MATHEMATICAL ANALYSIS I (DIFFERENTIAL CALCULUS) FOR ENGINEERS AND BEGINNING MATHEMATICIANS

SEVER ANGEL POPESCU

Department of Mathematics and Computer Sciences, Technical University of Civil Engineering Bucharest, B-ul Lacul Tei 124, RO 020396, sector 2, Bucharest 38, ROMANIA.

E-mail address: angel.popescu@gmail.com

To my family.

To those unknown people who by hard and honest working make possible our daily life of thinking.

Contents

Preface

I start this preface with some ideas of my former Teacher and Master, senior researcher I, corresponding member of the Romanian Academy, Dr. Doc. Nicolae Popescu (Institute of Mathematics of the Romanian Academy).

Question: What is Mathematics?

Answer: It is the art of reasoning, thinking or making judgements. It is difficult to say more, because we are not able to exactly define the notion of a "table", not to say Math! In the greek language "mathema" means "knowledge". Do you think that there is somebody who is able to define this last notion? And so on... Let us do Math, let us apply or teach it and let us stop to search for a definition of it!

Q: Is Math like Music?

A: Since any human activity involves more or less need of reasoning, Mathematics is more connected with our everyday life then all the other arts. Moreover, any description of the natural or social phenomena use mathematical tools.

Q: What kind of Mathematics is useful for an engineer?

A: Firstly, the basic Analysis, because this one is the best tool for strengthening the ability of making correct judgements and of taking appropriate decisions. Formulas and notions of Analysis are at the basis of the particular language used by the engineering topics like Mechanics, Material Sciences, Elasticity, Concrete Sciences, etc. Secondly, Linear Algebra and Geometry develop the ability to work with vectors, with geometrical object, to understand some specific algebraic structures and to use them for applying some numerical methods. Differential Equations, Calculus of Variations and Probability Theory have a direct impact in the scientific presentation of all the engineering applications. Computer Science cannot be taught without the basic knowledge of the above mathematical topics. Mathematics comes from reality and returns to it.

Q: How can we learn Math such that this one not becomes abstract, annoying, difficult, etc.?

A: There is only one way. Try to clarify and understand everything, step by step, from the simplest notions up to the more complicated ones. Without gaps! Try to work with all the new notions, definitions, theorems, by looking at appropriate simple examples and by doing appropriate exercises. Do not learn by heart! This is the most useless thing you can do in trying to become a scientist, an engineer or an economist! Or anything else!

Math becomes nice and easy to you if it is presented in a lively way and if you make some efforts to come closer and closer to it. If you hate it from the beginning, don't say that it is difficult!

The present course of Mathematical Analysis covers the Differential Calculus part only.

It is assumed that students have the basic skills to compute simple limits, differentials and the integrals of some elementary functions. My teaching experience of almost 30 years at the Technical University of Civil Engineering Bucharest made me clear that the Math syllabus for engineering courses is not only a "part" from the syllabus of the faculties of mathematics. Engineering teaching should have at its basis very "concrete" facts. Mathematics for engineers should be very live. Student should realize that such type of Math came from "practice", returns to it and, what is most important, it helps a lot to make rational "models" for some specific phenomena. Besides this point of view, we have not to forget that the most important tool of an engineer, economist, etc. is his (her) power of reasoning. And this power of reasoning can be strengthened by mathematical training.

My opinion is that some motivations and drawings are always very useful in the complicated process of making "easy" and "nice" the mathematical teaching.

I consider that it is better to start with the notion of a real number, which reflects a measurement. Then to consider sequences, series, functions, etc.

In Chapter I tried to put together some notions and ideas which have more features in common. We end every chapter with some problems and exercises. In some places you will find more detailed examples and worked problems, in others you will find fewer. At any moment I have in my mind a beginner student and not a moment a professional in Math. My last goal in this was "the art of teaching Math for engineers" and not "the art of solving sophisticated Math problems". We should be very careful that a good Math teaching means "not multa, sed multum" (C. F. Gauss, in Latin). Gauss wanted to say that the

quality is more important then the quantity, "not much and superficial, but fewer and deep". We have computers which are able to supply us with formulas, with complicated and long computations but, up to now, they are not able to learn us the deep and the original creative work. They are useful for us, but the last decision is better to be ours. The deep "feeling" of an experienced engineer is as important as some long computations of a computer. If we consider a computer to be only a "tool" is OK. But, how to obtain this "feeling"? The answer is: a good background (including Math training) + practice + the capacity of doing things better and better.

I tried to use as proofs for theorems, propositions, lemmas, etc. the most direct, simple and natural proofs that I know, such that the student be able to really understand what the statement wants to say. The mathematical "tricks" and the simplifications by using more abstract mathematical machinery are not so appropriate in teaching Math at least for the non mathematical community. This is why we (teachers) should think twice before accepting a new "shorter" way. My opinion is that student should begin with a particular case, with an example, in order to understand a more general situation. Even in the case of a definition you should search for examples and "counterexamples", you should work with them to become "a friend" of them... .

I am grateful to many people who helped me directly or indirectly. The long discussions with some of my colleagues from the Department of Mathematics and Computer Sciences of the Technical University of Civil Engineering Bucharest enlightened me a lot. In particular, the teaching skill, the knowledge and the enthusiasm of Prof. Dr. Gavriil Păltineanu impressed and encouraged me in writing this course. He is always trying to really improve the way of Math Analysis teaching in our university and he helped me with many useful advices after reading this course.

Many thanks go to Prof. Dr. Octav Olteanu (University Politehnica Bucharest) for many useful remarks on a previous version of this course.

To be clear and to try to prove "everything" I learned from Prof. Dr. Mihai Voicu, who was previously teaching this course for many years.

The friendly climate created around us by our departmental chiefs (Prof. Dr. ing. Nicoleta Rădulescu, Prof. Dr. Gavriil Păltineanu, Prof. Dr. Romică Trandafir, etc.) had a great contribution to the natural development of this project.

I thank to my assistant professor Marilena Jianu for many corrections made during the reading of this material.

A special thought goes to the late Dr. Ion Petrică who (many years ago) had the "feeling" that I could write a "popular" book of Math Analysis with the title "Analysis is easy, isn't it?".

The last, but not the least, I express my gratitude to my wife for helping me with drawings and for a lot of patience she had during my writing of this book.

I will be very grateful to all the readers who will send me their remarks on this course to the e-mail address: angel.popescu@gmail.com, in order to improve everything in future editions.

Prof. Dr. Sever Angel Popescu
Bucharest, January, 2009.

CHAPTER 1

The real line.

1. The real line. Sequences of real numbers

To measure is a basic human activity. To measure time, temperature, velocity, etc., reduces to measure lengths of segments on a line. For this, we need a fixed point O on a straight line (d) and a "witness" oriented segment $[OA_1]$ $(A_1 \neq O)$, i.e. a unitary vector $\overrightarrow{OA_1}$ (see Fig.1.1). Here, unitary means that always in our considerations the length of the segment $[OA_1]$ will be considered to have 1 meter. The pair $(O, \overrightarrow{i})$, where $\overrightarrow{i} = \overrightarrow{OA_1}$ is called a *Cartesian* (from the French mathematician R. Descartes, the father of the Analytical Geometry, what shortly means to study figures by means of numbers) *coordinate system (or a frame of reference).* We assume that the reader has a practical knowledge of the *digits* $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ which represent (in Fig.1.1) the points $O, A_1, A_2, ..., A_9$. Let us now consider the point B on the line (d) such that the length $\left|\overrightarrow{A_9B}\right|$ of the vector $\overrightarrow{A_9B}$ is 1 meter and $B \neq A_8$, i.e. $\overrightarrow{A_9B} = \overrightarrow{OA_1}$ as FREE vectors.

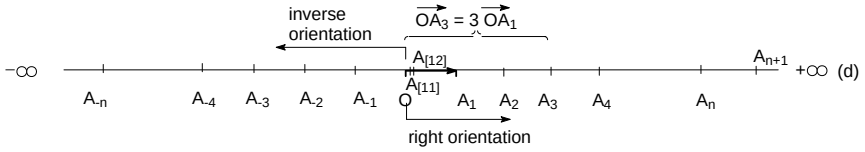


Fig. 1.1

Our intention is to associate a sequence of digits to the point B. Here appears a first great idea of an anonymous inventor who denoted B by A_{10}, this means one group of ten units (a unit is one $\overrightarrow{OA_1}$) and 0 (nothing) from the next similar group. For instance, A_{64} is the point on (d) which is between the points A_{60} and A_{70} such that it marks 6 groups of ten units + 4 units from the 7-th group. Now A_{269} marks 2 groups of hundreds + 6 groups of tens + 9 units, ... and so on. In this way we can represent on the *real line* (d) any quantity which is a multiple of a unity (for instance 130 km/h if the unity is 1 km/h). The idea of grouping in units, tens, hundreds, thousands, etc. supply

us with an addition law for the set of the so called "natural numbers": $0, 1, 2, ...9, 10, 11, ..., 99, 100, 101, ...$ We denote this last set by $\mathbb{N}$.

For instance, let us explain what happens in the following addition:

$$\begin{array}{cccc} 3 & 6 & 8 & + \\ & 9 & 7 & \\ \hline 4 & 6 & 5 & \end{array} \tag{1.1}$$

First of all let us see what do we mean by 368. Here one has 3 groups of one hundred each + 6 groups of one ten each + 8 units (i.e. 8 times $\overrightarrow{OA_1}$). We explain now the result 465 $(= 368 + 97)$: 8 units + 7 units is equal to 15 units. This means 5 units and 1 group of ten units. This last 1 must be added to $6 + 9$ and we get 16 groups of ten units each. Since 10 groups of 10 units means a group of 1 hundred, we must write 6 for tens and add to 3 this last 1. So one gets 4 for hundreds. We say that a point A on the line (d) is "*less*" than the point B on the same line if the point B is on the right of A and not equal to it. Assume now that A is represented by the sequence of digits $\overline{a_n a_{n-1}...a_0}$ (a_0 units, a_1 tens, etc.) and B by the sequence $\overline{b_m b_{m-1}...b_0}$. Here we suppose that a_n and b_m are distinct of 0 and that $n \geq m$. Otherwise, we change A and B between them. Think now at the way we defined these sequences! If $n \geq m$, A must be on the right of B or identical to it. If $n > m$ then A is greater than B. If $n = m$, but $a_n > b_n$, again A is greater than B. If $n = m$, $a_n = b_n$, but $a_{n-1} < b_{n-1}$, then B is greater than A. If $n = m$, $a_n = b_n$, $a_{n-1} = b_{n-1}$, we compare a_{n-2} with b_{n-2} and so on. If all the corresponding terms of the above sequences are equal one to each other (and $n = m$) we have that A is identical with B. If for instance, $n = m$, $a_n = b_n$, $a_{n-1} = b_{n-1}, ..., a_k = b_k$, but $a_{k-1} > b_{k-1}$ we must have $A > B$ (A is greater than B). Here in fact we described what is called the "lexicographic order" in the set of finite sequences (define it!). If $A \geq B$ one can subtract B from A as it follows in this example:

$$\begin{array}{cccc} 3 & 6 & 8 & - \\ & 9 & 7 & \\ \hline 2 & 7 & 1 & \end{array} \tag{1.2}$$

This operation is as natural as the addition. Namely, 8 units minus 7 units is 1 unit. Since we cannot subtract 9 tens from 6 tens, we "borrow" 1 hundred = 10 tents from 3. So, now 10 tens + 6 tens = 16 tens minus 9 tens is equal to 7 tens. It remains 2 hundreds from which we subtract 0 hundreds and obtain 2 hundreds. Instead of 10 tens we write $10 \times 10 = 10^2$ units, etc. Thus, any natural number

$A = \overline{a_n a_{n-1}...a_0}$ (we identified here the name of the point with its corresponding sequence of digits) can be uniquely written as:

$$A = a_0 + 10a_1 + 10^2 a_2 + ... + 10^n a_n \tag{1.3}$$

This is also called the representation of A in the base (of numeration) 10. If instead of grouping units, tens, hundreds, etc., in groups of 10, we group them in groups of 2 for instance, we obtain the writing of same point A in base 2, etc. Why our ancestors chose $10, ...$ we do not know! Maybe because we have 10 fingers...!!

Hence, the subtraction is not defined for any pair A, B. This means that $A - B$ does not belong to $\mathbb{N}$ for any pair A, B. For instance, $3 - 4$ is not in $\mathbb{N}$, but it is in $\mathbb{Z}$! The algebraists say that $\mathbb{N}$ is a monoid and $\mathbb{Z}$ is a group (see any advanced Algebra course), relative to the addition. We can also introduce a multiplication in $\mathbb{Z}$. First of all, if n, m are in $\mathbb{N}$ and both are not zero (otherwise we put $n \cdot m = 0$), we define $n \cdot m \overset{not}{=} nm$ by $n + n + ... + n$, m times. For extending this operation to $\mathbb{Z}$, we put by definition $(-n)m = n(-m) = -(nm)$, for any pair n, m of $\mathbb{N}$. The algebraists say that $\mathbb{Z}$ is a ring relative to the addition and this last defined multiplication (see the Algebra course). We use here freely the elementary basic properties of the addition and multiplication. For instance, $5 \cdot (7 - 9) = 5 \cdot 7 - 5 \cdot 9$, because of the distributive property.

We also have a dynamic interpretation of the set $\mathbb{N}$. 0 is for O. 1 is for the extremity A_1 of the vector $\overrightarrow{OA_1}$. 2 is for the extremity of the vector $\overrightarrow{OA_2}$ which is twice the vector $\overrightarrow{OA_1}$, etc. We must remark that we just have chosen "an orientation" on the line (d), namely, we started our above construction "from O to the right", not "to the left". So, on (d) one has two orientations: the *direct* one, "to the right" and the *inverse* one, "to the left". If we construct everything again, "on the left" (by symmetry) we get the set of negative integers: $-1, -2, -3,...$. The whole set $\mathbb{Z} = \{..., -3, -2, -1, 0, 1, 2, 3, ...\}$ is called the set of *integers.*

By "*Arithmetic*" we mean all the properties of $\mathbb{N}$ (or $\mathbb{Z}$) derived from the "algebraic" operations of addition and multiplication. A *prime number* p is a natural number distinct of 1, which cannot be written as a product $p = nm$, where n and m are natural numbers, both distinct of 1 (or of p). For instance, $2, 3, 5, 7, 11, 13, 17, ...$ are prime numbers. Any natural number n greater than 1 is either a prime number or it can be decomposed into a finite product of prime numbers (Euclid). Indeed, if n is not a prime number, there are $n_1, n_2,$ natural numbers such that $n = n_1 n_2$, where $n_1, n_2 < n$. We go on with the same procedure for n_1

and n_2 instead of n, etc., up to the moment when $n = p_1p_2p_3...p_k$, where all $p_1, p_2, ..., p_k$ are prime numbers. Maybe some of them are equal one to the other so, we can write $n = q_1^{m_1}q_2^{m_2}...q_h^{m_h}$, where $q_1, q_2, ..., q_h$ are distinct primes.

THEOREM 1. *(The Fundamental Theorem of Arithmetic) Any natural number n greater than 1 is either a prime number or it can be uniquely written as $n = q_1^{m_1}q_2^{m_2}...q_h^{m_h}$, where $q_1, q_2, ..., q_h$ are distinct prime numbers.*

All the other basic results in number theory are directly or indirectly connected with this main result. For instance, Euclid proved that the set of all prime numbers is infinite. Indeed, if it was not so, let $q_1, q_2, ..., q_N$ be all the distinct primes. Then, let us consider the natural number $m = q_1q_2...q_N + 1$. It is either a prime number or it is divisible by a prime number p. Since $q_1, q_2, ..., q_N$ are all the prime numbers, this p must be equal to a q_j for a $j \in \{1, 2, ..., N\}$. Then 1 is divisible by q_j, a contradiction (Why?). Thus, our assumption is false, i. e. the set of prime numbers is infinite. The most delicate hypotheses and results in Mathematics are connected with this set.

Recall that a function $f : X \to Y$, where X and Y are arbitrary sets, is said to be *injective (or one-to-one)* if for any pair of distinct elements a and b from X, their images $f(a)$ and $f(b)$ are distinct in Y. f is *surjective* (*or onto... Y*) if any element y of Y is the image of an element x of X, i. e. $y = f(x)$. Injective + surjective means *bijective*. If f is bijective we simply say that it is "a *bijection*" between the sets X and Y. Or that they have "*the same cardinal*". For instance, $\mathbb{N}$ and $\mathbb{Z}$ have the same cardinal because $f : \mathbb{N} \to \mathbb{Z}$, $f(0) = 0$, $f(2n) = -n$ and $f(2n-1) = n$, for $n = 1, 2, ...$ is a bijection (Why?).

Generally, if a set A has the same cardinal with $\mathbb{N}$ we say that it is *countable*. If a set B has the same cardinal with a set of the form $\{1, 2, ..., n\}$ we say that it is *finite* and that it has n elements, or that its cardinal is n. Why a set A cannot be finite and countable at the same time?

Any countable set A can be represented like a sequence: $a_0 = f(0)$, $a_1 = f(1)$, $a_2 = f(2), ...$ where $f : \mathbb{N} \to A$ is a bijection between $\mathbb{N}$ and A (see the definition of countability!). Conversely, any set A which can be represented like a sequence is countable, i.e. it is the image of the natural number set $\mathbb{N}$ through a bijection f (prove this!). Hence, we define *"a sequence"* in a set A by a function $g : \mathbb{N} \to A$. Usually we denote $g(n)$ by a_n and write the sequence g as $a_0, a_1, a_2, ..., a_n, ...$ or simply as $\{a_n\}$, where a_n is said to be the *general term* of the sequence g. Here, for instance, a_5 is called the term of *rank* 5 of the sequence g.

A sequence $\{b_m\}$ is called a "*subsequence*" of the sequence $\{a_n\}$ if there is a sequence $k_1 < k_2 < ... < k_n < ...$ of natural numbers such that for any $m \in \mathbb{N}$, b_m is equal to a_{k_m}. For instance $\{b_k = 2k\}$, $k = 0, 1, 2, ...$ is a subsequence of $\mathbb{N} = \{0, 1, 2, ...\}$. But the sequence $\{0, 1, 2, 2, 2, ...\}$ is NOT a subsequence of $\mathbb{N}$ (Why?). Yes, the set $\{0, 1, 2\}$ IS a subset of $\mathbb{N}$, but not ...a subsequence! Can $\mathbb{N}$ be a subsequence of $\mathbb{Z}$?

Now our question is: "How do we represent 2 kg and a quarter on the line (d)?" More exactly, to the point C on (d) which is the extremity of a vector $\overrightarrow{OC}$, obtained by taking $\overrightarrow{OA_1}$ twice + a quarter from the same vector $\overrightarrow{OA_1}$, what kind of sequence of digits $0, 1, 2, ..., 9$ could we associate? Let us divide the segment $[OA_1]$ into 10 equal parts and let us associate the symbol 0.1 to the extremity $A_{[11]}$ of the vector $\overrightarrow{OA_{[11]}}$ which is the 10-th part of $\overrightarrow{OA_1}$. In the same way we construct $A_{[12]}$, $A_{[13]}, ..., A_{[19]}$ and their corresponding symbols 0.2, $0.3, ..., 0.9$. We continue by dividing the segment $[OA_{[11]}]$ into 10 equal parts and obtain the new symbols 0.01, $0.02, ..., 0.09$, etc. We say that $0.1 = \frac{1}{10}$, $0.01 = \frac{1}{100}$, and so on. For instance, the sequence (or the number) 23.0145 represents the point E on (d) obtained in the following way. To the vector $\overrightarrow{OA_{23}}$ we add: $\frac{1}{100}\overrightarrow{OA_1} + \frac{4}{1000}\overrightarrow{OA_1} + \frac{5}{10000}\overrightarrow{OA_1}$. The resultant vector is $\overrightarrow{OE}$, etc. If one works (by symmetry) on the left of O, one gets the "negative" numbers of the form: $-\overline{a_n a_{n-1}...a_0}.b_1b_2...b_m$, where a_i and b_j are digits from the set $\{0, 1, 2, ...9\}$. This last number can be written as:

$$-(10^n a_n + 10^{n-1}a_{n-1} + ... + a_0 + \frac{b_1}{10} + \frac{b_2}{10^2} + ... + \frac{b_m}{10^m})$$

$$= -\frac{\overline{a_n a_{n-1}...a_0 b_1 b_2...b_m}}{10^m} \tag{1.4}$$

Here appeared fractions like $\frac{a}{b}$, where a and b are natural numbers and $b \neq 0$. We suppose that the reader is familiar with the operations of addition, subtraction, multiplication and division with such fractions. If $a \in \mathbb{Z}$ and $b = 10^m$, from this discussion, we have the geometrical meaning of the fraction $\frac{a}{b}$. We also call any fraction, a number. What is the geometrical meaning of $\frac{4}{7}$? Take again the vector $\overrightarrow{OA_1}$ and divide it into 7 equal parts. Let $\overrightarrow{OG}$ be the 7-th part of $\overrightarrow{OA_1}$. Then $4\overrightarrow{OG} = \overrightarrow{OH}$ and H will be the point which corresponds to the number $\frac{4}{7}$. The Greeks said that the number $\frac{4}{7}$ is obtained when we want to measure a segment $[ON]$ with another segment $[OM]$ and if we can find a third segment $[OP]$ such that $[ON] = 4[OP]$ and $[OM] = 7[OP]$, i.e.

$\frac{[ON]}{[OM]} = \frac{4}{7}$. A representation of a number (for instance a fraction) as $\pm \overline{a_n a_{n-1}...a_0}.b_1 b_2...b_m...$ is called a *decimal representation* (or a decimal fraction). Let us try to find a decimal representation for the fraction $\frac{4}{7}$. The idea is to write $\frac{4}{7}$ as $\frac{1}{10} \cdot \frac{40}{7}$. Then, $40 = 5 \cdot 7 + 5$ implies $\frac{40}{7} = 5 + \frac{5}{7}$, where $\frac{5}{7} < 1$. Hence $\frac{4}{7} = \frac{5}{10} + \frac{1}{10} \cdot \frac{5}{7}$. Now we do the same for $\frac{5}{7}$. Namely, $\frac{5}{7} = \frac{1}{10} \cdot \frac{50}{7} = \frac{1}{10}(7 + \frac{1}{7})$, so

$$\frac{4}{7} = \frac{1}{10}[5 + \frac{1}{10}(7 + \frac{1}{7})] = \frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^2} \cdot \frac{1}{7}.$$

Write now

$$\frac{1}{7} = \frac{1}{10} \cdot \frac{10}{7} = \frac{1}{10}(1 + \frac{3}{7}).$$

So

$$\frac{4}{7} = \frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^3}(1 + \frac{3}{7}) = \frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^3} + \frac{1}{10^3} \cdot \frac{3}{7}.$$

Since the remainders obtained by dividing natural numbers by 7 can be $0, 1, 2, 3, 4, 5$, or 6, in the sequence $\frac{4}{7}, \frac{5}{7}, \frac{1}{7}, \frac{3}{7}$, ..., at least one of the fraction must appear again after at most 7 steps. Thus, let us go on! Write

$$\frac{3}{7} = \frac{1}{10} \cdot \frac{30}{7} = \frac{1}{10}(4 + \frac{2}{7}).$$

So

$$\frac{4}{7} = \frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^3} + \frac{4}{10^4} + \frac{1}{10^4} \cdot \frac{2}{7}.$$

But

$$\frac{2}{7} = \frac{1}{10} \cdot \frac{20}{7} = \frac{1}{10}(2 + \frac{6}{7}) = \frac{2}{10} + \frac{1}{10^2} \cdot \frac{60}{7} = \frac{2}{10} + \frac{1}{10^2}(8 + \frac{4}{7}).$$

So

$$\frac{4}{7} = \frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^3} + \frac{4}{10^4} + \frac{2}{10^5} + \frac{8}{10^6} + \frac{1}{10^6} \cdot \frac{4}{7}.$$

But

$$\frac{4}{7} = \frac{1}{10} \cdot \frac{40}{7} = \frac{1}{10}(5 + \frac{5}{7}).$$

Hence

$$\frac{4}{7} = \frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^3} + \frac{4}{10^4} + \frac{2}{10^5} + \frac{8}{10^6} + \frac{5}{10^7} + ... \tag{1.5}$$

Since the digit 5 appears again, we must have:

$$\frac{4}{7} = 0.5714285714285... \stackrel{not}{=} 0.(571428).$$

We say that $\frac{4}{7}$ is a simple periodical decimal fraction. Here we meet with an "infinite" sum, i.e. with a series:

$$0.(571428) = \frac{5}{10}(1 + \frac{1}{10^6} + ...) + \frac{7}{10^2}(1 + \frac{1}{10^6} + ...) + ...$$
$$= (\frac{5}{10} + \frac{7}{10^2} + \frac{1}{10^3} + \frac{4}{10^4} + \frac{2}{10^5} + \frac{8}{10^6})(1 + \frac{1}{10^6} + \frac{1}{10^{12}} + ...).$$

But $1 + \frac{1}{10^6} + \frac{1}{10^{12}} + ...$ is an infinite geometrical progression with the first term 1 and the ratio $\frac{1}{10^6}$. The actual mathematical meaning of this infinite sum will be explained later.

The next question is if always one can measure a segment a by another segment b and obtain as a result a fraction $\frac{m}{n}$. Even Greeks discovered in Antiquity that this operation is not always possible. For instance, if one wants to measure the diagonal d of a square with the side a of the same square we obtain a new number $\frac{d}{a}$ such that $\left(\frac{d}{a}\right)^2 = 2$ (apply Pythagoras' Theorem). If $\frac{d}{a}$ was a fraction $\frac{m}{n}$, where $m, n \in \mathbb{N}$, $n \neq 0$ and m, n have no common divisor except 1, then $m^2 = 2n^2$ and 2 would be a divisor of m, i.e. $m = 2m'$. Thus, $2m'^2 = n^2$ and then n would also have 2 as a divisor, a contradiction. Usually such a number $\frac{d}{a}$ is denoted by $\sqrt{2}$ because its square is 2. Such numbers were not accepted by Greeks as being "real" numbers ! But $\sqrt{2}$ can be represented on the real line (d). It is the point U which denotes the extremity of a vector $\overrightarrow{OU}$ such that its length is equal to the length of the diagonal of a square of side 1 (= the length of $\overrightarrow{OA_1}$). Any fraction is called a *rational number* and any other number (like $\sqrt{2}$) is called an *irrational number*. $\sqrt{2}$ is an *algebraic number* because it is a root of an equation with rational coefficients ($X^2 - 2 = 0$). We say that a number is a *real number* if it is the result of a measurement, i.e. it can be associated with a point of the real line (d). Up to now we know that NOT all real numbers can be represented by ordinary fractions (like $\sqrt{2}$). We shall indicate below a natural way to associate to any point of the line (d) a decimal fraction, usually infinite. Recall that to the point A_n ($\overrightarrow{OA_n} = n\overrightarrow{OA_1}$) we associated a natural number n (given as a finite sequence of digits). The symmetric point of A_n relative to the origin O was denoted by A_{-n} (see Fig.1.1). Our intuition says that any point M belongs to a segment of the type $[A_n, A_{n+1})$, where n here can be positive or nonpositive (i.e. $n \in \mathbb{Z}$). We want to associate to the point M its *coordinate* x_M i.e. a decimal number in the interval $[n, n+1)$ = the set of all the real numbers (known or unknown up to now!) which are greater or equal to n and less than $n+1$ (relative to the above lexicographic order). So $\bigcup_{n\in\mathbb{Z}} [A_n, A_{n+1})$ = all the points of

(d). But this last assertion cannot be mathematically proved using only previous simpler results! It is called the *Archimedes' Axiom*. In the language of the real numbers it says that any such number r belongs to an interval of the type $[n, n+1)$. This n is called the integral part of r and it is denoted by $[r]$. For instance, $[3.445] = 3$, but $[-3.445] = -4$, because $-3.445 \in [-4, -3)$. So, our point M belongs to an interval of the type $[A_n, A_{n+1})$ for ONLY one $n = \pm\overline{a_k a_{k-1}...a_0}$, where a_i are digits. Let us divide the segment $[A_n, A_{n+1})$ into 10 equal parts by 9 points B_1, B_2, ..., B_9, such that:

$$[A_n, A_{n+1}) = [A_n \overset{not}{=} B_0, B_1) \cup [B_1, B_2) \cup ... \cup [B_9, A_{n+1} \overset{not}{=} B_{10}).$$

To these points we obviously associate the following rational numbers: $B_1 \to n + 0.1$,

$$B_2 \to n + 0.2, ..., B_9 \to n + 0.9.$$

Since $M \in [A_n, A_{n+1})$, M belongs to one and only to one subsegment $[B_i, B_{i+1})$, where $i \in \{0, 1, ..., 9\}$. By definition we take as the first decimal of x_M to be this last digit $b_1 = i$. If M is just B_i we have $x_M = \pm\overline{a_k a_{k-1}...a_0}.b_1$. If M is on the right of B_i the actual x_M will be greater then the rational number $\overline{a_k a_{k-1}...a_0}.b_1$ and we continue our above division process. Namely, instead of $[A_n, A_{n+1})$ we take $[B_i, B_{i+1})$ that M belongs to and divide this last interval into 10 equal parts by the points $C_0 = B_i$, C_1, ..., C_9 and $C_{10} = B_{i+1}$. There is only one j such that $M \in [C_j, C_{j+1})$. By definition, the second decimal of x_M is $b_2 = j$. If $M = C_j$, then $x_M = \pm\overline{a_k a_{k-1}...a_0}.b_1 b_2$ and x_M would be a rational number. If NOT, then we go on with the segment $[C_j, C_{j+1})$ instead of $[B_i, B_{i+1})$, etc. If at a moment M will be the left edge of an interval obtained like above, then x_M will have a finite decimal representation, i.e. it will be a rational number. If M will never be in this situation, then x_M can or cannot be a rational number. For instance, the point P which corresponds to the fraction $\frac{4}{7}$ is in this last position but, ... it is represented by a fraction, so x_P is a rational number. The point V which corresponds to $\sqrt{2}$ is in the same position as P, but x_V is not a rational number as we proved above. The segments constructed above, are contained one into the other:

$$[A_n, A_{n+1}) \supset [B_i, B_{i+1}) \supset [C_j, C_{j+1}) \supset ....$$

If M is not the left edge of no one of these segments, then their intersection is exactly M (Why?).

In general, the following question arises. If one has a tower of closed segments

$$[T_1, U_1] \supset [T_2, U_2] \supset ... \supset [T_n, U_n] \supset ...$$

on the real line (d), their intersection is empty or not? Our intuition says that it could not be empty for ever! But,... there is no mathematical proof for this! This is way this last assertion is an axiom, called the *Cantor's Axiom*. Now we can call a real number r any decimal fraction (finite or not) of the type:

$$r = \pm\overline{a_k a_{k-1}...a_0}.b_1 b_2...b_m... \tag{1.6}$$

We can write this "number" as a sum of some special type of fractions

$$r = \pm\left(10^k a_k + ... + 10a_1 + a_0 + \frac{b_1}{10} + \frac{b_2}{10^2} + ... + \frac{b_m}{10^m} + ...\right) \tag{1.7}$$

Using this last representation, it is not difficult to define the usual elementary operations of addition, subtraction, multiplication, and division for the set $\mathbb{R}$ of all the real numbers (do it and find a natural explanation for the rules you learned in the high school!-You must also use the fact that $r = \lim\limits_{m\to\infty} r_m$, where

$$r_m = \pm\left(10^k a_k + ... + 10a_1 + a_0 + \frac{b_1}{10} + \frac{b_2}{10^2} + ... + \frac{b_m}{10^m}\right)$$

and the usual operations with convergent sequences). The algebraists say that $\mathbb{R}$ together with the addition and multiplication is a field (see the exact definition of a field in any Algebra course and verify this last assertion!). Because of the fact that the real numbers are nothing else than a representation of the points of the real line (together with a Cartesian reference frame on it!), the Archimedes's and the Cantor's axioms work on $\mathbb{R}$. They can be expressed in the following way (in language of numbers...):

AXIOM 1. *(Archimedes's Axiom) For any real number r there is one and only one integer number n such that $n \leq r < n+1$.*

AXIOM 2. *(Cantor's Axiom) Let $a_1 \leq a_2 \leq, ..., \leq a_n \leq, ...$ and $b_1 \geq b_2 \geq, ..., \geq b_n \geq, ...$ be two sequences of real numbers such that for any n one has that $a_n \leq b_n$. Then there is at least one real number r between a_n and b_n for any $n \in \mathbb{N}$. If in addition, the difference $b_n - a_n$ becomes smaller and smaller to zero, whenever n becomes larger and larger, then this real number r is unique (in fact, this last assertion is not an axiom !).*

Hence, the real numbers can always be seen like points on a real line (d). If we change the line and (or) the Cartesian reference frame we clearly obtain different sets of real numbers. But,...all these fields of real numbers are *isomorphic* like ordered fields. This means that for any two such fields R_1 and R_2 there is at least one bijection $f : R_1 \to R_2$

such that $f(x+y) = f(x) + f(y)$, $f(xy) = f(x)f(y)$ (f preserves the algebraic structure of fields) and $f(x) \leq f(y)$, whenever $x \leq y$ (f preserves the order introduced above). Here $x, y \in R_1$. In fact, it is not difficult to construct such a bijection. If we take $x \in R_1$, it is the decimal representation of a point X on the first real line (d_1). But always one can construct a natural bijection g between the points of (d_1) and the points of (d_2) which carries the Cartesian coordinate system of the first line into the coordinate system of the second line. Now we take for $f(x)$ the real number which corresponds to the point $g(X)$ of the second line (prove that this construction works).

From now on we fix a field $\mathbb{R}$ of real numbers and we assume that the reader knows the usual elementary rules of operating in this $\mathbb{R}$. It is of a great benefit if one always think of a real number as being a point on a fixed real line (d). So, ... draw everything or almost everything! This is why we say a point instead of a number and a number instead of a point!

We realize that the "practical" representation of an irrational number on the real line (d) is impossible! This means that you will never find a finite algorithm to do this. Because the point on (d) which corresponds to such an irrational number is obtained as the intersection of an infinite number of closed intervals, each of them contained into another one. Since the length of these intervals becomes smaller and smaller up to zero, practically we can approximate the real position of that point by one of the two ends of such a "very small" interval.

We must remark that the correspondence between the points of the real line (d) and the decimal representations is not a bijection. For instance, $0.999... = 1$. But,... the correspondence between the points of the real line (d) and the real numbers is a bijection! (Descartes' bijection).

Let us come back and recall that the set of natural numbers

$$\mathbb{N} = \{0, 1, ..., 9, 10, 11, ..., 20, 21, ..., n, ...\}$$

can be naturally embedded in the ring of integers

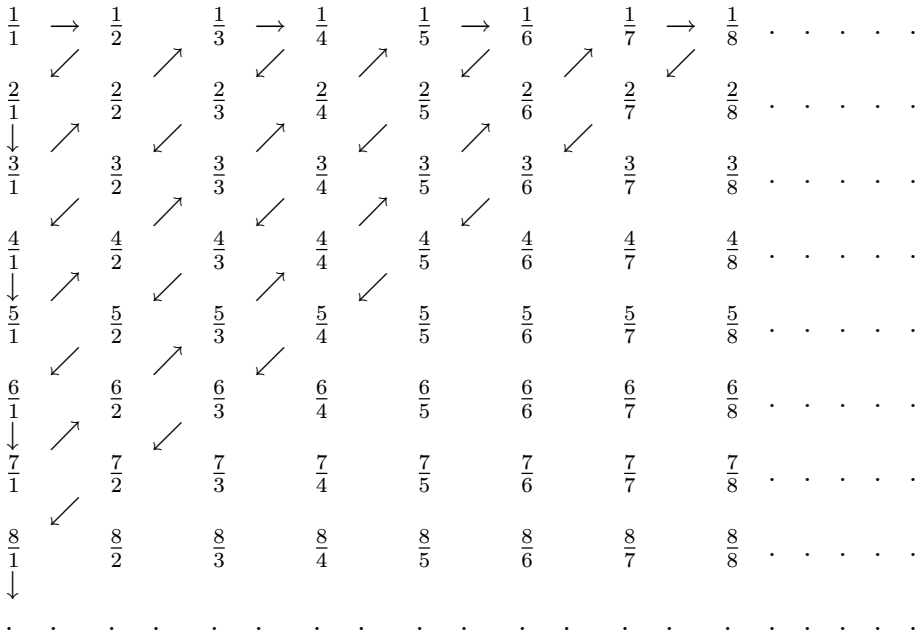
$$\mathbb{Z} = \{0, 1, -1, 2, -2, ..., n, -n, ...\},$$

where n is a natural number. This embedding preserves the usual operations of addition and multiplication. Both sets $\mathbb{N}$ and $\mathbb{Z}$ are clearly countable because they are naturally represented like sequences. What is the difference between $\mathbb{N}$ and $\mathbb{Z}$? The equation $X - 3 = 0$ has a solution in $\mathbb{N}$, $x = 3$, whereas the equation $X + 3 = 0$ has NO solution in $\mathbb{N}$, but it has the solution $x = -3$ in $\mathbb{Z}$. The next step is to see that the general linear equation of the form $aX + b = 0$, where $a, b \in \mathbb{Z}$,

may have no solution in $\mathbb{Z}$. For instance, $2X+1=0$ has no solution in $\mathbb{Z}$, but its solution is the fraction $\frac{-1}{2}=-\frac{1}{2}$ which is a rational number. Let us denote by $\mathbb{Q}$ the field of rational numbers and see that any integer number m can be represented as a rational number: $m=\frac{m}{1}$. So, $\mathbb{N}\subset\mathbb{Z}\subset\mathbb{Q}\subset\mathbb{R}$, since any rational number is a particular real number by the definition of a real number.

