above proposition shows that if m functions $f_1, \ldots, f_m$ are functionally dependent on A, then the rank of the corresponding Jacobi matrix is less than m, which equals the number of functions, at any $x \in A$. In practice, we frequently need information about the converse implication, in order to establish the *existence* of a functional dependence (we stressed on "existence" because finding the concrete dependence F is too complicated). In this respect we mention the following:

4.23. Theorem. Let us take $f_1, \ldots, f_m \in \mathbf{C}_{\mathbb{R}}^1(A)$, where $A \subseteq \mathbb{R}^p$ is an open set, and $m \leq p$. If there exists $x_0 \in A$ and $V \in \mathscr{V}(x_0)$ such that

$$\text{rank } \mathbf{J}_{(f_1,...,f_m)}(x) = r < m$$

holds at all $x \in V$, then r of the given functions are functionally independent. The other $m - r$ functions are functionally dependent on the former ones on a neighborhood of x_0.

Proof. To make a choice, let us suppose that

$$\Delta = \frac{D(f_1,...,f_r)}{D(x_1,...,x_r)}(x_0) \neq 0$$

According to the above proposition 4.20, the functions $f_1, \ldots, f_r$ are functionally independent on a neighborhood of x_0. So, it remains to show that the other functions $f_{r+1}, \ldots, f_m$ depend on $f_1, \ldots, f_r$ in a neighborhood of x_0. In fact, we may remark that according to theorem 4.6, the system

$$\begin{cases} f_1(x_1,...,x_p) - y_1 = 0 \\ \quad\vdots \\ f_r(x_1,...,x_p) - y_r = 0 \end{cases}$$

has a unique solution

$$\begin{cases} x_1 = \varphi_1(x_{r+1},...,x_p;y_1,...,y_r) \\ \quad\vdots \\ x_r = \varphi_r(x_{r+1},...,x_p;y_1,...,y_r) \end{cases}$$

in a neighborhood of the point $(x_1^0, \ldots, x_p^0; y_1^0, \ldots, y_r^0)$. Consequently, in such a neighborhood, and for all $j = 1, \ldots, r$, the following equalities hold:

$$f_j(\varphi_1(x_{r+1},\ldots,x_p;y_1,\ldots,y_r), \ldots, \varphi_r(x_{r+1},\ldots,x_p;y_1,\ldots,y_r), x_{r+1},\ldots,x_p) - y_j = 0 .$$

Deriving these relations relative to x_k, $k = r + 1, \ldots, p$, it follows that

$$\sum_{i=1}^{r} \frac{\partial f_j}{\partial x_i} \cdot \frac{\partial \varphi_i}{\partial x_k} + \frac{\partial f_j}{\partial x_k} = 0, \; j = 1,...,r \qquad (1)$$

Now, by replacing $x_1,\ldots,x_r$ in f_s, where $s > r$, we obtain

$$f_s(x_1,\ldots,x_r;x_{r+1},\ldots,x_p) =$$
$$= f_s(\varphi_1(x_{r+1},\ldots,x_p;y_1,\ldots,y_r),\ldots,\varphi_r(x_{r+1},\ldots,x_p;y_1,\ldots,y_r),x_{r+1},\ldots,x_p) =$$
$$= F_s(x_{r+1},\ldots,x_p;y_1,\ldots,y_r).$$

The assertion of the theorem is proved if we show that F_s does not depend on $x_{r+1},\ldots,x_p$. Aiming at this result we show that all the derivatives

$$\frac{\partial F_s}{\partial x_k} = \frac{\partial f_s}{\partial x_k} + \sum_{i=1}^{r} \frac{\partial f_s}{\partial x_i}\cdot\frac{\partial \varphi_i}{\partial x_k} \tag{2}$$

vanish on a neighborhood of x_0 for all $s = r+1,\ldots,m$, and $k = r+1,\ldots,p$.

In fact, the hypothesis concerning the rank of the Jacobian leads to

$$\begin{vmatrix} \frac{\partial f_1}{\partial x_1}\cdots\frac{\partial f_1}{\partial x_r} & \frac{\partial f_1}{\partial x_k} \\ \vdots \quad\quad \vdots & \vdots \\ \frac{\partial f_r}{\partial x_1}\cdots\frac{\partial f_r}{\partial x_r} & \frac{\partial f_r}{\partial x_k} \\ \frac{\partial f_s}{\partial x_1}\cdots\frac{\partial f_s}{\partial x_r} & \frac{\partial f_s}{\partial x_k} \end{vmatrix} = \begin{vmatrix} \frac{\partial f_1}{\partial x_1}\cdots\frac{\partial f_1}{\partial x_r} & 0 \\ \vdots \quad\quad \vdots & \vdots \\ \frac{\partial f_r}{\partial x_1}\cdots\frac{\partial f_r}{\partial x_r} & 0 \\ \frac{\partial f_s}{\partial x_1}\cdots\frac{\partial f_s}{\partial x_r} & \frac{\partial F_s}{\partial x_k} \end{vmatrix} = 0,$$

where the second form of this determinant is obtained by using (1) and (2). This means that $\Delta\cdot\frac{\partial F_s}{\partial x_k} = 0$ on a neighborhood of x_0. Since $\Delta \neq 0$, we obtain $\frac{\partial F_s}{\partial x_k} = 0$, i.e. F_s does not depend on x_k. Because $s = r+1,\ldots,m$ and $k = r+1,\ldots,p$ are arbitrary, it follows that

$$f_s(x_1,\ldots,x_p) = F_s(f_1(x),\ldots,f_r(x)),$$

i.e. $f_{r+1},\ldots,f_m$ depend on $f_1,\ldots,f_r$. ◇

4.24. <u>Corollary</u>. The functions $f_1,\ldots,f_m : A\to\mathbb{R}$, where $A\subseteq\mathbb{R}^m$ is an open set (notice that $p = m$!) are functionally dependent on A, if and only if

$$\frac{D(f_1,\ldots,f_m)}{D(x_1,\ldots,x_m)}(x) = 0$$

at any $x\in A$.

No proof is necessary since this assertion is a direct consequence of proposition 4.20 and theorem 4.22.

In practice, it is also useful to notice that m functions of p variables are always dependent if $m > p$.

PROBLEMS § IV.4.

1. Give geometrical interpretation to the construction realized in the proof of theorem 4.3, on a figure corresponding to $F(x, y) = x^2 + y^2 - 1$
Hint. Compare to example 4.1. Intersect the paraboloid $z = x^2 + y^2 - 1$ with planes of equations $x = x_0$ and $y = y_0$.

2. Evaluate the derivatives $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$, and find the extreme values of the function y, implicitly defined by $(x^2 + y^2)^3 - 3(x^2 + y^2) + 1 = 0$.
Solution. $y' = -\frac{x}{y}$; $y'' = -(x^2 + y^2)\, y^{-3}$.

3. Find the derivatives $\frac{dz}{dx}$ and $\frac{dz}{dy}$, and study the extreme values of the function $z(x, y)$, which is implicitly defined by $x^2 - 2y^2 + 3z^2 - yz + y = 0$.
Hint. Either use the formula (**) in remark 4.4, or differentiate the given equation. From $2x\,dx - 4y\,dy + 6z\,dz - y\,dz - z\,dy + dy = 0$ we deduce

$$dz = \frac{2x}{y - 6z}dx + \frac{1 - 4y - z}{y - 6z}dy.$$

4. Find $\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}$ if $u + v = x + y$ and $xu + yv = 1$.
Hint. Use theorem 4.6. Derive the given equations relative to x

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} = 1$$

$$u + x\frac{\partial u}{\partial x} + y\frac{\partial v}{\partial x} = 0$$

to obtain $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}$. Similarly, we calculate $\frac{\partial u}{\partial y}, \frac{\partial v}{\partial y}$.

Another method is based on the differentials of the given conditions

$$du + dv = dx + dy$$

$$xdu + udx + ydv + vdy = 0,$$

which provide du and dv.

5. Calculate the derivative $\frac{\partial^2 z}{\partial x \partial y}(1,-2)$, where $z(x, y)$ is implicitly defined by the equation $x^2 + 2y^2 + 3z^3 + xy - z - 9 = 0$, and $z(1, -2) = 1$.

6. The system

$$\begin{cases} x^2 + y^2 - 2z^2 = 0 \\ 2x^3 + y^3 - 3z^3 = 0 \end{cases}$$

implicitly defines y and z as functions of x in a neighborhood of (1, 1, 1). Calculate $y^{/}$, $z^{/}$, $y^{//}$ and $z^{//}$ at $x_0 = 1$.

7. Show that if $f(x, y, z) = 0$, then $\frac{\partial x}{\partial y} \cdot \frac{\partial y}{\partial z} \cdot \frac{\partial z}{\partial x} = -1$.

Hint. Express the partial derivatives by $f_x^{/}$, $f_y^{/}$ and $f_z^{/}$.

8. Show that if $F\left(\frac{x}{z}, \frac{y}{z}\right) = 0$, then $x\frac{\partial z}{\partial x} + y\frac{\partial z}{\partial y} = z$.

9. Find the point of extreme ordinate on the curve $x^3 + 8y^3 - 6xy = 0$.

10. Find the extreme values of the function $z = z(x, y)$, implicitly defined by $x^2 + y^2 + z^2 - 4z = 0$, and give geometrical interpretation to the result.
Hint. The given condition is the equation of the sphere of center (0, 0, 2).

11. Find the points of extremum for the function $f(x, y, z) = xy + xz + yz$ in the domain $D = \{(x, y, z) \in \mathbb{R}^3 : xyz = 1, x > 0, y > 0, z > 0\}$.

Hint. The Lagrange function is $F(x, y, z) = xy + xz + yz + \lambda(xyz - 1)$. The system $\frac{\partial F}{\partial x} = 0, \frac{\partial F}{\partial y} = 0, \frac{\partial F}{\partial z} = 0$, $g = 0$ gives $x = 1$, $y = 1$, $z = 1$, $\lambda = -2$. The second order differential is $d^2F(1, 1, 1) = -(dxdy + dydz + dxdz)$, but from $dg(1, 1, 1) = 0$ we obtain $dz = -dx - dy$. Replacing dz in d^2F, we obtain $d^2F(1, 1, 1) = dx^2 + dxdy + dy^2$, which is positively defined. Consequently, f has a minimum at the point (1, 1, 1).

Another method consists in studying the Hessean of $F(x, y, z(x, y))$, where z is explicitly given by $g = 0$, namely $z = \frac{1}{xy}$.

12. Find the extreme values of the function

$$f : \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i = 1, \ldots, n, n > 1\} \to \mathbb{R} ,$$

of values $f(x) = x_1 \cdot x_2 \cdot \ldots \cdot x_n$, under the restriction $x_1 + \ldots + x_n = S$, where S is a constant. Use the result to deduce that

$$\frac{x_1 + \ldots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} .$$

Hint. Use the Lagrange function $F(x) = x_1 \cdot x_2 \cdot ... \cdot x_n - \lambda(x_1 + \ldots + x_n - S)$ to find (see the above problem) $x_k = \frac{S}{n}$ for all $k = 1, \ldots, n$ and $\lambda = \left(\frac{S}{n}\right)^{n-1}$.
Further, evaluate the second order differential

$$d^2 f_{(x_0)} = \left(\frac{S}{n}\right)^{n-2} 2 \sum_{1=i<j}^{n} dx_i dx_j = \left(\frac{S}{n}\right)^{n-2} \left[\left(\sum dx_i\right)^2 - \sum dx_i^2\right]$$

and because the restriction gives $\sum dx_i = 0$, we have $d^2 f_{(x_0)} < 0$.

13. Determine the greatest and the smallest values attained by the explicit function $z = x^3 + y^3 - 3xy$ in the region $0 \le x \le 2,\ -1 \le y \le 2$.
Solution. The greatest value $z = 13$ is attained at the boundary point (2, –1). The smallest value $z = -2$ is taken at both (1, 1), which is an internal point, and at (0, –1), which belongs to the boundary.

14. Seek the extreme points of the function $f(x, y) = x^2 + y^2 - 3x - 2y + 1$ on the set $K = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$.
Hint. The only stationary point is $\left(\frac{3}{2}, 1\right) \notin \overset{\circ}{K}$. Besides the method of Lagrange function containing the equation of $\mathrm{Fr}\, K$, a geometric solution is possible if we remark that $f(x, y)$ involves the Euclidean distance between (x, y) and the stationary point of f.

15. Let $y = y(x)$, $x \in \mathbb{R}_+^*$, be a solution of the equation

$$ax^2 \frac{d^2 y}{dx^2} + bx \frac{dy}{dx} + cy = \ln x .$$

Write the equation of $u = u(t)$, where $x = e^t$, and $y(e^t) = u(t)$.
Hint. Derive u as a composed function, i.e.

$$\frac{du}{dt} = \frac{dy}{dx} \cdot \frac{dx}{dt}, \ \text{and} \ \frac{d^2 u}{dt^2} = \frac{d^2 y}{dx^2}\left(\frac{dx}{dt}\right)^2 + \frac{dy}{dx} \cdot \frac{d^2 x}{dt^2}$$

It remains to replace $y^{/}$ and $y^{//}$ in the given equation.

16. Let $f, g, h \in \mathbf{C}_{\mathbb{R}}^1(\mathbb{R})$. Find a number $\alpha \in \mathbb{R}$ such that the functions

$$u(x, y, z) = f(\alpha x + 2y - z)$$
$$v(x, y, z) = g(-x - 2y + 2z)$$
$$w(x, y, z) = h(x + 3y - 2z)$$

are functionally dependent, and write the respective dependence. Particular case: $f(t) = t^2$, $g(t) = \sin t$, and $h(t) = e^t$.
Hint. According to corollary 4.24, u, v, w are functionally dependent iff

$$\Delta = \frac{D(u,v,w)}{D(x,y,z)}(x,y,z) = \begin{vmatrix} \alpha & 2 & -1 \\ -1 & -2 & 2 \\ 1 & 3 & -2 \end{vmatrix} f'g'h' = 0 \ .$$

The case when $f' = 0$ is trivial because it leads to $f(t) = c = constant$, and this constant can be replaced in the other functions. If $I \subseteq \mathbb{R}$ be an interval on which $f' \neq 0$, $g' \neq 0$, $h' \neq 0$, then we can speak of f^{-1}, g^{-1}, h^{-1} on $f(I)$, $g(I)$, and respectively $h(I)$. In this case $\Delta = 0$ is assured by $\alpha = \frac{1}{2}$, when between the lines of Δ there exists the relation (geometrical interpretation!)

$$x + 3y - 2z = (\frac{1}{2}x + 2y - z) - \frac{1}{2}(-x - 2y + 2z) \ .$$

Because $\frac{1}{2}x + 2y - z = f^{-1}(u(x, y, z))$ and $-x - 2y + 2z = g^{-1}(v(x, y, z))$, the above (linear) relation takes the form

$$w = h(f^{-1}(u) - \frac{1}{2} g^{-1}(v)).$$

For the particularly mentioned functions, we may work on $I = (0, \frac{\pi}{2})$. The dependence becomes $w = exp\,(\sqrt{u} - \frac{1}{2} arc\,sin\,v)$.

17. Let us consider $f, g, h \in \mathbf{C}_{\mathbb{R}}^{1}(\mathbb{R})$, and define

$$u(x, y, z) = f(\frac{x-y}{y-z}),\ v(x, y, z) = g(\frac{y-z}{z-x}), \text{ and } w(x, y, z) = h(\frac{z-x}{x-y}) \ .$$

Show that u, v, w are functionally dependent on a domain $D \subseteq \mathbb{R}^3$, and find their dependence.

Hint. We may take the domain $D = \{(x, y, z) \in \mathbb{R}^3 : x > y, y > z\}$, where in addition $f'(\frac{x-y}{y-z}) \neq 0$, $g'(\frac{y-z}{z-x}) \neq 0$, and $h'(\frac{z-x}{x-y}) \neq 0$. The functional relation follows from

$$\frac{x-y}{y-z} \cdot \frac{y-z}{z-x} \cdot \frac{z-x}{x-y} = 1 \ .$$

§ IV.5. COMPLEX FUNCTIONS

In this section we study the derivable complex functions, which depend on one complex variable. The basic notions of derivative and differential obey the general rules of section IV.2. Although they closely resemble the corresponding notions for a real function, there are many specific features, on which we stress by using special terms.

5.1. <u>Definition</u>. We say that function $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{C}$, is *derivable* (in complex sense, or $\mathbb{C}$-*derivable*) at a point $z_0 \in D$ if there exists the limit

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} \overset{not.}{=} f'(z_0).$$

If so, this limit is called *derivative* of f at z_0.

5.2. <u>Examples</u>. 1. The *power function*, $f(z) = z^n$, where $n \in \mathbb{N}^*$, is derivable at every $z_0 \in \mathbb{C}$, and its derivative resembles that of a real power, namely

$$f'(z_0) = \lim_{z \to z_0} \frac{z^n - z_0^n}{z - z_0} = \lim_{z \to z_0} \sum_{k=1}^{n} z^{n-k} z_0^{k-1} = n\, z_0^{n-1} .$$

Because the general method of introducing functions is based on power series, we need rules of deriving in such series (see later IV.5.13).

2. If a real function of a complex variable is derivable at z_0, then $f'(z_0)$ is necessarily null. In fact, let us note $z = z_0 + t$, and $\zeta = z_0 + it$, where $t \in \mathbb{R}$, and suppose that f is derivable at z_0. Then there exist the limits

$$\lim_{t \to 0} \frac{f(z) - f(z_0)}{t} \overset{not.}{=} \tilde{f}(z_0), \text{ and } \lim_{t \to 0} \frac{f(\zeta) - f(z_0)}{it} = -i\, \tilde{\tilde{f}}(z_0).$$

Because both $\tilde{f}(z_0)$, $\tilde{\tilde{f}}(z_0) \in \mathbb{R}$, and $\mathbb{R} \cap \{i\,\mathbb{R}\} = \{0\}$, we obtain $f'(z_0) = 0$.

To illustrate the great difference between the real and complex analysis, we may compare particular real and complex functions. For example, the real function $f : \mathbb{C} \to \mathbb{R}$, of values $f(z) = \mathrm{Re}^2(z) + \mathrm{Im}^2(z)$, is derivable at the point $z_0 = 0$ only, while the function $\varphi : \mathbb{R}^2 \to \mathbb{R}$, which takes the same values $\varphi(x, y) = x^2 + y^2$, is differentiable on the whole plane.

3. Because the limit of a complex function reduces to the limits of the real and imaginary parts, we may express the derivative of a complex function of one real variable as a derivative of a vector function. In fact, to each complex function $f : I \to \mathbb{C}$, where $I \subseteq \mathbb{R}$, and $f(t) = P(t) + i\,Q(t)$, there corresponds a vector function $F : I \to \mathbb{R}^2$, of the same components P and Q, i.e. $F(t) = (P(t), Q(t))$. According to the above definition, the derivative of f is $f'(t) = P'(t) + i\,Q'(t)$, while $F'(t) = (P'(t), Q'(t))$.

The geometric interpretations of F' in terms of tangent to a plane curve remain valid in the case of f'.

5.3. Remarks. (a) We remind that, differently from the real case, to be the *domain* of a function, *D* shall be an open and connected set. Consequently, z_0 is an interior point of *D*, hence the evaluation of $f^{/}(z_0)$ involves a lot of *directions* and "*ways*" to realize $z \to z_0$ (compare to the derivatives from the *left*, and from the *right*, of a function depending on one real variable). Whenever we intend to put forward the uniqueness of this limit in spite of its infinitely many reductions to one-directional limits, we may use terms of more historical connotation, derived from the French *monogène* (also met in Romanian). However, to get the exact meaning of the limit in the above definition, we have to recall its detailed formulation in terms of neighborhoods, ε and δ, etc. (see §§ I.4, III.2, etc.).