THEOREM 2. *The rational number field $\mathbb{Q}$ is also a countable set.*

PROOF. It will be enough to represent the positive elements of $\mathbb{Q}$ as a subsequence of a sequence (Why?-Use the same trick like in the case of the countability of $\mathbb{Z}$). Look now carefully to the following infinite table



and to the arrows which indicate "the next term" in the sequence. This sequence covers ALL the entries of this table and any positive rational number is an element of this sequence, i.e. $\mathbb{Q}_+$ can be viewed as a subsequence of this last sequence. Thus $\mathbb{Q}_+$ is countable. Since $\mathbb{Q}=\mathbb{Q}_-\cup\{0\}\cup\mathbb{Q}_+$, $\mathbb{Q}$ is also countable. □

Recall that a real number r is a "disjoint union" of two sequences of digits with $+$ or $-$ in front of it:

$$r=\pm\overline{a_ka_{k-1}...a_0}.b_1b_2...b_n... \tag{1.8}$$

The first sequence is always finite: $a_k, a_{k-1}, ..., a_0$. After its last digit a_0 (the units digit) we put a point "." . Then we continue with the digits of the second sequence: $b_1, b_2, ..., b_n, ...$. As we saw above, this last sequence can be infinite. If this last sequence is finite, i.e. if from a moment on $b_{n+1} = b_{n+2} = ... = 0$, we say that r is a *simple rational number*. Any simple rational number is a fraction of the form $\frac{a}{10^n}$ where $a \in \mathbb{Z}$ and $n \in \mathbb{N}$. If r is not a simple rational number, it can be canonically approximated by the simple rational numbers

$$r_n = \pm\overline{a_k a_{k-1}...a_0}.b_1 b_2...b_n,$$

for $n = 1, 2, ...$ This means that when n becomes larger and larger, the absolute value

(1.9)

$$error_n = |r - r_n| = 0.\underbrace{00...0}_{n-times} b_{n+1}b_{n+2}... = \frac{1}{10^{n+1}}(b_{n+1} + \frac{b_{n+2}}{10} + \frac{b_{n+3}}{10^2} + ...)$$

becomes closer and closer to 0. Indeed,

$$\frac{1}{10^{n+1}}(b_{n+1} + \frac{b_{n+2}}{10} + \frac{b_{n+3}}{10^2} + ...) \leq \frac{1}{10^{n+1}}(9 + \frac{9}{10} + \frac{9}{10^2} + ...) = \frac{1}{10^n}$$

and, since $\frac{1}{10^n} < \frac{1}{n}$ (prove it!), one gets that $|r - r_n| \to 0$ (tends to 0), when $n \to \infty$ (the values of n become larger and larger).

REMARK 1. *Hence, in any interval (a, b), $a \neq b$, a, b real numbers, one can find an infinite numbers of simple rational numbers (prove it!).*

But, what is the mathematical model for the fact that a sequence $\{x_n\}$, $n = 0, 1, ...$ tends to 0 (i.e. $|x_n|$ becomes closer and closer to 0, when n becomes larger and larger $(n \to \infty)$)?

DEFINITION 1. *We say that a sequence $\{x_n\}, n = 0, 1, ...$ is convergent to 0 (or tends to 0), when n tends to ∞ $(n \to \infty)$, if for any positive (small) real number $\varepsilon > 0$, there is a natural number N_ε (depending on ε) such that $|x_n| < \varepsilon$ for any $n \geq N_\varepsilon$. We simply write this: $x_n \to 0$, or, more formally: $\lim_{n\to\infty} x_n = 0$, or, less formally: $\lim x_n = 0$. We also say that a sequence $\{x_n\}$, $n = 1, 2, ...$ is convergent to a real number x (or that x is the limit of $\{x_n\}$; write $\lim_{n\to\infty} x_n = x$) if the difference sequence $\{x_n - x\}$, $n = 1, 2, ...$ is convergent to 0, or, if the "distance" $|x_n - x|$ between x_n and x becomes smaller and smaller as $n \to \infty$. This is equivalent to saying that for any positive (small) real number ε, all the terms of the sequence $\{x_n\}, n = 0, 1, ...,$ except a finite number of them, belong to the open interval $(x - \varepsilon, x + \varepsilon)$. Such an interval, centered at x and of "radius ε", is called an ε-neighborhood of x.*

THEOREM 3. *Let $\{x_n\}$ be a convergent sequence. Then its limit is a unique real number.*

PROOF. Let us assume that x and x' are two distinct limits of the sequence $\{x_n\}$ and let ε be a positive small real number such that $\varepsilon < |x - x'|$. Since both x and x' are limits of the sequence $\{x_n\}$, for n large enough, one must have $|x_n - x| < \frac{\varepsilon}{4}$ and $|x' - x_n| < \frac{\varepsilon}{4}$. Now

$$\varepsilon < |x' - x| = |x' - x_n + x_n - x| \leq |x' - x_n| + |x_n - x| < \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2},$$

or $\varepsilon < \frac{\varepsilon}{2}$, a contradiction! So, any two limits of the sequence $\{x_n\}$ must be equal! □

In (1.9) we have in fact that any real number r can be approximated by its simple rational number components (or approximates) r_n, i.e. $\lim r_n = r$. We say that the set of simple rational numbers is *dense* in $\mathbb{R}$. In particular, $\mathbb{Q}$ is dense in $\mathbb{R}$. Let m be a fixed nonzero natural number and let Q_m be the set of fractions of the form $\frac{a}{m^n}$, where a runs in $\mathbb{Z}$ and n runs in $\mathbb{N}$. Then any real number r is a limit of elements from Q_m, i.e. Q_m is dense in $\mathbb{R}$ (prove it!-write r in the basis m, instead of 10).

We just used above that the sequence $\{\frac{1}{n}\}$, $n = 1, 2, ...$ is convergent to 0. Our intuition says that if we divide the unity vector $\overrightarrow{OA}_1$ (see Fig.1.1) into n equal parts, the length $\frac{1}{n}$ of one of them becomes smaller and smaller. But,...why? What is the mathematical explanation for this?

THEOREM 4. *The sequence $\{\frac{1}{n}\}$ is convergent to* 0.

PROOF. We apply Definition 1. Let $\varepsilon > 0$ be a small positive real number and, by using the Archimedes's Axiom, let N_ε be the unique natural number such that $\frac{1}{\varepsilon} \in [N_\varepsilon - 1, N_\varepsilon)$. So, for any $n \geq N_\varepsilon$, one has that $\frac{1}{\varepsilon} < N_\varepsilon \leq n$, i.e. $\frac{1}{n} < \varepsilon$. □

REMARK 2. *The absolute value or the modulus $|r|$ of the real number r from (1.8) is simply*

$$\overline{a_k a_{k-1} ... a_0}.b_1 b_2 ... b_n ...,$$

i.e. r without minus if it has one. For instance, $|-3.14| = 3.14 = |3.14|$. Since the function dist, which associates to any pair of real number (x, y) the nonnegative real number $|x - y|$, i.e. $dist(x, y) = |x - y|$, has the following basic properties (prove them!):

i) $dist(x, y) = 0$, if and only if $x = y$,

ii) $dist(x, y) = dist(y, x)$,

iii) $dist(x, y) \leq dist(x, z) + dist(z, y)$ (the triangle inequality),

for any x, y, z in $\mathbb{R}$, we say that $dist(x, y) = |x - y|$ is the distance between x and y and that $\mathbb{R}$ together with this distance function dist is a metric space.

Another example of a metric space is the Cartesian plane xOy with the distance function between two points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ given by the formula:

$$dist(M_1, M_2) = \left|\overrightarrow{M_1M_2}\right| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

i.e. the length of the segment $[M_1M_2]$. Here we can see why the property iii) was called "the triangle property" (be conscious of this by drawing a triangle in plane...!).

Now, what is the difference between the rational number field $\mathbb{Q}$ and the real number field $\mathbb{R}$? The first one is that $\mathbb{Q}$ is countable and, as the following result says, $\mathbb{R}$ is not countable, so the subset of irrational numbers is "greater" than the subset of rational numbers.

THEOREM 5. *(Cantor's Theorem). The set $\mathbb{R}$ is not countable, i.e. one can NEVER represent the whole set of the real numbers as a sequence.*

PROOF. Let r be like in (1.8). It is enough to prove that the set S of all the sequences $\{b_1, b_2, ..., b_n, ...\}$, where b_n is a digit, is not countable. Suppose on the contrary, namely that S can be represented like a sequence of ... sequences: $S = \{B_1, B_2, ..., B_n, ...\}$, where

$$B_n = \{b_{n1}, b_{n2}, b_{n3}, ..., b_{nn}, ...\},$$

and b_{nj} are digits. In order to obtain a contradiction, it is enough to construct a new sequence of digits, which is distinct of any B_i for $i = 1, 2, ...$. Let $C = \{c_1, c_2, ..., c_n, ...\}$ with the following property: $c_n = b_{nn} + 1$, if $b_{nn} \neq 9$ and $c_n = 0$, if $b_{nn} = 9$. Now, let us see that C is not in S. Assume that $C = B_k$ for a $k \in \{1, 2, ...\}$. By the definition of c_k, this last one cannot be equal to b_{kk}, thus the k-th term of C is not equal to the k-th term of B_k and so, $C \neq B_k$, a contradiction! Hence $C \notin S$. So S cannot be represented like a sequence. □

It is not difficult to prove that the subset of $\mathbb{R}$ which consists of all the algebraic elements over $\mathbb{Q}$ (roots of polynomials with coefficients in $\mathbb{Q}$) is countable. So, $\mathbb{R}$ contains an uncountable subset of *transcendental numbers* (numbers which are not algebraic). In fact we know very few of them, e, π, $e^{\sqrt{2}}$, etc. A real number which is not rational is called an irrational number. Since any interval (a, b) is in a one-to-one correspondence onto the interval $(0, 1)$ ($f : (0, 1) \to (a, b)$, $f(t) = a + (b - a)t$ is a bijection between $(0, 1)$ and (a, b)) and since tan :

$(-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ is a bijection between $(-\frac{\pi}{2}, \frac{\pi}{2})$ and $\mathbb{R}$, there is a bijection between $\mathbb{R}$ and any nontrivial interval (a, b), does not matter as small as this last interval is.

REMARK 3. *Hence, (a, b) with $a \neq b$ is not countable. Thus, in (a, b) one can find an infinite number of irrational numbers and even an infinite number of transcendental numbers (why?-explain step by step!).*

Can we solve any equation in $\mathbb{R}$? The answer is no! Even the simple equation $X^2 + 1 = 0$, with the coefficients in $\mathbb{Z}$ has no real solution. Why? Because $x = 0$ is not a solution and, if $x \neq 0$, then x^2 is positive (see the multiplication rule of signs!). So, $x^2 + 1$ is greater than 1, thus it cannot be zero. In order to solve this last equation we need to enlarge $\mathbb{R}$ up to another field $\mathbb{C}$, the complex number field. Its algebraic structure is the following. Take the 2-dimensional real vector space $V = \mathbb{R} \times \mathbb{R}$ with the componentwise addition and the componentwise scalar multiplication. Then we introduce a "strange" multiplication:

$$(a, b)(c, d) \stackrel{def}{=} (ac - bd, ad + bc). \tag{1.10}$$

It is not difficult to prove that V together with this multiplication becomes a field in which $(0, 1)^2 = (-1, 0)$, identified with the real number -1, because $a \to (a, 0)$ is a canonical embedding of $\mathbb{R}$ into V. This new field is usually denoted by $\mathbb{C}$. It is clear that $\pm(0, 1)$ are the solutions of the equation $X^2 + 1 = 0$. What is amazing is that C. F. Gauss proved that any polynomial with coefficients in $\mathbb{C}$ has all its roots in $\mathbb{C}$. The algebraists say that $\mathbb{C}$ is *algebraically closed* (it cannot be enlarged by adding to it new roots of polynomials with coefficients in it). Later, Frobenius proved that there is no other superfield of $\mathbb{R}$, which has a finite dimension over it, but $\mathbb{C}$ (which has dimension 2 over $\mathbb{R}$). Here dimension means the dimension of $\mathbb{C}$ as a vector space over $\mathbb{R}$. Since any $z = a + ib$, where $i = (0, 1)$ and a, b are unique real numbers, $\{(1, 0), (0, 1)\}$ is a basis in $\mathbb{C}$. So the dimension of $\mathbb{C}$ over $\mathbb{R}$ is 2.

Let us now come back to our problem relative to the differences between $\mathbb{Q}$ and $\mathbb{R}$. Since $\mathbb{Q}$ is a subfield of $\mathbb{R}$, the Archimedes Axiom also works on $\mathbb{Q}$. But, what about Cantor's Axiom? We know that $\sqrt{2}$ is not in $\mathbb{Q}$. Let us consider the (infinite) decimal representation of $\sqrt{2}$:

$$\sqrt{2} = 1.41b_3b_4...b_n... \tag{1.11}$$

and let us denote by $x_n = 1.41b_3b_4...b_n$, the corresponding n-th simple rational number of $\sqrt{2}$. It is clear that the sequence $\{x_n\}$ is an increasing sequence which converges to $\sqrt{2}$. Let us also consider the following decreasing sequence $\{y_n\}$ of simple rational numbers, convergent to the same $\sqrt{2}$. $y_1 = 1.5$, $y_2 = 1.42, ...$, $y_n = 1.41b_3b_4...b_{n-1}c_nb_{n+1}b_{n+2}...$, where $c_n = b_n + 1$, if $b_n \neq 9$ and $c_n = b_n = 9$, if $b_n = 9$. It is easy to see that the intersection of all the closed intervals $[x_n, y_n]$, $n = 1, 2, ...$, in $\mathbb{Q}$, is empty in $\mathbb{Q}$ (since the intersection in $\mathbb{R}$ is exactly $\sqrt{2}$, which is not in $\mathbb{Q}$). Hence the Cantor axiom does not work for the ordered field $\mathbb{Q}$.

In this last counterexample we needed some tricks, so it will be desirable to have an equivalent statement to the Cantor's Axiom. For this we introduce two important new notions, namely the notion of the *least upper bound* (LUB) and the notion of the *greatest lower bound* (GLB) of a given subset of $\mathbb{R}$. We do everything for the LUB and we leave to the reader to translate all of these in the case of the GLB.

Let A be a nonempty subset in $\mathbb{R}$. A real number z is called an *upper bound* for A if any element a of A is less or equal to z. A *least upper bound* (LUB) for A is (if it does exist!) the least possible z which is an upper bound for A. For instance, the LUB of $A = [0, 7)$ is 7 and the GLB of A is 0. We cannot have two distinct LUB for the same subset A (Why?). If A is (upper) unbounded (i.e. if for any natural number n there is at least one element b of A such that $b > n$), then A has no upper bound in $\mathbb{R}$ and as a logical consequence it has no LUB in $\mathbb{R}$. For instance, $A = [0, \infty)$ has no upper bound in $\mathbb{R}$, but 0 is the GLB of A. $\mathbb{R}$ and $\mathbb{Z}$ have neither an LUB nor a GLB in $\mathbb{R}$.

Usually, the LUB of a subset A is denoted by $\sup A$ (the supremum of A) and the GLB of a subset B is denoted by $\inf B$ (infimum of B).

THEOREM 6. *(LUB test) Let A be a subset of $\mathbb{R}$. Then c is the LUB of A if and only if for any small positive real number $\varepsilon > 0$, there are an element a of A such that $c - \varepsilon < a \leq c$ and an upper bound z of A with $c \leq z < c + \varepsilon$. This is equivalent to saying that any ε-neighborhood of c must simultaneously contain an element a of A and an upper bound z of A (Why?).*

PROOF. Let us suppose that $c = \sup A$. Assume that we found an $\varepsilon > 0$ such that all the elements of A are less or equal to $c - \varepsilon$. So $c - \varepsilon$ is an upper bound of A less than c, a contradiction, because, by definition, c is the least upper bound of A. Hence, there is at least one $a \in A$ in the interval $(c-\varepsilon, c]$. If all the upper bounds of A were greater or equal to $c + \varepsilon$, then c would not be the least upper bound of A and we would obtain again a contradiction.

Conversely, let us assume that c is a real number with the property described in the statement of the above theorem. If c were not $\sup A$, we have two options: 1) c is not an upper bound of A, i.e. there is at least one a greater than c. Taking now $\varepsilon = a - c$ and using our hypothesis for this particular $\varepsilon > 0$, we get an upper bound z of A in the interval $[c, c + \varepsilon = a)$, i.e. z is less than a. This is in contradiction with the fact that z is an upper bound of A. Hence 1) cannot appear. It remains only the second option: 2) c is an upper bound of A, but it is not the least, namely there is another upper bound y which is less than c. Take now $\varepsilon = c - y > 0$ and use again the hypothesis of the theorem for this new ε. So, one can find an element b of A in the interval $(c-\varepsilon = y, c]$. Thus, b is greater than y, which was considered to be an upper bound of A. Again a contradiction! Therefore, the second option is also impossible and the proof is complete. □

The LUB test is very useful because it supply us with some important results.

THEOREM 7. *The following statements are logically equivalent: i) The Cantor Axiom (see Axiom 2) works in* $\mathbb{R}$*, ii) Any upper bounded subset* A *of* $\mathbb{R}$ *has a LUB in* $\mathbb{R}$ *and, iii) Any lower bounded subset* B *of* $\mathbb{R}$ *has a GLB in* $\mathbb{R}$*.*

PROOF. First of all let us see that ii) and iii) are equivalent. Let us prove for instance that ii)$\Rightarrow$ iii). For the lower bounded subset B of $\mathbb{R}$ let us put $-B = \{x \in \mathbb{R} : -x \in B\}$, the symmetric subset of B with respect to the origin O (on the real line (d)). It is not difficult to see that the new subset $-B$ is upper bounded in $\mathbb{R}$ and so, from ii) it has a LUB b in $\mathbb{R}$. We leave the reader (eventually using Theorem 6) to prove that $-b$ is the GLB of B in $\mathbb{R}$.

We leave as an exercise for the reader to prove that iii)$\Longrightarrow$ i).

Now we prove that i)$\Longrightarrow$ ii). Let b_0 be an upper bound of A and let a_0 be an element of A. It is clear that $a_0 \leq b_0$. If $a_0 = b_0$ we have nothing more to prove because the LUB of A will be this common value $c = a_0 = b_0$. Assume that a_0 is less than b_0 an let us divide the closed interval $[a_0, b_0]$ into two equal closed subintervals by the mid point c_0. By the "essential choice" we mean to choose the subinterval $[a_0, c_0]$ if c_0 is an upper bound for A, or to choose the subinterval $[c_0, b_0]$ if there is at least one element $a_1' \in A$ in the second subinterval, $[c_0, b_0]$. After we have performed "the essential choice", let us denote by $[a_1, b_1]$ either the subinterval $[a_0, c_0]$ in the first choice, or the subinterval $[c_0, b_0]$ in the case of the second choice. In both situations $a_1 \in A$, b_1 is an upper bound of A and $a_0 \leq a_1 \leq b_1 \leq b_0$. Now we take the interval $[a_1, b_1]$,

divide it into two equal parts and repeat the "essential choice" for this new interval $[a_1, b_1]$, find $a_2 \in A$ and b_2 an upper bound of A with

$$a_0 \leq a_1 \leq a_2 \leq b_2 \leq b_1 \leq b_0$$

and so on. We obtain two sequences: an increasing one and a decreasing one in the following position:

$$a_0 \leq a_1 \leq ... \leq a_n \leq ... \leq b_n \leq ... \leq b_1 \leq b_0,$$

such that the distance $dist(a_n, b_n) = \frac{dist(a_0,b_0)}{2^n}$. In particular,

$$dist(a_n, b_n) \to 0,$$

whenever $n \to \infty$. Now we can apply the Cantor Axiom and find a unique point c belonging to all the intervals $[a_n, b_n]$ for any $n = 1, 2, ...$, i. e. $\lim a_n = \lim b_n = c$ (Why?). We prove now that this c is exactly $\sup A$. Let us now apply the LUB test (see Theorem 6). Take an ε and let us consider the ε-neighborhood $(c-\varepsilon, c+\varepsilon)$. Since $\lim a_n = \lim b_n = c$, there is an $n \in \{1, 2, ...\}$ such that $[a_n, b_n] \subset (c - \varepsilon, c + \varepsilon)$. But, by the above construction, $a_n \in A$ and b_n is an upper bound of A. So, by the criterion of Theorem 6, we get that $c = \sup A$.

ii)$\Longrightarrow$ i) Let $\{a_n\}$ and $\{b_n\}$ be two sequences of real numbers such that

$$a_0 \leq a_1 \leq ... \leq a_n \leq ... \leq b_n \leq ... \leq b_1 \leq b_0.$$

The subset $A = \{a_0, a_1, ..., a_n, ...\}$ is upper bounded in $\mathbb{R}$ by any term of the second sequence $\{b_n\}$. From ii) we have that A has a LUB $c = \sup A$ and $c \leq b_n$ for any $n = 0, 1, ...$. Since c is in particular an upper bound of A, one also has that $a_n \leq c \leq b_n$ for any $n = 0, 1, ...$. Hence the Cantor Axiom works on $\mathbb{R}$. □

A sequence is said to be *monotonous* if it is either an increasing or a decreasing sequence. For instance, $x_n = \frac{1}{n^2+1}$ and $y_n = -\frac{1}{n^2+1}$ are monotonous sequences.

REMARK 4. *Let us now introduce two symbols: 1) ∞, which is considered to be greater than any real number r, $r + \infty = \infty$, $\infty + \infty = \infty$, and 2) $-\infty$, which is considered to be less then any real number r, $r + (-\infty) = -\infty$, $-\infty - (\infty) = -\infty$, $r \cdot \infty = \infty$, if $r > 0, r \cdot \infty = -\infty$, if $r < 0$. Moreover, $r \cdot (-\infty) = -\infty$ if $r > 0$ and $r \cdot (-\infty) = \infty$, if r is negative. In the same logic,*

$$\infty \cdot \infty = (-\infty) \cdot (-\infty) = \infty, (-\infty) \cdot \infty = -\infty = \infty \cdot (-\infty), \frac{r}{\pm\infty} = 0, etc.$$

The operations $0 \cdot (\pm\infty), \infty - \infty, \frac{0}{0}$ and $\frac{\infty}{\infty}$ are not permitted. We denote by $\overline{\mathbb{R}} = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$ and call it the accomplished (or completed)

real line. By definition, a neighborhood of ∞ is an open interval of the form (M, ∞) and a neighborhood of $-\infty$ is an interval of the form $(-\infty, L)$, where M, L are real numbers. For instance, in $\overline{\mathbb{R}}$ any subset of real numbers is bounded (upper or lower) and an unbounded (in $\mathbb{R}$) increasing sequence is said to be "convergent to ∞" (for example, $x_n = n^3 \to \infty$). But the sequence $y_n = (-1)^n n$ is bounded in $\overline{\mathbb{R}}$ but it is not "convergent" there (Why?). Usually, if a sequence of real numbers is "convergent to ∞" in $\overline{\mathbb{R}}$, we say that it is divergent in $\mathbb{R}$. Sometimes, by abuse, we write $\lim_{n\to\infty} x_n = \infty$ when the sequence $\{x_n\}$ is unbouded and increasing. If $\{x_n\}$ is a sequence in $\overline{\mathbb{R}}$ and if $L(\{x_n\})$ is the set of all the limits of all the convergent subsequences of $\{x_n\}$, we denote by $\limsup\{x_n\}$, the $\sup L(\{x_n\})$ and by $\liminf\{x_n\}$, the $\inf L(\{x_n\})$. For instance, for the sequence $x_n = \sin(\frac{2n+1}{2}\pi) = (-1)^n$, $\limsup x_n = 1$ and $\liminf x_n = -1$ (prove this!).

THEOREM 8. *a) Let $\{x_n\}$ be an increasing sequence in $\mathbb{R}$. Then $\limsup x_n$ exist in $\overline{\mathbb{R}}$ and the sequence is convergent to $\limsup x_n$ in $\overline{\mathbb{R}}$. If $\{x_n\}$ is also upper bounded in $\mathbb{R}$, then $\limsup x_n$ is its limit in $\mathbb{R}$ too, i.e. $\lim x_n = \limsup x_n$. b) Let $\{y_n\}$ be a decreasing sequence in $\mathbb{R}$. Then $\liminf x_n$ always exist in $\overline{\mathbb{R}}$ and the sequence is convergent to $\liminf x_n$ in $\overline{\mathbb{R}}$. If $\{x_n\}$ is also lower bounded in $\mathbb{R}$, then $\liminf x_n$ is also in $\mathbb{R}$ and so $\lim x_n = \limsup x_n$.*

PROOF. We prove only a) and we think that b) is a good exercise for the reader. If $\{x_n\}$ is upper unbounded then, for any real number M, there is at least one n with $x_n \geq M$. Since $\{x_n\}$ is an increasing sequence, $x_{n+p} \geq x_n$ for any $p = 1, 2...$. So, outside the neighborhood (M, ∞) of ∞ we have only a finite number of terms of our sequence, i.e. $x_n \to \infty$, which is at the same time $\limsup x_n$ (Why?). If $\{x_n\}$ is upper bounded, then, using Theorem 7, we get that $c = \limsup x_n$ is a real number. Take now an ε-neighborhood $(c-\varepsilon, c+\varepsilon)$ of c. Since c is the LUB of the set $\{x_n\}$, we can apply Theorem 6 and find an x_m in the interval $(c-\varepsilon, c]$. Since the sequence is increasing, $x_{m+1}, x_{m+2}, ...$ are in the same interval (Why?). So, outside this interval one has at most a finite number of terms of our sequence, i.e. $x_n \to c$ (see Definition 1). □

Let us come back to the approximation of $\sqrt{2} = 1.41b_3b_4...b_n...$ (see (1.11)) by the increasing sequence $x_n = 1.41b_3b_4...b_n$, $n = 1, 2, ...$ of simple rational numbers. This last sequence $\{x_n\}$ is a sequence in $\mathbb{Q}$ but its limit $\sqrt{2}$ is not in $\mathbb{Q}$. However, this sequence has an interesting property. If we fix an $n \in \mathbb{N}$, and if we consider the terms $x_n, x_{n+1}, x_{n+2}, ...x_{n+p}$, we see that the distance between x_n and x_{n+p}

goes to 0 independently of $p \in \mathbb{N}$, but dependently of n. This means that from a rank N on the distance $dist(x_l, x_m)$ becomes smaller and smaller $(l, m \geq N)$. Indeed,

$$dist(x_n, x_{n+p}) = 0.\underbrace{00...0}_{n-times} b_{n+1}b_{n+2}...b_{n+p} \leq 0.\underbrace{00...0}_{n-times} 999... = \frac{1}{10^n} \to 0$$

independently on p, i.e. for any small real number $\varepsilon > 0$, there is a rank N_ε such that whenever $n \geq N_\varepsilon$ one has that $dist(x_n, x_{n+p}) < \varepsilon$, for any $p = 1, 2, ...$.

DEFINITION 2. *Let $\{x_n\}$ be a sequence of real numbers. We say that $\{x_n\}$ is a Cauchy sequence or a fundamental sequence if for any small positive real number $\varepsilon > 0$. there is a rank N_ε (depending on ε) such that $|x_{n+p} - x_n| < \varepsilon$ for any $n \geq N_\varepsilon$ and for any $p = 1, 2, ...$ This means that $|x_{n+p} - x_n| \to 0$, when $n \to \infty$, independently on p.*

For instance, the above sequence $x_n = 1.41b_3b_4...b_n$, $n = 1, 2, ...$ is a Cauchy sequence of rational numbers which is not convergent in $\mathbb{Q}$, but which is convergent in $\mathbb{R}$, its limit being the real number $\sqrt{2}$. This is why we say that $\mathbb{Q}$ is not "complete".

DEFINITION 3. *In general, a metric space X with its distance dist (see Remark 2) is said to be complete if any Cauchy sequence $\{x_n\}$ with terms in X is convergent to a limit x of X.*

Let us consider the following sequence

$$x_n = \frac{\cos 1}{2} + \frac{\cos 2}{2^2} + \frac{\cos 3}{2^3} + ... + \frac{\cos n}{2^n},$$

where the arcs are measured in radians. Let us prove that this last sequence is a Cauchy sequence. For this, let us evaluate the distance

$$dist(x_n, x_{n+p}) = |x_{n+p} - x_n| =$$

$$= \left| \frac{\cos(n+1)}{2^{n+1}} + \frac{\cos(n+2)}{2^{n+2}} + ... + \frac{\cos(n+p)}{2^{n+p}} \right| <$$

$$< \frac{1}{2^{n+1}}(1 + \frac{1}{2} + \frac{1}{2^2} + ...) = \frac{1}{2^n}.$$

This last equality comes from the definition of the infinite geometrical progression

$$1 + \frac{1}{2} + \frac{1}{2^2} + ... \overset{def}{=} \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^n}\right) = \lim_{n\to\infty} \frac{1 - \frac{1}{2^{n+1}}}{1 - \frac{1}{2}} = 2$$

So $dist(x_n, x_{n+p})$ tends to 0 independently of p, because $\frac{1}{2^n}$ goes to 0, whenever $n \to \infty$, independently of p. Indeed, for a small $\varepsilon > 0$, let us find the first natural number N_ε such that $\frac{1}{2^{N_\varepsilon}} < \varepsilon$. Applying $\log_2$ we get $N_\varepsilon > -\log_2 \varepsilon$, so $N_\varepsilon = [-\log_2 \varepsilon] + 1$. Now, if $n \geq N_\varepsilon$,

$$dist(x_n, x_{n+p}) < \frac{1}{2^n} \leq \frac{1}{2^{N_\varepsilon}} < \varepsilon,$$

independently on p.

THEOREM 9. *Any convergent sequence $\{x_n\}$ to x is also a Cauchy sequence. Thus, the class of Cauchy sequences "appears" to be larger then the class of convergent sequences.*

PROOF. We simply verify Definition 2. Let ε be a positive small real number and let N_ε be a rank (dependent on ε) such that $|x_n - x| < \frac{\varepsilon}{2}$ for any $n \geq N_\varepsilon$ (see Definition 1 with $\frac{\varepsilon}{2}$ instead of ε). So,

$$|x_{n+p} - x_n| = |x_{n+p} - x + x - x_n| \leq |x_{n+p} - x| + |x_n - x| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for any $n \geq N_\varepsilon$. Hence our convergent sequence is also a Cauchy sequence. □

A basic result in Mathematics was discovered by Cauchy: "Any fundamental sequence of real numbers is convergent to a real number, i.e. $\mathbb{R}$ is a "complete metric space".

To prove this important result we need some specific properties of the Cauchy sequences.

THEOREM 10. *Any Cauchy sequence $\{x_n\}$ is bounded, i.e. there is a positive real number M such that $|x_n| \leq M$ for any $n = 0, 1, ...$ or, equivalently, if there is an interval $[A, B]$ in $\mathbb{R}$ such that all the terms of the sequence $\{x_n\}$ belong to this interval, i.e. $x_n \in [A, B]$ for any $n = 0, 1, ...$ (Why this equivalence?).*

PROOF. Take an arbitrary positive real number, for instance 2. Since $\{x_n\}$ is a Cauchy sequence, there is a rank N such that whenever $n \geq N$, $|x_{n+p} - x_n| < 2$ for any $p = 1, 2...$ (see Definition 2). In particular, $|x_{N+p} - x_N| < 2$, or $x_{N+p} \in (x_N - 2, x_N + 2)$ for any $p \in \mathbb{N}$. So, outside this last interval one may have at most x_0, $x_1, ..., x_{N-1}$ as terms of our sequence. Take now $A = \min\{x_0, x_1, ..., x_{N-1}, x_N - 2\}$ and $B = \max\{x_0, x_1, ..., x_{N-1}, x_N + 2\}$. It is easy to see that all the terms of the sequence $\{x_n\}$ belong to the interval $[A, B]$. If one takes now $M = \max\{|A|, |B|\}$, then $x_n \in [-M, M]$, or $|x_n| \leq M$ for any $n = 0, 1, ...$. □

Here is a strange property of the Cauchy sequences.

THEOREM 11. *If a Cauchy sequence $\{x_n\}$ contains at least one subsequence $\{x_{k_n}\}$, $(k_0 < k_1 < k_2 < ... < k_n < ...)$ which is convergent to x, then the whole sequence $\{x_n\}$ is convergent to the same x. Therefore, all the other subsequences of $\{x_n\}$ are convergent to x.*

PROOF. Let ε be a small positive real number. Since $\{x_{k_n}\}$ is convergent to to x whenever $n \to \infty$, for n large enough, let us assume that for $n \geq N'$, one has

$$|x_{k_n} - x| < \frac{\varepsilon}{2}. \tag{1.12}$$

Since $\{x_n\}$ is a Cauchy sequence, for n large enough, suppose $n \geq N''$, one has that

$$|x_{n+p} - x_n| < \frac{\varepsilon}{2}, \tag{1.13}$$

for any $p = 1, 2, ...$. Let now N be a natural number greater than N' and than N'', at the same time. Let n be a fixed natural number greater than N and let us choose k_m such that it is greater than this fixed n and m itself is greater than N. So, $k_m = n + p$, for a natural number $p\ (= k_m - n)$. From (1.13) we get that

$$|x_{k_m} - x_n| < \frac{\varepsilon}{2}, \tag{1.14}$$

because $n > N > N''$. From (1.12) one has that

$$|x_{k_m} - x| < \frac{\varepsilon}{2}, \tag{1.15}$$

because $m > N > N'$. Now,

$$|x_n - x| = |x_n - x_{k_m} + x_{k_m} - x| \leq |x_{k_m} - x_n| + |x_{k_m} - x| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

And this is true for any $n > N$. Hence, the sequence $\{x_n\}$ is convergent to x. We leave to the reader to convince himself (or herself) that if a sequence $\{x_n\}$ is convergent to a real number x, then any subsequence of it is also convergent to the same x. □

We prove now a basic property of a bounded infinite subset A of real numbers. For this we give a definition.

DEFINITION 4. *We say that a subset A of real numbers has the point (real number) x as a limit point if there is a sequence $\{a_n\}$, with distinct terms a_n from A, which is convergent to x.*

For instance, 0 is a limit point of

$$A = \{1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...\}$$

and of the interval $[0,1]$. But 0 is NOT a limit point of the set $B = \{0,1,2\}$ (Why?). $\mathbb{N}$ and $\mathbb{Z}$ have no limit points in $\mathbb{R}$! (Why?). Find all the limit points of $\mathbb{Q}$ in $\mathbb{R}$! (Hint: the whole $\mathbb{R}$ is the set of all the limit points of $\mathbb{Q}$, why?)

THEOREM 12. *(Cesaro-Bolzano-Weierstrass Theorem). Any infinite and bounded subset A of $\mathbb{R}$ has at least one limit point in $\mathbb{R}$, i.e. there is an $x \in \mathbb{R}$ and a nonconstant sequence $\{a_n\}$ with $a_n \in A$ for any $n = 0, 1, ...$, such that $a_n \to x$.*

PROOF. Since A is bounded, there is a closed interval $[a_0, b_0]$ ($a_0, b_0 \in \mathbb{R}$) which contains A. Let us divide this last interval into two equal closed subintervals and let denote by $[a_1, b_1]$ that subinterval which contains an infinite number of elements of A. Let x_1 be in $[a_1, b_1]$ and in A, i.e. $x_1 \in [a_1, b_1] \cap A$. Let us divide now the interval $[a_1, b_1]$ into two equal closed subintervals and let us choose that one $[a_2, b_2]$ which contains an infinite number of elements from A. Let x_2 be in $A \cap [a_2, b_2]$ and $x_2 \neq x_1$. We continue to construct subintervals $[a_3, b_3]$, $[a_4, b_4]$, ..., $[a_n, b_n]$, ... and elements x_n of $A \cap [a_n, b_n]$, such that $x_n \notin \{x_1, x_2, ..., x_{n-1}\}$ for any $n = 3, 4, ..., n, ...$. Since the length of the interval $[a_n, b_n]$ is $\frac{l}{2^n}$, where l is $b_0 - a_0$, the length of the initial interval, we can use Cantor Axiom (Axiom 2) and find a unique real number x in the common intersection $\bigcap_{n=0}^{\infty} [a_n, b_n]$ of all the intervals $[a_n, b_n]$. Since x_n and x are in $[a_n, b_n]$, $dist(x_n, x) \leq \frac{l}{2^n}$ so, $x_n \to x$ (see Definition 1). Because x_n, $n = 1, 2, ...$ are distinct elements of A, one has that x is a limit point of A and the theorem is completely proved. □

THEOREM 13. *(Cauchy test 1). Any fundamental (Cauchy) sequence in $\mathbb{R}$ is convergent in $\mathbb{R}$, i.e. $\mathbb{R}$ is a complete metric space. This means that in $\mathbb{R}$ there is no difference between the set of convergent sequences and the set of Cauchy sequences (In $\mathbb{Q}$ there is!-Why?)*

PROOF. Let $\{y_n\}$ be a fundamental sequence in $\mathbb{R}$. If $\{y_n\}$ has only a finite distinct terms then, from a rank on, the sequence becomes a constant sequence, so it would be convergent to the value of the constant terms. Let us assume that $\{y_n\}$ has an infinite number of distinct terms, i.e. that the set $A = \{y_n\}$ is infinite. Since A is bounded (see Theorem 10) and infinite, it has a limit point y (see Theorem 12), i.e. there is a nonconstant subsequence $\{y_{k_n}\}$, $n = 1, 2, ...$ of the sequence $\{y_n\}$, which is convergent to y. We apply now Theorem 11 and find that the whole sequence $\{y_n\}$ is convergent to y. □

This theorem has not only a great theoretical importance, but a practical one too. For instance, take again the sequence

$$x_n = \frac{\cos 1}{2} + \frac{\cos 2}{2^2} + \frac{\cos 3}{2^3} + ... + \frac{\cos n}{2^n}.$$

We proved that $\{x_n\}$ is a Cauchy sequence. Now, we know (see Theorem 13) that it is also a convergent sequence to an unknown limit (we cannot express this limit as a decimal fraction!) x. Knowing that $x_n \to x$ is a very good situation! For a large n we can approximate x with x_n. But this last one can be easily computed with an usual computer. So, we have a good idea about the limit. Moreover, the Cauchy test 1 is useful to check if a sequence is convergent or not. For instance, the sequence $\{a_n\}$ is recurrently defined: $a_0 = 0$, $a_n = \sqrt{2 + a_{n-1}}$ for $n = 1, 2, ...$. Let us prove that it is a Cauchy sequence. Indeed,

$$a_n - a_{n-1} = \sqrt{2 + a_{n-1}} - \sqrt{2 + a_{n-2}} = \tag{1.16}$$

$$\frac{a_{n-1} - a_{n-2}}{\sqrt{2 + a_{n-1}} + \sqrt{2 + a_{n-2}}} < \frac{1}{2}(a_{n-1} - a_{n-2}).$$

We can apply (1.16) $(n-1)$-times and find

$$a_n - a_{n-1} < \frac{1}{2}(a_{n-1} - a_{n-2}) < \frac{1}{2^2}(a_{n-2} - a_{n-3}) < ... < \frac{1}{2^{n-1}}(a_1 - a_0).$$

So,

$$a_{n+p} - a_n = a_{n+p} - a_{n+p-1} + a_{n+p-1} - a_{n+p-2} + ... + a_{n+1} - a_n <$$

$$< (\frac{1}{2^{n+p-1}} + \frac{1}{2^{n+p-2}} + ... + \frac{1}{2^n})(a_1 - a_0) <$$

$$< \frac{1}{2^n}(1 + \frac{1}{2} + \frac{1}{2^2} + ...)(a_1 - a_0) = \frac{1}{2^{n-1}}(a_1 - a_0).$$

Here we just used that

$$1 + \frac{1}{2} + \frac{1}{2^2} + ... \stackrel{def}{=} \lim_{n\to\infty}(1 + \frac{1}{2} + ... + \frac{1}{2^n}) = \lim \frac{1 - \frac{1}{2^{n+1}}}{1 - \frac{1}{2}} = 2.$$

Since $\{a_n\}$ is an increasing sequence (Why?), one has that

$$|a_{n+p} - a_n| < \frac{1}{2^{n-1}}(a_1 - a_0),$$

so, $|a_{n+p} - a_n|$ can be made as small as we want when $n \to \infty$, independently on p. Thus, $\{a_n\}$ is a Cauchy sequence (see Definition 2). Hence $\{a_n\}$ is convergent to a limit l (see Cauchy test 1). As we shall

see in the following theorem (Theorem 14), we can apply the "operation" lim to the equality: $a_n = \sqrt{2 + a_{n-1}}$ and find: $l = \sqrt{2 + l}$, or $l = 2$. Therefore, $\lim_{n\to\infty} a_n = 2$.