(b) The connection between derivability and differentiability of a complex function is similar to that of real functions. In fact, according to the general definition of differentiability (see § IV 2), a complex function $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{C}$, is *differentiable* at a point $z_0 \in D$ if there exists a linear (and continuous) function $L_{z_0} : \mathbb{C} \to \mathbb{C}$ such that

$$\lim_{h\to 0} \frac{f(z_0+h) - f(z_0) - L_{z_0}(h)}{h} = 0.$$

Similarly to the case of a real function, the general form of such a linear function is $L_{z_0}(h) = c \cdot h$, for some $c \in \mathbb{C}$. Consequently, *f* is differentiable at z_0 if and only if it is derivable at this point, and $c = f^{/}(z_0)$. For historic reasons we may note $L_{z_0} = df_{z_0}$, $df_{z_0}(h) = f^{/}(z_0)h$, $df = f^{/}dz$, etc. as for real functions of a real variable.

(c) The applications of the differential to the approximation theory are also similar to the real case. More exactly, we obtain an approximate value of $f(z_0+h)$ if we write the differentiability in the form

$$f(z_0+h) \cong f(z_0) + f^{/}(z_0) \cdot h.$$

5.4. Geometric interpretation. If the function $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{C}$, is *derivable* at the point $z_0 \in D$, and $f^{/}(z_0) \neq 0$, then locally (i.e. in 'small" neighborhoods of z_0) it realizes a dilation of factor $\left| f^{/}(z_0) \right|$ and a rotation of angle $\arg f^{/}(z_0)$. To justify this interpretation, let us note the increments $h = \Delta z$ and $f(z_0+h) - f(z_0) = \Delta Z$, and write the approximation rule from above in the form $\Delta Z \cong f^{/}(z_0) \cdot \Delta z$. The *local* character of this property means that for every imposed error, we can find a radius $\delta > 0$, such that $\cong$ be accepted as equality whenever $|\Delta z| < \delta$. If so, then $|\Delta Z| = \left| f^{/}(z_0) \right| \cdot |\Delta z|$ and $\arg \Delta Z = \arg f^{/}(z_0) + \arg \Delta z$. It remains to interpret $|\cdot|$ and arg.

The derivative of f has strong connections with the partial derivatives of the real and imaginary parts P = Re f, and Q = Im f, first of all:

5.5. <u>Theorem</u>. If the function $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{C}$, is *derivable* at the point $z_0 = x_0 + iy_0 \in D$, then P = Re f and Q = Im f, are derivable at the point $(x_0, y_0) \in D$, now considered in $\mathbb{R}^2$, and the following relations hold

$$\begin{aligned} \frac{\partial P}{\partial x}(x_0, y_0) &= \frac{\partial Q}{\partial y}(x_0, y_0) \\ \frac{\partial P}{\partial y}(x_0, y_0) &= -\frac{\partial Q}{\partial x}(x_0, y_0) . \end{aligned} \qquad \text{(C-R)}$$

The abbreviation (C-R) comes from *Cauchy* and *Riemann*, who have discovered and used these equations for the first time.

<u>Proof</u>. We may realize the limit from definition 5.1 in two particular ways, namely I. $h \overset{not.}{=} \Delta z = \Delta x$, and II. $h \overset{not.}{=} \Delta z = i\,\Delta y$ (as in the figure below).

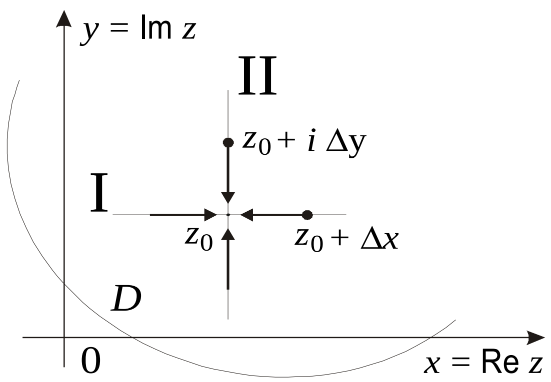


Fig. IV.5.1.

In the first case, the quotient $\dfrac{f(z_0 + \Delta z) - f(z_0)}{\Delta z}$ equals

$$\frac{P(x_0 + \Delta x, y_0) - P(x_0, y_0)}{\Delta x} + i\,\frac{Q(x_0 + \Delta x, y_0) - Q(x_0, y_0)}{\Delta x} ,$$

while in the second one it becomes

$$\frac{P(x_0, y_0 + \Delta y) - P(x_0, y_0)}{i\,\Delta y} + i\,\frac{Q(x_0, y_0 + \Delta y) - Q(x_0, y_0)}{i\,\Delta y} .$$

Taking $h \to 0$, the existence of $f'(z_0)$ implies the existence of the partial derivatives of P and Q at (x_0 , y_0). In addition, the equality of the two expressions of $f'(z_0)$, namely

$$\frac{\partial P}{\partial x}(x_0, y_0) + i\,\frac{\partial Q}{\partial x}(x_0, y_0) = \frac{\partial Q}{\partial y}(x_0, y_0) - i\,\frac{\partial P}{\partial y}(x_0, y_0)$$

proves the Cauchy-Riemann relations. ◇

5.6. <u>Corollary</u>. If the function $f = P + i\,Q : D \to \mathbb{C}$, is *derivable* at the point $z_0 = x_0 + iy_0 \in D \subseteq \mathbb{C}$, then its derivative is calculable by the formulas

$$f'(z_0) = \frac{\partial P}{\partial x}(x_0, y_0) + i\,\frac{\partial Q}{\partial x}(x_0, y_0) = \frac{\partial Q}{\partial y}(x_0, y_0) - i\,\frac{\partial P}{\partial y}(x_0, y_0) =$$

$$= \frac{\partial Q}{\partial y}(x_0, y_0) + i\,\frac{\partial Q}{\partial x}(x_0, y_0) = \frac{\partial P}{\partial x}(x_0, y_0) - i\,\frac{\partial P}{\partial y}(x_0, y_0).$$

These formulas appear in the proof of the above theorem.

Simple examples show that the C-R conditions do not imply derivability:

5.7. <u>Example</u>. Let us define $f : \mathbb{C} \to \mathbb{C}$ by $f(z) = (1 + i)\,\varphi(z)$, where

$$\varphi(z) = \begin{cases} 1 & if\ \mathrm{Re}z \cdot \mathrm{Im}z \neq 0 \\ 0 & if\ \mathrm{Re}z \cdot \mathrm{Im}z = 0\,. \end{cases}$$

The Cauchy-Riemann conditions hold at $z_0 = 0$ because the real and imaginary parts of f constantly vanish on the axes. However, if we take the increments along different directions of equations $\Delta y = m\,\Delta x$, then we see that f is not derivable at the origin.

The following theorem gives an answer to the question "what should we add to the Cauchy-Riemann conditions to assure derivability?"

5.8. <u>Theorem</u>. Let us consider a function $f = P + i\,Q : D \to \mathbb{C}$, and a point $z_0 = x_0 + iy_0 \in D \subseteq \mathbb{C}$. If P and Q are differentiable at (x_0, y_0) and satisfy the Cauchy-Riemann conditions, then f is derivable at z_0.

<u>Proof</u>. We note $h = (x - x_0, y - y_0)$, and we express the differentiability of P and Q by $A(h) \underset{h\to 0}{\to} 0$ and $B(h) \underset{h\to 0}{\to} 0$, where

$$A(h) = \frac{\left| P(x,y) - P(x_0,y_0) - \left[\frac{\partial P}{\partial x}(x_0,y_0)(x - x_0) + \frac{\partial P}{\partial y}(x_0,y_0)(y - y_0) \right] \right|}{\|h\|},$$

$$B(h) = \frac{\left| Q(x,y) - Q(x_0,y_0) - \left[\frac{\partial Q}{\partial x}(x_0,y_0)(x - x_0) + \frac{\partial Q}{\partial y}(x_0,y_0)(y - y_0) \right] \right|}{\|h\|}.$$

Using the C-R conditions, we may write A and B by $\frac{\partial P}{\partial x}$ and $\frac{\partial Q}{\partial x}$ only. Because $\|h\| = |z - z_0|$, we have

$$\left| \frac{f(z) - f(z_0) - \left[\frac{\partial P}{\partial x}(x_0,y_0) + i\,\frac{\partial Q}{\partial x}(x_0,y_0) \right] \cdot [z - z_0]}{z - z_0} \right| \le A(h) + B(h).$$

Consequently, f is differentiable at (x_0, y_0). ◇

5.9. Remarks. (*i*) We may replace the differentiability of P and Q in the above theorem by harder conditions, e.g. by the continuity of their partial derivatives (in accordance to theorem IV.3.5).
(*ii*) Conversely to theorem 5.8, the differentiability of P and Q follows from the derivability of f . In fact, both $A(h)$ and $B(h)$ are less than

$$\left| \frac{f(z) - f(z_0) - \left[\frac{\partial P}{\partial x}(x_0, y_0) + i \frac{\partial Q}{\partial x}(x_0, y_0) \right] \cdot [z - z_0]}{z - z_0} \right| .$$

In other terms, the derivability of f is equivalent to the differentiability of P and Q, plus the Cauchy-Riemann conditions.
(*iii*) The form of the Cauchy-Riemann conditions, theorem 5.8, and other results from above, is determined by the use of Cartesian coordinates in the definition and target planes. More exactly, the correspondence $Z = f(z)$ was meant as $z = x + iy \xrightarrow{f} Z = X + iY$, where $X = P(x, y)$ and $Y = Q(x, y)$. In practice we sometimes meet representations of z and Z in other coordinates, especially polar (see problem 5 at the end of the section).

So far we have studied derivability at a single point. Similarly to the real analysis, this local property can be extended to a global one, which refers to functions derivable at each point of a domain. The specific notion is:

5.10. Definition. We say that function $f : D \to \mathbb{C}$, where $D \subseteq \mathbb{C}$, is *globally derivable* on D (or simply *derivable* on D) iff it is 1:1 (i.e. univalent) and derivable at each point of D. If so, we note the derivative of f on D by f'.

5.11. Remark. There are plenty of terms and variants of presenting the global derivability. For example, the 1:1 condition is sometimes omitted, but tacitly included in the hypothesis that the target space is $\mathbb{C}$. This is the case of the functions $\sqrt[n]{\ }$, *Ln*, *Arcsin*, etc., which are not globally derivable because they a multivalent, i.e. they take values in $\mathcal{P}(\mathbb{C})$. Some authors (frequently including Romanian) use French terms, e.g. "holomorphic" for global derivability, "meromorphic" for a quotient of "holomorphic" functions, "entire" for functions derivable on the whole $\mathbb{C}$, etc.

The analytic method of defining functions (see § II.4, etc.) turns out to be very advantageous in the construction of globally derivable functions. First of all we need information about the convergence of the derived series. If we derive term by term in a power series, then we obtain another power series, hence the problem is to correlate the two radiuses of convergence:

5.12. Lemma. If we derive term by term in a power series $\sum a_n z^n$, then the derived series, i.e. $\sum n\, a_n z^{n-1}$, has the same radius of convergence.

Proof. The essential case is $R = 1/\omega$, where $\omega \overset{not.}{=} \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} \in (0,\infty)$. We have to show that $\omega^{/} \overset{not.}{=} \overline{\lim_{n\to\infty}} \sqrt[n]{n|a_n|}$ there exists, it belongs to $(0,\infty)$ too, and $R = R^{/}$, where $R^{/} = 1/\omega^{/}$. We recall that the superior limit means:

(I) $\underset{\varepsilon>0}{\forall}\ \underset{n_0\in\mathbb{N}}{\exists}$ *such that* $[(\forall) n > n_0 \Rightarrow \sqrt[n]{|a_n|} < \omega + \varepsilon]$, and

(II) $\underset{\varepsilon>0}{\forall}\ \underset{m\in\mathbb{N}}{\exists}$ *such that* $\sqrt[m]{|a_m|} > \omega - \varepsilon$.

On the other hand, we know that $\lim_{n\to\infty} \sqrt[n]{n} = 1$, hence

(I$^{/}$) $\underset{\varepsilon>0}{\forall}\ \underset{n_1\in\mathbb{N}}{\exists}$ *such that* $[(\forall) n > n_1 \Rightarrow \sqrt[n]{n} < 1+\varepsilon]$, and

(II$^{/}$) we have $\sqrt[m]{m} > 1$ at each $m \in \mathbb{N}\setminus\{0,1\}$.

Because the above inequalities are essential for small ε, we may restrain $\varepsilon \in (0,1)$, and fix a number k such that $\omega + 1 + \varepsilon < \omega + 2 < k$. If such an ε is given, then we find n_0 from (I), and n_1 from (I$^{/}$), which correspond to ε/k, and we note $n^* = \max\{n_0, n_1\}$. From (I) and (I$^{/}$) we deduce that

(I*) $\underset{\varepsilon>0}{\forall}\ \underset{n^*\in\mathbb{N}}{\exists}$ *such that* $[(\forall) n > n^* \Rightarrow \sqrt[n]{n|a_n|} < (\omega + \frac{\varepsilon}{k})(1 + \frac{\varepsilon}{k}) < \omega + \varepsilon]$.

Similarly, multiplying the inequalities from (II) and (II$^{//}$), we obtain

(II*) $\underset{\varepsilon>0}{\forall}\ \underset{m\in\mathbb{N}}{\exists}$ *such that* $\sqrt[m]{m|a_m|} > \omega - \varepsilon$.

The conditions (I*) and (II*) show that $\omega^{/}$ exists, and $\omega^{/} = \omega$. ◇

5.13. <u>Theorem</u>. The sum of a power series is globally derivable on the disk of convergence, and its derivative is obtained by deriving each term.
Proof. The claimed property is qualitative hence it does not depend on the center of the power series. To simplify the formulas, we suppose that $z_0 = 0$. Since the case $R = 0$ is trivial, we take $R > 0$ (see the figure below).

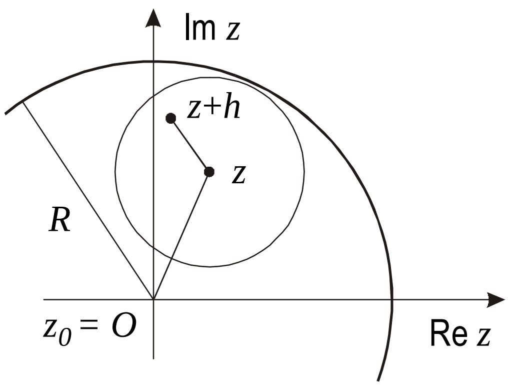


Fig. IV.5.2.

Consequently, our object is the function $f : D \to \mathbb{C}$, where

$$D = S(0, R) = \{z \in \mathbb{C} : |z| < R\}$$

is the disk of convergence of a power series $\sum a_n z^n$, and $f(z) = \sum_{n=0}^{\infty} a_n z^n$.

Because each limit is unique, the values $f(z)$ are uniquely determined, i.e. function f is univalent. It remains to prove the derivability on D.

If we fix $z \in D$, and take $h \neq 0$ such that $z + h \in D$ too, then we may write

$$\frac{f(z+h) - f(z)}{h} = \frac{1}{h} \sum_{n=0}^{\infty} a_n \left[(z+h)^n - z^n \right] = \sum_{n=0}^{\infty} a_n g_n(h),$$

where $g_n(h) \overset{not.}{=} (z+h)^{n-1} + (z+h)^{n-2} z + ... + (z+h) z^{n-2} + z^{n-1}$. It is easy to see that the function series $\sum a_n g_n$ fulfils the hypotheses of a theorem similar to II.3.13, formulated for complex functions. More exactly, g_n are polynomials, hence continuous functions, the convergence of the series is uniform in a neighborhood of z. Consequently, the limit $h \to 0$ preserves the above equality, where $\lim_{h \to 0} g_n(h) = n z^{n-1}$ is immediate; the existence of the limit of the series $\sum a_n g_n$ shows that f is derivable at z, and

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} \overset{def.}{=} f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1}.$$

The convergence of the derived series follows from Lemma 5.12. ◇

5.14. <u>Analytic prolongation</u>. An important problem appears relative to the domain of definition of a function depending on the used method. If we define it by a power series, then the domain is a disk, but the general form of the domain of definition is not circular. For example, the function

$$f(z) = \frac{1}{1-z}$$

is defined on $D = \mathbb{C} \setminus \{1\}$, while the analytic definition

$$f(z) = 1 + z + z^2 + ... + z^n + ...$$

makes sense only in the disk $D(0, 1) = \{z \in \mathbb{C} : |z| < 1\}$, where the geometric series is convergent. For fairness we mention two extreme cases when this difference disappears (see also problem 6, and other examples), namely:

1) $R = \infty$, since we have accepted to interpret $\mathbb{C}$ as a disk, and
2) $R < \infty$, but f is not definable outside the disk $D(z_0, R)$.

It is easy to see that some developments of the same function around other points, different from z_0, may overpass the initial disk of convergence (see Fig. IV.5.3). In the example from above, if we choose $z_1 = i/2$, then the development around this point will be a new power series, namely

$$\frac{1}{1-z}=\frac{1}{1-\frac{i}{2}-(z-\frac{i}{2})}=\frac{1}{1-\frac{i}{2}}\cdot\frac{1}{1-\frac{z-\frac{i}{2}}{1-\frac{i}{2}}}=\frac{1}{1-\frac{i}{2}}\sum_{n=0}^{\infty}\left(\frac{z-\frac{i}{2}}{1-\frac{i}{2}}\right)^n .$$

This is a geometric series, which is convergent iff $\left|z-\frac{i}{2}\right|<\left|1-\frac{i}{2}\right|=\frac{\sqrt{5}}{2}$, and has the same values as the former one in the common part of the domains.

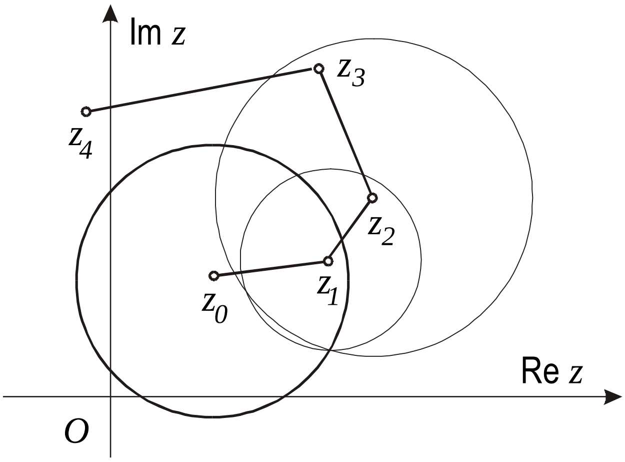


Fig. IV.5.3.

In general, we may repeat this process of "immediate" prolongation along various sequences of points $\{z_0, z_1, z_2, \ldots\}$ to extend the initial function on larger and larger domains. Without going into details (see specialized treatises like [CG], [G-S], [SS], etc.), we mention some of the main terms:

5.15. Definition. By *element of analytic function* we understand a function f_0, whose values are the sum of a power series, i.e. $f_0(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n$.

Its domain of definition, D_0, is the disk of convergence of this series, i.e.

$$D_0=S(z_0,R)=\{z\in\mathbb{C}:|z-z_0|<R\}.$$

The power series of f_0 around a point $z_1\in D_0\setminus\{z_0\}$ is called (*immediate*) *prolongation* of f_0. The greatest domain D, to which f_0 can be extended by all possible repeated prolongations is named *domain of analyticity*. The resulting function $f: D\to\mathbb{C}$, is called *analytic function* generated by the element of function f_0.

Each point of D is said to be *ordinary* (or *regular*) point of f, while the points of the frontier of D are named *singular*.

5.16. Remark. As a result of iterated prolongations, we naturally recover zones where the function has been previously defined. There is no guaranty that the new power series takes the same values as the previous ones on these zones. In practice, we always have to identify the case and distinguish between *univalent* and *multivalent* functions.