Now, we describe some compatibilities of the "operation" lim (which associates to a convergent sequence its limit), with the algebraic operations "$+$", "$-$", "$\cdot$", "$\div$", with the order relation "$\leq$", with the functions x^m, $\sqrt[m]{x}$, $\exp x$, $\ln x$, a^x, $\log_a, a > 0$, $\sin x$, $\cos x$, $\tan x$, $\cot x$ and with their compositions. This means, ... with all the elementary functions. We recall a basic definition:

DEFINITION 5. *Let (X, d_1) and (Y, d_2) be two metric spaces and let $f : X \to Y$ be a mapping defined on X with values in Y. We say that f is continuous at $x \in X$ (with respect to these metric space structures) if for any convergent sequence $\{x_n\}$ in X, $\{x_n\} \to x$, i.e. $d_1(x_n, x) \to 0$ as $n \to \infty$, one has that the corresponding sequence of the images, $\{f(x_n)\}$ is convergent to $f(x)$ in Y, i.e. $d_2(f(x_n), f(x)) \to 0$, when $n \to \infty$. If f is continuous at any x of X, we say that f is continuous in X.*

All the elementary functions (polynomials, rational functions, power functions, exponential and logarithmic functions, trigonometric functions and their compositions) are continuous on their definition domains. To prove this, it is not always so easy. For instance, what do we mean by $3^{\sqrt{2}}$? First of all, we define $3^{\frac{1}{m}}$, $m = 1, 2, ...$, by the unique positive real root of the equation $X^m - 3 = 0$. Then we define $3^{\frac{n}{m}} \overset{def}{=} \left(3^{\frac{1}{m}}\right)^n$. By $3^{-\frac{5}{7}}$ we understand $\frac{1}{3^{\frac{5}{7}}}$. Then, we approximate $\sqrt{2}$ with an increasing sequence $\{r_n\}$ of rational numbers, i.e. $r_n \to \sqrt{2}$ and $r_n < r_{n+1}$ for any $n = 1, 2, ...$ As we know, we simply take for r_n the rational number $1.b_1b_2...b_n$, i.e. we get out all the decimals of $\sqrt{2}$ from the $(n+1)$-th decimal on. Now, by definition, $3^{\sqrt{2}} = \lim_{n\to\infty} 3^{r_n}$. To prove the existence of this limit is not an easy task. It is sufficient to prove that the sequence $\{3^{r_n}\}$ is a Cauchy sequence. But,... even this one is difficult! So, the proof of the continuity of the power function $x \to 3^x$ is not so easy at all! This is why we tacitly assume that all the elementary functions are continuous.

THEOREM 14. *Let $\{x_n\}$ and $\{y_n\}$ be two convergent sequences to x and to y respectively. Then:*

a) $\{x_n \pm y_n\} \to x \pm y$,

b) $\{x_n y_n\} \to xy$,

c) If y_n and y are not zero for any $n = 0, 1, ...$, then $\{\frac{x_n}{y_n}\} \to \{\frac{x}{y}\}$.

d) If $x_n \leq y_n$ for any $n = 0, 1, ...$, then $x \leq y$,

e) $\{(x_n)^m\} \to x^m$ *for any fixed natural number* m,

f) $\sqrt[m]{x_n} \to \sqrt[m]{x}$ *if* m *is odd and, for* $x_n \geq 0$, $\sqrt[m]{x_n} \to \sqrt[m]{x}$ *for any natural number* m,

g) $\{\exp x_n\} \to \exp x$ *and, if* $x_n > 0$, *then* $\{\ln x_n\} \to \ln x$,

h) $\{a^{x_n}\} \to a^x$ *and, if* $x_n > 0, \{\log_a x_n\} \to \log_a x$ *for any fixed* $a > 0$,

i) $\sin x_n \to \sin x$, $\cos x_n \to \cos x$, $\tan x_n \to \tan x$, $\cot x_n \to \cot x$,

PROOF. (partially) *a)* Let us prove for instance that $\{x_n + y_n\} \to x + y$. For this, let us evaluate the difference:

$$|x_n + y_n - (x + y)| = |(x_n - x) + (y_n - y)| \leq |x_n - x| + |y_n - y|.$$

But $|x_n - x| \to 0$ and $|y_n - y| \to 0$, so their sum tends to 0 too (Why?). Thus, $|x_n + y_n - (x + y)|$ also goes to 0.

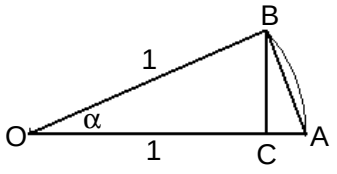
d) Assume that $x > y$ and take $c = \frac{x-y}{2}$. Let us consider the open intervals: $I = (y - c, y + c)$ and $J = (x - c, x + c)$. Since $x_n \to x$ and $y_n \to y$, for a large n one can find $x_n \in J$ and $y_n \in I$. But any element of I is less than any element of J. Hence $y_n < x_n$ and we obtain a contradiction, because, for any n, one has in the hypothesis of *d)* that $x_n \leq y_n$.

i) Let us prove for instance that $\sin x_n \to \sin x$, whenever $x_n \to x$. First of all we remark that $|\sin \alpha| = \sin |\alpha|$ for any $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$. Since $x_n \to x$, one can take n large enough such that $x_n - x \in (-\frac{\pi}{2}, \frac{\pi}{2})$. If α is measured in radians and $\alpha \in (-\frac{\pi}{2}, \frac{\pi}{2})$ then, an easy geometrical construction (see Fig.1.2) tell us that $\sin |\alpha| \leq |\alpha|$.

Let us use now some trigonometry:

$$|\sin x_n - \sin x| = 2 \left| \sin \frac{x_n - x}{2} \cos \frac{x_n + x}{2} \right| \leq 2 \cdot \left| \frac{x_n - x}{2} \right| = |x_n - x|,$$

so $|\sin x_n - \sin x| \to 0$, whenever $x_n \to x$. □



$|BC| = \sin|\alpha| \leq |BA| <$ lenght (arcBA) $= |\alpha|$

Fig. 1.2

COROLLARY 1. *Let* $f : A \to B$ *and* $g : B \to C$ *(*A, B, C *are subsets in* $\mathbb{R}$*) be two functions with the following property: If* $f(x_n) \to f(x)$ *and* $g(y_n) \to g(y)$ *for ANY convergent sequences* $\{x_n\}$ *to* x *and* $\{y_n\}$

to y, then $(g \circ f)(x_n) \to (g \circ f)(x)$. The functions f and g considered here are continuous on their definition domains in the sense of Definition 5. So, the composition between two continuous functions is also a continuous function. Moreover,the sum, the difference, the product and the quotient of two continuous functions is also a continuous function.

PROOF. Since f and g are continuous (see the definition in the statement of the theorem) then, $x_n \to x$ implies $f(x_n) \to f(x)$ (continuity of f). Since g is continuous, $g(f(x_n)) \to g(f(x))$, i.e. $(g \circ f)(x_n) \to (g \circ f)(x)$. Thus $g \circ f$ is also continuous. The other statements are easy consequences of some of the previous statements of the above theorem (prove them!). □

2. Sequences of complex numbers

Let $\mathbb{C}$ be the complex number field. Since any element z of $\mathbb{C}$ is a pair $z = (x, y)$ of two real numbers and since the element $i = (0, 1)$ has the property that $i(y, 0) = (0, y)$ (see the multiplication rule defined in (1.10)), we can write $z = x + iy$, where we identify $(x, 0)$ and $(y, 0)$ with x and y respectively. Let us fix a Cartesian coordinate system $\{O; \mathbf{i}, \mathbf{j}\}$ in a plane (P). Here $\mathbf{i}$ and $\mathbf{j}$ are orthogonal versors and they give the directions and the orientations of the Ox-axis and Oy-axis respectively. Since any vector $\overrightarrow{OM}$, where M is an arbitrary point in the plane (P), can be uniquely written as: $\overrightarrow{OM} = x\mathbf{i} + y\mathbf{j}$, where $x, y \in \mathbb{R}$, we call x and y the coordinates of the point M. Write $M(x, y)$. The association $z = x + iy \longleftrightarrow M(x, y)$ give rise to a geometrical representation of the complex number field $\mathbb{C}$. This is way we always call $\mathbb{C}$, the complex plane. The distance d between two complex numbers $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is simply the distance between their corresponding points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ respectively, i.e.

$$d(z_1, z_2) \overset{def}{=} \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

It is not difficult to check the three properties of a distance function for this d.

A sequence $\{z_n\}$ of complex numbers is said to be convergent to z if the numerical sequence of real numbers $\{d(z_n - z)\}$ is convergent to 0. For instance, $z_n = \frac{1}{n} + (1 + \frac{1}{n})^n i$ is convergent to ei because

$$d(z_n, ei) = \sqrt{(\frac{1}{n} - 0)^2 + [(1 + \frac{1}{n})^n - e]^2} \to 0.$$

The sequence $\{z_n\}$ is said to be fundamental (or Cauchy) if for any $\varepsilon > 0$, there is a natural number N_ε (depending of ε) such that $d(z_{n+p}, z_n) < \varepsilon$ for any $n \geq N_\varepsilon$ and for any $p = 1, 2, \ldots$.

The following result reduces the study of the convergence of a sequence $z_n = x_n + iy_n$ in $\mathbb{C}$ to the study of the convergence of the real and imaginary part $\{x_n\}$ and $\{y_n\}$ respectively.

THEOREM 15. *Let $\{z_n = x_n + y_n i\}$ be a sequence of complex numbers (here x_n and y_n are real numbers). Then the sequence $\{z_n\}$ is convergent to the complex number $z = x + yi$ if and only if $x_n \to x$ and $y_n \to y$ as sequences of real numbers.*

PROOF. One has the following double implications:

$$z_n \to z \Leftrightarrow d(z_n, z) = \sqrt{(x_n - x)^2 + (y_n - y)^2} \to 0 \Leftrightarrow x_n - x \to 0$$

and $y_n - y \to 0$ (simultaneously), i.e. if and only if $x_n \to x$ and $y_n \to y$. □

The sequence $z_n = 3 + (2n \sin \frac{1}{n})i$ tends to $3 + 2i$ because $3 \to 3$ and $2n \sin \frac{1}{n} = 2\frac{\sin \frac{1}{n}}{\frac{1}{n}} \to 2$.

THEOREM 16. *Relative to the distance d, the complex number field $\mathbb{C}$ is complete, i.e. any Cauchy sequence $\{z_n\}$ of $\mathbb{C}$ is convergent to a complex number z.*

PROOF. Let $z_n = x_n + y_n i$, where x_n and y_n are real numbers. Since $\{z_n\}$ is a Cauchy sequence if and only if $d(z_{n+p}, z_n)$ is as small as we want when n is large enough, independent on $p = 1, 2, ...$ and since

$$d(z_{n+p}, z_n) = \sqrt{(x_{n+p} - x_n)^2 + (y_{n+p} - y_n)^2},$$

one sees that $|x_{n+p} - x_n|$ and $|y_{n+p} - y_n|$ are simultaneously small enough whenever n is large enough, independent on p. But this is equivalent to saying that $\{x_n\}$ and $\{y_n\}$ are both Cauchy sequences. Since $\mathbb{R}$ is complete (see Theorem 13), $\{x_n\}$ is convergent to a real number x and $\{y_n\}$ is convergent to another real number y. Let us put $z = x + yi$. Applying now Theorem 15 we get that z_n is convergent to z. □

We say that a subset A of $\mathbb{C}$ is bounded if there is a sufficiently large ball $B(0, r) = \{z \in \mathbb{C} \mid |z| = d(0, z) < r\}$, with centre at 0 and of radius $r > 0$, such that $A \subset B(0, r)$. We also have for $\mathbb{C}$ a Bolzano-Weierstrass type theorem. Namely, any infinite bounded sequence $\{z_n\}$ of complex numbers has a convergent subsequence. If we add a symbol ∞ to $\mathbb{C}$ with similar properties like the infinite ∞ for $\mathbb{R}$, we get $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the Riemann sphere. It is easy to see that in $\overline{\mathbb{C}}$ any sequence has a convergent subsequence. Because of this last property, we say that $\overline{\mathbb{C}}$ and $\overline{\mathbb{R}}$ are the "compactifications" of $\mathbb{C}$ and of $\mathbb{R}$ respectively.

Generally, in a metric space (A, d) a subset M is said to be compact if any sequence of M has at least a convergent subsequence with its limit in M. For instance, any closed interval $[a, b]$ is a compact subset of $\mathbb{R}$ (because of Bolzano-Weierstrass Theorem). A subset C of $\mathbb{C}$ is said to be closed if for any sequence $\{z_n\}$ of elements in C, which is convergent to z in $\mathbb{C}$, its limit z is also in C. Then, the compact subsets of $\mathbb{C}$ are exactly the closed and bounded subsets of $\mathbb{C}$ (have you any idea to prove this?-try a similar idea like that one from the real line situation!)

3. Problems

1. Prove that the following subsets of $\mathbb{R}$ have the same cardinal:

a) $A = (0, 1)$ and $B = \mathbb{R}$, b) $A = (0, 1]$ and $B = \mathbb{R}$, c) $A = (-\infty, a)$ and $B = \mathbb{R}$, d) $A = (0, 1)$ and $B = (a, b)$, e) $A = (a, \infty)$ and $B = (0, 1]$, f) $A = \mathbb{Q} \cap [0, 3]$ and $B = \mathbb{Q} \cap [-7, 3]$.

2. Prove that $\sup(A + B) = \sup A + \sup B$ and, if $A, B \subset [0, \infty)$, then $\sup(A \cdot B) = \sup A \cdot \sup B$, where $A + B = \{x + y \mid x \in A,\, y \in B\}$ and $A \cdot B = \{xy \mid x \in A,\, y \in B\}$. Define $\inf A$ and prove the same equalities for inf instead of sup .

3. Construct $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ and prove that any sequence of elements in $\overline{\mathbb{R}}$ has a convergent subsequence in $\overline{\mathbb{R}}$. Prove that if a sequence $\{x_n\}$ is convergent in $\overline{\mathbb{R}}$, then it has only one limit point, namely the limit of the sequence. Find the limit points for the sequence $a_n = \cos \frac{n\pi}{3}$, $n = 0, 1, 2, \ldots$. Recall that $x \in M$ is a limit point of a subset A of a metric space (M, d) if there is a nonconstant sequence $\{x_n\}$ of elements from A, which is convergent to x.

4. Prove that if $\frac{a_{n+1}}{a_n} \to l$, where $a_n > 0$ for any n, then $\sqrt[n]{a_n} \to l$. Apply this result to compute the limit: $\lim \sqrt[n]{\frac{(2n)!}{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (4n+1)}}$, whenever $n \to \infty$.

5. Prove that the set $\mathbb{R} \setminus \mathbb{Q}$ of irrational numbers is not countable. Prove that it has the same cardinal as the cardinal of $\mathbb{R}$ (i.e. there is a bijection between $\mathbb{R} \setminus \mathbb{Q}$ and $\mathbb{R}$).

6. Prove that the length of the diagonal of a square which has the side a rational number, is not a rational number.

7. Are $\sqrt[3]{5}$ and $\sqrt[7]{3}$ rational numbers? Are they algebraic numbers?

8. Prove that the metric space $([0, 1), d)$, where $d(x, y) = |x - y|$, is not a complete metric space, i.e. there is at least a Cauchy sequence $\{x_n\}$, $x_n \in [0, 1)$, which has no limit in $[0, 1)$. Prove that this limit must be 1.

9. Define the notion of "boundedness" in a general metric space. Is Cesaro's Lemma (any infinite bounded sequence has at least a convergent subsequence) true in a general metric space? Find a simple counterexample.

10. Why a decreasing sequence always has a limit in $\overline{\mathbb{R}}$? If instead of $\overline{\mathbb{R}}$ you put $\overline{\mathbb{Q}} = \mathbb{Q} \cup \{-\infty, \infty\}$, is the last statement also true?

11. Prove that the Archimedes' Axiom is equivalent to the fact that $\lim_{n\to\infty} \frac{1}{n} = 0$. If instead of this last limit we put $\lim_{n\to\infty} \frac{2n+3}{3n-2} = \frac{2}{3}$, does our statement work too?

CHAPTER 2

Series of numbers

1. Series with nonnegative real numbers

We know to add a finite number of real numbers $a_1, a_2, ..., a_n$:

$$s_n = (... ((a_1 + a_2) + a_3) + ...) + a_{n-1}) + a_n)$$

For instance,

$$s_4 = 7 + 3 + (-4) + 5 = 10 + (-4) + 5 = 6 + 5 = 11.$$

However, we have just met infinite sums when we discussed about the representation of a real number as a decimal fraction. For instance,

$$s = 3.3444... = 3.3(4) = 3 + \frac{3}{10} + \frac{4}{10^2} + \frac{4}{10^3} + ... =$$

$$= \lim_{n\to\infty}(3 + \frac{3}{10} + \frac{4}{10^2} + \frac{4}{10^3} + ... + \frac{4}{10^n}) =$$

$$= \frac{33}{10} + \frac{4}{10^2}\lim_{n\to\infty}\frac{1 - \frac{1}{10^{n-1}}}{1 - \frac{1}{10}} = \frac{301}{90}.$$

Generally, if m and n are digits, then

$$0.m(n) = \frac{\overline{mn} - m}{90}$$

(Prove it!).

Since such infinite sums (called series) appear in many applications of Mathematics, we start here a systematic study of them.

DEFINITION 6. *Let $\{a_n\}$ be a sequence of real numbers. The infinite sum*

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + ... + a_n + ... \tag{1.1}$$

is by definition the value (if this one exists) of the limit $s = \lim_{n\to\infty} s_n$, where $s_n = a_0 + a_1 + ... + a_n$ is called the partial sum of order n. The new mathematical object defined in (1.1) is said to be the series of general term a_n and of sum s (if the limit exists). If s exists we say that the

series (1.1) is convergent. If the limit does not exist we say that the series (1.1) is divergent.

For instance, the series

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = \lim_{n\to\infty}(1 + \frac{1}{2} + \frac{1}{2^2} + ... + \frac{1}{2^n}) = 2$$

is convergent to 2, or its sum is 2, whereas the series $\sum\limits_{n=0}^{\infty} n = \infty$, or $\sum\limits_{n=0}^{\infty}(-1)^n$ are divergent. The last divergent series is said to be oscillatory because its partial sums have the values 0 or 1, i.e. it oscillates between the distinct values $\{0, 1\}$.

THEOREM 17. *Let x be a real number. The geometrical series $\sum\limits_{n=0}^{\infty} x^n$ is convergent (and its sum is $\frac{1}{1-x}$) if and only if $|x|$ is less then* 1.

PROOF. By Definition 6,

$$\sum_{n=0}^{\infty} x^n = \lim_{n\to\infty}(1 + x + x^2 + ... + x^n) = \lim_{n\to\infty} \frac{1 - x^{n+1}}{1 - x}.$$

Since $\lim\limits_{n\to\infty} x^{n+1}$ exists and is finite if and only if $|x| < 1$ (when the limit is 0), the series $\sum\limits_{n=0}^{\infty} x^n$ is convergent if and only if $|x| < 1$. In this last case, its sum is $s = \lim\limits_{n\to\infty} \frac{1-x^{n+1}}{1-x} = \frac{1}{1-x}$. For instance, if $x = 1$, then the series becomes $1+1+1+... = \infty$ (in $\overline{\mathbb{R}}$). If $x > 1$, then $\lim\limits_{n\to\infty} x^{n+1} = \infty$. If $x \leq -1$, then the sequence $\{x^{n+1}\}$ has no limit at all (why?) so $\lim\limits_{n\to\infty} \frac{1-x^{n+1}}{1-x}$ also does not exist. □

THEOREM 18. *(The Cauchy general test) A series $\sum\limits_{n=0}^{\infty} a_n$ is convergent if and only if the sequence of partial sums $\{s_n\}$ is a Cauchy sequence, i.e. for any small real number $\varepsilon > 0$, there is a natural number N_ε such that*

$$|a_{n+1} + a_{n+2} + ... + a_{n+p}| < \varepsilon$$

for any $n \geq N_\varepsilon$ and for any $p = 1, 2, ...$

PROOF. We only use the fact that $\mathbb{R}$ is complete, i.e. that the sequence $\{s_n\}$ is convergent if and only if it is a Cauchy sequence. □

COROLLARY 2. *(The zero test) If the sequence* $\{a_n\}$ *does not tend to zero, then the series* $\sum_{n=0}^{\infty} a_n$ *is divergent. Or, if the series* $\sum_{n=0}^{\infty} a_n$ *is convergent, then* $a_n \to 0$.

PROOF. If the series $\sum_{n=0}^{\infty} a_n$ was convergent, then the sequence of partial sums $\{s_n\}$ would be a Cauchy sequence (see Theorem 18). Thus, for n large enough, $a_n = s_n - s_{n-1}$ becomes smaller and smaller, i.e. $a_n \to 0$. In fact, we do not need the previous theorem. Indeed, let $s = \sum_{n=0}^{\infty} a_n$ and write $a_n = s_n - s_{n-1}$. Then, $\lim a_n = s - s = 0$. □

For instance, $\sum_{n=0}^{\infty} \left(\frac{n+1}{n}\right)^n$ is divergent, because $a_n = \left(\frac{n+1}{n}\right)^n \to e \neq 0$.

THEOREM 19. *(The renouncement test) Let us consider the series:* $\sum_{n=0}^{\infty} a_n$ *and* $\sum_{n=N}^{\infty} a_n = a_N + a_{N+1} + ...$ *(we just got out the terms* $a_0, a_1, ..., a_{N-1}$ *in the previous series). Then these two series have the same nature (i.e. they are convergent or divergent) at the same time. Moreover, if they are convergent, then* $s = s' + a_0 + a_1 + ... + a_{N-1}$, *where* $s = \sum_{n=0}^{\infty} a_n$ *and* $s' = \sum_{n=N}^{\infty} a_n$.

PROOF. Let n be large enough ($n \geq N$) and let $s_n = a_0 + a_1 + ... + a_{N-1} + a_N + ... + a_n$. If we denote $s'_n = a_N + ... + a_n$, then s'_n is the partial sum of order n of the series s'. It is clear that $s_n = s'_n + a_0 + a_1 + ... + a_{N-1}$ and that the sequences $\{s_n\}$ and $\{s'_n\}$ are convergent or divergent at the same time (prove it!). Now, in the last equality, let us make $n \to \infty$. We get: $s = s' + a_0 + a_1 + ... + a_{N-1}$ and the proof is completed. □

Let $\sum_{n=0}^{\infty} a_n$ be a series with

$$a_n = n, \text{ if } n \leq 100 \text{ and } a_n = \frac{1}{3^n}, \text{ if } n > 100.$$

The question is:"What is the nature of this series?" So we must decide if our series is convergent or not. Let us renounce the terms $a_0, a_1, ..., a_{100}$ in the initial series. We get a new series

$$\sum_{n=101}^{\infty} \frac{1}{3^n} = \frac{1}{3^{101}}(1 + \frac{1}{3} + \frac{1}{3^2} + ...).$$

Let us use now Theorem 17 and find that

$$\sum_{n=0}^{\infty} a_n = 0 + 1 + ... + 100 + \frac{1}{3^{101}} \frac{1}{1 - \frac{1}{3}} = \frac{100 \cdot 101}{2} + \frac{1}{2 \cdot 3^{100}}.$$

THEOREM 20. *(The boundedness test) Let* $\sum_{n=0}^{\infty} a_n$ *be a series with nonnegative terms* ($a_n \geq 0$). *Then the series is convergent if and only if the partial sums sequence* $\{s_n\}$, $s_n = a_0 + a_1 + ... + a_n$, *is bounded.*

PROOF. Let us assume that the series $\sum_{n=0}^{\infty} a_n$ is convergent, i.e. the sequence $\{s_n\}$ is convergent. Since any convergent sequence is bounded (see also Theorem 10), one has that $\{s_n\}$ is bounded.

Conversely, we suppose that $\{s_n\}$ is bounded. Since $a_n \geq 0$, $s_n \leq s_{n+1}$, i.e. the sequence $\{s_n\}$ is increasing. But Theorem 8 says that an increasing and bounded sequence $\{s_n\}$ is convergent to its superior limit $\limsup s_n$. Thus the series $\sum_{n=0}^{\infty} a_n$ is convergent to this $\limsup s_n$, i.e. its sum $s = \limsup s_n$. □

THEOREM 21. *(The integral test) Let c be a fixed real number and let* $f : [c, \infty) \to [0, \infty)$ *be a decreasing continuous function (see Definition 5). Let* n_0 *be a natural number greater or equal to c. For any* $n \geq n_0$ *let* $a_n = f(n)$ *and let* $A_n = \int_{n_0}^{n} f(x)dx$ *for* $n \geq n_0$. *Then the series* $\sum_{n=n_0}^{\infty} a_n$ *is convergent if and only if the sequence* $\{A_n\}$ *is convergent (it is sufficient to be bounded-why?).*

PROOF. Suppose that the series $\sum_{n=n_0}^{\infty} a_n = \sum_{n=n_0}^{\infty} f(n)$ is convergent. Since in Fig.2.1 $s_n = f(n_0) + ... + f(n)$ is exactly the sum of the hatched and of the double hatched areas and since the integral $A_n = \int_{n_0}^{n} f(x)$ dx is equal to the area under the graphic of $y = f(x)$ which corresponds to the interval $[n_0, n]$, then $A_n \leq s_n$. Since $\sum_{n=n_0}^{\infty} a_n$ is convergent, the sequence $\{s_n\}$ is bounded, thus the sequence $\{A_n\}$ is bounded.

Conversely, let us assume that the sequence $\{A_n\}$ is bounded. Look again at Fig.2.1! We see that the double hatched area is just equal to $a_{n_o+1} + a_{n_0+2} + ... + a_{n+1} = s_{n+1} - a_{n_0}$. Since this double hatched area is less then the area $A_{n+1} = \int_{n_0}^{n+1} f(x)\ dx$, one has that the sequence $\{s_{n+1} - a_{n_0}\}$ is bounded. Hence the sequence $\{s_n\}$ is also bounded

(why?). Now, Theorem 20 tells us that the series $\sum_{n=n_0}^{\infty} a_n$ is convergent.

□

Why we say that if $\lim_{n\to\infty} f(x) \neq 0$, then the above series is divergent?

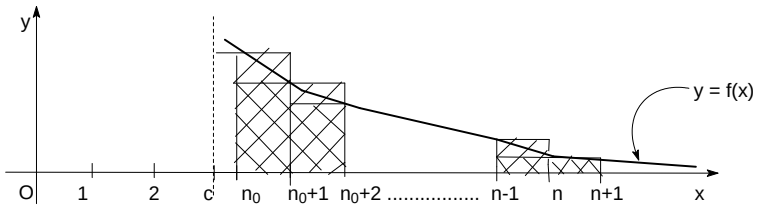


Fig. 2.1

The integral test is very useful in practice. Suppose that somebody is interested in the nature of the series $\sum_{n=2}^{\infty} \frac{1}{n\ln(n)}$. Let us apply the integral test and consider the associated decreasing continuous function

$$f : [2, \infty) \to [0, \infty), f(x) = \frac{1}{x \ln x}$$

(we simply put x instead of n in $a_n = \frac{1}{n\ln(n)}$ for $n \geq 2$). Since

$$A_n = \int_2^n \frac{1}{x \ln x} dx = \ln(\ln(x))|_2^n = \ln(\ln n) - \ln(\ln(2)) \to \infty,$$

A_n is unbounded, thus our series is divergent (see Theorem 21).

In the last 150 years one of the most interesting function in Mathematics, which was highly considered, is the Zeta function of Riemann. "Zeta" comes from the Greek letter ζ. The notation of this function was firstly used by the great German mathematician B. Riemann. Its analytic expression is:

$$\zeta(\alpha) = \sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}, \alpha \in \mathbb{R} \tag{1.2}$$

This famous function is usually defined by a series. Thus, the maximal domain of definition for this function is exactly the set of all $\alpha \in \mathbb{R}$ with the property that the numerical series $\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}$ is convergent. We call this last set, the *set of convergence* of our series. In the following, using the integral test, we find the convergence set for the *Riemann (zeta) series* $\sum_{n=1}^{\infty} \frac{1}{n^{\alpha}}$.

THEOREM 22. *(Riemann zeta series) The Riemann zeta series is convergent if and only if $\alpha > 1$. This means that the real definition domain of the function ζ is the interval $(1, \infty)$.*

PROOF. Let us take in Theorem 21 $f(x) = \frac{1}{x^\alpha}$ for $x \geq 1$. Since

$$A_n = \int_1^n \frac{1}{x^\alpha} dx = \frac{1}{1-\alpha}[n^{-\alpha+1} - 1] \text{ if } \alpha \neq 1$$

and $A_n = \ln n$, if $\alpha = 1$, then A_n is bounded if and only if $\alpha > 1$(why?). Now, Theorem 21 says that the Riemann series $\sum\limits_{n=1}^{\infty} \frac{1}{n^\alpha}$ is convergent if and only if $\alpha > 1$. □

The sum

$$s = 1 + \frac{1}{2} + \frac{1}{3} + ... = \sum_{n=1}^{\infty} \frac{1}{n} = \zeta(1) = \infty,$$

because the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^\alpha}$ is divergent for $\alpha = 1$, thus the sequence of partial sums

$$s_n = 1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n}$$

is strictly increasing and unbounded. Hence $s = \lim s_n = \infty$. The Theorem 22 says that the series

$$\zeta(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + ...$$

is convergent. So it can be approximated by

$$s_N = 1 + \frac{1}{2^2} + \frac{1}{3^2} + ... + \frac{1}{N^2}$$

for N large enough. We call the series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ the *harmonic series*. It is very important in Analysis. Sometimes the following test is useful.

THEOREM 23. *(The Cauchy's compression test) Let $\{a_n\}$ be a decreasing sequence of nonnegative real numbers. Then the series $\sum\limits_{n=0}^{\infty} a_n$ and $\sum\limits_{n=0}^{\infty} 2^n a_{2^n}$ have one and the same nature, i.e. they are simultaneous convergent or divergent.*

PROOF. Let $s_k = \sum\limits_{n=0}^{k} a_n$ and $S_m = \sum\limits_{n=0}^{m} 2^n a_{2^n}$ be the k-th and the m-th partial sums of the first and of the second series respectively.

Let us fix k and let us take a m such that $k \leq 2^m - 1$. Then,

$$s_k = a_0 + a_1 + ... + a_k \leq a_0 + a_1 + ... + a_{2^m-1} = a_0 + a_1 + (a_2 + a_3) +$$

$$+(a_4 + a_5 + a_6 + a_7) + ... + (a_{2^{m-1}} + a_{2^{m-1}+1} + a_{2^{m-1}+2} + ... + a_{2^m-1}) \leq$$

$$\leq a_0 + a_1 + 2a_2 + 2^2 a_{2^2} + ... + 2^{m-1} a_{2^{m-1}} = a_0 + S_{m-1},$$

So

$$s_k \leq a_0 + S_{m-1} \tag{1.3}$$

Now, if the series $\sum_{n=0}^{\infty} 2^n a_{2^n}$ is convergent, then the increasing sequence $\{S_m\}$ is bounded. The inequality (1.3) says that the sequence $\{s_k\}$ is also bounded, thus the series $\sum_{n=0}^{\infty} a_n$ is convergent (see Theorem 20). If $\sum_{n=0}^{\infty} a_n$ is divergent, then the sequence $\{s_k\}$ is unbounded. From (1.3) we see that the sequence $\{S_m\}$ is also unbounded, so the series $S = \sum_{n=0}^{\infty} 2^n a_{2^n}$ is divergent.

Assume now that m is fixed and let us take k such that $k \geq 2^m$. Then

$$s_k = a_0 + a_1 + ... + a_k \geq a_0 + a_1 + ... + a_{2^m} =$$

$$= a_0 + a_1 + a_2 + (a_3 + a_4) + (a_5 + a_6 + a_7 + a_8) +$$

$$...+(a_{2^{m-1}}+a_{2^{m-1}+1}+...+a_{2^m}) \geq a_0+\frac{1}{2}a_1+a_2+2a_4+2^2a_8+...+2^{m-1}a_{2^m}$$

$$\geq \frac{1}{2}(a_1 + 2a_2 + 2^2 a_{2^2} + ... + 2^m a_{2^m}) = \frac{1}{2} S_m,$$

thus,

$$s_k \geq \frac{1}{2} S_m \tag{1.4}$$

If the series $\sum_{n=0}^{\infty} a_n$ is convergent, then the sequence $\{s_k\}$ is bounded and, using (1.4), we get that the sequence $\{S_m\}$ is also bounded (why?). Hence, the series $\sum_{n=0}^{\infty} 2^n a_{2^n}$ is convergent (why?). If $\sum_{n=0}^{\infty} 2^n a_{2^n}$ is divergent, then the sequence $\{S_m\}$ tends to ∞ (why?) so, from (1.4), we

get that the sequence $\{s_k\}$ also goes to ∞ and thus, the series $\sum_{n=0}^{\infty} a_n$ is also divergent. Now the theorem is completely proved. □

We can use this test to find again the result on the Riemann zeta function $\zeta(\alpha) = \sum_{n=0}^{\infty} \frac{1}{n^\alpha}$ (see Theorem 22). Indeed, here $a_n = \frac{1}{n^\alpha}$ and $a_{2^n} = \frac{1}{2^{n\alpha}} = \left(\frac{1}{2^\alpha}\right)^n$. The series

$$\sum_{n=0}^{\infty} 2^n \left(\frac{1}{2^\alpha}\right)^n = \sum_{n=0}^{\infty} \left(\frac{1}{2^{\alpha-1}}\right)^n$$

is obviously convergent if and only if $\alpha > 1$ (see Theorem 17). Thus, from the Cauchy compression test, we get that the Riemann series is convergent if and only if $\alpha > 1$.

Now, let us find all the values of $\alpha \in \mathbb{R}$ such that the series $\sum_{n=2}^{\infty} \frac{1}{n(\log_7 n)^\alpha}$ is convergent. If in $\frac{1}{n(\log_7 n)^\alpha}$we put instead of n, 2^n and if we multiply the result by 2^n, we get the series

$$\sum_{n=2}^{\infty} 2^n \frac{1}{2^n(\log_7 2^n)^\alpha} = \frac{1}{(\log_7 2)^\alpha} \sum_{n=2}^{\infty} \frac{1}{n^\alpha}.$$

Thus, the nature of our series is the same like the nature of the Riemann series. Therefore, our series is convergent if and only if $\alpha > 1$.

Another useful convergence test is the following:

THEOREM 24. *(The comparison test) Let* $\sum_{n=0}^{\infty} a_n$ *and* $\sum_{n=0}^{\infty} b_n$ *be two series with* $a_n \geq 0$, $b_n \geq 0$ *and* $a_n \leq b_n$ *for* $n = 0, 1, 2, \ldots$. *a) If the series* $\sum_{n=0}^{\infty} b_n$ *is convergent, then the series* $\sum_{n=0}^{\infty} a_n$ *is also convergent. b) If the series* $\sum_{n=0}^{\infty} a_n$ *is divergent, then the series* $\sum_{n=0}^{\infty} b_n$ *is also divergent.*

PROOF. Since $a_n \leq b_n$ for $n = 0, 1, 2, ...$, then

$$s_n = a_0 + a_1 + ... + a_n \leq b_0 + b_1 + ... + b_n \stackrel{def}{=} u_n,$$

the partial n-th sum of the series $\sum_{n=0}^{\infty} b_n$. a) If the series $\sum_{n=0}^{\infty} b_n$ is convergent, the sequence $\{u_n\}$ is bounded. Hence the sequence $\{s_n\}$ is also bounded, and so the series $\sum_{n=0}^{\infty} a_n$ is convergent (see Theorem 20). b) If the series $\sum_{n=0}^{\infty} a_n$ is divergent, then the sequence $\{s_n\}$ is unbounded

(see Theorem 20). Hence the sequence $\{u_n\}$ is unbounded (why?), so the series $\sum_{n=0}^{\infty} b_n$ is divergent. □

For instance, the series $\sum_{n=0}^{\infty} \frac{1}{n^2+7}$ is convergent because $\frac{1}{n^2+7} < \frac{1}{n^2}$ and because the series $\sum_{n=0}^{\infty} \frac{1}{n^2} = Z(2)$ is convergent (see Theorem 22).

The comparison test is also useful in proving the following basic convergence test (see Theorem 25).

First of all we remark that the natural way to add two series is the following

$$\sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty} (a_n + b_n). \tag{1.5}$$

It is easy to see that if the both series are convergent, then the resulting series on the right is also convergent (prove it!). If a_n, b_n are nonnegative then, if at least one series is divergent, the series on the right in (1.5) is also divergent (prove it!). In general this is not true. For instance, $\sum_{n=0}^{\infty} n + \sum_{n=0}^{\infty} (-n) = 0$!

Now, if λ is a real number, by definition,

$$\lambda \sum_{n=0}^{\infty} a_n = \sum_{n=0}^{\infty} \lambda a_n$$

If $\lambda = -1$, we can define the subtraction:

$$\sum_{n=0}^{\infty} a_n - \sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} (-b_n).$$

For $\lambda \neq 0$, the series $\sum_{n=0}^{\infty} a_n$ and $\lambda \sum_{n=0}^{\infty} a_n$ have the same nature (prove it!). Pay attention to the following wrong calculation:

$$\sum_{n=2}^{\infty} \frac{1}{n+1} - \sum_{n=2}^{\infty} \frac{1}{n-1} = -2 \sum_{n=0}^{\infty} \frac{1}{n^2-1}$$

The series on the right side is convergent, but on the left side we have $\infty - \infty$, an undetermined operation, so it cannot be equal to a determined one!