One of the most rigorous ways to avoid the multivalence consists in considering the domain of analyticity as a manifold with multiple *leafs* (or *branches*). For example, the domain of the function $\sqrt[n]{\ }$ has n leafs, such that each turn about the origin (which is the single common point of these branches) leads us to the "next" leaf. Consequently, whenever we take a point $z \in \mathbb{C}$, we have to specify the index k ($k = 0, 1, \ldots, n-1$) of the leaf for which z belongs to. The corresponding value is the well known

$$\left(\sqrt[n]{z}\right)_{(k)} = \sqrt[n]{|z|}\left[\cos\frac{\arg z + 2k\pi}{n} + i\sin\frac{\arg z + 2k\pi}{n}\right].$$

Beside this complicated study of such manifolds (see [SL], [SS], [BN], etc.), we can solve plenty of practical problems on another way, namely to transform the multivalent function, which takes values in $\mathscr{P}(\mathbb{C})$, into more univalent functions, which are defined on customary domains in $\mathbb{C}$, and take unique values. This method is based on the so-called "cuts":

5.17. <u>Definition</u>. Let $f: D \to \mathscr{P}(\mathbb{C})$, where $D \subseteq \mathbb{C}$, be a multivalent function.

If the point $\varsigma \in D \cup \mathrm{Fr}\, D$ has the property that f is multivalent on arbitrary neighborhood $V \in \mathscr{V}(\varsigma)$, then it is called *critical* point (or *multivalent singularity*) of f.

Each restriction $f\big|_A : A \to \mathbb{C}$, where $A \subseteq D$, which is continuous on A (and self-evidently univalent, since it ranges in $\mathbb{C}$), is named *univalent branch* (or *univalent determination*) of f. If A is obtained by removing a curve $\mathcal{C}$ from D, i.e. $A = D \setminus \mathcal{C}$, then $\mathcal{C}$ is called *cut* of D.

5.18. <u>Examples</u>. The most frequent multivalent functions are: Arg, $\sqrt[n]{\ }$, *Ln*, the complex power, and the inverse trigonometric functions. All of them have the origin of the complex plane as a critical point. The half-line $\mathbb{R}_-$ is customarily used as a cut of $D = \mathbb{C} \setminus \{0\}$. The effect of this cut application is the elimination of the complete turns about the origin, which avoids the possibility of passing from one branch to another.

Of course, the combined functions, which involve some of the simple multivalent examples from above, have more complicated sets of critical points. For example, the function $f: \mathbb{C} \setminus \{1, -1\} \to \mathbb{C}$, of values

$$f(z) = Ln\frac{z-1}{z+1}$$

has the critical points $z_1 = 1$ and $z_2 = -1$. We may cut along two half-lines

$$\mathcal{C}_1 = \{z = x \in \mathbb{R} : x \le -1\} \cup \{z = x \in \mathbb{R} : x \ge 1\}.$$

The analysis of f shows that another possible cut is $\mathcal{C}_2 = [-1, 1]$, as well as many other curves of endpoints -1 and $+1$ (see problem 7 at the end).

5.19. <u>Classification of the singular points</u>. During the first stage of the analysis, we have to establish whether the singular point is

- *Isolated* (in the set of all singular points), or
- *Non-isolated*.

Further on, the isolated singular points can be

- *Multivalent* (i.e. *critical*), or
- *Univalent*.

Among univalent isolated singular points we distinguish

- *Poles*, and
- *Essential singularities*

More exactly, z_0 is a *pole* of f if there is some $p \in \mathbb{N}^*$, for which there exists $\lim_{z \to z_0} (z - z_0)^p f(z)$, and it is finite. The smallest natural number with this property is called *order* of the pole z_0. In the contrary case, when there is no such a number $p \in \mathbb{N}^*$, we say that z_0 is an *essential* singular point.

A singular point can be non-isolated in the following cases:

- It is an *accumulation* point of a sequence of singular points
- It belongs to a curve consisting of singular points
- It is adherent to a set of singular points, which has a positive area.

For example, $z_0 = 0$ is univalent isolated singular point for the function $f(z) = \frac{1}{z}$, and critical point (i.e. multivalent isolated) for $g(z) = \sqrt{z}$. To be more specific, $z_0 = 0$ is a pole of order p of the functions

$$\varphi_1(z) = \frac{1}{z^p},\ \varphi_2(z) = \frac{1}{\sin^p z}, \text{ etc.}$$

and essential singularity for the functions

$$\psi_1(z) = \exp(1/z),\ \psi_2(z) = \sin(1/z), \text{ etc.}$$

The same $z_0 = 0$ is non-isolated singular point for the function

$$h(z) = \frac{1}{\sin \frac{1}{z}},$$

since it is the limit of the sequence $\left(\frac{1}{n\pi} \right)_{n \in \mathbb{N}^*}$, which consists of (isolated) singular points. A line of singular points appears when no prolongation is available outside the disk of convergence, e.g. $\sum_{n=0}^{\infty} z^{n!}$ in problem 6. The examples of domains with positive area, which consist of singular points, are much more complicated, and we will skip this topic here; however, we mention that a remarkable contribution in this field is due to the Romanian mathematician Dumitru Pompeiu (about 1905, see [SS], [CG], etc).

The above classification includes the point at infinity.

Another major topic in the theory of the derivable functions concerns the properties of their real and imaginary parts. We obtain simpler formulation of the main theorem if we ask D to be particularly connected:

5.20. <u>Definition</u>. We say that the domain $D \subseteq \mathbb{R}^2$ is *connected by segments* if there exists a point $M_0 = (x_0, y_0) \in D$ such that

$$[(x_0, y_0), (x, y_0)] \cup [(x, y_0), (x, y)] \subseteq D$$

holds for all $M = (x, y) \in D$, where [. , .] denotes a line segment.

Alternatively, we may refer to the *dual* broken line

$$[(x_0, y_0), (x_0, y)] \cup [(x_0, y), (x, y)],$$

or to concatenations of such curves.

A comparison to the line integral is advisable (see later § VI.3, where we replace the broken lines by a single segment $[M_0, M]$, and we say that D is a *star-like* domain).

5.21. <u>Theorem</u>. Let $f = P + i\,Q : D \to \mathbb{C}$ be a function, for which the real and imaginary parts have continuous partial derivatives of the second order, i.e. $P, Q \in C^2_{\mathbb{R}}(D)$. If $f = P + i\,Q : D \to \mathbb{C}$ is *derivable* on $D \subseteq \mathbb{C}$, then P and Q are harmonic functions on this domain, i.e. they fulfill the Laplace equation $\Delta P = 0$, $\Delta Q = 0$, at each point $(x, y) \in D$.

Conversely, if the function $P : D \to \mathbb{R}$ is harmonic on the domain $D \subseteq \mathbb{R}^2$, which is *connected by segments*, then there exists a function $f : D \to \mathbb{C}$, derivable in the complex sense, such that $P = \operatorname{Re} f$.

A similar property holds for Q .

<u>Proof</u>. If f is derivable, then according to Theorem 5.5, P and Q satisfy the Cauchy-Riemann conditions. Since $P, Q \in C^2_{\mathbb{R}}(D)$, we may derive one more time in these relations, and we obtain

$$\frac{\partial^2 P}{\partial x^2}(x, y) = \frac{\partial^2 Q}{\partial x \partial y}(x, y)$$

$$\frac{\partial^2 P}{\partial y^2}(x, y) = -\frac{\partial^2 Q}{\partial y \partial x}(x, y)$$

at each $(x, y) \in D$. The continuity of the mixed derivatives of Q assures their equality (see the Schwarz' Theorem IV.3.10), hence

$$\Delta P \overset{not.}{=} \frac{\partial^2 P}{\partial x^2} + \frac{\partial^2 P}{\partial y^2} = 0 .$$

Appropriate derivations in the C-R conditions lead to $\Delta Q = 0$.

Conversely, let us suppose that a harmonic function $P \in C^2_{\mathbb{R}}(D)$ is given, and we have to point out a derivable function f, for which $P = \operatorname{Re} f$. What we need is $Q = \operatorname{Im} f$, hence we start the proof by a constructive step, in which we claim that the function $Q : D \to \mathbb{R}$, of values

$$Q(x,y) = -\int_{x_0}^{x} \frac{\partial P}{\partial y}(t, y_0)\,dt + \int_{y_0}^{y} \frac{\partial P}{\partial x}(x, s)\,ds, \qquad (*)$$

fulfils the requirements. First of all, Q is correctly constructed since D is connected by segments, and P has continuous partial derivatives. The main part of the proof refers to the Cauchy-Riemann conditions, so we evaluate

$$\frac{\partial Q}{\partial x}(x,y) = -\frac{\partial P}{\partial y}(x,y_0) + \int_{y_0}^{y} \frac{\partial^2 P}{\partial x^2}(x,s)\,ds =$$

$$= -\frac{\partial P}{\partial y}(x,y_0) - \int_{y_0}^{y} \frac{\partial^2 P}{\partial s^2}(x,s)\,ds = -\frac{\partial P}{\partial y}(x,y_0) - \frac{\partial P}{\partial s}(x,s)\Big|_{s=y_0}^{s=y} =$$

$$= -\frac{\partial P}{\partial y}(x,y).$$

The other derivative in (*) immediately gives

$$\frac{\partial Q}{\partial y}(x,y) = \frac{\partial P}{\partial x}(x,y).$$

According to theorem IV.5.8, the function $f = P + i\,Q$ is derivable. ◇

5.22. <u>Remarks</u>. (*i*) Slight modifications of the above proof are necessary if D is connected by other types of segments, or Q is the given function, and P is the asked one. The key is a good adaptation of the formula (*), which is explained in Fig.IV.5.4 from below.

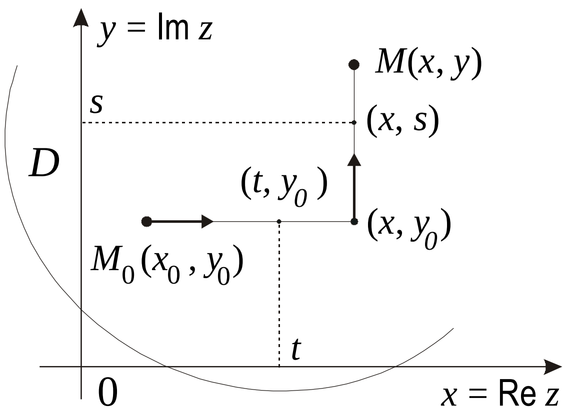


Fig. IV.5.4.