THEOREM 25. *(The limit comparison test) Let* $\sum_{n=0}^{\infty} a_n$ *and* $\sum_{n=0}^{\infty} b_n$ *be two numerical series of real numbers such that* $a_n \geq 0$ *and* $b_n > 0$

for any $n = 0, 1, 2, ...$ *Suppose that the sequence* $\left\{\frac{a_n}{b_n}\right\}$ *is convergent to* $l \in \mathbb{R} \cup \{\infty\}$. *Then, a) if* $l \neq 0, \infty$, *both series have the same nature (they are convergent or not) at the same time, b) if* $l = 0$, $\sum_{n=0}^{\infty} b_n$ *convergent implies* $\sum_{n=0}^{\infty} a_n$ *convergent and, c) if* $l = \infty$, $\sum_{n=0}^{\infty} b_n$ *divergent implies* $\sum_{n=0}^{\infty} a_n$ *divergent. This is why the series* $\sum_{n=0}^{\infty} b_n$ *is called a* witness *series.*

PROOF. a) Since $l \neq 0, \infty$, $l > 0$, so there is an $\varepsilon > 0$ such that $l - \varepsilon > 0$. Since $\lim_{n\to\infty} \frac{a_n}{b_n} = l$, there is a natural number N (depending on ε) with $l - \varepsilon < \frac{a_n}{b_n} < l + \varepsilon$ for any $n \geq N$. Because of the last double inequality and since $b_n > 0$, one can write

$$(l - \varepsilon)b_n < a_n < (l + \varepsilon)b_n, \tag{1.6}$$

for any $n \geq N$. Now, if for instance, $\sum_{n=0}^{\infty} a_n$ is convergent (this means that the series $\sum_{n=N}^{\infty} a_n$ is also convergent from Theorem 19) then, using the inequality $(l - \varepsilon)b_n < a_n$ and the comparison test (Theorem 24) we get that the series $(l - \varepsilon) \sum_{n=N}^{\infty} b_n$ is convergent. Since $l - \varepsilon \neq 0$ we finally obtain that the series $\sum_{n=N}^{\infty} b_n$ is convergent, i.e. the series $\sum_{n=0}^{\infty} b_n$ is convergent (see the renouncement test). If this last series is convergent, using the second inequality, $a_n < (l + \varepsilon)b_n$, from (1.6), one gets that the first series $\sum_{n=0}^{\infty} a_n$ is convergent (complete the reasoning!). b) If $l = 0$, take an $\varepsilon > 0$ and take a natural number N_1 (depending on ε) such that for any $n \geq N_1$ we have $0 \leq \frac{a_n}{b_n} < \varepsilon$ or $a_n < \varepsilon b_n$. If the series $\sum_{n=0}^{\infty} b_n$ is convergent, then the series $\varepsilon \sum_{n=N_1}^{\infty} b_n$ is also convergent, so the series $\sum_{n=N_1}^{\infty} a_n$ is convergent (see the comparison test). Using again the renouncement test we get that the series $\sum_{n=0}^{\infty} a_n$ is convergent. c) If $l = \infty$, take a positive real number $M > 0$ and take a natural number N_2 (depending on M) such that for $n \geq N_2$, $\frac{a_n}{b_n} > M$, or

$a_n > Mb_n$. Now, if the series $\sum_{n=0}^{\infty} b_n$ is divergent, then the series $\sum_{n=N_2}^{\infty} b_n$ is also divergent (see Theorem 19). Use the inequality $a_n > Mb_n$ to obtain that the series $\sum_{n=N_2}^{\infty} a_n$ is divergent (see the comparison test). Using again the renouncement test we get that the series $\sum_{n=0}^{\infty} a_n$ is divergent. □

Let us decide if the series $\sum_{n=0}^{\infty} \frac{\sqrt[3]{n}}{n^2+4}$ is convergent or not. We intend to use the limit comparison test with $a_n = \frac{\sqrt[3]{n}}{n^2+4}$ and $b_n = \frac{1}{n^\alpha}$. We try to find an α such that the limit $l = \lim_{n\to\infty} \frac{a_n}{b_n}$ be finite and nonzero. If we can do this, such an α is unique. Its value is called the "Abel degree" of the function $f(x) = \frac{\sqrt[3]{x}}{x^2+4}$. So,

$$l = \lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{n^{\alpha+\frac{1}{3}}}{n^2(1+\frac{4}{n^2})} \neq 0, \infty$$

$(= 1)$ if and only if $\alpha + \frac{1}{3} = 2$, i.e. $\frac{5}{3} > 1$. Since the series $\sum_{n=1}^{\infty} \frac{1}{n^{\frac{5}{3}}} = Z(\frac{5}{3})$ is convergent (see the Riemann Zeta series), from the limit comparison test one has that the series $\sum_{n=1}^{\infty} \frac{\sqrt[3]{n}}{n^2+4}$ is convergent. Applying again the renouncement test we get that our initial series $\sum_{n=0}^{\infty} \frac{\sqrt[3]{n}}{n^2+4}$ is convergent.

Let us put in a systematic manner all the reasonings in this last example.

THEOREM 26. *(The α-comparison test) Let $\sum_{n=0}^{\infty} a_n$ be a series with nonnegative terms ($a_n \geq 0$). We assume that there is a real number α, such that the following limit does exist: $\lim_{n\to\infty} n^\alpha a_n = l \in \mathbb{R} \cup \{\infty\}$. a) If $l \neq 0, \infty$ then, the series $\sum_{n=0}^{\infty} a_n$ is convergent if and only if $\alpha > 1$. b) If $l = 0$ and $\alpha > 1$, then our series $\sum_{n=0}^{\infty} a_n$ is convergent. c) If $l = \infty$ and $\alpha \leq 1$, then the series $\sum_{n=0}^{\infty} a_n$ is divergent and equal to ∞.*

PROOF. It is enough to take $b_n = \frac{1}{n^\alpha}$ in the Theorem 25 (do everything slowly, step by step!). □

Let us apply this last test to the following situation. For a large N (> 100, for instance), can we use the approximation

$$\sum_{n=0}^{\infty} \frac{n^3+7n+1}{\sqrt{n^9+2n+2}} \approx \sum_{n=0}^{N} \frac{n^3+7n+1}{\sqrt{n^9+2n+2}}?.$$

We can do this if and only if our series is convergent (why?). In order to see if our series is convergent or not, let us consider the limit:

$$\lim_{n\to\infty} n^{\alpha} \frac{n^3+7n+1}{\sqrt{n^9+2n+2}} = \lim_{n\to\infty} \frac{n^{\alpha+3}(1+\frac{7}{n^2}+\frac{1}{n^3})}{n^{\frac{9}{2}}\sqrt{1+\frac{2}{n^8}+\frac{2}{n^9}}} = \lim_{n\to\infty} \frac{n^{\alpha+3}}{n^{\frac{9}{2}}}.$$

But, this last limit is neither 0 nor ∞, if and only if $\alpha+3=\frac{9}{2}$, or $\alpha=\frac{3}{2}$ (why?). Since in this case $\alpha>1$ and the limit l is 1, we apply the α-comparison test (Theorem 26) and find that our initial series is convergent. Hence the above approximation works!

A very useful test is the ratio test or D'Alembert test.

THEOREM 27. *(the ratio test) Let* $\sum_{n=0}^{\infty} a_n$ *be a series with positive terms.*

a) If there is a real number λ *such that* $0<\lambda<1$ *and* $\frac{a_{n+1}}{a_n} \leq \lambda$ *for any* $n \geq N$, *where* N *is a fixed natural number, then the series is convergent. This is equivalent to say that* $\limsup \frac{a_{n+1}}{a_n} < 1$.

b) If $\frac{a_{n+1}}{a_n} \geq 1$ *for any* $n \geq M$, *where* M *is a fixed natural number, then the series is divergent.*

c) If $\limsup \frac{a_{n+1}}{a_n} = 1$, *and if* $\frac{a_{n+1}}{a_n}$ *is not equal to* 1 *from a rank on, then, in general, we cannot decide if the series is convergent or not (in this situation use more powerful tests, for instance the "Raabe-Duhamel Test").*

PROOF. a) Let us put $n = N, N+1, N+2, ...$ in the inequality $\frac{a_{n+1}}{a_n} \leq \lambda$. We find:

$$a_{N+1} \leq \lambda a_N, a_{N+2} \leq \lambda a_{N+1} \leq \lambda^2 a_N, ..., a_{N+m} \leq \lambda^m a_N, ...$$

Hence,

$$a_N + a_{N+1} + a_{N+2} + ... + a_{N+m} + ... \leq$$

$$\leq a_N(1+\lambda+\lambda^2+...+\lambda^m+...) = a_N \frac{1}{1-\lambda}.$$

So any partial sum of the series $\sum_{n=N}^{\infty} a_n$ is bounded. Since $a_n \geq 0$, the series $\sum_{n=N}^{\infty} a_n$ is convergent (Theorem 20). The renouncement test says that the whole series $\sum_{n=0}^{\infty} a_n$ is also convergent.

b) If $\frac{a_{n+1}}{a_n} \geq 1$ for any $n \geq M$, then

$$a_M + a_{M+1} + ... + a_{M+m} + ... \geq a_M + a_M + ... + a_M + ... = \infty,$$

so the series $\sum_{n=0}^{\infty} a_n$ is divergent (explain everything slowly, step by step!).

c) For instance, the harmonic series $\sum_{n=1}^{\infty} \frac{1}{n}$ is divergent, but

$$\limsup_{n\to\infty} \frac{\frac{1}{n+1}}{\frac{1}{n}} = 1.$$

This last property is also true for the series $\sum_{n=1}^{\infty} \frac{1}{n^2}$, but this last series is convergent! This is why we cannot say anything in general if one can find numbers of the form $\frac{a_{n+1}}{a_n} < 1$ as close as we want to 1. □

REMARK 5. *The condition from a) of Theorem 27 is equivalent to saying that* $\limsup \frac{a_{n+1}}{a_n} < 1$ *(why?). If the sequence* $\left\{\frac{a_{n+1}}{a_n}\right\}$ *is convergent to* l*, then the Theorem 27 is more exactly. Namely, in this last case, the series* $\sum_{n=0}^{\infty} a_n$ *is convergent if* $l < 1$*, it is divergent if* $l > 1$ *and if* $l = 1$ *we cannot say anything (prove it!).*

For instance, the series $\sum_{n=0}^{\infty} \frac{2^n}{n!}$ is convergent because $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 0 < 1$ (see Remark 5).

Usually, if $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = 1$, we try to apply the following "more powerful" test.

THEOREM 28. *(The Raabe-Duhamel test) Let* $\sum_{n=0}^{\infty} a_n$ *be a series with positive terms.*

a) If there is a real number $\lambda \in (1, \infty)$ *and a natural number* N *such that* $n\left(\frac{a_n}{a_{n+1}} - 1\right) \geq \lambda$ *for any* $n \geq N$*, then the series is convergent.*

b) If $n\left(\frac{a_n}{a_{n+1}} - 1\right) < 1$ *for* $n \geq M$*, where* M *is a fixed natural number, then the series is divergent.*

c) Assume that the following limit exists, $\lim_{n\to\infty} n\left(\frac{a_n}{a_{n+1}}-1\right)=l\in\mathbb{R}\cup\{\infty\}$. *Then, if* $l>1$, *the series is convergent, if* $l<1$, *the series is divergent and if* $l=1$, *we cannot decide on the nature of this series.*

One can find a proof of this result in [**Nik**], or in [**Pal**]. See also Problem 11 of this chapter.

Let us find the nature of the series

$$\sum_{n=1}^{\infty}\frac{1\cdot 3\cdot 5\cdot ...\cdot(2n+1)}{2\cdot 4\cdot 6\cdot ...\cdot 2n}\cdot\frac{1}{2n+3}.$$

Since

$$\frac{a_{n+1}}{a_n}=\frac{(2n+3)^2}{(2n+2)(2n+5)}\to 1,$$

let us apply Raabe-Duhamel test. Since

$$n\left(\frac{a_n}{a_{n+1}}-1\right)=\frac{2n^2+n}{(2n+3)^2}\to\frac{1}{2}<1,$$

the series is divergent.

THEOREM 29. *(The Cauchy root test) Let* $\sum_{n=0}^{\infty}a_n$ *be a series with nonnegative terms.*

a) If there is a real number $\lambda\in(0,1)$ *such that* $\sqrt[n]{a_n}\le\lambda$ *for* $n\ge N$, *where* N *is a fixed natural number, then the series is convergent.*

b) If $\sqrt[n]{a_n}\ge 1$ *for all* $n\ge M$, *where* M *is a fixed natural number, then the series is divergent.*

c) Assume that the following limit exists, $\lim_{n\to\infty}\sqrt[n]{a_n}=l\in\mathbb{R}\cup\{\infty\}$. *Then, if* $l<1$, *the series is convergent, if* $l>1$, *the series is divergent and if* $l=1$, *we cannot decide on the nature of this series.*

PROOF. a) The condition $\sqrt[n]{a_n}\le\lambda$ for $n\ge N$ implies

$$a_N+a_{N+1}+...+a_{N+m}+...\le a_N\lambda^N(1+\lambda+...+\lambda^m+...)=$$

$$=a_N\frac{\lambda^N}{1-\lambda}<\frac{a_N}{1-\lambda},$$

so, the partial sums of the series $\sum_{n=N}^{\infty}a_n$ are bounded. Hence the series $\sum_{n=N}^{\infty}a_n$ is convergent (see Theorem 20). From the renouncement test we derive that the series $\sum_{n=0}^{\infty}a_n$ is convergent.

b) The condition $\sqrt[n]{a_n} \geq 1$ for $n \geq M$, implies $a_n \geq 1$ for an infinite number of terms, so $\{a_n\}$ does not tend to zero. Hence the series is divergent (see Corollary 2).

c) Take $\varepsilon > 0$ such that $l + \varepsilon < 1$. Since $\sqrt[n]{a_n} \to l$, there is a natural number N such that if $n \geq N$, $\sqrt[n]{a_n} < l + \varepsilon$. Apply now a) and find that the series is convergent. If $l > 1$, there is a rank M from which on $\sqrt[n]{a_n} \geq 1$ for $n \geq M$ and so, the series is divergent (see b)). If $l = 1$, there are some cases in which the series is convergent and there are other cases in which the series is divergent. For instance, the series $\sum\limits_{n=1}^{\infty} \frac{1}{n^2}$ is convergent and $l = \lim\limits_{n\to\infty} \sqrt[n]{\frac{1}{n^2}} = 1$ (since $\sqrt[n]{n} \to 1$; prove this! Hint:

$$\alpha_n = \sqrt[n]{n} - 1 \Longrightarrow n = (1 + \alpha_n)^n = 1 + n\alpha_n + \frac{n(n-1)}{2}\alpha_n^2 + ... >$$

$$> \frac{n(n-1)}{2}\alpha_n^2 \Longrightarrow \alpha_n < \sqrt{\frac{2}{n-1}},$$

so, $\alpha_n \to 0$. But the series $\sum\limits_{n=1}^{\infty} \frac{1}{n}$ is divergent and $l = \lim\limits_{n\to\infty} \sqrt[n]{\frac{1}{n}} = 1$. $\square$

The series $\sum\limits_{n=0}^{\infty} \frac{1}{(2+n)^n}$ is convergent because $\sqrt[n]{a_n} = \frac{1}{2+n} \leq \frac{1}{2}$ for any $n = 0, 1, ...$ (we just applied the Cauchy Root Test, a)). We can also apply the Comparison Test: $\frac{1}{(2+n)^n} < \frac{1}{n^2}$ for any $n = 1, 2, ...$, etc.

REMARK 6. *A natural question arises: what is the connection (if there is one!) between the ratio test and the root test? To explain this we need a powerful result from the calculus of the limits of sequences. This is the famous Cesaro-Stolz Theorem: Let $\{a_n\}$ be an arbitrary sequence and let $\{b_n\}$ be an increasing and unbounded sequence of positive numbers such that the sequence $\left\{\frac{a_{n+1}-a_n}{b_{n+1}-b_n}\right\}$ is convergent to $l \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. Then $\frac{a_n}{b_n} \to l$. A direct consequence of this result is the Cesaro Theorem: Let $\{c_n\}$ be a convergent to l sequence. Then the "means" sequence $\left\{\frac{c_0+c_1+...c_{n-1}}{n}\right\}$ is also convergent to l (prove it as an application of the Cesaro-Stolz Theorem). We prove now that for a sequence $\{a_n\}$ of positive numbers, such that the limit of the sequence $\left\{\frac{a_{n+1}}{a_n}\right\}$ does exist in $\overline{\mathbb{R}}$, then $\left\{\frac{a_{n+1}}{a_n}\right\} \to l$ if and only if $\{\sqrt[n]{a_n}\} \to l$. Suppose that $\left\{\frac{a_{n+1}}{a_n}\right\} \to l$, then $\ln a_{n+1} - \ln a_n \to \ln l$, or $\frac{\ln a_{n+1} - \ln a_n}{(n+1)-n} \to \ln l$. From the Cesaro-Stolz Theorem we get that $\frac{\ln a_n}{n} = \ln \sqrt[n]{a_n} \to \ln l$, or $\sqrt[n]{a_n} \to l$. Conversely, assume that $\{\sqrt[n]{a_n}\} \to l$ and that $\left\{\frac{a_{n+1}}{a_n}\right\} \to l'$.*

From the first implication, one has that $l = l'$ and the statement is completely proved.

Suppose we have a series $\sum_{n=0}^{\infty} a_n$ with $a_n > 0$ for any $n > N$, such that $\left\{\frac{a_{n+1}}{a_n}\right\} \to 1$. We cannot decide on the nature of this series. Remark 6 says that it is not a good idea to try to apply the Cauchy Root Test because this one also cannot decide if the series is convergent or not.

2. Series with arbitrary terms

Up to now we just considered (in principal) series with nonnegative terms. If the number of positive or negative terms in a series are finite, to decide the nature of this series, it is sufficient to get out those terms and thus to obtain a new series with all its term positive or negative (see the renouncement test). If $a_n \leq 0$ in a series $\sum_{n=0}^{\infty} a_n$, we consider the new series $\sum_{n=0}^{\infty}(-a_n) = -\sum_{n=0}^{\infty} a_n$ and apply the results obtained in the previous section. For instance, $\sum_{n=0}^{\infty} -\frac{1}{n^3} = -\sum_{n=0}^{\infty} \frac{1}{n^3}$ is convergent, because $\sum_{n=0}^{\infty} \frac{1}{n^3}$ is convergent (it is the value of the Riemann series for $\alpha = 3 > 1$). A numerical series $\sum_{n=0}^{\infty} a_n$ is said to have arbitrary terms if the sign of its terms a_n may be positive, negative or zero, but not all (or a finite number of them) are of the same sign. We also call such a series a *general series*. The Cauchy general test (see Theorem 18) and the zero test are the only tests we know (up to now) on general series. Here is another important one.

THEOREM 30. *(The Abel-Dirichlet test) Let $\{a_n\}$ be a decreasing to zero ($a_n \to 0$) sequence of nonnegative ($a_n \geq 0$) real numbers. Let $\sum_{n=0}^{\infty} b_n$ be a series with bounded partial sums (i.e. there is a real number $M > 0$ such that for $s_n = b_0 + b_1 + ... + b_n$, one has $|s_n| < M$, where $n = 0, 1, ...$). Then the series $\sum_{n=0}^{\infty} a_n b_n$ is convergent.*

PROOF. We intend to apply the Cauchy general test (Theorem 18). Let us denote $S_n = a_0 b_0 + a_1 b_1 + ... + a_n b_n$ the n-th partial sum of the

series $\sum_{n=0}^{\infty} a_n b_n$ and let us evaluate

$$|S_{n+p} - S_n| = |a_{n+1}b_{n+1} + ... + a_{n+p}b_{n+p}| =$$

$$= |a_{n+1}(s_{n+1} - s_n) + a_{n+2}(s_{n+2} - s_{n+1}) + ... + a_{n+p}(s_{n+p} - s_{n+p-1})| =$$

$$|-a_{n+1}s_n + (a_{n+1} - a_{n+2})s_{n+1} + ... \ + (a_{n+p-1} - a_{n+p})s_{n+p-1} + a_{n+p}s_{n+p}|$$

(2.1)

$$\leq a_{n+1}\,|s_n| + (a_{n+1} - a_{n+2})\,|s_{n+1}| + ... + (a_{n+p-1} - a_{n+p})\,|s_{n+p-1}| + a_{n+p}\,|s_{n+p}|\,.$$

Let $\varepsilon > 0$ be a small positive real number. In the last row of (2.1) we put instead $|s_j|$, $j = n, n+1, ..., n+p$, the greater number M. So we get

(2.2)

$$|S_{n+p} - S_n| \leq M(a_{n+1} + a_{n+1} - a_{n+2} + a_{n+2} - a_{n+3} + ... + a_{n+p-1} - a_{n+p} + a_{n+p})$$
$$= 2Ma_{n+1}$$

Since $\{a_n\}$ tends to 0 as $n \to \infty$, there is a natural number N (which depend on ε) such that for any $n \geq N$, on has that $2Ma_{n+1} < \varepsilon$. Since $|S_{n+p} - S_n| \leq 2Ma_{n+1}$ (see (2.2)), we get that $|S_{n+p} - S_n| < \varepsilon$ for any $n \geq N$. This means that the sequence $\{S_n\}$ is a Cauchy sequence, i.e. the series $\sum_{n=0}^{\infty} a_n b_n$ is convergent (see Theorem 18) and our theorem is completely proved. □

The following test is a direct consequence of the Abel-Dirichlet test.

COROLLARY 3. *(The Leibniz test) Let $\{a_n\}$ be a decreasing to zero ($a_n \to 0$) sequence of nonnegative ($a_n \geq 0$) real numbers. Then the series*

$$\sum_{n=1}^{\infty}(-1)^{n-1}a_n = a_1 - a_2 + a_3 - ...$$

is convergent.

For instance, applying this test, we get that the series $\sum_{n=1}^{\infty}(-1)^n \frac{n+1}{n^2+3} = -\sum_{n=1}^{\infty}(-1)^{n-1}\frac{n+1}{n^2+3}$ is convergent (do it!).

A famous example is the *standard alternate series*

$$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + ... \tag{2.3}$$

This series is a general series (why?) and it is convergent. Indeed, $\left\{a_n = \frac{1}{n}\right\}$ is a decreasing to zero sequence with nonnegative terms so, we can apply the Leibniz test and find that the series is convergent.

DEFINITION 7. *(absolute convergence) A series* $\sum\limits_{n=0}^{\infty} a_n$ *is said to be absolutely convergent if the series of moduli* $\sum\limits_{n=0}^{\infty} |a_n|$ *is convergent.*

For instance, the series $\sum\limits_{n=1}^{\infty}(-1)^n \frac{1}{n^2}$ is convergent (why?) and absolutely convergent, but the series $\sum\limits_{n=0}^{\infty}(-1)^n \frac{1}{n}$ is convergent (why?) and it is not absolutely convergent, because the harmonic series $\sum\limits_{n=1}^{\infty} \frac{1}{n} = Z(1) = \infty$ (see the Riemann series). A series which is convergent, but not absolutely convergent, is called *semiconvergent*.

The following result says that the notion of absolutely convergence is stronger then the notion of (simple) convergence.

THEOREM 31. *Any absolute convergence series* $\sum\limits_{n=0}^{\infty} a_n$ *is also (simple) convergent.*

PROOF. We use again the Cauchy General Test (see Theorem 18). Let $s_n = a_0 + a_1 + ... + a_n$ be the n-th partial sum of the initial series $\sum\limits_{n=0}^{\infty} a_n$ and let $S_n = |a_0| + |a_1| + ... + |a_n|$ be the n-th partial sum of the series $\sum\limits_{n=0}^{\infty} |a_n|$. Let us evaluate

$$|s_{n+p} - s_n| = |a_{n+1} + a_{n+2} + ... + a_{n+p}| \leq \tag{2.4}$$

$$|a_{n+1}| + |a_{n+2}| + ... + |a_{n+p}| = |S_{n+p} - S_n| .$$

Let $\varepsilon > 0$ be a small positive real number and let N be a sufficiently large natural number such that for any $n \geq N$ one has $|S_{n+p} - S_n| < \varepsilon$ for any $p = 1, 2, ...$ (since $\{S_n\}$ is a Cauchy sequence). From (2.4) we have that $|s_{n+p} - s_n| \leq |S_{n+p} - S_n|$, so $|s_{n+p} - s_n| \leq \varepsilon$ for any $n \geq N$ and for any $p = 1, 2, ...$. But this means that the sequence $\{s_n\}$ is a Cauchy sequence. Hence the series $\sum\limits_{n=0}^{\infty} a_n$ is convergent (see Theorem 18). □

For instance, the series $\sum_{n=1}^{\infty} \frac{\sin(5n)}{n^2}$ is convergent because it is absolutely convergent. Indeed, since $\left|\frac{\sin(5n)}{n^2}\right| \leq \frac{1}{n^2}$ and since the series $\sum_{n=1}^{\infty} \frac{1}{n^2} = Z(2)$ is convergent (see the Riemann series), the Comparison Test says that the series of moduli $\sum_{n=1}^{\infty} \frac{|\sin(5n)|}{n^2}$ is convergent, i.e. the initial series $\sum_{n=1}^{\infty} \frac{\sin(5n)}{n^2}$ is convergent.

REMARK 7. *(see* [**Nik**] *or* [**Pal**]*) We saw above that any absolutely convergent series is convergent, but the converse is not true. Cauchy proved that in any absolutely convergent series one can change the order of the terms in the infinite sum (by any permutation) and the sum of the series remains the same. On the contrary, Riemann proved that for a semiconvergent series* $\sum_{n=0}^{\infty} a_n$ *and for any number* $A \in \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, *one can find a permutation of the terms of the series* $\sum_{n=0}^{\infty} a_n$ *such that its sum becomes exactly A. Two absolutely convergent series can be multiplied by the usual polynomial multiplication rule*

$$\sum_{n=0}^{\infty} a_n \cdot \sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty} c_n, \text{ where } c_n = a_0 b_n + a_1 b_{n-1} + ... + a_n b_0,$$

and the resulting product series is again absolutely convergent (Mertaens).

REMARK 8. *If instead of series with real numbers we consider a series with complex numbers* $\sum_{n=0}^{\infty} z_n$, *where* $z_n = x_n + iy_n$, $x_n, y_n \in \mathbb{R}$ *for any* $n = 0, 1, 2, ...$, *we say that such a series is convergent to its sum* $s = u + iv$, $u, v \in \mathbb{R}$ *if the sequence of partial sums*

$$s_n = z_0 + z_1 + ... + z_n = (x_0 + x_1 + ... + x_n) + i(y_0 + y_1 + ... + y_n)$$

is convergent to s, i.e.

$$|s - s_n| = \sqrt{[u - (x_0 + x_1 + ... + x_n)]^2 + [v - (y_0 + y_1 + ... + y_n)]^2} \to 0,$$

when $n \to \infty$. *This is equivalent to saying that both series with real numbers,* $\sum_{n=0}^{\infty} x_n$ *(the real part) and* $\sum_{n=0}^{\infty} y_n$ *(the imaginary part) are convergent to u and v respectively. Hence,* $\sum_{n=0}^{\infty} z_n = \sum_{n=0}^{\infty} x_n + i \sum_{n=0}^{\infty} y_n$ *and the calculus with complex series reduces to the calculus with real series.*

Practically, in general, it is difficult to decide if both the "real part" and the "imaginary part" are convergent. For instance, let us consider the series

$$s=\sum_{n=0}^{\infty}\frac{(1+i)^n}{n!}=\sum_{n=0}^{\infty}\frac{\sqrt{2^n}\left(\frac{1}{\sqrt{2}}+i\frac{1}{\sqrt{2}}\right)^n}{n!}=\sum_{n=0}^{\infty}\frac{\sqrt{2^n}\left(\cos\frac{\pi}{4}+i\sin\frac{\pi}{4}\right)^n}{n!}$$

Let us use now the Moivre formula and find:

$$s=\sum_{n=0}^{\infty}\frac{\sqrt{2^n}\cos n\frac{\pi}{4}}{n!}+i\sum_{n=0}^{\infty}\frac{\sqrt{2^n}\sin n\frac{\pi}{4}}{n!}.$$

Since

$$\left|\frac{\sqrt{2^n}\cos n\frac{\pi}{4}}{n!}\right|\leq\frac{\sqrt{2^n}}{n!}$$

and since

$$\lim_{n\to\infty}\frac{\frac{\sqrt{2^{n+1}}}{(n+1)!}}{\frac{\sqrt{2^n}}{n!}}=0,$$

the series $\sum_{n=0}^{\infty}\frac{\sqrt{2^n}\cos n\frac{\pi}{4}}{n!}$ *is absolutely convergent, so it is convergent (why?-precise the theorems that we used!). In the same way we prove that the imaginary part series* $\sum_{n=0}^{\infty}\frac{\sqrt{2^n}\sin n\frac{\pi}{4}}{n!}$ *is also convergent. An easier way to prove the convergence of the complex series* $s=\sum_{n=0}^{\infty}\frac{(1+i)^n}{n!}$ *is the following. It is not difficult to prove that an absolutely convergent series* $\sum_{n=0}^{\infty}z_n$ *(i.e.* $\sum_{n=0}^{\infty}|z_n|$ *is convergent) is also convergent (see the proof of Theorem 31). In our case,*

$$\left|\frac{(1+i)^n}{n!}\right|=\frac{(|1+i|)^n}{n!}=\frac{\sqrt{2^n}}{n!}.$$

So, the series $\sum_{n=0}^{\infty}|z_n|=\sum_{n=0}^{\infty}\frac{\sqrt{2^n}}{n!}$ *is convergent (use the ratio test), i.e. the series* $s=\sum_{n=0}^{\infty}\frac{(1+i)^n}{n!}$ *is absolutely convergent. Hence, it is convergent. If a series* $\sum_{n=0}^{\infty}z_n$ *is not absolutely convergent, the general way to study it is to write it as:*

$$\sum_{n=0}^{\infty}z_n=\sum_{n=0}^{\infty}x_n+i\sum_{n=0}^{\infty}y_n$$

and to study separately the real series $\sum_{n=0}^{\infty} x_n$ *and* $\sum_{n=0}^{\infty} y_n$. *If both of them are convergent, the initial series is also convergent. If at least one of them is divergent, the series* $\sum_{n=0}^{\infty} z_n$ *is divergent (why?).*

3. Approximate computations

Usually, whenever one cannot exactly compute the sum of a convergent series $s = \sum_{n=0}^{\infty} a_n$, one approximate s by its n-th partial sum $s_n = a_0 + a_1 + ... + a_n$, for sufficiently large n. For instance,

$$s = \sum_{n=1}^{\infty} \frac{1}{n^2} \approx s_{1000} = \frac{1}{1^2} + \frac{1}{2^2} + ... + \frac{1}{1000^2}.$$

The difference $\varepsilon_n = |s - s_n|$ is called the *(absolute) error of order* n in our process of approximation. It is clear enough why we are interested in the evaluation of this error. Since the series is convergent, $\varepsilon_n \to 0$, when n becomes large enough. Given a small positive real number $\varepsilon > 0$, the problem is to find an n (very small if it is possible!) which depend on ε, such that the error $\varepsilon_n < \varepsilon$. For instance, if $\varepsilon = \frac{1}{10^3}$, we say that "$s$ is approximated by s_n with 3 exact decimals".

We study this problem in two cases.

Case 1 Let $s = \sum_{n=0}^{\infty} a_n$ be a series with positive terms ($a_n > 0$, $n = 0, 1, ...$) and let $\alpha \in (0, 1)$ such that $\frac{a_{n+1}}{a_n} \leq \alpha$ for $n \geq N$ (remember yourself the Ratio Test). The series is convergent (see Theorem 27). Let now k be a natural number greater or equal to N. Let us evaluate the error $\varepsilon_k = s - s_k$:

$$\varepsilon_k = a_{k+1} + a_{k+2} + ... \leq \alpha a_k + \alpha^2 a_k + ... = \frac{\alpha}{1-\alpha} a_k \tag{3.1}$$

We see that if $\varepsilon > 0$ is an arbitrary small positive real number, always one can find a least $k \in \mathbb{N}$ such that $\frac{\alpha}{1-\alpha} a_k < \varepsilon$. Since $\varepsilon_k \leq \frac{\alpha}{1-\alpha} a_k$, for this k one also has: $\varepsilon_k < \varepsilon$. If we want a small k, we must find a small $\alpha \in (0, 1)$ such that for a small N (0 if it is possible), we have $\frac{a_{n+1}}{a_n} \leq \alpha$ for $n \geq N$.

Let us compute the value of $\sum_{n=0}^{\infty} \frac{1}{n!}$ (we shall see later that it is exactly e, the base of the Neperian logarithm) with 2 exact decimals. Since $\frac{a_{n+1}}{a_n} = \frac{1}{n+1} \leq \frac{1}{2}$ for $n \geq 1$,

$$\varepsilon_k = s - s_k \leq \frac{\frac{1}{2}}{1 - \frac{1}{2}} \frac{1}{k!} = \frac{1}{k!}.$$

Let us find the least k such that $\frac{1}{k!} < \varepsilon = \frac{1}{10^2}$. By trials, $k = 1, 2, ...,$ we find $k = 5$. So

$$s \approx s_5 = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \frac{1}{5!} = 2.71666...,$$

i.e. we obtained the value of e with 2 exact decimals, $e \approx 2.71$.

Let $s = \sum a_n$ be a series with nonnegative terms ($a_n \geq 0$, $n = 0, 1, ...$) and let $\alpha \in (0, 1)$ such that $\sqrt[n]{a_n} \leq \alpha$ for $n \geq N$ (remember yourself the Cauchy Root Test). The series is convergent (see Theorem 29). Let now k be a natural number greater or equal to N. Let us evaluate the error $\varepsilon_k = s - s_k$. Prove that $\varepsilon_k \leq \frac{\alpha^{k+1}}{1-\alpha}$. Use this estimation to find the value of $s = \sum\limits_{n=1}^{\infty} \frac{1}{n^{n^2}}$ with 3 exact decimals.

Case 2 Suppose now that we want to approximate the value of an alternate series, $s = \sum\limits_{n=1}^{\infty} (-1)^{n-1} a_n$, where $\{a_n\}$ is a decreasing sequence with nonnegative terms and $a_n \to 0$. The Leibniz test (see Corollary 3) says that our series is convergent. Since

$$s_{2n} = s_{2n-2} + (a_{2n-1} - a_{2n}) \geq s_{2n-2}$$

and since

$$s_{2n+1} = s_{2n-1} - (a_{2n} - a_{2n+1}) \leq s_{2n-1},$$

one has:

$$s_2 \leq s_4 \leq s_6 \leq ... \leq s_{2n} \leq ... \leq s \leq ... \leq s_{2n+1} \leq ... \leq s_3 \leq s_1. \tag{3.2}$$

So,

$$0 \leq s - s_{2n} \leq s_{2n+1} - s_{2n} = a_{2n+1}$$

and

$$0 \leq s_{2n+1} - s \leq s_{2n+1} - s_{2n+2} = a_{2n+2}.$$

Hence

$$\varepsilon_n = |s - s_n| \leq a_{n+1} \tag{3.3}$$

i.e. the absolute error is less or equal to the modulus of the first neglected term. Here, in fact we have another proof of the Leibniz Test (see Theorem 3). This one is independent of the Abel-Dirichlet Test (Theorem 30). It uses only Cantor Axiom (Axiom 2) (where?).

Let us compute $s = \sum\limits_{n=1}^{\infty} (-1)^{n-1} \frac{1}{(n!)^2}$ with 2 exact decimals. We use the estimation (3.3) and force with

$$a_{n+1} = \frac{1}{[(n+1)!]^2} < \frac{1}{10^2}$$

for $n \geq 3$, so

$$s \approx s_3 = \frac{1}{1} - \frac{1}{4} + \frac{1}{36} = 0.777... = 0.(7)$$

4. Problems

1. Compute the sum of the following series:

a) $\sum_{n=2}^{\infty} \ln\left(1 - \frac{1}{n^2}\right)$; b) $\sum_{n=1}^{\infty} \frac{2^{n-1}+3^n}{5^{n+1}}$; c) $\sum_{n=1}^{\infty} \frac{1}{n(n+2)}$; d) $\sum_{n=1}^{\infty} \frac{1}{n(n+1)(n+2)}$;

e) $\sum_{n=1}^{\infty} \frac{1}{(n+2)(n+4)}$; f) $\sum_{n=1}^{\infty} (-1)^n \frac{1+2^{n-1}}{3^{n-2}}$;

2. Decide if the following series are convergent or not:

a) $\sum_{n=0}^{\infty} \frac{2^n}{n!}$; b) $\sum_{n=1}^{\infty} \frac{1\cdot 4\cdot 7\cdot \ldots \cdot (1+3n)}{1\cdot 5\cdot 9\cdot \ldots \cdot (1+4n)} \frac{1}{n}$; c) $\sum_{n=0}^{\infty} (-1)^n \frac{1}{n!}$; d) $\sum_{n=0}^{\infty} \frac{2^n+1}{2^{n+1}+1}\alpha^n$, $\alpha \geq 0$ (discussion on α); e) $\sum_{n=1}^{\infty} n\left(\frac{2\alpha-1}{2}\right)^n$ (discussion on $\alpha \in \mathbb{R}$); f) $\sum_{n=1}^{\infty} \frac{(-1)^n}{10^n n!}$; g) $\sum_{n=1}^{\infty} \frac{2\cdot 7\cdot 12\cdot \ldots \cdot [2+5(n-1)]}{3\cdot 8\cdot 13\cdot \ldots \cdot [3+5(n-1)]}$; h) $\sum_{n=0}^{\infty} \frac{(\alpha+2)^n}{2^n+3^n}$, (discussion on $\alpha \geq 0$); i) $\sum_{n=1}^{\infty} \frac{1}{n}(2\lambda - 1)^n$, (discussion on $\lambda \in \mathbb{R}$); j) $\sum_{n=1}^{\infty} \frac{(4\alpha-5)^n}{n\cdot 5^n}$, $\alpha \geq 2$ (discussion on α); k) $\sum_{n=0}^{\infty} \frac{1}{\sqrt[3]{n^\alpha+2}}$ (discussion on α); l) $\sum_{n=1}^{\infty} \frac{2^n}{1\cdot 3\cdot 5\cdot \ldots \cdot (2n-1)}(2\alpha-1)^n$, (discussion on $\alpha \geq 1$); m) $\sum_{n=1}^{\infty} \frac{1}{\sqrt[3]{4n+1} - \sqrt[3]{4n-1}}$; n) $\sum_{n=1}^{\infty} 3^{\ln n}$; o) $\sum_{n=1}^{\infty} \frac{2(n!)}{(2n)!}$; p) $\sum_{n=0}^{\infty} \frac{(-1)^n}{n!}(1+3^n)$; r) $\sum_{n=0}^{\infty} \frac{2^{n-2}}{3^{n+1}+1}$; s) $\sum_{n=1}^{\infty} \frac{5n+1}{6n-2}\alpha^n$ (discussion on $\alpha \geq 0$).

3. Find the Abel's degree of the expression $E = \frac{\sqrt[3]{n^5}+2\sqrt[5]{n^3}+n+3}{\sqrt{n+2}-\sqrt{n}}$, $n \in \mathbb{N}$.

4. Use the α-Comparison Test to decide if the series $\sum_{n=1}^{\infty} \sin\left(\frac{1}{\sqrt[3]{n}+1}\right)$ is convergent or not.

5. Find all $x \in \mathbb{R}$ such that the series $\sum_{n=0}^{\infty} \frac{\sqrt{n^2+1}}{\sqrt{n+1}}x^n$ to be convergent. What about all $x \in \mathbb{C}$ such that the same series is convergent?

6. Find all z in $\mathbb{C}$ such that the following series are absolutely convergent.

a) $\sum_{n=0}^{\infty} \frac{z^n}{n!}$; b) $\sum_{n=1}^{\infty} \frac{(z-i)^n}{n}$; c) $\sum_{n=0}^{\infty} nz^n$; d) $\sum_{n=0}^{\infty} (z - 3i + 2)^n$;

7. Draw the set $M = \left\{ x \in \mathbb{R} \mid \sum_{n=1}^{\infty} (-1)^n \frac{x^n}{n3^n} \text{ is convergent} \right\}$ on the real line.

8. Draw the set $U = \left\{ z \in \mathbb{C} \mid \sum_{n=1}^{\infty} (-1)^n \frac{z^n}{n3^n} \text{ is convergent} \right\}$ in the complex plane.

9. Compute $\sum_{n=1}^{\infty} (-1)^n \frac{1}{n^2}$ with 2 exact decimals.

10. Compute $\sum_{n=1}^{\infty} \frac{2^n}{n!}$ with one exact decimal.

11. Prove the Raabe-Duhamel test. Hint:

a) Write:

$$Na_N - (N+1)a_{N+1} \geq (\lambda - 1)a_{N+1}$$
$$(N+1)a_{N+1} - (N+2)a_{N+2} \geq (\lambda - 1)a_{N+2}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$(N+p)a_{N+p} - (N+p+1)a_{N+p+1} \geq (\lambda - 1)a_{N+p+1}$$

Sum these inequalities on columns and get:

$$Na_N - (N+p+1)a_{N+p+1} \geq (\lambda - 1)\left[a_{N+1} + a_{N+2} + a_{N+3} + ... + a_{N+p+1}\right]$$

So

$$\frac{Na_N}{\lambda - 1} \geq a_{N+1} + a_{N+2} + a_{N+3} + ... + a_{N+p+1}$$

for any $p = 1, 2, ...$ Hence, the partial sums of our initial series are bounded. Thus the series is convergent.

b) Since $na_n < (n+1)a_{n+1}$ for $n \geq M$, the limit $\lim_{n\to\infty} na_n$ is greater than 0. So, using the α-comparison test for $\alpha = 1$, we get that our initial series is divergent (why?).

c) Apply a) and b).

12. Compute $\sum_{n=1}^{\infty} \frac{1}{n^n}$ with 3 exact decimals (use the approximate computation with the Root Test).

CHAPTER 3

Sequences and series of functions

1. Continuous and differentiable functions

Recall that a metric space is a set X with a distance d on it. A distance d on X is a function which associates to any pair (x, y) of X a nonnegative real number $d(x, y)$ with the following properties:

d1. $d(x, y) = 0$ if and only if $x = y$.

d2. $d(x, y) = d(y, x)$ for any x and y in X.

d3. $d(x, y) \leq d(x, z) + d(z, y)$ for any x, y and z in X.

See also the Remark 2. We usually denote by (X, d) a metric space X with a distance d on it. The standard example of a metric space is $(\mathbb{R},\ d)$, where $d(x, y) = |x - y|$. We say that $x_n \to x$ in (X, d) if the numerical sequence $\{d(x_n, x)\}$ tends to zero, i.e. if the distance between x_n and x becomes smaller and smaller to zero as $n \to \infty$. We define again the basic notion of continuity.

DEFINITION 8. *(continuity of a function at a point) Let (X, d), (X', d') be two metric spaces, let $f : X \to X'$ be a function defined on X with values in X' and let x be a fixed element in X. We say that f is continuous at x if for any sequence $\{x_n\}$ which converges to x, we have that $f(x_n) \to f(x)$. For instance, if $X = X' = \mathbb{R}$, with the usual distance, f is continuous at a point x if the graphic of f is not "broken (or interrupted)" at x (see Fig.3.1). All the elementary functions (polynomials, rational functions, power functions, exponential functions, logarithmic functions, trigonometric functions) and their compositions are continuous on their definition domains, i.e. in any point of their definition domains (see also the Theorem 14). Hence, the continuity is essentially a "local" property, i.e. its definition shows the behavior of the function f at a given point x.*

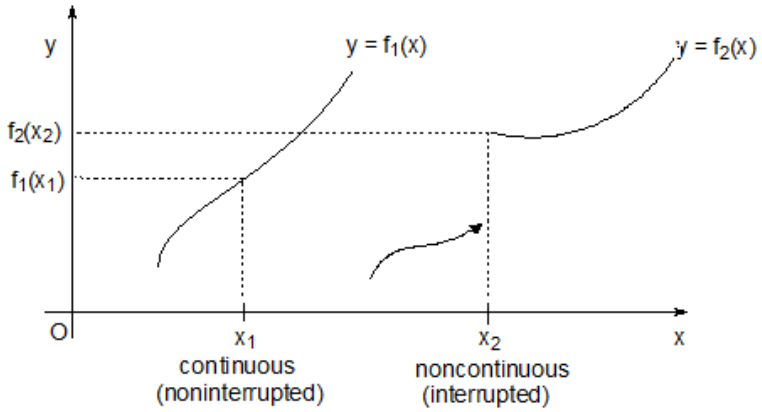


Fig. 3.1

For instance, a) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{x^3+1}{x^2+1}$ is continuous on the whole $\mathbb{R}$. Indeed, let a be a fixed point in $\mathbb{R}$ and let $\{a_n\}$ be a sequence convergent to a. Then, using the basic properties of the convergent sequences relative to the elementary algebraic operations $(+, -, \cdot, :$, see the Theorem 14), we find that

$$f(a_n) = \frac{a_n^3 + 1}{a_n^2 + 1} \to \frac{a^3 + 1}{a^2 + 1} = f(a),$$

i.e. the function f is continuous at a, for any $a \in \mathbb{R}$. Hence f is continuous on $\mathbb{R}$. Now, if we compose the function $\ln x$ (which is continuous on $(0, \infty)$) with $f(x)$ we get a new continuous function $g(x) = \ln \frac{x^3+1}{x^2+1}$ on $(-1, \infty)$ (why?).

REMARK 9. *We need in this chapter another basic "local" notion, namely the notion of differentiability of a function f at a given point a. Recall that a subset A of $\mathbb{R}$ is said to be open if for any point a of A, there is a small positive real number ε, such that the interval $(a-\varepsilon, a+\varepsilon)$ (the "ball" with centre at a and of radius ε, usually called the ε-neighborhood of a) is completely included in A (define the notion of an open subset in a metric space (X, d); instead of ε-neighborhoods use open balls $B(a, \varepsilon) = \{x \in X : d(x, a) < \varepsilon\}$, etc.). A subset B of $\mathbb{R}$ is said to be closed if its complementary $\mathbb{R} \setminus B$ is an open subset (B is closed in an arbitrary metric space (X, d) if $X \setminus B$ is open in X). For instance, $(-\infty, 1)$ is open and $[-3, 7]$ is closed. If $X = (-1, 7)$,*

with the induced distance of $\mathbb{R}$*, then* $[0,7)$ *is closed in* X*, but NOT in* $\mathbb{R}$ *(why?). It is not difficult to prove that a subset* B *is closed if and only if for any sequence* $\{b_n\} \to b$*, with all* b_n *in* B*, one has that* $b \in B$ *(prove it!). For instance, if* $f : X \to \mathbb{R}$ *is a continuous function defined on a metric space* (X,d) *and if* λ *is a real number, then the set* $B_\lambda = \{x \in X : f(x) \geq \lambda$ *(or* $\leq \lambda$*, or* $= \lambda)\}$ *is closed in* X*. Indeed, let* $\{b_n\}$ *be a sequence of elements in* B*, which is convergent to an element* b *in* X*. Since* f *is continuous,* $f(b_n) \to f(b)$*. Because* $b_n \in B$*,* $f(b_n) \geq \lambda$ *for any* $n = 0, 1, ...$ *Then* $f(b) \geq \lambda$ *(otherwise,* $f(b) < \lambda$ *and, from a rank* N *on,* $f(b_n) < \lambda$*, for* $n \geq N$ *(why?-see the definition of the limit* $f(b_n) \to f(b)$*!)), a contradiction i.e.* b *itself is in* B *and so* B *is a closed subset in* X*.*

DEFINITION 9. *Let* A *be an open subset of* $\mathbb{R}$ *(for instance an open interval* (c,d)*), let* $f : A \to \mathbb{R}$ *be a function defined on* A *with values real numbers and let* a *be a fixed point in* A*. We say that* f *is differentiable at* a *if the following limit exists (and it is a real number):*

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a} \stackrel{def}{=} f'(a) \tag{1.1}$$

The limit of a function $g : A \to \mathbb{R}$ in a limit point b (it is the limit of at least one sequence of elements from A) of A is a unique number $l \in \overline{\mathbb{R}}$ such that for any nonconstant sequence $\{b_n\}$, $b_n \in A$ which is convergent to b, one has that $g(b_n) \to l$. We shortly write $\lim\limits_{x \to b} g(x) = l$. Not always a function g has a limit at a given limit point b. For instance, the function $sign : \mathbb{R} \to \{-1, 0, 1\}$,

$$sign(x) = \begin{cases} -1, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ 1, & \text{if } x > 0 \end{cases} \tag{1.2}$$

has the limit $l = -1$ at any point $a < 0$, has the limit $l = 1$ at any point $a > 0$ and at 0 it has no limit at all (prove this!).

We recall that the limit "on the left" of a function $f : A \to \mathbb{R}$, $A \subset \mathbb{R}$, A an open subset, at a point a of A is a number l_l such that for any sequence $\{x_n\}$, $x_n < a$, which is convergent to a, one has that $l_l = \lim f(x_n)$. If we take x_n "on the right" of a, we get the notion of the limit l_r "on the right" of f at a. A function f has the limit l at a if and only if $l_l = l_r = l$ (prove it!).

It is clear enough that a continuous function f at a point $a \in A$ has the limit $l = f(a)$ at a (why?). In fact, a function $f : A \to \mathbb{R}$ is continuous at a point $a \in A$ if and only if it has a limit l at a and if that one is exactly $l = f(a)$ (prove it!).

We call the number $f'(a)$ from (1.1) the *derivative of f at a*. The linear function $df(a) : \mathbb{R} \to \mathbb{R}$, $df(a)(x) = f'(a) \cdot x$ is called the (first) *differential of f at a*. This is simply a *dilation (or a homotety)* of modulus $f'(a)$ of the real line $\mathbb{R}$. If the function f is differentiable at any point a of A, we say that f is *differentiable (or has a derivative) on A*. In this last case, the new function $a \rightsquigarrow f'(a)$, where a runs on A, is called the (first) derivative of f. It is denoted by f'. We know (see any elementary course in Calculus for the different rules in computing derivatives!) that almost all the elementary functions (described above) and their compositions (recall the chain rule: $(f \circ g)'(a) = f'(g(a)) \cdot g'(a)$) are differentiable on their definition domains. "Almost" because of some exceptions like $f(x) = \sqrt{x}$, $f : [0, \infty) \to \mathbb{R}$. Since $f'(x) = \frac{1}{2\sqrt{x}}$, the derivative of f does not exists at $a = 0$. Indeed, $\lim\limits_{x \to 0,\ x>0} \frac{\sqrt{x}-0}{x} = \infty$! One can interpret the derivative of a function f at a point a, either as "the velocity" of f at a or as the slope of the tangent line at a to the graphic of f (why?). Not all the continuous functions at a given point a are also differentiable at a (see Fig.3.2). But a differentiable function f at a given point a is continuous. Indeed, let $x_n \to a$. $\lim\limits_{x_n \to a} \frac{f(x_n)-f(a)}{x_n-a} = f'(a)$ (see Definition 9 and what follows) says that only the nondeterministic case $\frac{0}{0}$ could give a finite number $f'(a)$. Hence, $f(x_n) \to f(a)$, i.e. f is continuous at a.

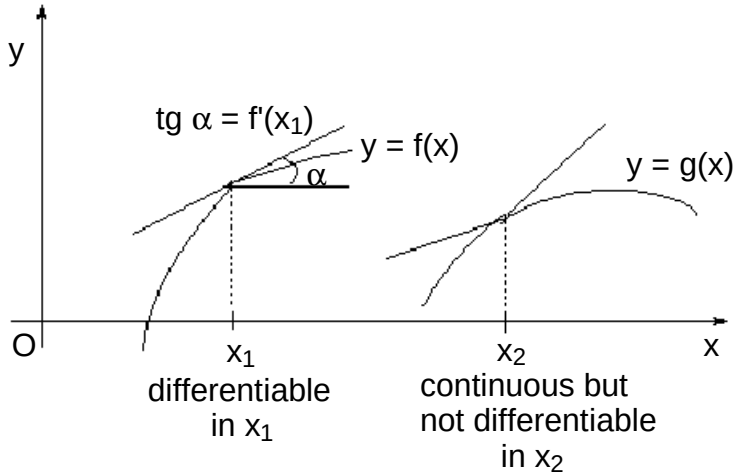


Fig. 3.2

Let C be a set and let $f : C \to \mathbb{R}$ be a function defined on C with values in $\mathbb{R}$. We say that f is *bounded* if its image $f(C) = \{f(x) : x \in C\}$ is a bounded subset in $\mathbb{R}$. This means that there is a positive real number $M > 0$ such that $|f(x)| < M$ (i.e. $-M < f(x) < M$) for any $x \in C$. Equivalently, if $C \subset \mathbb{R}$, then f is bounded if the graphic of it is contained into the band bounded by the horizontal lines: $y = -M$ and $y = M$

A fundamental property of continuous functions is the following:

THEOREM 32. *(Weierstrass boundedness theorem) Let $f : [a, b] \to \mathbb{R}$ be a continuous function defined on the closed and bounded interval $[a, b]$. Then f is bounded, $M \stackrel{def}{=} \sup f([a, b]) = f(c)$ and $m \stackrel{def}{=} \inf f([a, b]) = f(d)$, where $c, d \in [a, b]$. This means that the least upper bound ($\sup f([a, b])$) and the greatest lower bound ($\inf f([a, b])$) of the bounded set $f([a, b])$ are realized at c and at d respectively.*

PROOF. a) Let us prove that $M = \sup f([a, b]) < \infty$. Suppose on the contrary, namely that $M = \infty$. Then, there is at least one sequence $\{x_n\}$ of elements from $[a, b]$ such that $f(x_n) \to \infty$. Since $\{x_n\}$ is bounded, we can apply the Cesaro-Bolzano-Weierstrass Theorem (see Theorem 12) and find a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ which is convergent to an $x_* \in [a, b]$ (here we use the fact that $[a, b]$ is closed, how?). Since f is continuous, one has that $f(x_{n_k}) \to f(x_*)$ when $k \to \infty$. But $f(x_n) \to \infty$ and the uniqueness of the limit implies that $f(x_*) = \infty$, a contradiction (why?). Hence f is upper bounded. In the same way we can prove that f is lower bounded (do it!).

b) Let us prove now that $M = f(c)$ for a c in $[a, b]$. Since M is the least upper bound, for any natural number n we can find an element $y_n \in [a, b]$ such that

$$M - \frac{1}{n} \leq f(y_n) \leq M \text{ (why?)} \tag{1.3}$$

The sequence $\{y_n\}$ is bounded and nonconstant (why?). Applying again the Cesaro-Bolzano-Weierstrass Theorem, one can find a subsequence $\{y_{n_k}\}$ of $\{y_n\}$ which is convergent to an element $c \in [a, b]$ (because the interval is closed). Since f is continuous, $f(y_{n_k}) \to f(c)$, when $k \to \infty$. Making $k \to \infty$ in the inequality $M - \frac{1}{n_k} \leq f(y_{n_k}) \leq M$ and using the definition of a subsequence ($n_1 < n_2 < ...$), we get that $M = f(c)$. To prove that $m = f(d)$, $d \in [a, b]$, we work in the same manner (do it!). □

THEOREM 33. *(Darboux) Let $f : [a, b] \to \mathbb{R}$ be a continuous function defined on the closed and bounded interval $[a, b]$. Let $M = \sup f([a, b])$ and let $m = \inf f([a, b])$. Then the image of the interval $[a, b]$ through f*

is exactly the closed interval $[m, M]$. More general, a continuous function carries intervals into intervals.

PROOF. Let λ be an element in $[m, M]$. We want to find an element z in $[a, b]$ such that $f(z) = \lambda$. If λ is equal to m or to M, we can take $z = d$ or c (from Theorem 32) respectively. So, we can assume that $\lambda \in (m, M)$ and that f is not a constant function (in this last case the statement of the theorem is obvious). We define two subsets of the interval $[a, b]$:

$$A_1 = \{x \in [a, b] : f(x) \geq \lambda\}$$

and

$$A_2 = \{x \in [a, b] : f(x) \leq \lambda\}.$$

If $A_1 \cap A_2$ is not empty, take z in this intersection and the proof is finished. Suppose on the contrary, namely that $A_1 \cap A_2 = \varnothing$. Since λ cannot be either m or M, A_1 and A_2 are not empty (why?). Now, $[a, b] = A_1 \cup A_2$ (why?) and, since f is continuous, A_1 and A_2 are closed in $\mathbb{R}$ (see Remark 9). In order to obtain a contradiction, we shall prove that it is not possible to decompose (to write as a union, or to cover) an interval $[a, b]$ into two disjoint closed and nonempty subsets. Indeed, let $c_2 = \sup A_2$. Since f is continuous, $f(c_2) \leq \lambda$ (why?-remember the definition of the least upper bound and of the continuity!) i.e. $c_2 \in A_2$. If $c_2 \neq b$, then the subset $S_1 = \{x \in A_1 : x > c_2\}$ is not empty (why?). Take now $c_1 = \inf S_1$. Since A_1 is closed, $c_1 \in A_1$ (why?). If $c_1 > c_2$, take $h \in (c_2, c_1)$. This $h \in [a, b]$ and it cannot be either in A_1 or in A_2 (why?). Since $c_1 \geq c_2$, the unique possibility for c_1 is to be equal to c_2. But then, $c = c_1 = c_2 \in A_1 \cap A_2 = \varnothing$, a contradiction! Hence, $c_2 = \sup A_2 = b$. Take now $d_2 = \inf A_2$. Since A_2 is closed, one has that $d_2 \in A_2$. If $d_2 \neq a$, then the subset $S_2 = \{x \in A_1 : x < d_2\}$ is not empty (why?). Take now $d_1 = \sup S_2$. Since A_1 is closed, $d_1 \in A_1$ (why?). If $d_1 < d_2$, take again $g \in (d_1, d_2)$ and this last one cannot be either in A_1 or in A_2. Hence $d_1 = d_2 \overset{not}{=} d$ and this one must be in $A_1 \cap A_2$, a contradiction! So, $d_2 = a$, i.e. $\inf A_2 = a$ and $\sup A_2 = b$, thus $A_2 = [a, b]$. Since A_1 is not empty and it is included in $[a, b]$, $A_1 \subset A_2$, and we get again a new and the last contradiction! Hence $A_1 \cap A_2$ cannot be empty and the proof of the theorem is over. □

We agree with the reader that the proof of this last theorem is too long! But,...it is so clear and so elementary! Trying to understand and to reproduce logically the above proof is a good exercise for strengthen your power of concentration and not only!

THEOREM 34. *Let I be an open interval on the real line and let $f : I \to \mathbb{R}$, be a continuous function defined on I with real values.*

1) Assume that there are two points b and d in I $(b < d)$ such that the values $f(b)$ and $f(d)$ are nonzero and have distinct signs. Then, there is a point c in the interval (b, d) at which the value of f is zero, i.e. $f(c) = 0$. 2) Now suppose that at $a \in I$ the value $f(a) > 0$ (or $f(a) < 0$). Then there is an ε-neighborhood $(a - \varepsilon, a + \varepsilon) \subset I$, such that $f(x) > 0$ (or $f(x) < 0$) for any $x \in (a - \varepsilon, a + \varepsilon)$.

PROOF. 1) We can simply apply Theorem 33. Indeed, since $f(I)$ is an interval (Theorem 33), the segment generated by $f(b)$ and $f(d)$ is completely contained in $f([b, d])$. Since $f(b)$ and $f(d)$ have distinct signs, 0 is between them, so, $0 \in f([b, d])$, or $0 = f(c)$ for a $c \in [b, d]$. 2) Suppose that $f(a) > 0$. Let us assume contrary, i.e. for all small possible ε we can find in $(a - \varepsilon, a + \varepsilon)$ at least on number x_ε (an x which depends on ε) such that $f(x_\varepsilon) \leq 0$. Take for such epsilons the values

$$1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...,$$

and find $x_{\frac{1}{n}} \in (a - \frac{1}{n}, a + \frac{1}{n})$ with $f(x_{\frac{1}{n}}) \leq 0, n = 1, 2, ...$. Since f is continuous at a and since the sequence $\{x_{\frac{1}{n}}\}$ tends to a (why?), one has that $f(x_{\frac{1}{n}}) \to f(a)$. But $f(x_{\frac{1}{n}})$ are all nonpositive, so $f(a)$ is nonpositive, a contradiction! Hence, there is at least one ε small enough such that for any x in $(a - \varepsilon, a + \varepsilon)$, $f(x) > 0$. The case $f(a) < 0$ can be similarly manipulated (do it!). □

DEFINITION 10. *Let (X, d) be a metric space and let I be an interval on the real line $\mathbb{R}$ (a subset I of $\mathbb{R}$ is said to be an interval if for any pair of numbers r_1, $r_2 \in I$ and any real number r with $r_1 \leq r \leq r_2$, one has that $r \in I$). Practically, we think of a curve in X as being the image in X of an interval I through a continuous function $h : I \to X$. More exactly, we denote the couple (I, h) by a small greek letter γ and say that γ is a curve in X. If A and B are two "points" (elements) in X, we say that a curve $\gamma = (I, h)$ connects A and B if there are $a, b \in I$ such that $A = h(a)$ and $B = h(b)$. By an (closed) arc $[AB]$ in X we mean the image in X of a closed interval $[a, b]$ of $\mathbb{R}$ through a continuous function $h : [a, b] \to X$, i.e. $[A, B] = \{x \in X :$ there is $c \in [a, b]$ with $h(c) = x\}$.*

EXAMPLE 1. *a) Let $\{O; \mathbf{i}, \mathbf{j}, \mathbf{k}\}$ be a Cartesian coordinate system in the vector space V_3 of all free vectors in our 3-D space (identified with $\mathbb{R}^3$). Any point M in $\mathbb{R}^3$ has 3 coordinates: $M(x, y, z)$, where $\overrightarrow{OM} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$, $x, y, z \in \mathbb{R}$. Let $A(a_1, a_2, a_3)$ and $B(b_1, b_2, b_3)$ be two*

points in $\mathbb{R}^3$. *The usual segment* $[A,B]$ *is a closed arc which connect the points* A *and* B. *Indeed, let* $h:[0,1]\to\mathbb{R}^3$, $h(t)=(a_1+t(b_1-a_1),a_2+t(b_2-a_2),a_3+t(b_3-a_3))$, *be the usual continuous parameterization of the segment* $[A,B]$:

$$\begin{cases} x=a_1+t(b_1-a_1) \\ y=a_2+t(b_2-a_2) \\ z=a_3+t(b_3-a_3) \end{cases},t\in[0,1]$$

Here $\gamma=([0,1],h)$ *is a curve in* $\mathbb{R}^3$. *This function* h *describes a composition between the dilation of moduli* b_1-a_1,b_2-a_2,b_3-a_3, *along the* Ox, Oy, *and* Oz *axes respectively, and the translation* $\mathbf{x}\to\mathbf{a}+\mathbf{x}$, *of center* $\mathbf{a}=(a_1,a_2,a_3)$.

b) Let $C=\{(x,y)\in\mathbb{R}^2:(x-a)^2+(y-b)^2=r^2\}$ *be the circle with center at* (a,b) *and radius* r. *The parametrization of* C

$$\begin{cases} x=a+r\cos t \\ y=b+r\sin t \end{cases},t\in[0,2\pi]$$

give rise to a curve $\gamma=([0,2\pi],h)$, *where* $h(t)=(a+r\cos t,b+r\sin t)$. *In fact,* h *describes the continuous deformation process of the segment* $[0,2\pi]\subset\mathbb{R}$ *into the circle* C *in the metric space* $\mathbb{R}^2$.

Definition 11. *A subset* A *of a metric space* (X,d) *is said to be connected if any pair of two points* M_1 *and* M_2 *of* A *can be connected by a continuous curve* $\gamma=(I,h)$, $h:I\to X$.

Corollary 4. *The connected subsets in* $\mathbb{R}$ *are exactly the intervals of* $\mathbb{R}$ *(for proof use the Darboux Theorem 33).*

For instance, $A=[0,1]\cup[5,8]$ is not connected because it is not an interval (4 is between 0 and 8, but it is not in A!).

Remark 10. *A subset* S *of* $\mathbb{R}^3$ *is said to be convex if for any pair of points* $A,B\in S$, *the whole segment* $[A,B]$ *is included in* S. *For instance, the parallelepipeds, the spheres, the ellipsoids, etc., are convex subsets of* $\mathbb{R}^3$. *The union between two tangent spheres is connected but it is not convex! (why?). It is clear that any convex subset of* $\mathbb{R}^3$ *is also a connected subset in* $\mathbb{R}^3$ *(prove it!).*

Definition 12. *Let* $f:A\to\mathbb{R}$ *be a function defined on an open subset* A *of* $\mathbb{R}$ *with values in* $\mathbb{R}$. *A point* a *of* A *is a local maximum point of* f *if there is an* ε*-neighborhood of* a, $(a-\varepsilon,a+\varepsilon)\subset A$, *such that* $f(x)\leq f(a)$ *for any* $x\in(a-\varepsilon,a+\varepsilon)$. *The value* $f(a)$ *of* f *at* a *is called a local extremum (maximum) for* f. *A point* b *of* A *is said to be a local minimum point for* f *if there is an* η*-neighborhood of* b, $(b-\eta,b+\eta)\subset A$, *such that* $f(x)\geq f(b)$ *for any* $x\in(b-\eta,b+\eta)$. *The value* $f(b)$ *of* f

at b is called a local extremum (minimum) for f. A local maximum point or a local minimum point is called a local extremum point. The local extrema of f on A are all the local maxima and the local minima of f in A. The (global) maximum of f on A is $\max f(A)$ *($\in \overline{\mathbb{R}}$). The (global) minimum of f on A is* $\min f(A)$ *($\in \overline{\mathbb{R}}$) (see Fig.3.3).*

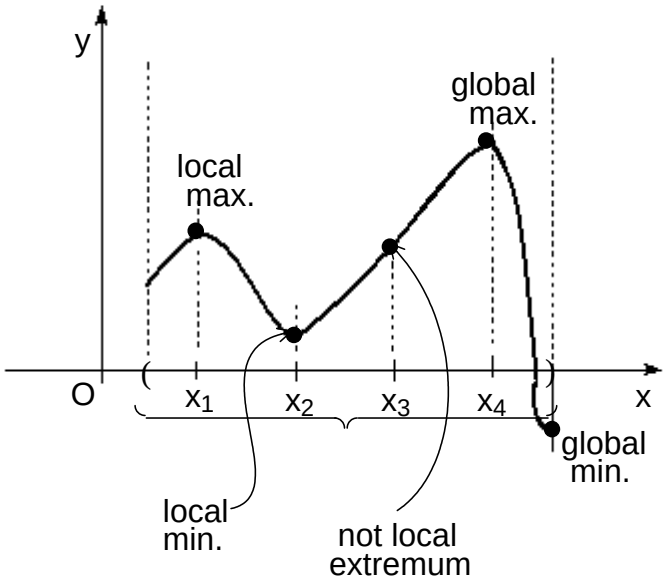


Fig. 3.3

A *critical* (or *stationary*) point $c \in A$ for a differentiable function $f : A \to \mathbb{R}$ on A is a root of the equation $f'(x) = 0$, i.e. $f'(c) = 0$. For instance, $c = 2$ is a stationary point for $f(x) = (x-2)^3$, $f : \mathbb{R} \to \mathbb{R}$, but it is not an extremum point for f (why?). The next result clarifies the converse situation.

THEOREM 35. *(1-D Fermat's Theorem) Let a be a local extremum (local maximum or local minimum) point for a function $f : A \to \mathbb{R}$ (A is open). Assume that f is differentiable at a. Then $f'(a) = 0$, i.e. a is a critical point of f. Practically, this statement says that for a differentiable function f we must search for local extrema between the critical points of f, i.e. between the solutions of the equation $f'(x) = 0$, $x \in A$.*

PROOF. Suppose that a is a local maximum point for f, i.e. there is a small $\varepsilon > 0$ such that $(a-\varepsilon, a+\varepsilon) \subset A$ and $f(x) \leq f(a)$ for any

x in $(a-\varepsilon, a+\varepsilon)$ (if a is a local minimum point, one proceeds in the same way, do it!). Look now at the formula:

$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a}=f'(a)! \tag{1.4}$$

If $x\in(a-\varepsilon,a+\varepsilon)$ and $x<a$, since $f(x)\leq f(a)$, one has that $f'(a)\geq 0$ (why?). Now, if $x\in(a-\varepsilon,a+\varepsilon)$, but $x>a$, again since $f(x)\leq f(a)$, one gets that $f'(a)\leq 0$. Both inequalities give us that $f'(a)=0$ and the Fermat's theorem for a function of one variable is proved. □

However, the Fermat's Theorem works only at the points at which our function is differentiable. For instance, $f(x)=|x|$ has at $x=0$ a local (even a global) minimum (why?), but it is not differentiable at this point (why?). The moral is that we must consider separately the points at which a function is not differentiable and see (using the definition only!) if these points are or not local extremum points for our function.

THEOREM 36. *(Rolle Theorem) Let $f:[a,b]\to\mathbb{R}$ $(a<b)$ be a continuous function. Assume that f is differentiable on the open subinterval (a,b) and that $f(a)=f(b)$. Then there is at least one point $c\in(a,b)$ such that $f'(c)=0$.*

PROOF. Let us apply the Weierstrass boundedness theorem (Theorem 32) and find $m=\inf f([a,b])$ and $M=\sup f([a,b])$ as real numbers. If $m=M$, then our function is a constant function and so, $f'(x)=0$ for any x in (a,b). Hence we assume that $m\neq M$. So the number $f(a)=f(b)$ cannot be simultaneously equal to m and M. Suppose for instance that $f(a)=f(b)\neq M$. Thus, a c with $M=f(c)$, $c\in[a,b]$ (see the Weierstrass boundedness theorem) cannot be either a or b, i.e. $c\in(a,b)$. Therefore, this c is a local maximum for f. Use now Fermat's Theorem and find that $f'(c)=0$. □

For instance, if $f(x)=x^4-16$, $x\in[-1,1]$, then $f(-1)=f(1)=-15$ and $f'(x)=0$ supplies us with a unique solution $c=0$. The continuity at the ends of the interval $[a,b]$ is necessary, as we can see in the following example. Let us take

$$f(x)=\begin{cases} x, \text{if } x\in[0,1) \\ 0, \text{if } x=1 \end{cases}, x\in[0,1].$$

This function is defined on $[0,1]$, it is differentiable on $(0,1)$ and $f(0)=f(1)$, but its derivative $f'(x)=1$ has no zero on $(0,1)$.

2. Sequences and series of functions

We know to measure the length $\|\mathbf{a}\| = \sqrt{a_1^2 + a_2^2 + a_3^2}$ of a vector $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ of V_3, the 3-dimensional vector space of all free vectors (here $a_1, a_2, a_3 \in \mathbb{R}$ are the coordinates of $\mathbf{a}$). The function $\mathbf{a} \rightsquigarrow \|\mathbf{a}\|$, which associates to a vector $\mathbf{a}$ its length $\|\mathbf{a}\|$, has the following basic properties:

$$n1.\ \|\mathbf{a}\| = 0, \text{ if and only if } \mathbf{a} = \mathbf{0},$$

$$n2.\ \|\mathbf{a} + \mathbf{b}\| \leq \|\mathbf{a}\| + \|\mathbf{b}\|,$$

for any $\mathbf{a}, \mathbf{b} \in V_3$,

$$n3.\ \|\lambda\mathbf{a}\| = |\lambda|\,\|\mathbf{a}\| \text{ for any } \lambda \in \mathbb{R} \text{ and } \mathbf{a} \in V_3. \tag{2.1}$$

If instead of V_3 we take any real vector space V together with a mapping like above, $x \to \|x\| \in [0, \infty)$, $x \in V$, which fulfils the analogous requirements $n1$, $n2$ and $n3$ from (2.1), we get the general notion of a *normed space* $(V, \|.\|)$.

Definition 13. *Let V be an arbitrary real vector space and let $f \rightsquigarrow \|f\|$ be a mapping which associates to any element f of V a nonnegative real number $\|f\|$. If this mapping satisfies the following properties:*

$$ns1.\ \|f\| = 0, \text{ if and only if } f = 0, f \in V,$$

$$ns2.\ \|f + g\| \leq \|f\| + \|g\|,$$

for any $f, g \in V$ and,

$$ns3.\ \|\lambda f\| = |\lambda|\,\|f\| \text{ for any } \lambda \in \mathbb{R} \text{ and } f \in V,$$

we say that the pair $(V, \|.\|)$ is a normed space and the mapping $x \rightsquigarrow \|x\|$ (the norm of x) is called a norm application (function) or simply a norm on V.

For instance, the norm of a matrix $A = (a_{ij})$, $i = 1, 2, ..., n$; $j = 1, 2, ..., m$, is

$$\|A\| = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2}.$$

The mapping $A \rightsquigarrow \|A\|$ satisfies the properties of a norm (prove it!) on the vector space of all $n \times m$ matrices. In addition, one can prove (not so easy!) that

$$ns4.\ \|AB\| \leq \|A\| \, \|B\| \tag{2.2}$$

for any two matrices $n \times m$ and $m \times p$ respectively.

REMARK 11. *It is easy to see that a normed space $(V, \|.\|)$ is also a metric space with the induced distance d, where $d(x, y) = \|x - y\|$ (prove this!). For instance, $\{x_n\} \to x$ if and only if $\|x_n - x\| \to 0$ as $n \to \infty$.*

If we consider now a bounded function $f : A \to \mathbb{R}$ defined on an arbitrary set A with real values, we can define the norm ("length") of f by the formula: $\|f\| = \sup |f(A)|$, where $|f(A)| = \{|f(a)| : a \in A\}$ is the absolute value of the image of A through f, or simply the modulus of the image of f. This norm is also called the sup-*norm*.

THEOREM 37. *Let $\mathcal{B}(A) = \{f : A \to \mathbb{R}$, f bounded$\}$ be the vector space of all bounded functions defined on a fixed set A. Then the mapping $f \rightsquigarrow \|f\|$ is a norm on $\mathcal{B}(A)$ with the additional property:*

$$n4.\ \|fg\| \leq \|f\| \, \|g\|$$

for any $f, g \in \mathcal{B}(A)$. Moreover, any Cauchy sequence $\{f_n\}$ with respect to this norm is a convergent sequence in $\mathcal{B}(A)$.

PROOF. Let us prove for instance $ns2$. Since

$$|f(a) + g(a)| \leq |f(a)| + |g(a)| \leq$$

$$\leq \sup\{|f(a)| : a \in A\} + \sup\{|g(a)| : a \in A\},$$

taking sup on the left side (it exists, because it is upper bounded by a constant quantity), we get the property $n2.$: $\|f + g\| \leq \|f\| + \|g\|$. The property $n4.$ can be proved in the same manner (do it!). The other properties are obvious (prove them with all details!). Let us prove the last statement. Since

$$|f_{n+p}(x) - f_n(x)| \leq \sup\{|f_{n+p}(x) - f_n(x)| : x \in A\} = \|f_{n+p} - f_n\|,$$

for a fixed x in A, the numerical sequence $\{f_n(x)\}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, i.e. any Cauchy sequence in $\mathbb{R}$ has a (unique) limit in $\mathbb{R}$, let us associate to x the limit $\lim\limits_{n\to\infty} f_n(x)$, denoted by $f(x)$, i.e. a real number which depends on x. We shall prove that this new function $f : A \to \mathbb{R}$:1) is bounded, i.e. belongs to $\mathcal{B}(A)$ and 2) it is the limit of the sequence $\{f_n\}$ in $\mathcal{B}(A)$, relative to the sup-norm. For

2) let us take a small $\varepsilon > 0$ and let us find a rank N which depends on ε such that

$$\|f_{n+p} - f_n\| < \varepsilon \tag{2.3}$$

for any $n \geq N$ and for any $p = 1, 2, ...$ Since $f_n(x) \to f(x)$ for any fixed x in A and since

$$|f_{n+p}(x) - f_n(x)| \leq \|f_{n+p} - f_n\| < \varepsilon$$

for any $n \geq N$ and any p, let us make p large enough, i.e. $p \to \infty$ in the last inequality. We get $|f(x) - f_n(x)| \leq \varepsilon$ (why?) for $n \geq N$ and for any x in A. Take now sup on the left and get:

$$\|f - f_n\| \leq \varepsilon \tag{2.4}$$

for any $n \geq N$. Hence $f_n \overset{\|.\|}{\to} f$. We make $n = N$ in (2.4) and write

$$|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| \leq \|f - f_N\| + \|f_N\| \leq \varepsilon + \|f_N\| .$$

Take now sup on the left and we get:

$$\|f\| \leq \varepsilon + \|f_N\| ,$$

i.e. f is bounded and so, $f_n \overset{\|.\|}{\to} f$ in $\mathcal{B}(A)$. □

Definition 14. *Let $\{f_n\}$ be a sequence of bounded functions on A and let f be another bounded function on A. We say that the sequence $\{f_n\}$ is uniformly convergent to f (write $f_n \overset{uc}{\to} f$) if the sequence of numbers $\{\|f_n - f\|\}$ is convergent to 0. If for any fixed $x \in A$ the sequence of numbers $\{f_n(x)\}$ is convergent to $f(x)$, we say that the sequence of functions $\{f_n\}$ is simply (or pointwise) convergent to f ($f_n \overset{sc}{\to} f$). Since $|f_n(x) - f(x)| \leq \|f_n - f\|$, the uniform convergence implies the simple convergence (why?-give details!).*

The notion of uniform convergence is stronger then the notion of simple convergence. For instance, let

$$f_n(x) = x^n, x \in [0, 1].$$

Here $A = [0, 1]$ and, for $x \in [0, 1)$, $\lim_{n\to\infty} f_n(x) = 0$ (why?). For $x = 1$, $\lim_{n\to\infty} f_n(1) = 1$. So, the pointwise limit function $f(x) = 0$, if $0 \leq x < 1$ and $f(1) = 1$. Hence, the sequence of functions $\{f_n\}$ is pointwise convergent to this f. Let us evaluate now

$$\|f_n - f\| = \sup\{|f_n(x) - f(x)| : x \in [0, 1]\} = 1.$$

Hence $\|f_n - f\| = 1$ does not tend to 0! So, the sequence of functions is not uniformly convergent.