In addition, we may formally obtain formula (*) by integrating dQ, i.e.

$$Q(x,y) = \int_{(x_0,y_0)}^{(x,y)} dQ = \int_{(x_0,y_0)}^{(x,y)} \frac{\partial Q}{\partial x}dx + \frac{\partial Q}{\partial y}dy =$$

$$= -\int_{(x_0,y_0)}^{(x,y)} \frac{\partial P}{\partial y}dx + \frac{\partial P}{\partial x}dy,$$

where the integral is realized along the broken line, and the C-R conditions are supposed to be valid.

(*ii*) Formula (*) is essential in practical problems. It allows finding Q up to a constant. More exactly, the concrete computation in (*) produces three types of terms, namely:

- Terms in (x, y), which form $Q(x, y)$;
- Terms in (x_0, y) and (x, y_0), which must disappear;
- Terms in (x_0, y_0), which form $Q(x_0, y_0)$ = constant.

We stress on the fact that the terms in (x_0, y) and (x, y_0) <u>must</u> disappear, and the final result in (*) takes the form

$$-\int_{x_0}^{x} \frac{\partial P}{\partial y}(t, y_0)\, dt + \int_{y_0}^{y} \frac{\partial P}{\partial x}(x, s)\, ds = Q(x, y) - Q(x_0, y_0).$$

In other terms, Q (and implicitly f) are determined up to a constant, which is sometimes established using a condition of the form $f(z_0) = Z_0$.

(*iii*) In more complicated problems, instead of P (respectively Q) we have a relation satisfied by these functions. In this case it is very useful to know many particular functions (see problems 8, 9, etc.).

(*iv*) In a typical problem, e.g. P is given and we find Q, we obtain the solution, i.e. function f , in the form $P(x, y) + i\, Q(x, y)$. Whenever we have to write the answer as $f(z)$, the following formal rule is recommended

$$f(z) = P(z, 0) + i\, Q(z, 0).$$

(*v*) Formula (*) and the other results concerning the properties of P and Q in a derivable function $f = P + i\, Q$ strongly depend on the chosen type of coordinates, namely Cartesian. Whenever a practical problem asks, we may reconsider the same topic in other coordinates, and put the main ideas from above in an appropriate formalism.

The geometric interpretation of the derivability at a point can be naturally extended to the global derivability, in terms of particular transformations.

5.23. <u>Definition</u>. We say that the function $T : D \to \mathbb{R}^2$, where $D \subseteq \mathbb{R}^2$, is a *conformal* transformation of D if it preserves the angles.

More exactly, the notion of angle refers to smooth curves, respectively to the tangent vectors to such curves. To transform smooth curves into smooth ones, we tacitly use the hypothesis $T \in C^1_{\mathbb{R}^2}(D)$. The specific property is to leave the size of the angle between corresponding curves unchanged.

We recall that the complex functions represent plane transformations.

5.24. <u>Examples</u>. We represent the elementary geometric transformations of the plane by the following complex functions:

- $Z = z + b$ represents a translation of vector b, where $b \in \mathbb{C}$;
- $Z = z\, e^{i\theta}$ is a rotation of angle θ, where $\theta \in [0, 2\pi)$;
- $Z = r\, z$ is a dilation / contraction of center 0 and factor $r > 0$;
- $Z = \bar{z}$ means symmetry relative to the real axis ;
- $Z = 1/\bar{z}$ is an inversion relative to the unit circle.

By composing such transformations we obtain the *linear* complex function

$$Z = a\,z + b,$$

where $a, b \in \mathbb{C}$, $a \neq 0$, and the *circular* (or *homographic*) function

$$Z = \frac{az+b}{cz+d} \ ,$$

where $ad \neq bc$, and $z \neq -d/c$ if $c \neq 0$.

The derivability of f makes it conformal transformation in the plane:

5.25. Theorem. Let $f: D \to \mathbb{C}$ be a derivable function on the domain $D \subseteq \mathbb{C}$. If $f'(z) \neq 0$ at each $z \in D$, then f realizes a conformal transformation of this domain.

Proof. Let γ_1 and γ_2 be two smooth curves in D, which are concurrent in the point $z = x + iy$. The angle α between these curves is defined by

$$\cos \alpha = \langle \vec{\mathbf{t}}_1, \vec{\mathbf{t}}_2 \rangle ,$$

where $\vec{\mathbf{t}}_1$ and $\vec{\mathbf{t}}_2$ are the unit tangent vectors to the curves γ_1 and γ_2.

Since f is a derivable function, the images $\Gamma_1 = f(\gamma_1)$ and $\Gamma_2 = f(\gamma_2)$ are also smooth curves (see Fig.IV.5.5.), which form the angle ω.

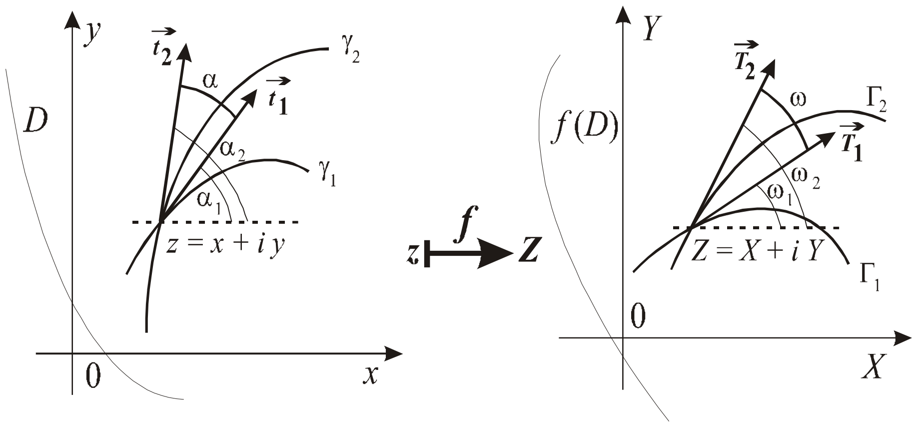


Fig.IV.5.5.

The assertion of the theorem reduces to the equality $\alpha = \omega$, for arbitrary γ_1 and γ_2 in D. The proof is based on the relation $\Delta Z \cong f'(z_0) \cdot \Delta z$, which was already used to obtain the geometric interpretation 5.4 of the local action of a derivable function. More exactly, we recall that

$$\arg \Delta Z \cong \arg f'(z) + \arg \Delta z ,$$

where "$\cong$" can be successfully replaced by " = " in small neighborhoods of the point z. If we take the increment Δz along γ_1, then $\arg \Delta z \cong \alpha_1$, and for the corresponding image Γ_1, $\arg \Delta Z \cong \omega_1$. Consequently, we obtain

$$\omega_1 \cong \arg f'(z) + \alpha_1 ,$$

with equality for small increments of z.

Similarly, along γ_2 we have

$$\omega_2 \cong \arg f^{/}(z) + \alpha_2 .$$

The subtraction of these relations leads to the desired equality, because $\alpha = \alpha_2 - \alpha_1$, and $\omega = \omega_2 - \omega_1$. ◇

5.26. Remarks. Besides this theorem, in practice we frequently have to use other results concerning the conformal mappings. Without going into the details of the proof, we mention several theorem of this type:

- (The converse of Theorem 5.25) Each conformal mapping of a plane domain is realized by a derivable complex function f, or by $\bar{f}$.
- (The principle of correspondent frontiers) Let the domain $D \subseteq \mathbb{C}$ be simply connected, i.e. each closed curve from D has its interior in D, and let $\gamma = \mathrm{Fr}\, D$ be a piecewise smooth curve. Let also $f: D \to \mathbb{C}$ be a derivable function on D, which is continuous on $D \cup \gamma = \overline{D}$. If γ and $\Gamma = f(\gamma)$ are traced in the same sense, then f realizes a conformal correspondence between D and the interior of Γ, noted (Γ).
- (Riemann-Carathéodory theorem) Every simply connected domain, whose frontier has at least two points, allows a conformal mapping on the unit disk.

5.27. Applications. The derivable complex functions are frequently used in Mathematical Physics (see [KE], [HD], etc.). For example, since $P = \mathrm{Re}\, f$ is a *harmonic* function, it is appropriate to describe a potential, e.g. electrostatic. The *complex potential* $f = P + i\, Q$ is often preferred, because of its technical and theoretical advantages. In particular, Q is physically meaningful too, as a consequence of the orthogonality of the equipotential lines P = constant, and the lines of force Q = constant.

Other applications concern the heat conduction and the fluid flow. For example, we may obtain wing profiles from disks if we use conformal transformation like the Jukowski's function

$$Z = z + a^2 / z .$$

The advantage is that the streamlines around a disk are very simple, and the conformal transformation allows us to find out the streamlines around other profiles.

PROBLEMS § IV.5

1. Formulate and prove the rules of deriving the sum, product, quotient, and composition of complex functions of a complex variable.
Hint. The rules for real functions of a real variable remain valid.

2. Show that the functions *exp*, *Ln*, *sin*, *cos*, *sinh*, and *cosh* are derivable at each point of their domain of definition, and find the derivatives.
Hint. Identify the real and imaginary parts, and study the continuity of their partial derivatives (Theorem 5.8). Use either one formula of Corollary 5.6, and express the derivatives as in the real case.

3. Show that the function $f : \mathbb{C} \to \mathbb{C}$, of values

$$f(z) = \begin{cases} \frac{\mathrm{Re}\, z \ \mathrm{Im}\, z}{|z|}(1+i) & if \quad z \neq 0 \\ 0 & if \quad z = 0\,, \end{cases}$$

fulfils the C-R conditions at $z_0 = 0$, but it is not derivable at this point.
Hint. Because $P = \mathrm{Re}\, f$ and $Q = \mathrm{Im}\, f$ have the values

$$P(x,y) = Q(x,y) = \begin{cases} \frac{xy}{\sqrt{x^2+y^2}} & if \quad (x.y) \neq (0,0) \\ 0 & if \quad (x,y) = (0,0) \end{cases}$$

it is easy to show that

$$\frac{\partial P}{\partial x}(0,0) = \frac{\partial P}{\partial y}(0,0) = \frac{\partial Q}{\partial x}(0,0) = \frac{\partial Q}{\partial y}(0,0) = 0\,.$$

Study the derivatives on directions of equations $y = m\,x$.

4. Let $f = P + i\,Q : D \to \mathbb{C}$ be a function for which P and Q are partially derivable at a point $(x_0 , y_0) \in D$. Show that the C-R conditions hold if and only if $\frac{\partial f}{\partial \bar{z}}(x_0, y_0) = 0$. In particular, study the function $f(z) = \bar{z}$.
Hint. f depends on $\bar{z}$ via x and y, according to the formulas

$$x = \tfrac{1}{2}(z+\bar{z}) \ \ and \ \ y = \tfrac{1}{2i}(z-\bar{z})\ .$$

In the particular case, f is nowhere derivable.

5. Write the C-R conditions for a function $Z = f(z)$ in the cases:

1. z is expressed in polar coordinates, and Z in Cartesian coordinates;
2. z is expressed in Cartesian coordinates, and Z in polar coordinates;
3. Both z and Z are expressed in polar coordinates.

Hint. Change the coordinates by deriving composite functions in the C-R conditions in Cartesian coordinates.

1. If $z = r\,e^{it}$ and $Z = X + iY$, where $X = \mathfrak{P}(r,t)$ and $Y = \mathfrak{Q}(r,t)$, then

$$\frac{\partial \mathfrak{P}}{\partial r} = \frac{1}{r}\,\frac{\partial \mathfrak{Q}}{\partial t} \text{ and } \frac{\partial \mathfrak{Q}}{\partial r} = -\frac{1}{r}\,\frac{\partial \mathfrak{P}}{\partial t}\ ;$$

2. If $z = x + iy$ and $Z = R(x,y)\,e^{iT(x,y)}$, then the C-R conditions become

$$\frac{\partial R}{\partial x} = R\,\frac{\partial T}{\partial y} \text{ and } \frac{\partial R}{\partial y} = -\frac{\partial T}{\partial x}\ ;$$

3. If $z = r\,e^{it}$ and $Z = \mathfrak{R}(r,t)\,e^{i\mathfrak{T}(r,t)}$, then the C-R conditions take the form

$$\frac{\partial \mathfrak{R}}{\partial r} = \frac{\mathfrak{R}}{r}\,\frac{\partial \mathfrak{T}}{\partial t} \text{ and } \frac{\partial \mathfrak{R}}{\partial t} = -\mathfrak{R}r\,\frac{\partial \mathfrak{T}}{\partial r}\ .$$

6. Find the domain of definition of the function f, of values

$$f(z) = \sum_{n=0}^{\infty} z^{n!},$$

and show that no analytic prolongation is possible.

Hint. The power series has $R = 1$, hence the domain of f is the disk $D(0, 1)$. The impossibility to prolong f outside this disk follows from the property of this function to take high values when we get close to the circumference. In fact, if we take $z = r(\cos\alpha + i\sin\alpha)$, where $\alpha = \frac{p}{q}2\pi$, $p, q \in \mathbb{N}^*$, and $p < q$, then we may use q to decompose the sum and to obtain

$$|f(z)| = \left|\sum_{n=0}^{q-1} z^{n!} + \sum_{n=q}^{\infty} z^{n!}\right| \geq \left|\sum_{n=q}^{\infty} z^{n!}\right| - \left|\sum_{n=0}^{q-1} z^{n!}\right|.$$

Since $r < 1$, we have $\left|z^{n!}\right| = r^{n!} < 1$ for all $n < q$, hence $\left|\sum_{n=0}^{q-1} z^{n!}\right| < q$. On the other hand, for $n \geq q$ we obtain $z^{n!} = r^{n!} \in \mathbb{R}_+$, hence $\left|\sum_{n=q}^{\infty} z^{n!}\right| = \sum_{n=q}^{\infty} r^{n!}$. The last sum takes arbitrarily large values if r is close enough to 1, and so does $|f|$ too, according to the inequality $|f(z)| > \sum_{n=q}^{\infty} r^{n!} - q$. Finally, because

$$\{z = \cos\alpha + i\sin\alpha : \alpha = \tfrac{p}{q}2\pi\,;\, p, q \in \mathbb{N}\,;\, p < q\}$$

is a dense set in the circumference of the unit circle, this property of f also holds for "irrational directions" $\alpha = 2\pi\nu$, with $\nu \in [0,1) \setminus \mathbb{Q}$.

7. Find the critical points and indicate cuts for the functions:

(a) $\sqrt[3]{z+1}$; (b) *Ln* (z^2 +1); (c) $Ln\dfrac{z+i}{z-i}+\sqrt{z^2-1}$; (d) *Ln* (z – 1) + *Arcsin z*

Hint. Identify the real and imaginary parts of these functions, where the index of the branches is visible. For the logarithmic part in (c), we may include [– *i*, *i*] as a part of the cut (compare to Examples IV.5.18).

8. Find the derivable function $f = P + i\,Q$, if we know that:

(i) $P(x, y) = e^x \cdot \cos y$, and $f(0) = 1$;

(ii) $P(x, y) = 3xy^2 - x^3$, and $f(i) = 0$;

(iii) $P(x, y) = \dfrac{x}{x^2+y^2}$;

(iv) $Q(x, y) = e^x \cdot \cos y$, and $f(0) = 1$;

(v) $Q(x, y) = 1 - 3x^2 y + y^3$, and $f(0) = i$;

(vi) $Q(x, y) = \dfrac{-y}{x^2+y^2}$;

(vii) $P^2(x, y) - Q^2(x, y) = \sin x \cosh y$.

Hint. Use Theorem 5.21 and Remark 5. 22. In the case (*vii*), recognize that $\sin x \cosh y = \text{Re}(\sin z)$, and $P^2 - Q^2 = \text{Re}\, f^2$. The resulting function f is multi-valued, hence a cut of the plane is advisable.

9. Find the derivable function $f = P + i\,Q$, for which $f(1) = e$, and

$$P(x,y)\cos y + Q(x,y)\sin y = \frac{x e^x}{x^2+y^2}\ .$$

Hint. Remark that $e^{-z} = e^{-x}(\cos y - i \sin y)$, and

$$e^{-x}[P(x,y)\cos y + Q(x,y)\sin y] = \text{Re}\lfloor e^{-z} f(z)\rfloor\ .$$

Our previous experience, e.g. problem 8 from above, furnishes the relation $\dfrac{x}{x^2+y^2} = \text{Re}\left(\dfrac{1}{z}\right)$, hence $f(z) = \dfrac{1}{z} e^z$.

10. Show that the circular mappings preserve the family $\mathscr{C}$, of straight lines and circles in a plane. What circles are mapped into circles?

Hint. The general equation of a curve $\gamma \in \mathscr{C}$ is

$$A(x^2+y^2) + Bx + Cy + D = 0\,.$$

The elementary transformations contained in a homographic map preserve the form of this equation. In particular, the formulas of an inversion are

$$X = \frac{x}{x^2+y^2}\,,\ Y = \frac{y}{x^2+y^2}\,.$$

11. Show that we can determine a circular transformation by three pairs of correspondent points (in spite of its dependence on four parameters). Using this fact, find homographic transformations for which:

(a) The interior of the unit circle in transformed into $\{Z \in \mathbb{C} : |Z - i| > 2\}$;

(b) The upper half-plane is transformed into the unit circle;

(c) The (open) first quadrant is applied onto $\{Z \in \mathbb{C} : |Z| < 1, \operatorname{Im} Z > 0\}$.

Hint. At least one of the parameters *a*, *b*, *c*, *d* differs from zero, so that we may simplify and determine the remaining three parameters from three independent conditions. In the concrete cases, take three pairs of points on the corresponding frontiers, and respect the orientation (correlate to 5.26, or chose some proof-points in the transformed domains).

12. Find out the images of the circle $C(0, r)$ through the functions:

(a) $Z = \dfrac{z}{\bar{z}}$; (b) $Z = z^2$; (c) $Z = z + \dfrac{a^2}{z}$, $a \in \mathbb{R}_+$ (Jukowski).

Hint. Replace $z = r(\cos\theta + i \sin\theta)$, $\theta \in [0, 2\pi)$. The sought for images are: (a) The unit circle traced twice; (b) The circle of radius r^2, traced twice; (c) Either ellipse or hyperbola, depending on *r* and *a*.

13. Show that the function $Z = \sin z$ realizes a conformal transformation of the domain $D = \{z \in \mathbb{C} : -\pi < \operatorname{Re} z < \pi, \operatorname{Im} z > 0\}$ into the complex plane $\mathbb{C}$, cut along the line segment [–1, 1] and the negative imaginary axis.

Hint. Write $\sin z = \sin x \cosh y + i \sinh y \cos x$, and show that the function *sin* establishes a 1:1 (i.e. bijective) correspondence between *D* and

$$\mathbb{C} \setminus \{[-1, 1] \cup i\mathbb{R}_-\}.$$

Identify three linear parts in the Fr *D*, and find their images through this function.

14. A singular point $z_0 \in D$, of the function $f : D \to \mathbb{C}$, is said to be *apparent* (*illusory* or *eliminable*) if *f* allows a derivable prolongation to this point.

Show that $z_0 = 0$ is an apparent singular point of the functions:

$$\text{(a)}\ \frac{\sin z}{z};\quad \text{(b)}\ \frac{1 - \cos 2z}{2z^2};\quad \text{(c)}\ \frac{z}{e^z - 1}.$$

and specify the Taylor series around z_0 of the corresponding derivable prolongations.

Hint. Use the power series of *sin*, *cos*, and *exp*. In the third case we have to evaluate the coefficients of the quotient series.