REMARK 12. *(Weierstrass) Not always we must compute exactly the norm* $\|f_n - f\|$. *In fact, for the uniform convergence to* f *of the sequence* $\{f_n\}$, *it is sufficient to find a sequence of numbers* $\{\alpha_n\}$ *such that* $|f_n(x) - f(x)| \leq \alpha_n$ *for any* $x \in A$ *and for any* $n \geq N$ *(a fixed natural number) such that* $\{\alpha_n\} \to 0$ *(why?). For instance, take* $f_n(x) = \frac{\sin nx}{n}$. *Since for any fixed* $x \in \mathbb{R}$, $\left|\frac{\sin nx}{n}\right| \leq \frac{1}{n}$, *we have that* $f_n(x) \to 0$, *when* $n \to \infty$. *But the right side of this last inequality is independent on* x. *So we can take* $\alpha_n = \frac{1}{n}$ *and apply the above remark of Weierstrass. Hence* $f_n(x) = \frac{\sin nx}{n}$ *is uniformly convergent to* 0 *on* $\mathbb{R}$. *If instead of* $\sin nx$ *one takes any other bounded function* $g(x)$ *on an arbitrary interval* $I \subset \mathbb{R}$, *we get that* $f_n(x) = \frac{g(x)}{n}$ *is uniformly convergent to* 0 *on* I *(prove it!).*

In order to test the uniform convergence of a sequence of continuous functions we can use the following result.

THEOREM 38. *Let* (X, d) *be a metric space and let* $\{f_n\}$ *be a uniformly convergent sequence of bounded continuous functions defined on* X *with real or complex values. Let* f *be the limit function of* $\{f_n\}$. *Then the function* f *itself is a bounded and continuous function on* X.

PROOF. Recall that $\|f_n\| = \sup |f_n(X)| < \infty$ for any $n = 1, 2, ...$ (f_n is bounded). Let $\varepsilon > 0$ be a small positive real number and let N be a rank (a fixed natural number) such that

$$\|f - f_n\| < \varepsilon \text{ for any } n \geq N. \tag{2.5}$$

1) Let us prove that f is bounded on X. Take $n = N$ in (2.5), remember the basic property of the norm function (see Theorem 37) and write

$$\|f\| = \|(f - f_N) + f_N\| \leq \|f - f_N\| + \|f_N\| < \varepsilon + \|f_N\|.$$

Since f_N is bounded ($\|f_N\| < \infty$), we get that f is also bounded.

2) In order to prove the continuity of f at a fixed point a of X, let us take a sequence $\{a_k\}$ which is convergent to a, when $k \to \infty$. Since $\{f_n\}$ is uniformly convergent to f, there is a large number L such that $\|f - f_L\| < \frac{\varepsilon}{3}$. Since this f_L is continuous, there is a rank K such that for any $k \geq K$ one has

$$|f_L(a_k) - f_L(a)| < \frac{\varepsilon}{3}.$$

Now,

$$|f(a_k) - f(a)| = |f(a_k) - f_L(a_k) + f_L(a_k) - f(a)| \leq \tag{2.6}$$

$$\leq |f(a_k) - f_L(a_k)| + |f_L(a_k) - f(a)| \leq$$

$$\leq \sup\{|f(x) - f_L(x)| : x \in X\} + |f_L(a_k) - f(a)| =$$
$$= \|f - f_L\| + |f_L(a_k) - f(a)|$$

But,

$$|f_L(a_k) - f(a)| = |f_L(a_k) - f_L(a) + f_L(a) - f(a)| \leq \tag{2.7}$$
$$\leq |f_L(a_k) - f_L(a)| + |f_L(a) - f(a)| \leq \frac{\varepsilon}{3} + \sup\{|f_L(x) - f(x)| : x \in X\} =$$
$$= \frac{\varepsilon}{3} + \|f_L - f\|,$$

for any $k \geq K$ (here we just used the continuity of f_L). Combining the inequalities (2.6) and (2.7), we find

$$|f(a_k) - f(a)| \leq \|f - f_L\| + \frac{\varepsilon}{3} + \|f_L - f\| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

for any $k \geq K$. Hence $f(a_k) \to f(a)$, so f is continuous at a. □

This last result is useful whenever we want to prove that a sequence of continuous functions $\{f_n\}$ is NOT uniformly convergent. Namely, we construct the limit function $f(x) = \lim_{n\to\infty} f_n(x)$ for any fixed x. If the function $f(x)$ is not continuous, then, because of Theorem 38, we must conclude that $\{f_n\}$ cannot be uniformly convergent to f.

For instance, the sequence $f_n(x) = x^n$, $x \in [0, 1]$ is convergent to $f(x) = 0$ if $x \in [0, 1)$ and $f(1) = 1$. Since this last function is not continuous, our sequence cannot be uniformly convergent to f. It is only simply convergent to f.

Sometimes it is useful to integrate term by term a sequence of functions and see what happens with the limit function.

THEOREM 39. *Let $\{f_n\}$ be a sequence of continuous functions, which is uniformly convergent to a continuous (see Theorem 38) function f on the interval $[a, b]$. For any fixed $x \in [a, b]$ one defines $F_n(x) = \int_a^x f_n(t)dt$, $n = 0, 1, ...$ and $F(x) = \int_a^x f(t)dt$ be the canonical primitives of f_n and of f respectively on $[a, b]$. Then, the sequence $\{F_n\}$ is uniformly convergent to F on $[a, b]$. In particular, for $x = b$, we get a very useful relation:*

$$\lim_{n\to\infty} \int_a^b f_n(t)dt = \int_a^b \lim_{n\to\infty} f_n(t)dt. \tag{2.8}$$

PROOF. Let us evaluate

$$\|F_n - F\| = \sup\{|F_n(x) - F(x)|, x \in [a, b]\} \leq$$
$$\leq \sup\{\int_a^x |f_n(t) - f(t)| \, dt : x \in [a, b]\} \leq$$

$$\leq \|f_n - f\| \sup\{\int_a^x dt : x \in [a,b]\} = (b-a)\,\|f_n - f\|\,. \tag{2.9}$$

Now, since $\{f_n\}$ is uniformly convergent to f, the numerical sequence $\|f_n - f\|$ tends to zero. Hence, since 2.9 says that

$$\|F_n - F\| \leq \|f_n - f\|\,(b-a),$$

we have that $\|F_n - F\| \to 0$, i.e. $\{F_n\}$ is uniformly convergent to F on $[a,b]$. □

In the following we show how to use this result in practice.

Let us take the sequence of functions $f_n(x) = nxe^{-nx^2}$, $x \in [0,1]$. It is clear that this sequence is simply convergent to the continuous function $f(x) = 0$ for any x in $[0,1]$. Since f is continuous we cannot decide if our sequence is uniformly convergent or not, only by using Theorem 38. If the sequence were uniformly convergent, then, using the relation (2.8) we would get:

$$\lim_{n\to\infty} \int_0^1 nxe^{-nx^2}dx = \int_0^1 \lim_{n\to\infty} nxe^{-nx^2}dx = 0. \tag{2.10}$$

But

$$\int_0^1 nxe^{-nx^2}dx = -\frac{1}{2}e^{-nx^2}\,|_0^1 = -\frac{1}{2}[e^{-n} - 1] \to \frac{1}{2} \neq 0.$$

Hence, our assumption cannot be true. So, our sequence is not uniformly convergent on $[0,1]$.

REMARK 13. *In Theorem 39 we saw that a uniformly convergent sequence of continuous functions can be "termwisely" integrated. But what about their "termwise" derivatives? Can we "termwisely" differentiate a uniformly convergent sequence of differentiable functions? In general, we cannot, as the following example shows. Let* $f_n(x) = \frac{x^n}{n}$, $x \in [0,1]$. *Since* $\|f_n - 0\| = \sup\{\frac{x^n}{n} : x \in [0,1]\} = \frac{1}{n} \to 0$, *when* $n \to \infty$, *we find that* $\{f_n\}$ *is uniformly convergent to* $f(x) = 0$ *on* $[0,1]$. *But* $f_n'(x) = x^{n-1}$ *is not uniformly convergent on* $[0,1]$ *as we saw above.*

THEOREM 40. *If we want to differentiate "termwisely" the sequence* $\{f_n\}$ *of differentiable functions on* $[a,b]$, *the following conditions are sufficient: 1)* $\{f_n\}$ *is uniformly convergent to* f *on* $[a,b]$, *2)* $\{f_n'\}$ *is uniformly convergent to* g *on* $[a,b]$ *and 3)* $f_n \in C^1[a,b]$ *for any* $n = 0, 1, \ldots$. *Then* f *is also differentiable and* $f' = g$ *(*$\Rightarrow$ f *is also of class* C^1 *on* $[a,b]$*).*

PROOF. Indeed, using Theorem 39 for the sequence $f_n' \overset{uc}{\to} g$, one has that

$$F_n(x) = \int_a^x f_n'(t)dt = f_n(x) - f_n(a) \overset{uc}{\to} \int_a^x g(t)dt. \tag{2.11}$$

Since $f_n \overset{uc}{\to} f$ one has that $f(x) - f(a) = \int_a^x g(t)dt$ (why?). Let x_0 be a point in $[a,b]$. Since $\int_{x_0}^x g(t)dt = g(c_x)\cdot(x - x_0)$ (mean formula), where c_x is a point in the segment $[x_0, x]$,

$$\lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{x\to x_0} g(c_x) = g(x_0).$$

So, $f'(x_0)$ exists and it is equal to $g(x_0)$. Hence, $f' = g$ on $[a,b]$. □

DEFINITION 15. *Let $\{f_n\}$ be a sequence of functions defined on a subset A of $\mathbb{R}$. For every $n = 0, 1, ...$ we denote by*

$$s_n(x) = f_0(x) + f_1(x) + ... + f_n(x).$$

A series of functions f_n is an "infinite" sum

$$\sum_{k=0}^{\infty} f_k.$$

If the sequence of "partial sums" $\{s_n\}$ is simply convergent to the function s on A, we say that the series $\sum_{k=0}^{\infty} f_k$ is simply (pointwise) convergent to s (its sum) on A. If the sequence $\{s_n\}$ is uniformly convergent to s on A, we say that the series $\sum_{k=0}^{\infty} f_k$ is uniformly convergent to s (its sum) on A. In this last case, we simply write $s = \sum_{k=0}^{\infty} f_k$.

Let the series of functions

$$\sum_{k=0}^{\infty} x^k = \lim_{n\to\infty}(1 + x + x^2 + ... + x^n) = \lim_{n\to\infty} \frac{1 - x^{n+1}}{1 - x} = \frac{1}{1 - x},$$

for any $x \in (-1, 1)$. So, the (geometric) series $\sum_{k=0}^{\infty} x^k$ is simply (pointwise) convergent to $\frac{1}{1-x}$ on $(-1,1)$. Let us see if it is uniformly convergent on $(-1,1)$. For this, let us evaluate

$$\|s_n - s\| = \left\| \frac{1 - x^{n+1}}{1 - x} - \frac{1}{1 - x} \right\| =$$

$$= \left\| \frac{x^{n+1}}{1 - x} \right\| = \sup\left\{ \left| \frac{x^{n+1}}{1 - x} \right| : x \in (-1, 1) \right\} = \infty.$$

Hence, our series is not uniformly convergent on the whole interval $(-1,1)$ but,...it is uniformly convergent on every closed subinterval $[a,b]$ of $(-1,1)$. Indeed, in this case, if we denote by $c=\max\{|a|\,,\,|b|\}$, we get

$$\|s_n-s\|\leq\frac{c^{n+1}}{1-a}\to 0,\text{ when } n\to\infty,$$

because $c\in(0,1)$. Thus the series is uniformly convergent on $[a,b]$.

Sometimes, it is very difficult to evaluate "the error function" s_n-s. This is why we need some other tools for deciding if a series is uniformly convergent or not. A series of functions $\sum\limits_{k=0}^{\infty}f_k$ is said to be *absolutely uniformly convergent* if the series of the moduli of these functions $\sum\limits_{k=0}^{\infty}|f_k|$ is uniformly convergent. Recall that $|f|(x)\overset{def}{=}|f(x)|$. It is not difficult to see that an absolutely uniformly convergent series of functions $\sum\limits_{k=0}^{\infty}f_k$ is also uniformly convergent. Indeed, let $S_n=\sum\limits_{k=0}^{n}|f_k|$ and let $S=\sum\limits_{k=0}^{\infty}|f_k|$ be the sum of the series of moduli. Then

$$|s(x)-s_n(x)|=|f_{n+1}(x)+f_{n+2}(x)+...|\leq|f_{n+1}(x)|+|f_{n+2}(x)|+...$$

(why?)

$$=S(x)-S_n(x)\leq\sup\{|S(x)-S_n(x)|:x\in A\}=\|S-S_n\|\,.$$

Hence $|s(x)-s_n(x)|\leq\|S-S_n\|$ for any $x\in A$. Taking now sup on $x\in A$ we get that $\|s_n-s\|\leq\|S-S_n\|$. Since our series is absolutely uniformly convergent, then $\|S-S_n\|\to 0$, when $n\to\infty$. Using now the last inequality, we get that $\|s_n-s\|\to 0$, i.e. the initial series is uniformly convergent. A powerful and useful test for the absolute uniform convergence is the following test.

THEOREM 41. *(Weierstrass Test for series of functions) Let A be a subset of real numbers and let $\sum\limits_{k=0}^{\infty}f_k$ be a series of functions defined on A. Assume that $\|f_n\|$ can be upper bounded by $\alpha_n\in[0,\infty)$ ($|f_n(x)|\leq\alpha_n$ where x runs on A) for any $n=0,1,...$ and that the numerical series $\sum\limits_{k=0}^{\infty}\alpha_k$ is convergent. Then the series $\sum\limits_{k=0}^{\infty}f_k$ is absolutely uniformly convergent. In particular, it is also uniformly convergent.*

PROOF. Let us fix a small positive real number $\varepsilon>0$ and an $x\in A$. Let

$$S_n=|f_0|+|f_1|+...+|f_n|$$

be the n-th partial sum of the series $\sum_{k=0}^{\infty} |f_k|$. Since the numerical series $\sum_{k=0}^{\infty} \alpha_k$ is convergent, there is a rank N such that

$$\alpha_{n+1} + \alpha_{n+2} + ... + \alpha_{n+p} < \varepsilon$$

for any $n \geq N$ and for any natural number p.

Let us evaluate $|S_{n+p}(x) - S_n(x)|$:

$$|S_{n+p}(x) - S_n(x)| = |f_{n+1}(x)| + |f_{n+2}(x)| + ... + |f_{n+p}(x)| \leq \alpha_{n+1} + \alpha_{n+2} + ... + \alpha_{n+p} < \varepsilon. \tag{2.12}$$

From (2.12) we obtain that the sequence $\{S_n(x)\}$ is a Cauchy sequence of real numbers (see Definition 2). Since on the real line any Cauchy sequence is convergent (see Theorem 13) we get that the sequence $\{S_n(x)\}$ is convergent to a real number $S(x)$ (this means that this real number depends on x, i.e. it is changing if we change x, so it is a function of x). Come back now in (2.12) and make $p \to \infty$. We find that $|S(x) - S_n(x)| \leq \varepsilon$ for any $n \geq N$ and for any $x \in A$. If here, in the last inequality, we take sup on x, we finally get: $\|S - S_n\| \leq \varepsilon$ for any $n \geq N$. Hence, the series $\sum_{k=0}^{\infty} |f_k|$ is uniformly convergent to S (its sum). Thus, our initial series $\sum_{k=0}^{\infty} f_k$ is uniformly and absolutely convergent. □

The series of functions $\sum_{n=1}^{\infty} \frac{\arctan(nx)}{n^2}$ is absolutely uniformly convergent because $\left|\frac{\arctan(nx)}{n^2}\right| \leq \frac{\pi}{2} \cdot \frac{1}{n^2}$ and the numerical series $\sum_{n=1}^{\infty} \frac{\pi}{2} \cdot \frac{1}{n^2} = \frac{\pi}{2} \sum_{n=1}^{\infty} \frac{1}{n^2}$ is convergent (why?) (see the Weierstrass Test, Theorem 41).

Another very useful test is the Abel-Dirichlet Test for series of functions, a generalization of the test with the same name for numerical series.

THEOREM 42. *(Abel-Dirichlet Test for series of functions)*

Let $\{a_n(x)\}$, $\{b_n(x)\}$ be two sequences of functions defined on the same interval I of $\mathbb{R}$. We assume that $\|a_n\|$ is a decreasing to zero sequence and that the partial sums $s_n(x) = \sum_{k=0}^{n} b_n(x)$ of the series of functions $\sum_{k=o}^{\infty} b_n(x)$ are uniformly bounded, i.e. there is a positive real number $M > 0$ such that $\|s_n\| < M$ for any $n = 1, 2, ...$

Then the series of functions $\sum_{n=0}^{\infty} a_n(x)b_n(x)$ is (absolutely) uniformly convergent on the interval I.

PROOF. Let us come back to the Abel-Dirichlet's Test for numerical series and substitute the numbers a_n, b_n, s_n, S_n with the corresponding functions $a_n(x), b_n(x), s_n(x)$ and $S_n(x) = \sum_{k=0}^{n} a_k(x)b_k(x)$ respectively. We obtain (do it step by step!) that the sequence of functions $\{S_n(x)\}$ is uniformly Cauchy, i.e. for any $\varepsilon > 0$, there is a rank N_ε such that if $n \geq N_\varepsilon$ one has that

$$\|S_{n+p} - S_n\| < \varepsilon \tag{2.13}$$

for any $p = 1, 2, ...$ In particular,

$$|S_{n+p}(x) - S_n(x)| < \varepsilon$$

for any fixed x in I. So, the numerical sequence $\{S_n(x)\}$ is convergent to a number $S(x)$ which depend on x. Making $p \to \infty$ in (2.13) we get

$$|S(x) - S_n(x)| \leq \varepsilon$$

for any $n \geq N_\varepsilon$ and for any x in I. Take now sup on x and find that

$$\|S - S_n\| \leq \varepsilon$$

for any $n \geq N_\varepsilon$. This means that $\{S_n\}$ is uniformly convergent to S, i.e. our series of functions $\sum_{n=0}^{\infty} a_n(x)b_n(x)$ is uniformly convergent on the interval I. With some small changes in the proof, we find that this last series is absolutely uniformly convergent on I (do them!). □

Let us take the series of functions $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$ for $x \in [-1+\varepsilon, 1]$, where $0 < \varepsilon < 2$. Let us apply the Abel-Dirichlet Test for series of functions by taking $a_n(x) = \frac{x^n}{n}$ and $b_n(x) = (-1)^{n-1}$. We easily see that $\|a_n(x)\| = \frac{1}{n}$ and that the series $\sum_{n=1}^{\infty}(-1)^{n-1}$ has bounded partial sums. Hence our series $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n$, $x \in [-1+\varepsilon, 1]$, is absolutely and uniformly convergent.

The following question arises: can we integrate or differentiate term by term (termwise) a series of function $\sum\limits_{k=0}^{\infty} f_k$? Since everything reduces to the sequence of partial sums $s_n = f_0 + f_1 + ... + f_n$, we can apply the results from Theorem 39 and Theorem 40 and find:

THEOREM 43. *Let $\sum\limits_{n=0}^{\infty} f_n$ be a uniformly convergent series of continuous functions on the interval $[a, b]$, let s be its sum and let $F_n(x)$ be the canonical primitives of $f_n(t)$ on $[a, b]$: $F_n(x) = \int_a^x f_n(t)dt$, $n = 0, 1, ...$. Then the series of functions $\sum\limits_{n=0}^{\infty} F_n$ is uniformly convergent on $[a, b]$*

and $S(x)=\int_a^x s(t)dt$, *is its sum. So,*

$$\int_a^x \left(\sum_{n=0}^{\infty} f_n(t)\right) dt = \sum_{n=0}^{\infty} \int_a^x f_n(t)dt. \tag{2.14}$$

(this means that the integration symbol $\int$ *commutes with the symbol* $\sum$ *of a series). In particular, for* $x=b$, *we get a very useful formula:*

$$\int_a^b \left(\sum_{n=0}^{\infty} f_n(t)\right) dt = \sum_{n=0}^{\infty} \int_a^b f_n(t)dt. \tag{2.15}$$

If in addition, f_n *are functions of class* C^1 *on* $[a,b]$ *(*f_n *are differentiable and their derivatives are continuous on* $[a,b]$, *shortly write* $f_n \in C^1[a,b]$*) and if the series of derivatives,* $u=\sum\limits_{n=0}^{\infty} f_n'$ *is uniformly convergent on* $[a,b]$, *then* s *is differentiable on* $[a,b]$ *and* $s'=u$. *So, we can differentiate "term by term" (or termwise) the initial series of functions.*

In the first statement s is a continuous function on $[a,b]$ because of the basic Theorem 38. In this last theorem there is a requirement: f_n must be bounded. This is true because f_k are continuous and defined on a bounded and closed interval (see Theorem 32).

Let us study the following series of functions $\sum\limits_{n=0}^{\infty}(-1)^n x^n$ on $(-1,1)$. For any fixed x, one has the formula

$$1-x+x^2-...=\frac{1}{1+x}, x\in(-1,1), \tag{2.16}$$

the famous geometric series with ratio $-x$. Hence, our series is simply convergent on $(-1,1)$. It is not uniformly convergent on $(-1,1)$ but it is absolutely and uniformly convergent on any closed subinterval $[a,b]$ of $(-1,1)$ (apply the same reason as in the case of the infinite geometrical series). Let us derive an interesting and useful formula from (2.16). Let us fix an x_0 in $(-1,1)$ and take a,b such that $x_0\in[a,b]$, a or b is 0 (if $x_0<0$, take $b=0$, if $x_0\geq 0$, take $a=0$) and $[a,b]$ is included in $(-1,1)$. Since all conditions in Theorem 43 are fulfilled, we integrate term by term formula (2.16) and get

$$\int_0^{x_0}(1-t+t^2-...+(-1)^n t^n+...)dt=$$

$$=(t-\frac{t^2}{2}+\frac{t^3}{3}-...+(-1)^n\frac{t^{n+1}}{n+1}+...)\mid_0^{x_0}=$$

$$= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x_0^n}{n} = \int_0^{x_0} \frac{1}{1+t} dt = \ln(1+x_0).$$

Now, let us put instead of x_0 an arbitrary x in $(-1,1)$ and obtain

$$\ln(1+x) = \sum_{n=1}^{\infty} (-1)^{n-1} \frac{x^n}{n}, \text{ for any } x \in (-1,1). \tag{2.17}$$

The value of the alternate series $\sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n}$ is $\ln 2$ but, to prove this, one needs the continuity of the function on the right in the formula 2.17. And this is not so easy to be proved (see the Abel Theorem, Theorem 46).

Let us compute the sum of the series of functions $\sum_{n=0}^{\infty} nx^n$ on its maximal domain of definition. First of all, let us fix an x on the real line and try to find conditions for the convergence of the series $\sum_{n=0}^{\infty} nx^n$. Let us see where the series (numerical series this time!) is absolutely convergent. Applying the Ratio Test (Theorem 27) to the series of moduli $\sum_{n=0}^{\infty} n |x|^n$, we get $\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = |x|$. We know that if $|x| < 1$, the series is absolutely convergent, in particular it is convergent on $(-1,1)$. If $|x| > 1$, the series is divergent, because, in this case, the sequence $\{nx^n\}$ is not bounded (why?) so, it cannot be convergent to 0. For $x = 1$ or $x = -1$, the series is divergent. Hence, the definition domain of the function $s(x) = \sum_{n=0}^{\infty} nx^n$ is exactly $(-1,1)$. Let us compute $s(x)$.

$$s(x) = 1x + 2x^2 + 3x^3 + ... + nx^n + ... = x(1 + 2x + 3x^2 + ... + nx^{n-1} + ...)$$

$$= x(x + x^2 + ... + x^n + ...)' = x \cdot \left(\frac{x}{1-x}\right)' = \frac{x}{(1-x)^2}.$$

Here we used Theorem 43 to differentiate term by term the series $x + x^2 + ... + x^n + ... = \frac{x}{1-x}$ (why the hypotheses of this theorem are fulfilled?).

3. Problems

1. Find the convergence set and the limit for the following sequences of functions: a) $f_n(x) = x^n$; b) $f_n(x) = \frac{x}{n}$; c) $f_n(x) = \frac{n}{x+n}$, $x \in (0, \infty)$; d) $f_n(x) = \frac{nx}{1+n+x}$, $x \in [0,1]$; e) $f_n(x) = \frac{2nx}{1+n^2x^2}$, $x \in [1, \infty)$; f) $f_n(x) = \frac{x^2}{x^4+n^2}$, $x \in [1, \infty)$.

2. Say if the convergence of the above sequences (see Problem 1.) is uniform or not. Study the absolute uniform convergence of the same sequences.

3. Let $f_n(x) = \frac{nx}{1+n^2x^2}$, $x \in [0,1]$. Prove that $\{f_n\}$ is not uniformly convergent but $\int_0^1 f_n(x)dx \to \int_0^1 \lim\limits_{n\to\infty} f_n(x)dx$.

4. Prove that $f_n(x) = \frac{x}{1+n^2x^2}$, $x \in [-1,1]$ is uniformly convergent to $f(x)$ (find it!) but f_n' is not uniformly convergent to f'. Do the same for $f_n(x) = \frac{x^n}{n}$, $x \in [0,1]$.

5. Prove that the series of functions $\sum_{n=1}^{\infty}(x^n - x^{n-1})$ is uniformly convergent on $[0, 0.5]$, but not on $[0,1]$.

6. Is the series of functions $\sum_{n=1}^{\infty}\left(\sin\frac{x}{n+1} - \sin\frac{x}{n}\right)$ uniformly convergent on $\mathbb{R}$? But on $[0,1]$? But on $[a,b]$?

7. Prove that the following series of functions are absolutely and uniformly convergent on the indicated domain: a) $\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{x^2+n\sqrt{n}}, x \in \mathbb{R}$; b) $\sum_{n=1}^{\infty}\frac{(-1)^n 3^{-nx}}{x+2^n}, x \in [0,\infty)$; c) $\sum_{n=1}^{\infty}\frac{\sin nx}{n\sqrt{n}}, x \in \mathbb{R}$; d) $\sum_{n=1}^{\infty}\frac{1}{n^2+x^2}, x \in \mathbb{R}$; e) $\sum_{n=1}^{\infty}\frac{\sin nx}{\sqrt{x^2+n^4}}, x \in \mathbb{R}$.

8. Can we differentiate term by term the following series?

a) $\sum_{n=1}^{\infty}\exp(-nx)\sin nx$, $x \in [1,\infty)$; b) $\sum_{n=1}^{\infty}\frac{\sin(2^{\sqrt{n}}x)}{n^2 2^{\sqrt{n}}}, x \in \mathbb{R}$;

c) $\sum_{n=1}^{\infty}\frac{1}{n^2+x^2}$, $x \in \mathbb{R}$.

9. Find the image of the following functions:

a) $f(x) = -3x + 2$, $x \in [-3, 12]$;

b) $f(x) = 2x^2 + x - 5$, $x \in \mathbb{R}$;

c) $f(x) = x^3 - 3x + 2$, $x \in [-120, 120]$;

d) $f(x) = 3\sin 4x$, $x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$;

e) $f(x) = |\sin x - \cos 2x|$, $x \in [0, \pi]$;

f) $f(x) = |x^2 + 2x - 1| - 3$, $x \in (-\infty, 9]$.

10. Find the norm of the following functions: a) $f(x) = 2x - 5$, $x \in [-4, 7]$; b) $f(x) = 3\cos 5x$, $x \in [\pi, \infty)$; c) $f(x) = \ln(2x^2 + 3)$, $x \in [-2, 2]$; d) $f - g$, where $f(x) = 3x$ and $g(x) = 4x^2$, $x \in [0, 2]$.

CHAPTER 4

Taylor series

1. Taylor formula

Always the most elementary functions were considered to be polynomial functions. A polynomial function of degree n is a function defined on the whole real line by the formula:

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n,$$

where $a_0, a_1, ..., a_n$ are fixed real numbers and $a_n \neq 0$.

Many mathematicians tried and are trying to reduce the study of more complicated functions to polynomials.

It is clear enough that not all functions can be represented by a polynomial. For instance, the exponential function $f(x) = \exp(x) = e^x$ cannot be represented by a polynomial $P_n(x)$. Indeed, if

$$\exp(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

for $x \in (a, b)$, $a \neq b$, we differentiate n times and find: $\exp(x) = n! a_n$, a constant, which is not possible, because the exponential function is strictly increasing. Here we proved in fact that the exponential function cannot be represented by a polynomial in any small neighborhood of any point on the real line. The following problem appears in many applications. If x is very close to a fixed number a, i.e. if the difference $x - a$ is very small (is very close to zero!), can we represent a function f as an "infinite" polynomial in the variable $x - a$? This means

$$f(x) = a_0 + a_1 (x - a) + a_2 (x - a)^2 + ... \tag{1.1}$$

in a neighborhood $(a-\varepsilon, a+\varepsilon)$ of a. This would imply that our function is a function of class C^∞ , i.e. it has derivatives of any order. But this is not true for all functions. So, what can we hope is to "approximate" a function f in a small neighborhood of a point a with a polynomial of a given degree n in the variable $x - a$:

$$f(x) = a_0 + a_1 (x - a) + a_2 (x - a)^2 + ... + a_n (x - a)^n + R_n(x), \tag{1.2}$$

where $R_n(x)$ is a remainder which is a function of x (it also depends on f and on a!). This remainder is the error committed when we

approximate $f(x)$ by the polynomial

$$a_0 + a_1 (x-a) + a_2 (x-a)^2 + ... + a_n (x-a)^n.$$

This polynomial is called the *Taylor polynomial of order n at a.*

If $f(x)$ is a polynomial of degree n, we can represent f as in formula (1.2) with the remainder zero. Indeed, the set of $n+1$ binomials

$$\{1, x-a, (x-a)^2, (x-a)^3, ..., (x-a)^n\}$$

is linear independent in the vector space $\mathcal{P}_n$ of all polynomials of degree at most n, which has dimension $n+1$ over the real field (this comes directly from the definition of a polynomial-why?). Hence,

$$\{1, x-a, (x-a)^2, (x-a)^3, ..., (x-a)^n\}$$

is a basis in $\mathcal{P}_n$ and so, we always can uniquely find the constant elements a_0, a_1, $a_2, ..., a_n$ such that

$$f(x) = a_0 + a_1 (x-a) + a_2 (x-a)^2 + ... + a_n (x-a)^n. \tag{1.3}$$

In this last case we can compute the coefficients a_0, $a_1, ..., a_n$ by using the values of f and of its derivatives f', $f'', ...,$ $f^{(n)}$ at a. Indeed, let us make $x = a$ in the equality (1.3). We get $f(a) = a_0$. If one differentiates the same equality and makes $x = a$, one obtains $f'(a) = a_1$. Now, if we differentiate twice this equality (1.3), we get $f''(a) = 2a_2$, and so on. Take the k-th derivative in both sides in (1.3) and find $f^{(k)}(a) = k! a_k$ for any $k = 1, 2, ..., n$. Thus (1.3) becomes:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{1.4}$$

Generally, if the function f is not a polynomial of degree n, we formally can write (it is clear that f must be n-times differentiable):

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x), \tag{1.5}$$

where

$$R_n(x) = f(x) - f(a) - \frac{f'(a)}{1!}(x-a) - \frac{f''(a)}{2!}(x-a)^2 - ... - \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

The problem is to estimate this remainder. The famous Taylor formula gives a general estimation for this remainder.

THEOREM 44. *(Taylor formula) Let A be an open subset of $\mathbb{R}$ and let $f : A \to \mathbb{R}$ be a function defined on A with values in $\mathbb{R}$, which is $(n+1)$-times differentiable on A. Let us fix a point a in A and a natural number $p \neq 0$. Then, for any $x \in A$ such that the segment $[a, x]$*

is included in A, there is a point $c \in (a,x)$ with the following property: the remainder $R_n(x)$ from (1.5) has a representation of the form

$$R_n(x) = \left(\frac{x-a}{x-c}\right)^p \frac{(x-c)^{n+1}}{n!p} f^{(n+1)}(c) \tag{1.6}$$

This general form of the remainder was discovered by Schömlich. If $p = n+1$, we find the Lagrange form of the remainder

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \tag{1.7}$$

We see that this form is very similar to the general term form in (1.5). In fact, it is "the next" term after the n-th term $\frac{f^{(n)}(a)}{n!}(x-a)^n$ in which the value of $f^{(n+1)}$ is not computed at a, but at a close point $c \in [a,x]$ (here we do not mean that a is less then x!). Usually, the error made by approximating $f(x)$ with its Taylor polynomial $T_n(x)$ of order n,

$$T_n(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n, \tag{1.8}$$

is evaluated by the Lagrange form of the remainder $R_n(x)$. Since we have no supplementary information on the number c, we use the following upper bounded formula:

$$|R_n(x)| \leq \frac{|x-a|^{n+1}}{(n+1)!} \sup\{\left|f^{(n+1)}(z)\right| : z \in [a,x]\} \tag{1.9}$$

Since we frequently use Taylor formula with Lagrange remainder, we write it here in a complete form (together with this last form of the reminder)

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \tag{1.10}$$

PROOF. The proof of this theorem is not so natural. Let us assume that $x > a$. In this case, the segment $[a,x]$ is exactly the closed interval $[a,x]$. Let us denote in (1.5)

$$Q(x) = \frac{R_n(x)}{(x-a)^p}. \tag{1.11}$$

Thus, the formula (1.5) becomes:
(1.12)

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + (x-a)^p Q(x).$$

In order to obtain a representation for $Q(x)$, we consider an auxiliary function:
(1.13)

$$g(t) = f(t) + \frac{f'(t)}{1!}(x-t) + \frac{f''(t)}{2!}(x-t)^2 + ... + \frac{f^{(n)}(t)}{n!}(x-t)^n + (x-t)^p Q(x)$$

We obtained the expression of $g(t)$ by simply putting t instead of a, in (1.12). We apply now the Rolle's Theorem (Theorem 36) on the interval $[a, x]$. The function $g(t)$ is continuous and differentiable on $[a, x]$, $g(a) = f(x)$ (see 1.12) and $g(x) = f(x)$ so, $g(a) = g(x)$. Thus, there is a point $c \in (a, x)$ such that $g'(c) = 0$. Let us compute $g'(t)$:

$$g'(t) = f'(t) + \frac{f''(t)}{1!}(x-t) - \frac{f'(t)}{1!} + \frac{f'''(t)}{2!}(x-t)^2 - \frac{f''(t)}{1!}(x-t) + ...$$
$$+ \frac{f^{(n+1)}(t)}{n!}(x-t)^n - \frac{f^{(n)}(t)}{(n-1)!}(x-t)^{n-1} - p(x-t)^{p-1}Q(x),$$

So we get

$$g'(t) = \frac{f^{(n+1)}(t)}{n!}(x-t)^n - p(x-t)^{p-1}Q(x). \tag{1.14}$$

Make now $t = c$ in (1.14) and find

$$0 = g'(c) = \frac{f^{(n+1)}(c)}{n!}(x-c)^n - p(x-c)^{p-1}Q(x).$$

If here, instead of $Q(x)$ we put $\frac{R_n(x)}{(x-a)^p}$ (see (1.11)), we get

$$\frac{f^{(n+1)}(c)}{n!}(x-c)^n = p(x-c)^{p-1}\frac{R_n(x)}{(x-a)^p},$$

or

$$R_n(x) = \frac{(x-a)^p}{(x-c)^{p-1}}\frac{f^{(n+1)}(c)}{n!p}(x-c)^n = \frac{(x-a)^p}{(x-c)^p}\frac{f^{(n+1)}(c)}{n!p}(x-c)^{n+1},$$

i.e. formula (1.6). The other statements of the theorem are easily deduced from this last formula. □

REMARK 14. *A function $f(x)$ is a zero of another function $g(x)$ at a point a if $\lim_{x\to a}\frac{f(x)}{g(x)} = 0$. We write this as $f(x) = 0(g(x))$ at a.*

For instance, from (1.7) we see that the remainder $R_n(x)$ *is a zero of* $(x-a)^n$ *at* $x=a$, *i.e.* $R_n(x)=0((x-a)^n)$ *at* $x=a$.

If $a=0$, the formula (1.5) is called the *Mac Laurin formula*:

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+...+\frac{f^{(n)}(0)}{n!}x^n+R_n(x) \tag{1.15}$$

If we use the Lagrange form of the remainder (1.7), we get

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+...+\frac{f^{(n)}(0)}{n!}x^n+\frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1}, \tag{1.16}$$

where c is a real number between 0 and x. Since it is easier to manipulate Mac Laurin formulas for many functions which are defined on an interval (a,b) with $0\in(a,b)$ and since the translation $x\to x-a$ makes connections between Taylor formulas and Mac Laurin formulas, we prefer to deduce these last formulas for the basic elementary functions.

EXAMPLE 2. *(*$\exp(x)$*) Let* $f(x)=\exp(x)=e^x, x\in\mathbb{R}$. *Since the derivatives of* $\exp(x)$ *is* $\exp(x)$ *itself, the Taylor formula at* $a=0$ *(Mac Laurin formula) for* $\exp(x)$ *becomes*

$$\exp(x)=1+\frac{x}{1!}+\frac{x^2}{2!}+...+\frac{x^n}{n!}+\exp(c)\frac{x^{n+1}}{(n+1)!}, \tag{1.17}$$

where $c\in(0,x)$, *if* $x>0$, *or* $c\in(x,0)$, *if* $x<0$.

For instance, let us compute $\exp(0.03)$ with 2 exact decimals. Since $c\in(0,0.03)$, this means that

$$|R_n(0.03)|=\left|\exp(c)\frac{(0.03)^{n+1}}{(n+1)!}\right|<3\cdot\frac{(0.03)^{n+1}}{(n+1)!}<\frac{1}{100},$$

or

$$\frac{3^{n+2}}{100^{n+1}(n+1)!}<\frac{1}{100}\Leftrightarrow 3^{n+2}<100^n(n+1)!.$$

It is easy to prove this last inequality by mathematical induction for $n\geq 1$. So, $\exp(x)\cong 1+\frac{0.03}{1!}=1.03$, with 2 exact decimals. This is the method which computers use to (approximately) calculate $\exp(r)$ for a given real number r. Formula (1.17) can also be written as

$$\exp(x)=1+\frac{x}{1!}+\frac{x^2}{2!}+...+\frac{x^n}{n!}+0(x^n) \tag{1.18}$$

We can use this formula to compute nondeterministic limits. For instance, let us compute

$$\lim_{x\to 0}\frac{\exp(x^3)-1-x^3-\frac{x^6}{2}}{\exp(x^2)-1-x^2-\frac{x^4}{2}}=\frac{0}{0}.$$

In formula (1.18) we put instead of x, x^3 and $n=2$:

$$\exp(x^3)=1+x^3+\frac{x^6}{2}+0(x^6).$$

If we put now in (1.18) instead of x, x^2 and $n=3$, we get

$$\exp(x^2)=1+x^2+\frac{x^4}{2}+\frac{x^6}{6}+0(x^6).$$

Hence, our limit becomes

$$\lim_{x\to 0}\frac{0(x^6)}{\frac{x^6}{6}+0(x^6)}=\lim_{x\to 0}\frac{\frac{0(x^6)}{x^6}}{\frac{1}{6}+\frac{0(x^6)}{x^6}}=\frac{\lim\limits_{x\to 0}\frac{0(x^6)}{x^6}}{\frac{1}{6}+\lim\limits_{x\to 0}\frac{0(x^6)}{x^6}}=\frac{0}{\frac{1}{6}+0}=0.$$

In practice, we do not know in advance how many terms we must consider in numerator and in denominator such that the nondeterministic to be eliminated. So, it is a good idea to consider one or two terms more than the degree of the polynomial queue which induces the nondeterministic. In our example we write

$$\lim_{x\to 0}\frac{\exp(x^3)-1-x^3-\frac{x^6}{2}}{\exp(x^2)-1-x^2-\frac{x^4}{2}}=$$

$$=\lim_{x\to 0}\frac{(1+\frac{x^3}{1!}+\frac{x^6}{2!}+\frac{x^9}{3!}+...)-1-x^3-\frac{x^6}{2}}{(1+\frac{x^2}{1!}+\frac{x^4}{2!}+\frac{x^6}{3!}+\frac{x^8}{4!}+...)-1-x^2-\frac{x^4}{2}}=$$

$$=\lim_{x\to 0}\frac{\frac{x^9}{3!}+...}{\frac{x^6}{3!}+\frac{x^8}{4!}+...}=\lim_{x\to 0}\frac{\frac{x^3}{3!}+...}{\frac{1}{3!}+\frac{x^2}{4!}+...}=\frac{0}{\frac{1}{3!}}=0.$$

EXAMPLE 3. *(*$\sin(x)$*) Let* $f(x)=\sin(x)$, $x\in\mathbb{R}$. *Since* $[\sin(x)]'=\cos(x)$, $[\sin(x)]''=-\sin(x)$, $[\sin(x)]'''=-\cos(x)$ *and* $[\sin(x)]^{(4)}=\sin(x)$, *we obtain that* $[\sin(x)]^{(4k+1)}=\cos(x)$, $[\sin(x)]^{(4k+2)}=-\sin(x)$, $[\sin(x)]^{(4k+3)}=-\cos(x)$ *and* $[\sin(x)]^{(4k)}=\sin(x)$ *for any* $k=0,1,...$. *Now,* $\sin 0=0$, $\cos 0=1$ *and, applying formula (1.16), we get*

$$\sin(x)=\frac{x}{1!}-\frac{x^3}{3!}+\frac{x^5}{5!}-...+(-1)^n\frac{x^{2n+1}}{(2n+1)!}+0(x^{2n+1}). \tag{1.19}$$

It is more complicated to express the remainder in this case because the $(n+1)$*-derivative of* $\sin(x)$ *is either* $\pm\sin(x)$ *or* $\pm\cos(x)$. *Let us use the Mac Laurin formula for* $\sin(x)$ *in order to compute* $\sin(0.2)$ *with*

one exact decimal. Here 0.2 *means* 0.2 *radians. Now, the modulus of the remainder,* $|R_{2n+1}(x)|$ *is less or equal to* $\frac{1}{(2n+2)!}|x|^{2n+2}$. *So,*

$$|R_{2n+1}(0.2)| \leq \frac{1}{(2n+2)!}(0.2)^{2n+2},$$

and this last one must be less then $\frac{1}{10}$, *i.e.*

$$\frac{1}{(2n+2)!}2^{2n+2} < 10^{2n+1}$$

or

$$2^{2n+2} < (2n+2)!10^{2n+1}.$$

But this last one is true for any $n \geq 0$. *Hence,* $\sin(0.2) \simeq 0.2$ *with one exact decimal.*

EXAMPLE 4. *(*$\cos(x)$*) Let* $f(x) = \cos(x)$, $x \in \mathbb{R}$. *Like in Example 3 we easily deduce the following formula*

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + ... + (-1)^n\frac{x^{2n}}{(2n)!} + 0(x^{2n}). \tag{1.20}$$

EXAMPLE 5. *Let*

$$f(x) = \ln(1+x), x \in (-1, \infty).$$

Since

$$f'(x) = (1+x)^{-1}, f''(x) = -(1+x)^{-2}, f'''(x) = 2(1+x)^{-3}, ...$$

$$..., f^{(n)}(x) = (-1)^{n-1}(n-1)!(1+x)^{-n}, ...,$$

one has that $f(0) = 0$, $f'(0) = 1$, $f''(0) = -1$, $f'''(0) = 2, ...,$ $f^{(n)}(0) = (-1)^{n-1}(n-1)!, ...$. *So, the formula (1.16) becomes*

(1.21)

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... + (-1)^{n-1}\frac{x^n}{n} + (-1)^n\frac{(1+c)^{-n-1}}{n+1}x^{n+1},$$

where c *is a real number between* 0 *and* x. *Hence,*

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... + (-1)^{n-1}\frac{x^n}{n} + 0(x^n). \tag{1.22}$$

Let us compute $\ln(1.02)$ *with* 3 *exact decimals. Since*

$$\ln(1.02) = \ln(1+0.02) = 0.02 - \frac{(0.02)^2}{2} + \frac{(0.02)^3}{3} + ...$$

$$+(-1)^{n-1}\frac{(0.02)^n}{n} + (-1)^n\frac{(1+c)^{-n-1}}{n+1}(0.02)^{n+1},$$

where c is between 0 and 0.02, we must evaluate the modulus of the remainder and force this last upper bound to be less then $\frac{1}{1000}$,

$$\left|(-1)^n\frac{(1+c)^{-n-1}}{n+1}0.02^{n+1}\right| < \frac{2^{n+1}}{(n+1)100^{n+1}} < \frac{1}{1000}.$$

This last inequality is true for any $n \geq 1$. Thus, $\ln(1.02) \simeq 0.020$ with 3 exact decimals. Pay attention! It is not sure that 020 are the first three decimals of $\ln(1.02)$! What is sure is that $|\ln(1.02) - 0.02|$ is less then $0.001 = \frac{1}{1000}$ (this means "with 3 exact decimals!").

EXAMPLE 6. *(Binomial formula) Let $f(x) = (1+x)^\alpha$, where α is a fixed real number and $x > -1$. Since*

$$f'(x) = \alpha(1+x)^{\alpha-1}, f''(x) = \alpha(\alpha-1)(1+x)^{\alpha-2}, ...$$
$$..., f^{(n)}(x) = \alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)(1+x)^{\alpha-n}, ...,$$

one has that

$$f(0) = 1, f'(0) = \alpha, f''(0) = \alpha(\alpha-1), ...$$
$$..., f^{(n)}(0) = \alpha(\alpha-1)(\alpha-2)...(\alpha-n+1), ...$$

Now, formula (1.16) becomes

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + ...$$
$$... + \frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!}x^n +$$
$$+\frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n)(1+c)^{\alpha-n-1}}{(n+1)!}x^{n+1}, \tag{1.23}$$

where c is a real number between 0 and x.

Formula (1.23) can also be written as

$$(1+x)^\alpha = 1 + \frac{\alpha}{1!}x + \frac{\alpha(\alpha-1)}{2!}x^2 + ...+ \tag{1.24}$$
$$+\frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!}x^n + 0(x^n)$$

Let us use this formula to approximate the following expression $E = E(q) = \frac{1}{\sqrt{a+bq^2}}$, $a, b > 0$, by a polynomial of degree 2 (it is used in Physics for q small). In order to apply (1.23) we need to put our expression in the form $(1+x)^\alpha$. So,

$$E = (a+bq^2)^{-\frac{1}{2}} = a^{-\frac{1}{2}}(1+\frac{b}{a}q^2)^{-\frac{1}{2}}.$$

Let us take only $(1+\frac{b}{a}q^2)^{-\frac{1}{2}}$ and use (1.23) up to x^2, where $x=\frac{b}{a}q^2$ and $\alpha=-\frac{1}{2}$. We get

$$(1+\frac{b}{a}q^2)^{-\frac{1}{2}} \approx 1+(-\frac{1}{2})\frac{b}{a}q^2+\frac{(-\frac{1}{2})(-\frac{3}{2})}{2}\frac{b^2}{a^2}q^4,$$

Hence,

$$\frac{1}{\sqrt{a+bq^2}} \approx \frac{1}{\sqrt{a}}-\frac{b}{2a\sqrt{a}}q^2+\frac{3b^2}{8a^2\sqrt{a}}q^4.$$

If $\alpha=n$, a natural number, we obtain the famous binomial formula of Newton:

$$(1.25)\ \ (1+x)^n=1+\frac{n}{1!}x+\frac{n(n-1)}{2!}x^2+...+\frac{n(n-1)(n-2)...1}{n!}x^n,$$

because the remainder in (1.23) is zero. If instead of x we put $\frac{b}{a}$ in (1.25) we get

$$\frac{(a+b)^n}{a^n}=1+\binom{n}{1}\frac{b}{a}+\binom{n}{2}\frac{b^2}{a^2}+\binom{n}{3}\frac{b^3}{a^3}+...+\binom{n}{n}\frac{b^n}{a^n}.$$

Multiplying by a^n, we get:

(1.26)

$$(a+b)^n=a^n+\binom{n}{1}a^{n-1}b+\binom{n}{2}a^{n-2}b^2+\binom{n}{3}a^{n-3}b^3+...+\binom{n}{n}b^n.$$

Here, $\binom{n}{k}=\frac{n(n-1)(n-2)...(n-k+1)}{k!}=\frac{n!}{k!(n-k)!}$ means n objects taken k.

EXAMPLE 7. *The equilibrium position of a homogeneous weighted string, fixed at the ends, has a form given by the plane curve* $y=a\cdot ch(\frac{x}{b})$, *where* $ch(x)=\frac{\exp(x)+\exp(-x)}{2}$ *and* a,b *are real numbers. The function* $f(x)=ch(x)$ *is called the hyperbolic cosine of* x.

The derivative of the function $ch(x)$ *is* $sh(x)=\frac{\exp(x)-\exp(-x)}{2}$, *called the hyperbolic sine of* x. *Since the derivative of each of them is the other one, we easily get the formulas*

$$(1.27)\qquad sh(x)=\frac{x}{1!}+\frac{x^3}{3!}+\frac{x^5}{5!}+...+\frac{x^{2n+1}}{(2n+1)!}+0(x^{2n+1}),$$

$$(1.28)\qquad ch(x)=1+\frac{x^2}{2!}+\frac{x^4}{4!}+\frac{x^6}{6!}+...+\frac{x^{2n}}{(2n)!}+0(x^{2n}).$$

For instance, for x *small enough, we can approximate* $ch(x)$ *by the polynomial* $T_4(x)=1+\frac{x^2}{2!}+\frac{x^4}{4!}$. *For* $x=0.5$, $ch(0.5)\approx 1+\frac{0.25}{2}+\frac{0.0025}{24}$.

Taylor's and Mac Laurin's formulas have many applications in the local study of a function (or a curve).

COROLLARY 5. *(Lagrange formula) Let us write Taylor formula (1.10) for* $n = 0 : f(x) = f(a) + f'(c) \cdot (x - a)$, *where* c *is a number between* a *and* x. *If* $x = b > a$, *we get the classical Lagrange formula:* $f(b) = f(a) + f'(c) \cdot (b - a)$, *where* $c \in (a, b)$.

REMARK 15. *We can use Taylor formula (1.10) for study the shape of a function in a neighborhood of a point* a. *Suppose that*

$$f'(a) = f''(a) = ... = f^{(n-1)}(a) = 0$$

and $f^{(n)}(a) \neq 0$. *We also assume that* f *is of class* C^n *on an* ε*-neighborhood* $(a - \varepsilon, a + \varepsilon)$ *of* a. *Then*

$$f(x) - f(a) = \frac{f^{(n)}(c)}{n!}(x - a)^n, \tag{1.29}$$

where c *is between* a *and* x. *It is clear that the continuity of* $f^{(n)}(x)$ *at* a *implies that the sign of this last function on maybe a smaller subinterval* $(a - \delta, a + \delta)$ *of* $(a - \varepsilon, a + \varepsilon)$ *is constant and it is the same like the sign of* $f^{(n)}(a)$ *(see Theorem 34). Suppose that* $f^{(n)}(x) > 0$ *for any* $x \in (a - \delta, a + \delta)$. *Then, in (1.29),* $c \in (a - \delta, a + \delta)$ *and so, the sign of the difference* $f(x) - f(a)$ *depends exclusively on* n *and on the sign of* $f^{(n)}(a)$. *If* n *is even, and* $f^{(n)}(a) > 0$, *the difference* $f(x) - f(a)$ *is* > 0, *for any* $x \in (a - \delta, a + \delta)$, *thus* a *is a local minimum point for* f. *If* n *is even, but* $f^{(n)}(a) < 0$, *then the difference* $f(x) - f(a)$ *is* < 0, *for any* $x \in (a - \delta, a + \delta)$, *so* a *is a local maximum point for* f. *If* n *is odd, the point* a *is not an extremum point because the sign of* $(x - a)^n$ *changes (it is positive if* $x > a$ *and negative otherwise). For instance,* $f(x) = (x - 2)^5$ *has not an extremum at* $x = 2$.

Let A be an open subset of $\mathbb{R}$ and let $f : A \to \mathbb{R}$ be a function of class C^1 on A. This means that f is differentiable on A and its derivative f' is continuous on A. One also says that f is *smooth* on A. We say that f is *convex* at the point a of A if the graphic of f is above the tangent line of this graphic at a, on a small open ε-neighborhood U of a which is contained in A. If here we substitute the word "above" with the word "under", we get the definition of a *concave function* f at a point a. Since the equation of the tangent line of the graphic of the function f at a is:

$$Y = f(a) + f'(a)(X - a),$$

f is a convex function at a if and only if

$$f(x) \geq f(a) + f'(a)(x - a), \tag{1.30}$$

for any x in $U = (a - \varepsilon, a + \varepsilon) \subset A$.

COROLLARY 6. *Let the above f be a function of class C^2 on $U = (a-\varepsilon, a+\varepsilon)$. We assume that $f''(a) \neq 0$. Then f is convex at a if and only if $f''(a) > 0$.*

PROOF. Let x be a point in U and let us write the Taylor formula (1.10) for $n = 1$ at a on the segment $[a, x]$:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(c_x)}{2!}(x-a)^2, \tag{1.31}$$

where $c_x \in [a, x]$. If f is convex at a, then there is a small interval $U' = (a-\varepsilon', a+\varepsilon') \subset U$ such that (1.30) works on U'. Hence, for any x in U' one has that $f''(c_x) \geq 0$ in (1.31). Since f'' is continuous on U (see the fact that f is of class C^2 on U!) and since $c_x \to a$ whenever $x \to a$, one fas that $f''(a) \geq 0$. But we just assumed that $f''(a) \neq 0$, so $f''(a) > 0$. Conversely, if $f''(a) > 0$, then $f''(x) > 0$ on a whole neighborhood $U'' = (a-\varepsilon'', a+\varepsilon'') \subset U$. Thus $f''(c_x) > 0$ in (1.31) for any x in U''. So, (1.30) works on this U''. Therefore f is convex at a. □

We leave the reader to state and to prove a similar result for a concave function f at a.

2. Taylor series

Let us consider a function f of class C^∞ on an open subset A of $\mathbb{R}$. This means that f has derivatives of any arbitrary order on A. It is clear that all of these derivatives are continuous on A. Look at the formula (1.10) and push the remainder to ∞. We obtain the series of functions on the right side:

$$f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + ... \tag{2.1}$$

$$= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

This series of functions is called the *Taylor series associated to the function f at the point a*. If this series of functions is uniformly convergent and its sum is $f(x)$, we say that

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \tag{2.2}$$

is the *Taylor's expansion of f around the point a*. If the series on the right side is simple convergent and its sum is f on an ε-neighborhood

of a, we say that f *is analytic at* a. If f is analytic at any point of A we say that f is analytic on A. The series on the right in (2.2) is a particular case of a more general type of series of functions, namely, the power series. A power series is a series of functions of the form $\sum_{n=0}^{\infty} a_n(x-a)^n$, where $\{a_n\}$ is a sequence of real numbers and a is a fixed arbitrary number.

THEOREM 45. *Let $f : (c,d) \to \mathbb{R}$ be an indefinite differentiable function on an interval (c,d) ($f \in C^\infty(c,d)$) such that there is a positive real number M which verifies $\left|f^{(n)}(x)\right| \leq M$ for any $x \in (c,d)$ and for any $n = 0, 1, ...$ (we say that all the derivatives of f are uniformly bounded on (c,d)). Then the series $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$ is absolutely and uniformly convergent on (c,d) for any fixed a in (c,d). Moreover,*

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

for any fixed a in (c,d). The series on the right is absolutely uniformly convergent to f.

PROOF. Let us denote $L = d - c$, the length of the interval (c,d). We apply the Weierstrass Test (Theorem 41):

$$\left|\frac{f^{(n)}(a)}{n!}(x-a)^n\right| \leq \frac{M}{n!}L^n \text{ for any } x \in (c,d),$$

and the numerical series $\sum_{n=0}^{\infty} \frac{M}{n!}L^n$ is convergent (use the Ratio Test: $\frac{a_{n+1}}{a_n} = \frac{L}{n+1} \to 0 < 1$). Hence, the series $\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$ is absolutely and uniformly convergent. Let

$$s_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

Formula (1.10) gives us:

$$|f(x) - s_n(x)| = \left|\frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}\right| \leq \frac{M}{(n+1)!}L^{n+1}.$$

Taking sup we obtain $\|f - s_n\| \leq \frac{M}{(n+1)!}L^{n+1}$ and, since $\frac{M}{(n+1)!}L^{n+1} \to 0$ as $n \to \infty$ (prove it by using a numerical series!), we get that $\{s_n\}$ is uniformly convergent to f. In particular

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

□

EXAMPLE 8. *(Taylor series for the basic elementary functions)*
a) We know that

$$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ... + \frac{x^n}{n!} + \exp(c)\frac{x^{n+1}}{(n+1)!}.$$

Since all the derivatives of $\exp(x)$ *are uniformly bounded on any bounded interval* (a, b) *(why?) we can apply Theorem 45 and find that the series* $\sum_{n=0}^{\infty} \frac{1}{n!}x^n$ *is absolutely and uniformly convergent on any bounded interval* (a, b)*. In particular, we have the Taylor expansion*

$$\exp(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ... + \frac{x^n}{n!} + ... = \sum_{n=0}^{\infty} \frac{1}{n!}x^n, x \in \mathbb{R} \tag{2.3}$$

b) We leave the reader to deduce the following Taylor expansions:

$$\sin(x) = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - ... + (-1)^n\frac{x^{2n+1}}{(2n+1)!} + ... \tag{2.4}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!}x^{2n+1}, x \in \mathbb{R}$$

$$\cos(x) = 1 + \frac{x^2}{2!} - \frac{x^4}{4!} + \frac{x^6}{6!} - ... + (-1)^n\frac{x^{2n}}{(2n)!} + ... \tag{2.5}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!}x^{2n}, x \in \mathbb{R}$$

Since all the derivatives of $\sin x$ *and* $\cos x$ *are uniformly (independent of* x*) bounded (by* 1*) on* $\mathbb{R}$*, the series on the right side in the last two formulas are absolutely and uniformly convergent on any bounded interval of* $\mathbb{R}$ *(why not on the whole* $\mathbb{R}$*?).*
c)

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + ... + (-1)^{n-1}\frac{x^n}{n} + ... \tag{2.6}$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}x^n, x \in (-1, 1).$$

Since the n*-th derivative of* $f(x) = \ln(1 + x)$ *is*

$$f^{(n)}(x) = (-1)^{n-1}(n-1)!(1+x)^{-n}$$

it is not uniformly bounded on the whole interval $(-1,1)$ *(why? ... because* $\sup(1+x)^{-n}=\infty$ *there!). Even on any other small subinterval* $[a,b]$ *of* $(-1,1)$ *the derivatives of* $\ln(1+x)$ *are not uniformly bounded (because of* n*, this time!). Hence, we cannot apply the above Theorem 45. Let us look directly to the absolute value of the remainder in (1.21) when* $x\in(-1,1)$ *:*

$$\left|(-1)^n\frac{(1+c)^{-n-1}}{n+1}x^{n+1}\right|,$$

where c *belongs to the segment* $[0,x]^{\pm}$*, i.e.* $c\in[0,x]$*, or* $[x,0]$ *(for* $x<0$*). It is clear that if* $x\to-1$*,* c *may become closer and closer to* -1 *and the remainder cannot uniformly go to* 0*. But, if we take any subinterval* $[a,b]$ *of* $(-1,1)$*, then*

$$\sup_{x\in[a,b]}\left|(-1)^n\frac{(1+c)^{-n-1}}{n+1}x^{n+1}\right|\leq\frac{1}{n+1}\cdot\frac{M^{n+1}}{(1+m)^{n+1}},$$

where $M=\max\{|a|,|b|\}$ *and* $m=\min\{|a|,|b|\}$*. Thus, in this last case,*

$$\|\ln(1+x)-s_n\|=\sup_{x\in[a,b]}\left|(-1)^n\frac{(1+c)^{-n-1}}{n+1}x^{n+1}\right|\leq$$

$$\leq\frac{1}{n+1}\cdot\left[\frac{M}{1+m}\right]^{n+1}\to 0,$$

because $\frac{M}{1+m}<1$*. So,* $\{s_n(x)\}$ *is uniformly convergent to* $\ln(1+x)$*, relative to* x*, on* $[a,b]\subset(-1,1)$*.*

d)

$$(1+x)^{\alpha}=1+\frac{\alpha}{1!}x+\frac{\alpha(\alpha-1)}{2!}x^2+...$$

$$...+\frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!}x^n+...$$

or

$$(1+x)^{\alpha}=1+\sum_{n=1}^{\infty}\frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!}x^n,x\in(-1,1). \tag{2.7}$$

For the series on the right side we shall prove later (Ch.5, Abel Theorem, Theorem 46) that this one is absolutely and uniformly convergent on any closed subinterval $[a,b]$ *of* $(-1,1)$*. We leave the reader to try a direct proof for this last statement. For a fixed* x *in* $(-1,1)$ *the series in (2.7) is convergent (apply the Ratio Test). Thus, the series of functions is simple convergent on* $(-1,1)$*.*

3. Problems

1. Find the Mac Laurin expansion for the following functions. Indicate the convergence (or uniformly convergence) domain for each of them.

a) $f(x) = \frac{1}{4}(\exp(x) + \exp(-x) + 2\cos x)$; Hint: Use formula (2.3) for $\exp(x)$ and for $\exp(-x)$ (put $-x$ instead x!) and formula (2.5) for $\cos(x)$.

b) $f(x) = \frac{1}{2}\arctan(x) + \frac{1}{4}\ln\frac{1+x}{1-x}$; Hint: Compute

$$(\arctan(x))' = \frac{1}{1+x^2} = 1 - x^2 + x^4 - ...$$

and then integrate term by term; write then

$$\ln\frac{1+x}{1-x} = \ln(1+x) - \ln(1-x)$$

and use formula (2.6) twice.

c) $f(x) = x \cdot \arctan(x) - \ln\sqrt{1+x^2}$; Hint: Write

$$\ln\sqrt{1+x^2} = \frac{1}{2}\ln(1+x^2) = \frac{1}{2}(x^2 - \frac{x^4}{2} + \frac{x^6}{3} - ...).$$

d) $f(x) = \frac{1}{x^2-3x+2}$; Hint: Write $\frac{1}{x^2-3x+2} = \frac{A}{x-1} + \frac{B}{x-2}$, then, for instance

$$\frac{1}{x-2} = -\frac{1}{2}\frac{1}{1-\frac{x}{2}} = -\frac{1}{2}\left(1 + \frac{x}{2} + \frac{x^2}{2^2} + ... + \frac{x^n}{2^n} + ...\right).$$

e) $f(x) = \frac{5-2x}{6-5x+x^2}$; f) $f(x) = \ln(2-3x+x^2)$; Hint: $\ln(2-3x+x^2) = \ln(1-x) + \ln(2-x)$ and

$$\ln(2-x) = \ln 2 + \ln(1-\frac{x}{2}) = \ln 2 - \left(\frac{x}{2} + \frac{x^2}{2^2\cdot 2} + \frac{x^3}{2^3\cdot 3} + ...\right).$$

g) $f(x) = x\exp(-2x)$; Hint: in formula (2.3) put instead of x, $-2x$, etc.

h) $f(x) = \sin(3x) + x\cos(3x)$; i) $f(x) = \arcsin x$; Hint: Compute $f'(x) = (1-x^2)^{-\frac{1}{2}}$ and use the formula (2.7) with $-x^2$ instead of x and $\alpha = -\frac{1}{2}$.

j) $f(x) = \sin^3 x$; Hint: Write $\sin^3 x = \frac{3}{4}\sin x - \frac{1}{4}\sin 3x$ and use formula (2.4) twice.

2. Write as a series of the form $\sum_{n=0}^{\infty} a_n(x+3)^n$ the following functions (say where this representation is possible):

a) $f(x) = \sin(3x+2)$; Hint: Denote $x+3 = z$ (a new variable) and write $f(x)$ as a new function of z :

$g(z) = \sin(3(z-3)+2) = \sin(3z-7) = [\sin 3z]\cos 7 - [\cos 3z]\sin 7 =$

$$= [\cos 7]\left(3z - \frac{(3z)^3}{3!} + ...\right) - [\sin 7]\left(1 - \frac{(3z)^2}{2!} + ...\right);$$

now, come back to $f(x)$ by the substitution $z = x + 3$, etc.

b) $f(x) = \sqrt[3]{(3+2x)}$; c) $f(x) = \ln(5-4x)$; d) $f(x) = \exp(2x+5)$; e) $f(x) = \frac{1}{\sqrt{2-3x}}$; f) $f(x) = \frac{1}{x^2+3x+2}$.

3. Using Mac Laurin formulas, compute the following limits:

a)$\lim\limits_{x\to 0}\frac{\exp(x^3)-1+\ln(1+2x^3)}{x^3}$; b)$\lim\limits_{x\to 0}\frac{\ln(1+2x)-\sin 2x+2x^2}{x^3}$; c)$\lim\limits_{x\to 0}\frac{\sqrt[3]{1+3x}-x-1}{1-4x-\exp(-4x)}$;

d)$\lim\limits_{x\to 0}\frac{\cos x-\exp(-\frac{x^2}{2})}{x^4}$;

e) $\lim\limits_{x\to\infty}\left[x - x^2\ln\left(1+\frac{1}{x}\right)\right]$; Hint: Write $y = \frac{1}{x}$; now, $x \to \infty$ if and only if $y > 0$ and $y \to 0$; our limit becomes

$$\lim_{y\to 0}\left[\frac{1}{y} - \frac{1}{y^2}\ln(1+y)\right] = \lim_{y\to 0}\left[\frac{1}{y} - \frac{1}{y^2}\left(y - \frac{y^2}{2} + \frac{y^3}{3} - ...\right)\right] =$$

$$= \lim_{y\to 0}\left[\frac{1}{2} - \frac{y}{3} + ...\right] = \frac{1}{2}.$$

4. Using Taylor formula approximately compute: a)$\sqrt{1.07}$ with 2 exact decimal digits; b)$\exp(0.25)$ with 3 exact decimals; c)$\ln(1.2)$ with 3 exact decimals; d)$\sin 1^\circ$ with 5 exact decimals; Hint: $1^\circ = \frac{\pi}{180}$ radians; so,

$$\sin\frac{\pi}{180} \approx \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - ... + (-1)^n\frac{x^{2n+1}}{(2n+1)!},$$

where $x = \frac{\pi}{180}$ and n is chosen such that $|R_{2n+1}(x)|$, which is less then $\frac{1}{(2n+2)!}x^{2n+2}$, to be less than $\frac{1}{10^5}$. So, we force

$$\frac{1}{(2n+2)!}\left(\frac{\pi}{180}\right)^{2n+2} < \frac{1}{10^5}$$

and find such a n.

CHAPTER 5

Power series

1. Power series on the real line

We saw that Mac Laurin series are special cases of some particular series of functions $\sum_{n=0}^{\infty} a_n x^n$, where $\{a_n\}$ is a fixed numerical sequence. If one translates x into $x-a$, where a is a fixed real number, we obtain a more general series of functions, $\sum_{n=0}^{\infty} a_n (x-a)^n$. These ones are called *power series (with centre at a) on the real line*. If we put $y = x - a$ in this last series, we get $\sum_{n=0}^{\infty} a_n y^n$, i.e. a power series with centre at 0, but in the variable y. Such translations reduce the study of a general power series $\sum_{n=0}^{\infty} a_n (x-a)^n$ to a power series $\sum_{n=0}^{\infty} a_n x^n$ with centre at 0. The mapping $x \to \sum_{n=0}^{\infty} a_n x^n$ give rise to a function $S(x) = \sum_{n=0}^{\infty} a_n x^n$. The maximal definition domain $M_c = \{x \in \mathbb{R} : \sum_{n=0}^{\infty} a_n x^n$ is convergent$\}$ of this function S is called the *convergence set* of the series. At least $x = 0$ is an element of M_c ($S(0) = a_0$). Sometimes M_c reduces to the number 0. For instance, $S(x) = \sum_{n=0}^{\infty} n! x^n$ is convergent only at 0. Indeed, let us consider the series $\sum_{n=0}^{\infty} n! \, |x|^n$of moduli and apply the Ratio Test: $\lim\limits_{n\to\infty} \frac{a_{n+1}}{a_n} = \lim\limits_{n\to\infty} (n+1)\,|x| = \infty$, except $x = 0$. In fact, if $x \neq 0$, $\{n! x^n\}$ does not tend to 0 (why?). Sometimes $M_c = \mathbb{R}$, as in the case of the series $S(x) = \sum_{n=0}^{\infty} \frac{1}{n!} x^n = \exp(x)$.

In the following, we want to describe the general form of the convergence set of a power series $\sum_{n=0}^{\infty} a_n x^n$. Since the convergence set is the same if we get out a finite number of terms, we can assume that $a_n \neq 0$ for any $n = 0, 1, ...$ If for an infinite number of n the term a_n is 0, we can define the following number R by using the Cauchy-Hadamard formula (see Remark (16)). Thus, finally, we can suppose that $a_n \neq 0$ for any $n = 0, 1, ...$ The number

$$R = \frac{1}{\limsup\{\left|\frac{a_{n+1}}{a_n}\right|\}}$$

in $[0, \infty]$ (i.e. R can be also ∞) is called the *convergence radius* of the series $\sum_{n=0}^{\infty} a_n x^n$. Recall that $\limsup\{x_n\}$ is obtained in the following way. Take all the convergent subsequences (include the unbounded and increasing subsequences, i.e. subsequences which are "convergent" to

∞ in $\overline{\mathbb{R}}$) of the sequence $\{x_n\}$ and the greatest of all these limits of them is called $\limsup\{x_n\}$, the superior limit of the sequence $\{x_n\}$.

THEOREM 46. *(Abel Theorem) Let $\sum_{n=0}^{\infty} a_n x^n$ be a power series with real coefficients a_0, a_1, ..., a_n, ... and let $R = \frac{1}{\limsup\{\left|\frac{a_{n+1}}{a_n}\right|\}}$ in $[0, \infty]$ be its convergence radius.*

i) If $R \neq 0$, then the series S is absolutely convergent on the interval $(-R, R)$ and absolutely uniformly convergent on any closed interval $[-r, r]$, where $0 < r < R$. Moreover, the series is absolutely and uniformly convergent on any closed subinterval $[a, b]$ of $(-R, R)$. If $R \neq \infty$, the series S is divergent on $(-\infty, -R) \cup (R, \infty)$, so,

$$(-R, R) \subset M_c \subset [-R, R],$$

i.e. the convergence set of the series contains the open interval $(-R, R)$, it is contained in $[-R, R]$ and at $x = -R$, or at $x = R$ we must decide in each particular case if the series is convergent or not.

ii) If $R = 0$, then the series S is convergent only at $x = 0$, i.e. $M_c = \{0\}$.

iii) If $R \neq 0$, then the function $S : (-R, R) \to \mathbb{R}$ is of class C^∞ on $(-R, R)$, $S'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1}$ (termwise differentiation) and a primitive of S on $(-R, R)$ is $U(x) = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}$ (term by term integration). All these power series U, S, S', S'', S''', ..., $S^{(n)}$, ... and any other power series obtained from them by a termwise integration or differentiation process have the same convergence radius. Moreover, if the series $\sum_{n=0}^{\infty} a_n x^n$ is convergent at $x = R$, for instance, then the function $S : (-R, R] \to \mathbb{R}$, defined by $S(x) = \sum_{n=0}^{\infty} a_n x^n$ if $x \neq R$ and $S(R) = \sum_{n=0}^{\infty} a_n R^n$ is continuous on $(-R, R]$. With this last hypotheses fulfiled, we also have that the series $\sum_{n=0}^{\infty} a_n x^n$ is absolutely and uniformly convergent on each closed subinterval of the type $[-R + \varepsilon, R]$, where $\varepsilon > 0$ is a small ($\varepsilon < 2R$) positive real number. The same is true if we put $-R$ instead of R and if the numerical series $S(-R) = \sum_{n=0}^{\infty} a_n (-R)^n$ is convergent.

PROOF. The last statement will not be proved here. An elegant proof can be found in [**Pal**], Theorem 2.4.6.

i) Let us consider x as a fixed parameter (for the moment) and let us apply the Ratio Test to the series of moduli $\sum_{n=0}^{\infty} |a_n| |x|^n$. Let L be the limit

$$L = \limsup \left\{ \frac{|a_{n+1}| |x|^{n+1}}{|a_n| |x|^n} \right\} = \left[\limsup \left\{ \frac{|a_{n+1}|}{|a_n|} \right\} \right] |x| = \frac{|x|}{R}.$$

If $R = \infty$, then $L = 0 < 1$, so the series is absolutely convergent for any $x \in \mathbb{R}$. If $R = 0$, then $L = \infty$, except maybe the case when

$x = 0$. Hence, if $R = 0$, the series is convergent ONLY for $x = 0$, i.e. the statement of ii). Suppose now that $R \neq 0, \infty$. Then, whenever $L = \frac{|x|}{R} < 1$, or $x \in (-R, R)$, the series is absolutely convergent, in particular convergent (see Theorem 31). If $x \in (-\infty, -R) \cup (R, \infty)$, or $|x| > R$, then $L > 1$. Hence,

$$\limsup \left\{ \frac{|a_{n+1}| \, |x|^{n+1}}{|a_n| \, |x|^n} \right\} > 1.$$

This means that there is at least one subsequence $\left\{ \frac{\left|a_{n_k+1}\right| |x|^{n_k+1}}{\left|a_{n_k}\right| |x|^{n_k}} \right\}$ of $\left\{ \frac{|a_{n+1}||x|^{n+1}}{|a_n||x|^n} \right\}$ such that $\frac{\left|a_{n_k+1}\right| |x|^{n_k+1}}{\left|a_{n_k}\right| |x|^{n_k}} > 1$, i.e.

$$|a_{n_k+1}| \, |x|^{n_k+1} > |a_{n_k}| \, |x|^{n_k}$$

for any $k = 0, 1, \ldots$. Thus the sequence $\{a_n x^n\}$ cannot tend to 0 and so, the series $\sum_{n=0}^{\infty} a_n x^n$ cannot be convergent for such an x. Let now $x \in [-r, r]$, where $0 < r < R$. Since for $x = r < R$, the series $\sum_{n=0}^{\infty} |a_n| \, r^n$ is convergent ($r \in (-R, R)$, so the series $\sum_{n=0}^{\infty} a_n x^n$ is absolutely convergent, see i)). But, $|a_n x^n| \leq |a_n| \, r^n$ for any $n = 0, 1, \ldots$ implies that the series $\sum_{n=0}^{\infty} a_n x^n$ is absolutely and uniformly convergent (we apply here the Weierstrass Test Theorem 41) on $[-r, r]$. Since any interval $[a, b] \subset (-R, R)$ can be embedded in a symmetrical interval of the form $[-r, r] \subset (-R, R)$, we obtain that the series $\sum_{n=0}^{\infty} a_n x^n$ is absolutely and uniformly convergent on ANY closed subinterval $[a, b]$ of $(-R, R)$.

iii) It is easy to see that all the power series U, S', S'', ... have the same convergent radius R as the series S. Applying the Weierstrass test to each of them on an interval of the form $[-r, r] \subset (-R, R)$ and the theorems 39 and 40, we can prove easily the first statement of iii). □

Let us consider the power series

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n.$$

We know that this one is identical with $\ln(1 + x)$ on $(-1, 1)$. Let us find the convergence set M_c of it. The convergence radius is equal to

$$R = \frac{1}{\limsup\{\left|\frac{a_{n+1}}{a_n}\right|\}} = \frac{1}{\limsup\{\left|\frac{\frac{1}{n+1}}{\frac{1}{n}}\right|\}} = 1.$$

At $x=-1$, the series becomes

$$-\sum_{n=1}^{\infty}\frac{1}{n}=-\infty,$$

so the series is divergent at $x=-1$. Now, $S(1)=\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n}$ is the alternate series, which was proved to be convergent. Since both functions $S(x)$ and $\ln(1+x)$ are continuous at $x=1$ (prove it!-by using iii) of the Abel Theorem), one has that $S(1)=\ln 2$. From Abel Theorem we see that M_c is exactly $(-1,1]$. On this interval it is $\ln(1+x)$ but, the series does not exist outside of $(-1,1]$, while the function $\ln(1+x)$ does exist, for instance at $x=2$!