Index

D

E

F

G

H

I

J

K

BIBLIOGRAPHY

[AE] Arghiriade Emanoil, *Curs de algebră superioară, Vol. I, II,* EDP, Bucureşti, 1963

[A-M] Aramă Lia, Morozan Teodor, *Culegere de probleme de calcul diferențial şi integral*, Ed. Tehnică, Bucureşti, 1966

[BC] Brezinski C., *Accélération de la Convergence*, Springer-Verlag, Berlin, Heidelberg, New York, 1977, Lecture Notes Math. 584

[BN] Boboc Nicu, *Funcții complexe*, EDP, Bucureşti, 1969

[BN_1] Bourbaki Nicolas, *Eléments de mathématiques,* Livre III, *Topologie Générale*, Ed. Hermann, Paris, 1965

[BN_2] Bourbaki Nicolas, *Eléments de mathématiques,* Livres V-VI, Ed. Hermann, Paris, 1967

[B-S] Bălan Trandafir, Şterbeți Cătălin, *Analiză Complexă – Breviar teoretic şi culegere de probleme*, Ed. MJM, Craiova, 2003

[B-S-T] Beju I., Soós E., Teodorescu P.P., *Tehnici de calcul, Vol. I, II, III,* Ed. Tehnică, Bucureşti, 1976

[BT_1] Bălan Trandafir, *Matematici speciale – Ecuații diferențiale şi funcții speciale*, Reprogr. Univ. Craiova, 1982

[BT_2] Bălan Trandafir, *Capitole de matematici aplicate – Analiză Fourier*, Ed. Universitaria, Craiova, 1998

[BT_3] Bălan Trandafir, *Generalizing the Minkowskian Space-time*, Studii Cercet. Matem. Tom 44, Part I in Nr.2, p. 89-107; Part II in Nr.4, p. 267-284

[BT_4] Bălan Trandafir, *Guide to Mathematical Terms*, Ed. Radical, 1998

[C-A-B] Costinescu O., Amihăesei C., Bârsan T., *Topologie generală – Probleme*, EDP, Bucureşti, 1974

[C-E] Creangă I., Enescu I., *Algebre*, Ed. Tehnică, Bucureşti, 1973

[CG] Călugăreanu Gheorghe, *Elemente de teoria funcțiilor de o variabilă complexă*, EDP, Bucureşti, 1963

[CI] Colojoară Ion, *Analiză matematică*, EDP, Bucureşti, 1983

[C-I] Cerchez M., Iambor I., *Curbe plane*, Ed. Tehnică, Bucureşti, 1989

[CO] Costinescu Olga, *Elemente de topologie generală*, Ed. Tehnică, Bucureşti, 1969

[CR] Cristescu Romulus, *Analiză funcțională*, EDP, Bucureşti, 1970

[C-T] Craiu M., Tănase V., *Enaliză matematică*, EDP, Bucureşti, 1980

[CV] Crăciun C.V., *Contraexemple în analiza matematică*, Ed. Univ. Bucureşti, 1989

[DB] Demidovici Boris, *Problems in Mathematical Analysis*, Mir Publ. Moscow, 1989

[DE] Dobrescu Eugen, *Analiză matematică*, EDP, Bucureşti, 1975

[DJ] Dieudonné Jean, *Foundations of Modern Analysis*, Academic Press, New York, 1960

[FG] Fihtenholț G.M., *Curs de calcul diferențial şi integral, Vol.I, II, III,* Ed. Tehnică, Bucureşti, 1963-1965

[G-O] Gelbaum B.R., Olmsted J.M.H., *Counterexamples in Analysis,* Holden-Day Inc., San Francisco, London, Amsterdam, 1964 (trad. lb. rom., Ed. Ştiințifică, Bucureşti, 1973)

[G-S] Gaşpar Dumitru, Suciu Nicolae, *Analiză complexă*, Ed. Academiei Române, Bucureşti, 1999

[HD] Homentcovschi D., *Funcții complexe cu aplicații în ştiință şi tehnică,* Ed. Tehnică, Bucureşti, 1986

[HP] Halmos Paul, *Finite-dimensional vector spaces,* Springer-Verlag, New-York, Heidelberg, Berlin, 1974

[H-S] Hewitt E., Stromberg K., *Real and Abstract Analysis,* Springer-Verlag, Berlin, 1969

[I-P] Ilyin V.A., Poznyak E., *Fundamentals of Mathematical Analysis, Parts I, II*, Mir Publishers, Moscow, 1982

[KA] Kurosh A., *Higher Algebra,* Mir Publishers, Moscow, 1975

[K-A] Kantorovici L.V., Akilov G.P., *Analiză funcțională,* Ed. Ştiințifică şi Enciclopedică, Bucureşti, 1986

[KE] Kreysig E., Advanced Engineering Mathematics, John Wiley, New York, 1972

[K-F] Kolmogorov A.N., Fomin S.V., *Elements of Function Theory and Functional Analysis,* Nauka, Moscow, 1968

[KJ] Kelley John L., *General Topology*, D. van Nostrand Comp., 1957

[KK] Kuratowski K., *Introducere în teoria mulțimilor şi topologie,* Warszawa, 1955 (traducere din poloneză)

[K-K-M] Krasnov M.L., Kisselev A.I., Makarenko G.I., *Functions of a aomplex variable, Operational Calculus, and Stability Theory – Problems and Exercises*, Mir Publishers, Moscow, 1984

[KO] Konnerth O., *Greşeli tipice în învățarea analizei matematice*, Ed. Dacia, Cluj-Napoca, 1982

[KP] Kessler Peter, *Elemente de teoria mulțimilor şi topologie generală Culegere de exerciții şi probleme,* Ed. Secolul XXI, Craiova, 1996

[L-P] Leonte Alexandru, Predoi Maria, *Culegere de probleme de analiză matematică,* Repr. Univ. Craiova, 1981

[MAI] Markushevich A.I., *Teoria analiticeskih funktzii*, Moskva, 1950

[MC] Meghea C., *Introducere în analiza matematică*, Ed. Ştiințifică, Bucureşti, 1968

[ME] Mendelson Elliott, *Boolean Algebra and Switching Circuits*, McGraw Hill Book Comp., 1970

[MG] Marinescu G., *Tratat de analiză funcțională*, Ed. Academiei RSR, Bucureşti, 1972

[MM] Megan Mihail, *Iraţionalitatea şi transcendenţa numerelor* e *şi* π, Revista Matematică a elevilor, Timişoara, XVIII, Nr.1(1987)
[NC_1] Niculescu Constantin, *Analiza matematică pe dreapta reală. O abordare contemporană*, Ed. Universitaria, Craiova, 2002
[NC_2] Niculescu Constantin, *Probleme de analiză matematică*, Ed. Radical, Craiova, 1994
[NC_3] Niculescu Constantin, *Fundamentele analizei matematice - Analiza pe dreapta reală*, Ed. Acad. Române, Bucureşti, 1966
[NC_4] Niculescu Constantin, *Calculul integral al funcţiilor de mai multe variabile. Teorie şi Aplicaţii*, Ed. Universitaria, Craiova, 2002
[N-D-M] Nicolescu Miron, Dinculeanu N., Marcus S., *Manual de analiză matematică, Vol. I, II*, EDP, Bucureşti, 1963
[NM] Nicolescu Miron, *Funcţii reale şi elemente de topologie*, EDP, Bucureşti, 1968
[N-R] Negoiţă C.V., Ralescu D.A., *Mulţimi vagi şi aplicaţiile lor*, Ed. Tehnică, Bucureşti, 1974
[NS] Nikolski S.M., *A Course of Mathematical Analysis, Vol. I, II*, Mir Publishers, Moscow, 1981
[O-H-T] Olariu V., Halanay A., Turbatu S., *Analiză matematică*, EDP, Bucureşti, 1983
[OV] Olariu Valter, *Analiză matematică*, EDP, Bucureşti, 1981
[P-C-R_1] Predoi M., Constantinescu D., Racilă M., *Teme de calcul diferenţial*, Ed. Sitech, Craiova, 2000
[P-C-R_2] Predoi M., Constantinescu D., Racilă M., *Teme de calcul integral*, Ed. Sitech, Craiova, 2000
[PG] Preuβ G., *Allgemeine Topologie*, Springer-Verlag, 1972
[P-H-M] Popa C., Hiriş V., Megan M., *Introducere în analiza matematică prin exerciţii şi probleme*, Ed. Facla, Timişoara, 1976
[PM_1] Predoi Maria, *Analiză matematică*, Ed. Universitaria, 1994
[PM_2] Predoi Maria, *Sur la convergence quasi-uniforme*, Periodica Math. Hungarica, Vol. 10(1979), Nr.1
[PM_3] Predoi Maria, *Sur la convergence quasi-uniforme topologique*, An. Univ. Craiova, Nr.11(1983), p. 15-20
[P-S] Pólya G., Szegö G., *Aufgaben und Lehrsätze aus der Analysis, Vol. I, II*, Springer-Verlag, Berlin Heidelberg New-York, 1970
[PV] Piscounov V., *Calcul Differentiel et integral, Vol. I, II*, Editions de Moscou, 1972
[RI] Rus A. Ioan, *Principii şi aplicaţii ale teoriei punctului fix*, Ed. Dacia, Cluj-Napoca, 1979
[RM] Reghiş Mircea, *Elemente de teoria mulţimilor şi logică matematică*, Ed. Facla, Timişoara, 1981
[R-S] Rasiova H., Sikorski R., *The Mathematics of Metamathematics*, Polske Akad. Nauk, Warzyawa, 1963

[RW] Rudin W., *Functional Analysis*, McGraw Hill Company, New-York, 1973
[SG] Sirețchi Gh., *Calcul diferențial şi integral, Vol. I, II*, Ed. Ştiințifică şi Enciclopedică, Bucureşti, 1985
[ŞG] Şilov G.E., *Analiză matematică, Vol. I, II*, Ed. Ştiințifică, Bucureşti, 1983-1985
[SH] Schaefer Helmut, *Topological Vector Spaces*, New-York, 1966
[SO] Stănăşilă Octavian, *Analiză matematică*, EDP, Bucureşti, 1981
[SL] Schwartz Laurent, *Analyse Mathématique*, Hermann, Paris, 1967
[SS] Stoilow Simeon, *Teoria funcțiilor de o variabilă complexă*, Ed. Didactică şi Pedagogică, Bucureşti, 1962
[TK] Teleman K., *Geometrie diferențială locală şi globală*, Ed. Tehnică, Bucureşti, 1974
[UC] Udrişte Constantin, *Linii de câmp*, Ed. Tehnică, Bucureşti, 1988
[VB] Vulih Boris Z., *Introduction à l'analyse fonctionelle*, Nauka, 1973
[V-P] Vladimirescu Ion, Popescu Mariana, *Algebră Liniară şi Geometrie Analitică*, Ed. Universitaria, Craiova, 1994
[YK] Yosida K., *Functional Analysis*, Springer-Verlag, 1965