Let us now look at the binomial series

$$1+\sum_{n=1}^{\infty}\frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!}x^n,$$

where α is a fixed real parameter. Let us find the convergence radius of this series:

$$R=\frac{1}{\limsup\{\left|\frac{a_{n+1}}{a_n}\right|\}}=\lim_{n\to\infty,n>\alpha}\frac{n-\alpha}{n+1}=1 \tag{1.1}$$

If $x=-1$, the series is not convergent for any α. For instance, if $\alpha=-1$, then $\sum_{n=0}^{\infty}(-1)^n(-1)^n=\infty$. At $x=1$, $\sum_{n=0}^{\infty}(-1)^n$ is divergent. If α is a natural number k, then the series becomes a polynomial, so its convergence set is the whole $\mathbb{R}$. But,...the formula (1.1) and Abel Theorem say that... $M_c=\mathbb{R}\subset[-1,1]$!!! Somewhere must be a mistake! Indeed, since $a_{k+1}=a_{k+2}=...=0$, $\limsup\{\left|\frac{a_{n+1}}{a_n}\right|\}$ is nondeterministic, so the computation of R in (1.1) is wrong! We see that the convergence set $M_c(\alpha)$ of the binomial series strongly depends on α. We do not give here a complete discussion of $M_c(\alpha)$ as a function of α.

Let us find the convergence set for the following series of functions

$$S(x)=\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\left(\frac{1}{2x+1}\right)^n.$$

This is not a power series but, making the substitution $y=\frac{1}{2x+1}$, we obtain a power series $\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}y^n$ in the new variable y. The convergence radius of this last series is

$$R=\frac{1}{\limsup\{\left|\frac{a_{n+1}}{a_n}\right|\}}=\frac{1}{\lim_{n\to\infty}\frac{n^2}{(n+1)^2}}=1.$$

For $y = \pm 1$, the series is convergent (why?). So, the convergence set $M_{c,y}$ for the power series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} y^n$$

is $M_{c,y} = [-1, 1]$. Coming back to the variable x, we get that the initial series of functions

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \left(\frac{1}{2x+1} \right)^n$$

is convergent if and only of $-1 \leq \frac{1}{2x+1} \leq 1$, i. e.

$$x \in (-\infty, -1] \cup [0, \infty).$$

Hence, the set of all x in $\mathbb{R}$ such that the series

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n^2} \left(\frac{1}{2x+1} \right)^n$$

is convergent, i.e. the convergence set of this last series, is

$$(-\infty, -1] \cup [0, \infty).$$

REMARK 16. *(Cauchy-Hadamard) Another useful formula for computing the convergence radius R of a power series $\sum_{n=0}^{\infty} a_n x^n$ is the following Cauchy-Hadamard formula:*

$$R = \frac{1}{\limsup \sqrt[n]{|a_n|}} \tag{1.2}$$

This formula can be used even when an infinite number of a_n are zero. The proof of Abel's Theorem by using this formula for R is completely analogue to the proof of the same theorem given above. In this case one must use the Root Test (Theorem 29) instead of the Ratio Test as we did in proving Abel Theorem. If we start with the definition of R as it appears in formula Cauchy-Hadamard (1.2), we get the same interval of convergence $(-R, R)$ for our series $\sum_{n=0}^{\infty} a_n x^n$ (why?). Thus, the both formulas give rise to one and the same number.

Let us find the convergence set and the sum of the series of functions

$$\sum_{n=0}^{\infty} \frac{1}{2n+1} (3x+2)^{2n+1}.$$

This one is not a power series but,...we can associate to it a power series by the following substitution $y = 3x + 2$. Hence, we must study

the power series in y :

$$\sum_{n=0}^{\infty} \frac{1}{2n+1} y^{2n+1}.$$

Here $a_{2n+1} = \frac{1}{2n+1}$ and $a_{2n} = 0$ for any $n = 0, 1, \ldots$. In our case, it is not a good idea to apply Abel formula $R = \frac{1}{\limsup\{|\frac{a_{n+1}}{a_n}|\}}$ (why?). Let us apply Cachy-Hadamard formula (1.2):

$$R = \frac{1}{\limsup \sqrt[n]{|a_n|}} = 1,$$

because the sequence $\{\sqrt[n]{|a_n|}\}$ is the union between two convergent subsequences:

$$\{\sqrt[2n+1]{|a_{2n+1}|}\} = \{\sqrt[2n+1]{\frac{1}{2n+1}}\} \to 1$$

(why?) and

$$\{\{\sqrt[2n]{|a_{2n}|}\} = \{0\} \to 0$$

and so, $\limsup \sqrt[n]{|a_n|} = 1$. At $y = -1$ the series

$$\sum_{n=0}^{\infty} \frac{1}{2n+1} y^{2n+1}$$

becomes

$$-\sum_{n=0}^{\infty} \frac{1}{2n+1} = -\infty$$

(why?). At $y = 1$ the series is

$$\sum_{n=0}^{\infty} \frac{1}{2n+1} = \infty.$$

Hence, the convergence set for the power series in y is $(-1, 1)$ (see Abel Theorem 46). Now, if $T(y) = \sum_{n=0}^{\infty} \frac{1}{2n+1} y^{2n+1}$ for $y \in (-1, 1)$, one has:

$$T'(y) = \sum_{n=0}^{\infty} y^{2n} = \frac{1}{1-y^2} = \frac{1}{2} \cdot \frac{1}{1-y} + \frac{1}{2} \cdot \frac{1}{1+y}.$$

Thus,

$$T(y) = \frac{1}{2} \ln \frac{1+y}{1-y} + C.$$

But $C = 0$ because $T(0) = 0$. Let us come back to the series in x. The convergence set is

$$\{x \in \mathbb{R} : -1 < 3x + 2 < 1\} = (-1, -\frac{1}{3}).$$

Its sum is

$$S(x) = T(3x+2) = \frac{1}{2}\ln\left(-\frac{3x+3}{3x+1}\right)$$

for any $x \in (-1, -\frac{1}{3})$.

EXAMPLE 9. *(*arctan *series) Let us find the Mac Laurin expansion for* $f(x) = \arctan x$. *For this let us consider*

$$f'(x) = \frac{1}{1+x^2} = 1 - x^2 + x^4 - ... + (-1)^n x^{2n} + ...,$$

where $|x| < 1$ *(why?). Apply now Theorem 43 and termwisely integrate this last equality:*

$$\arctan x + C = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + ... + (-1)^n \frac{x^{2n+1}}{2n+1} + ..., \tag{1.3}$$

where $|x| < 1$. *For* $x = 0$ *we get* $C = 0$. *Since for* $x = 1$ *the series on the right is convergent and since the function*

$$S(x) = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + ... + (-1)^n \frac{x^{2n+1}}{2n+1} + ...$$

is continuous at $x = 1$ *(see Abel's Theorem, iii)), we get that*

$$\arctan 1 = \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + ... + (-1)^n \frac{1}{2n+1} + ... \tag{1.4}$$

Let us find the convergence set and the sum for the power series

$$\sum_{n=1}^{\infty} n(n+1)x^n.$$

The convergence radius is

$$R = \lim_{n\to\infty} \frac{n(n+1)}{(n+1)(n+2)} = 1$$

(why?). Since at $x = \pm 1$ the series is divergent ($n(n+1) \nrightarrow 0$), the convergence set is $M_c = (-1, 1)$. Let us integrate termwise (see Theorem 43) the above series for $x \in (-1, 1)$:

$$\int \left[\sum_{n=1}^{\infty} n(n+1)x^n\right] dx = \sum_{n=1}^{\infty} nx^{n+1} = \sum_{n=1}^{\infty} (n+2)x^{n+1} - 2\sum_{n=1}^{\infty} x^{n+1}.$$

But the series

$$\sum_{n=1}^{\infty} x^{n+1} = x^2 + x^3 + ... = \frac{x^2}{1-x}$$

(it is an infinite geometrical progression). So we get

$$\int \left[\sum_{n=1}^{\infty} n(n+1)x^n\right] dx = \sum_{n=1}^{\infty}(n+2)x^{n+1} - \frac{2x^2}{1-x}.$$

Let us integrate again this last equality

$$\int \left[\int \left[\sum_{n=1}^{\infty} n(n+1)x^n\right] dx\right] dx = \left(\sum_{n=1}^{\infty} x^{n+2}\right) + x^2 + 2x + 2\ln(1-x) =$$

$$= \frac{x^3}{1-x} + x^2 + 2x + 2\ln(1-x).$$

Coming back and differentiating twice, we get:

$$\sum_{n=1}^{\infty} n(n+1)x^n = \frac{2x}{(x-1)^3}, \text{ for } |x| < 1.$$

2. Complex power series and Euler formulas

In Chapter 2, Section 2, we introduced the metric space of complex number fields $\mathbb{C}$. In fact, $\mathbb{C}$ is a normed spaced with the norm given by the usual complex modulus $|z| = \sqrt{x^2+y^2}$, where $z = x+iy$, $x, y \in \mathbb{R}$ (prove the properties of the norm for this particular norm!). Since a sequence $\{z_n = x_n + iy_n\}$ is convergent to $z = x+iy$ in $\mathbb{C}$ if and only if both the real sequences $\{x_n\}$ and $\{y_n\}$ are convergent to x and to y respectively (see Theorems 1 and 16), the study of the numerical series with complex terms reduces to the study of the real numerical series. But this way is not so easy to put in practice. The best way is to use firstly the absolute convergence notion like in the case of series in a general normed space. Namely, let $s = \sum_{n=0}^{\infty} z_n$ be a series with complex numbers terms and let $S = \sum_{n=0}^{\infty} |z_n|$ be the real series of moduli. The following result is very useful in practice.

THEOREM 47. *If the series of moduli $S = \sum_{n=0}^{\infty} |z_n|$ is convergent (like a numerical real series with nonnegative terms), the initial series with complex terms $s = \sum_{n=0}^{\infty} z_n$ is convergent in $\mathbb{C}$.*

PROOF. Let $s_n = \sum_{k=0}^{n} z_k$ be the n-th partial sum of the series $s = \sum_{n=0}^{\infty} z_n$ and let $S_n = \sum_{k=0}^{n} |z_k|$ be the n-th partial sum of the series of moduli $S = \sum_{n=0}^{\infty} |z_n|$. Since

$$|s_{n+p} - s_n| \leq |z_{n+1}| + |z_{n+2}| + ... + |z_{n+p}| = S_{n+p} - S_n,$$

and since the series S is convergent (i.e. the sequence $\{S_n\}$ is a Cauchy sequence), one obtains that the sequences $\{s_n\}$ is a Cauchy sequence. Thus, it is convergent to a complex number s (the sum of the series

$\sum_{n=0}^{\infty} z_n$) in $\mathbb{C}$, because $\mathbb{C}$ is a complete metric space (see Theorem 16). □

The Cauchy Test and the zero Test also work in the case of a complex series (why?-Hint: $\mathbb{C}$ is a complete metric space-why?). Series of complex functions and power series are defined exactly in the same way like the analogous real case. However, in the complex case, the study of the convergence set of a series of function is more complicated than in the real case.

EXAMPLE 10. *(Complex geometrical series). Let us find the convergence set for the complex geometrical series*

$$s(z) = \sum_{n=0}^{\infty} z^n = 1 + z + z^2 + ...$$

Let us consider the series of moduli

$$S(|z|) = \sum_{n=0}^{\infty} |z|^n = \lim_{n\to\infty} \frac{1 - |z|^{n+1}}{1 - |z|}.$$

This limit exists if $|z| < 1$. Hence, the series is absolutely convergent if and only if $|z| < 1$. In particular, for $|z| < 1$, the series is convergent (see Theorem 47). Is the series convergent for a z with $|z| > 1$? Let us see ! If $|z| > 1$, the sequence $\{z^n\}$ goes to ∞ in $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, the Riemann sphere (why?), so, the series is divergent (see the zero Test). What happens if $|z| = 1$?, i.e. if z is a complex number on the circle of radius 1 and with centre at origin. If $z = 1$, the series is divergent. If $z \neq 1$, but $|z| = 1$, the sequence $\{z^n\}$ is never convergent to zero! (why?). Thus, the convergence set for the series $s(z) = \sum_{n=0}^{\infty} z^n$ is exactly the open disc $B(0;1) = \{z \in \mathbb{C} : |z| < 1\}$ in the complex plane $\mathbb{C}$.

To define the basic elementary complex functions one uses complex power series. For instance, the exponential complex function is defined by the formula

$$\exp(z) = 1 + \frac{z}{1!} + \frac{z^2}{2!} + ... + \frac{z^n}{n!} + ... = \sum_{n=0}^{\infty} \frac{z^n}{n!} \tag{2.1}$$

It is easy to prove (do it!) that this series is absolutely convergent on the whole complex plane $\mathbb{C}$ and absolutely uniformly convergent on any bounded subset of $\mathbb{C}$. One can prove that $\exp(z_1+z_2) = \exp(z_1)\exp(z_2)$ for any z_1, z_2 in $\mathbb{C}$ (see [**ST**] for instance).

The series on the right side of (2.1) is the natural extension of the Mac Laurin expansion of the real function $\exp(x)$ to the whole complex plane. Using this "trick" we can define other elementary complex functions:

$$\sin(z) \overset{def}{=} \frac{z}{1!} - \frac{z^3}{3!} + \frac{z^5}{5!} - ... + (-1)^n \frac{z^{2n+1}}{(2n+1)!} + ... \tag{2.2}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} z^{2n+1}, z \in \mathbb{C}$$

$$\cos(z) \overset{def}{=} 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \frac{z^6}{6!} + ... + (-1)^n \frac{z^{2n}}{(2n)!} + ... \tag{2.3}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} z^{2n}, z \in \mathbb{C}$$

$$\ln(1+z) \overset{def}{=} z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + ... + (-1)^{n-1} \frac{z^n}{n} + ... = \tag{2.4}$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} z^n, |z| < 1.$$

$$(1+z)^{\alpha} \overset{def}{=} 1 + \frac{\alpha}{1!} z + \frac{\alpha(\alpha-1)}{2!} z^2 + ... +$$

$$+ \frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!} z^n + ...,$$

so,

$$(1+z)^{\alpha} = 1 + \sum_{n=1}^{\infty} \frac{\alpha(\alpha-1)(\alpha-2)...(\alpha-n+1)}{n!} z^n, |z| < 1, \alpha \in \mathbb{C} \tag{2.5}$$

In the same way we can define any other complex function $f(z)$ if we know a Taylor expansion for the real function $f(x)$ (if this last one has real values and if it can be extended beyond the real line!). For instance, we know that

$$sh(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + ... + \frac{x^{2n+1}}{(2n+1)!} + ..., x \in \mathbb{R}.$$

We simply define the complex hyperbolic sine as

$$sh(z) \overset{def}{=} z + \frac{z^3}{3!} + \frac{z^5}{5!} + ... + \frac{z^{2n+1}}{(2n+1)!} + ..., z \in \mathbb{C}. \tag{2.6}$$

and

$$ch(z) \stackrel{def}{=} 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + ... + \frac{z^{2n}}{(2n)!} + ..., z \in \mathbb{C}. \tag{2.7}$$

We always have to check if the series on the right side is convergent on the extrapolated domain (for instance, we extrapolated $\mathbb{R}$ to $\mathbb{C}$). The restrictions of all these functions to their definition domains on the real line give rise to the well known real functions. For instance, $\ln(1+z)$, $|z| < 1$, restricted to $\mathbb{R}$ give rise to $\ln(1+x)$. This does not mean that we defined the function $\ln(z)$ for any $z \neq 0$! To define such a function, i.e. the inverse of the complex exponential function, is not an easy task, because it will be not an usual function, i.e. for a z we have more than one value of $\ln(z)$. This is because $\exp(z)$ is not injective at all. To see this we need some famous relations, the Euler formulas.

THEOREM 48. *(Euler relations) For any x a real number and for $i = \sqrt{-1}$ we have*

$$\exp(ix) = \cos(x) + i\sin(x), \tag{2.8}$$

$$\cos(x) = \frac{\exp(ix) + \exp(-ix)}{2} \tag{2.9}$$

and

$$\sin(x) = \frac{\exp(ix) - \exp(-ix)}{2i}.$$

PROOF. We simply use formula (2.1) to compute $\exp(ix)$:

$$\exp(ix) = 1 + \frac{ix}{1!} - \frac{x^2}{2!} - \frac{ix^3}{3!}... + \frac{(ix)^n}{n!} + ... = \cos(x) + i\sin(x).$$

If now we put instead of x, $-x$ in the formula (2.8), we get

$$\exp(-ix) = \cos(x) - i\sin(x), \tag{2.10}$$

because cosine is an even function and sine is an odd one. Adding formulas (2.8) and (2.10), we get the relation $\exp(ix) + \exp(-ix) = 2\cos(x)$. Now, subtract formula (2.10) from formula (2.8) and get the formula $\exp(ix) - \exp(-ix) = 2i\sin(x)$, etc. □

Let us justify now that the complex function $\exp(z)$ is not invertible, i.e. it cannot have like inverse an usual function. Using Euler formulas from the theorem we get that

$$\exp(2k\pi i) = \cos(2k\pi) + i\sin(2k\pi) = 1,$$

for any integer k. Thus one has an infinite number of complex numbers $\{2n\pi i\}$, $n = 0, \pm 1, \pm 2, ...$, at which the exponential function has value 1!. This is why the inverse of $\exp(z)$ is the multivalued function

$$Ln(z) = \ln|z| + i(\theta + 2k\pi), k = 0, \pm 1, \pm 2, ...$$

and θ is the argument of z, i.e. the unique real number in $[0, 2\pi)$ such that $z = |z|[\cos\theta + i\sin\theta]$, the trigonometric representation of z (prove this last equality by drawing...). It has a double infinite number of "branches", i.e. $Ln(z)$ is in fact the set

$$\{\ln^{(k)}(z) = \ln|z| + i(\theta + 2k\pi)\}, k = 0, \pm 1, \pm 2, ...$$

of usual functions. All of these functions have the same real part $\ln|z|$. For $k = 0$ we get the principal branch, $\ln(z) = \ln|z| + i\arg z$. Sometimes in books people work with this last expression for the complex logarithmic function, without mention this. We leave as an exercise for the reader to define the radical complex multiform function $\sqrt[n]{z}$ (it has only n branches!-find them!). One can start with the fact that $\sqrt[n]{z}$ is the inverse of the power n function $z \rightsquigarrow z^n$ and with the equality:

$$z^n = |z|^n[\cos n\theta + i\sin n\theta],$$

etc.

Euler's formulas from the above theorem are very useful in practice. For instance, the famous de Moivre formula

$$[\cos x + i\sin x]^n = \cos nx + i\sin nx$$

from trigonometry, can be immediately proved by using the basic properties of the complex exponential function : $\exp(z)\exp(w) = \exp(z+w)$ (try to prove it!), $(\exp z)^n = \exp(nz)$, where $z, w \in \mathbb{C}$, and n is an integer number. If one extends in a natural way (componentwise!) the integral calculus from real functions to functions of real variables but with complex values:

$$\int [f(x) + ig(x)]dx = \int f(x)dx + i\int g(x)dx,$$

one can compute in an easy way more complicated integrals. For instance, let us find a primitive for a very known family of functions $f(x) = \exp(ax)\cos(bx)$, where a, b are two fixed real numbers (parameters). Let us denote by $g(x) = \exp(ax)\sin(bx)$ (its partner!) and let us find a primitive for $f(x) + ig(x)$:

$$\int [\exp(ax)\cos(bx) + i\exp(ax)\sin(bx)]dx = \int \exp(ax)\exp(ibx)dx =$$

$$= \int \exp(ax + ibx)dx = \frac{\exp(ax + ibx)}{a + ib} =$$

$$= \frac{\exp(ax) \cdot [\cos(bx) + i \sin(bx)](a - ib)}{a^2 + b^2} =$$

$$= \exp(ax)\frac{a\cos(bx) + b\sin(bx)}{a^2 + b^2} + i\exp(ax)\frac{a\sin(bx) - b\cos(bx)}{a^2 + b^2}.$$

Hence,

$$\int \exp(ax)\cos(bx)dx = \exp(ax)\frac{a\cos(bx) + b\sin(bx)}{a^2 + b^2}$$

and

$$\int \exp(ax)\sin(bx)dx = \exp(ax)\frac{a\sin(bx) - b\cos(bx)}{a^2 + b^2}$$

(why?).

Another example of a nice application of Euler formulas is the following. Suppose we forgot the formula for $\sin 3x$ and of $\cos 3x$ in language of $\sin x$ and $\cos x$ respectively. Let us find it by writing

$$\cos 3x + i \sin 3x = \exp(i3x) =$$

(Euler formula)

$$= [\exp(ix)]^3 = [\cos x + i \sin x]^3 =$$

$$= \cos^3 x - 3\cos x \sin^2 x + i[3\cos^2 x \sin x - \sin^3 x].$$

Since two complex numbers are equal if their real and imaginary parts are equal, we get the formulas:

$$\cos 3x = \cos x[\cos^2 x - 3\sin^2 x] = \cos x[4\cos^2 x - 3],$$

$$\sin 3x = [3\cos^2 x \sin x - \sin^3 x] = \sin x[3 - 4\sin^2 x].$$

3. Problems

1. Find the convergence set and the sum for the following series of functions:

a)$\sum_{n=0}^{\infty}(3x + 5)^n$; b)$\sum_{n=0}^{\infty}(-1)^n(4x + 1)^n$; c)$\sum_{n=1}^{\infty}\frac{x^n}{n}$;
d)$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^n}{n}$; e)$\sum_{n=1}^{\infty} n(3x + 5)^n$; f)$\sum_{n=0}^{\infty}\frac{x^n}{(n+1)2^n}$;

2. Find the convergence set for the following series of functions:

a)$\sum_{n=1}^{\infty}\frac{1}{\left(1+\frac{1}{n}\right)^{n^2}}(x - 3)^n$; b)$\sum_{n=1}^{\infty}\frac{x^n}{n^2}$; c)$\sum_{n=0}^{\infty} n!x^n$; d)$\sum_{n=0}^{\infty}\frac{x^n}{n!}$;

e)$\sum_{n=1}^{\infty}\frac{x^n}{n^n}$; f)$\sum_{n=1}^{\infty}\frac{n^5}{5^n}x^n$; g)$\sum_{n=0}^{\infty}\frac{x^n}{2^n+3^n}$; h)$\sum_{n=1}^{\infty}\left(\frac{n+1}{n}\right)^{n^2}x^n$;
i)$\sum_{n=0}^{\infty}[1-(-2)^n]x^n$; j)$\sum_{n=0}^{\infty}(-1)^{n+1}3^nx^n$; k)$\sum_{n=1}^{\infty}\frac{1}{2n+1}\left(\frac{1+x}{1-x}\right)^n$;
l)$\sum_{n=1}^{\infty}(-1)^n\frac{2^n(x-5)^{2n}}{n^2}$; m)$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{(x-5)^{2n}}{n3^n}$ (find its sum);

3. Use the power series in order to compute the following sums: a)$\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1}{n}$; b)$\sum_{n=0}^{\infty}\frac{1}{(n+1)2^n}$; c)$\sum_{n=1}^{\infty}\frac{n}{2^n}$; (Hint: associate the power series

$$S(x)=\sum_{n=1}^{\infty}nx^n=x(1+2x+3x^2+...)=x(x+x^2+x^3+...)'=x\left(\frac{x}{1-x}\right)';$$

make then $x=\frac{1}{2}$).

CHAPTER 6

The normed space $\mathbb{R}^m$.

1. Distance properties in $\mathbb{R}^m$

Motivation Let $\{O; \mathbf{i}, \mathbf{j}\}$ be a Cartesian coordinate system in a plane $(\mathcal{P})$. To any point $M \in (\mathcal{P})$ we associate the position vector $\overrightarrow{OM}$. We know that there is a unique pair (x, y) of real numbers such that $\overrightarrow{OM} = x\mathbf{i} + y\mathbf{j}$. Here $\mathbf{i}, \mathbf{j}$ are two perpendicular versors with their origin in O. Usually one calls (x, y) the coordinates of M relative to the "basis" $\{\mathbf{i}, \mathbf{j}\}$. But we can view (x, y) as an element in $\mathbb{R} \times \mathbb{R} \stackrel{not}{=} \mathbb{R}^2$. If M' is another point in the same plane $(\mathcal{P})$ and if P is the unique point in $(\mathcal{P})$ such that $\overrightarrow{OM} + \overrightarrow{OM'} = \overrightarrow{OP}$, then the coordinates of P are $(x + x', y + y')$, where (x', y') are the coordinates of M'. Let α be a real number (scalar) and let us denote by $\overrightarrow{OM''}$ the vector $\alpha\overrightarrow{OM}$. Then, the coordinates of the point M'' are $(\alpha x, \alpha y) \in \mathbb{R}^2$. So, one can endow the cartesian product $\mathbb{R}^2$ with a natural algebraic structure of a real vector space with 2 dimensions (the number of the elements in any basis of it, in particular in the "canonical" basis $\{(1, 0), (0, 1)\}$, where $(1, 0)$ are the coordinates of the versor $\mathbf{i}$ and $(0, 1)$ are the coordinates of the versor $\mathbf{j}$). Hence, one can study the 2-dimensional dynamics only in the "abstract" space $\mathbb{R}^2$ (this is the basic idea of R. Descartes; the word "cartesian" comes from "Descartes", in Latin "Cartesius"; he invented a very useful tool for Engineering, namely the Analytic Geometry; here we work with numbers and equations instead of geometrical objects like lines, circles, parabolas, etc.). We call $\mathbb{R}^2$ the 2-dimensional space (2-D space). In the same way we can construct the 3-D space $\mathbb{R}^3$ or, more generally, the m-D space

$$\mathbb{R}^m = \underbrace{\mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R}}_{n-times} = \{\mathbf{x} = (x_1, x_2, ..., x_m) : x_j \in \mathbb{R}\}.$$

We recall that if $\mathbf{x} = (x_1, x_2, ..., x_m)$ and $\mathbf{y} = (y_1, y_2, ..., y_m)$ are two "vectors" in $\mathbb{R}^m$, then

$$\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, ..., x_m + y_m)$$

and

$$\alpha\mathbf{x} = (\alpha x_1, \alpha x_2, ..., \alpha x_m)$$

for any "scalar" $\alpha \in \mathbb{R}$ (componentwise operations). For instance, $(-7,3)+(6,0) = (-1,3)$ and $\sqrt{2}(-1,1) = (-\sqrt{2}, \sqrt{2})$. To do analysis in $\mathbb{R}^m$ means firstly to introduce a distance in $\mathbb{R}^m$. $\mathbb{R}^m$ has the "canonical basis"

$$\{(1,0,...,0),(0,1,0,...,0),...(0,0,...,0,1)\}$$

like a real vector space, so it has the dimension m over $\mathbb{R}$. It is more profitable to introduce first of all a "length" of a vector $\mathbf{x} = (x_1, x_2, ..., x_m)$ by the formula

$$\|\mathbf{x}\| \stackrel{def}{=} \sqrt{x_1^2 + x_2^2 + ... + x_m^2}. \tag{1.1}$$

The nonnegative real number $\|\mathbf{x}\|$ is called the *norm* or the *length* of $\mathbf{x}$. If $m = 1$, the norm of a real number x is its absolute value (modulus) $|x|$. If $m = 2$ and if $\mathbf{x} = (x_1, x_2)$ the norm $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$ is exactly the length of the diagonal of the rectangle $[OA_1MA_2]$, or the length of the resultant vector $\overrightarrow{OM} = \overrightarrow{OA}_1 + \overrightarrow{OA}_2$ (see Fig.6.1).

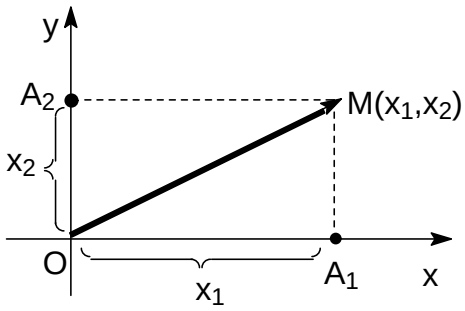


Fig. 6.1

In the 3-D space $\mathbb{R}^3$ the norm of $\mathbf{x} = (x_1, x_2, x_3)$ is $\sqrt{x_1^2 + x_2^2 + x_3^2}$ and it is exactly the length of the diagonal of the parallelepiped generated by $\overrightarrow{OA}_1$, $\overrightarrow{OA}_2$ and $\overrightarrow{OA}_3$ (see Fig.6.2).

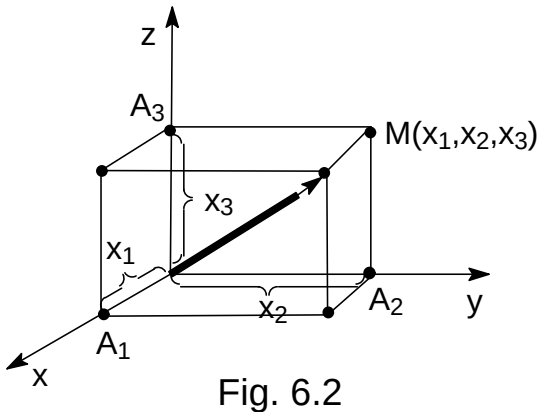


Fig. 6.2

EXAMPLE 11. *(the space-time representation) Let us consider the vector* $\mathbf{x} = (x_1, x_2, x_3, t) \in \mathbb{R}^4$, *where* (x_1, x_2, x_3) *are the coordinates of a point* $M(x_1, x_2, x_3)$ *in the 3-D space and* $t \geq 0$ *is the time when we "observe" the point* M. *Then*

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + t^2}.$$

EXAMPLE 12. *(the space of dynamics) Let us consider a moving point* M *on a trajectory* (γ) *in the 3-D space. The position of* M *is fixed by its coordinates* x_1, x_2, x_3. *Its velocity* $\mathbf{v}$ *is given by another 3 coordinates* $\dot{x}_1, \dot{x}_2, \dot{x}_3$, *the derivatives of the coordinates functions* $x_1(t), x_2(t), x_3(t)$ *at* M. *Thus, the "dynamic" state of* M *is described by the "vectors"*

$$\mathbf{x} = (x_1, x_2, x_3, \dot{x}_1, \dot{x}_2, \dot{x}_3) \in \mathbb{R}^6$$

and

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + x_3^2 + \dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}.$$

THEOREM 49. *The norm mapping*

$$\mathbf{x} \rightsquigarrow \|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + ... + x_m^2},$$

from $\mathbb{R}^m$ *to* $\mathbb{R}_+$, *has the following main properties: 1)* $\|\mathbf{x}\| = 0$ *if and only if* $\mathbf{x} = \mathbf{0}$; *2)* $\|\alpha\mathbf{x}\| = |\alpha| \|\mathbf{x}\|$ *for any* $\alpha \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^m$; *3)* $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, *for any* $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$.

PROOF. 1) and 2) are obvious (prove them!). To be clearer, let us prove 3) for $m = 2$ (for $m > 2$ one can use the Cauchy-Buniakovsky inequality, which can be found in any course of Linear Algebra!). Both sides in 3) are nonnegative, so the inequality is equivalent to

$$\|\mathbf{x} + \mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2 \|\mathbf{x}\| \|\mathbf{y}\|.$$

If $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$, one has

$$(x_1 + y_1)^2 + (x_2 + y_2)^2 \leq x_1^2 + x_2^2 + y_1^2 + y_2^2 + 2\sqrt{(x_1^2 + x_2^2)(y_1^2 + y_2^2)},$$

or, $x_1y_1 + x_2y_2 \leq \sqrt{(x_1^2 + x_2^2)(y_1^2 + y_2^2)}$. By squaring both sides we get

$$2x_1x_2y_1y_2 \leq x_2^2y_1^2 + x_1^2y_2^2,$$

or $0 \leq (x_2y_1 - x_1y_2)^2$. This last inequality is obvious. Moreover, from this last inequality, we can say that in 3) we have equality if and only if $x_2y_1 - x_1y_2 = 0$ or, if and only if $(x_1, x_2) = \lambda(y_1, y_2)$, i.e. $\mathbf{x}$ and $\mathbf{y}$ are collinear. □

The couple $(\mathbb{R}^m, \|.\|)$ is called a *normed space*. We know that in general, a normed space is a real vector space X with a norm mapping $\|.\|$ on it, which verifies the properties 1), 2) and 3) from Theorem 49. We recall that a normed space $(X, \|.\|)$ is also a metric space w.r.t. a canonically induced distance: $d(x, y) = \|x - y\|$ for any x, y in X. In the case of the normed space $(\mathbb{R}^m, \|.\|)$ the distance is given by the formula

$$d(\mathbf{x}, \mathbf{y}) = \|x - y\| = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \tag{1.2}$$

This distance is a very special one because it comes from the "scalar product"

$$< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{m} x_i y_i, \tag{1.3}$$

i.e. this last one induces the norm $\|\mathbf{x}\| =< \mathbf{x}, \mathbf{x} >= \sqrt{\sum_{i=1}^{m} {x_i}^2}$ on $\mathbb{R}^m$ and this norm gives rise exactly to our distance (1.2). As we know from the Linear Algebra course, the scalar product (1.3) endows $\mathbb{R}^m$ with a geometry. The length of a vector $\mathbf{x}$ is its norm $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^{m} {x_i}^2}$ and the cosine of the angle θ between two vectors $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{R}^m$ is defined as

$$\cos\theta = \frac{< \mathbf{x}, \mathbf{y} >}{\|\mathbf{x}\| \, \|\mathbf{y}\|}.$$

The fact that the quantity $\frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{x}\|\|\mathbf{y}\|}$ is always between -1 and 1 is exactly the famous Cauchy- Schwarz-Buniakowsky inequality

$$|< \mathbf{x}, \mathbf{y} >| \leq \|\mathbf{x}\| \, \|\mathbf{y}\| . \tag{1.4}$$

It can be proved only by using the basic properties of a scalar product (see any course in Linear Algebra).

Since $\mathbb{R}^m$ is a metric space relative to the distance d defined in (1.2) we can speak about the convergence of a sequence

$$\{\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, ..., x_m^{(n)})\}$$

from $\mathbb{R}^m$ to a vector $\mathbf{x} = (x_1, x_2, ..., x_m)$: we say that $\mathbf{x}^{(n)} \to \mathbf{x}$ if and only if $d(\mathbf{x}^{(n)}, \mathbf{x}) \to 0$, i.e. if and only if

$$\sqrt{\sum_{i=1}^{m}(x_i^{(n)} - x_i)^2} \to 0,$$

when $n \to \infty$. But, a sum of squares becomes smaller and smaller if and only if any square in the sum becomes smaller and smaller. Thus, we just obtained a part of the following basic result:

THEOREM 50. *(componentwise convergence). 1) A sequence*

$$\{\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, ..., x_m^{(n)})\}$$

of vectors from $\mathbb{R}^m$ is convergent to a vector $\mathbf{x} = (x_1, x_2, ..., x_m)$ if and only if for any $i = 1, 2, ..., m$, the numerical sequence $\{x_i^{(n)}\}$ is convergent to x_i, when $n \to \infty$. 2) A sequence

$$\{\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, ..., x_m^{(n)})\}$$

is a Cauchy sequence in $\mathbb{R}^m$ if and only if any "component" "i", $\{x_i^{(n)}\}$, is a Cauchy sequence in $\mathbb{R}$ for any $i = 1, 2, ..., m$. Since $\mathbb{R}$ is a complete metric space (see Theorem 13), we see that $\mathbb{R}^m$ is also a complete metric space.

PROOF. 1) was just proved before the statement of the theorem. For 2) let us consider a sequence $\{\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, ..., x_m^{(n)})\}$. It is a Cauchy sequence if for any $\varepsilon > 0$ we can find a rank N_ε such that if $n \geq N_\varepsilon$ one has that $d(\mathbf{x}^{(n+p)}, \mathbf{x}^{(n)}) < \varepsilon$ for any $p = 1, 2, ...$. This means that whenever n is large enough the distance $d(\mathbf{x}^{(n+p)}, \mathbf{x}^{(n)})$ is small enough, independent on p. But

$$d(\mathbf{x}^{(n+p)}, \mathbf{x}^{(n)}) = \sqrt{\sum_{i=1}^{m}(x_i^{(n+p)} - x_i^{(n)})^2}. \tag{1.5}$$

So, $\left|x_i^{(n+p)} - x_i^{(n)}\right|$ becomes small enough, independent on p whenever n is large enough. And this is true for any fixed $i = 1, 2...$. But this last remark says that the sequence $\{x_i^{(n)}\}$ is a Cauchy sequence for any fixed $i = 1, 2, ...$. Conversely, if all the sequences $\{x_i^{(n)}\}$ are Cauchy sequences for $i = 1, 2, ...$, then, in (1.5), all the differences

$\left|x_i^{(n+p)} - x_i^{(n)}\right|$ become smaller and smaller, independent of p, whenever n becomes large enough. Hence, the whole sum $\sum_{i=1}^m (x_i^{(n+p)} - x_i^{(n)})^2$ becomes smaller and smaller, independent of p, whenever $n \to \infty$, i.e. the sequence $\{\mathbf{x}^{(n)}\}$ is a Cauchy sequence in $\mathbb{R}^m$. The last statement becomes very easy now (why?). □

For instance, the sequence $\{(\frac{1}{n}, \frac{n+1}{n})\}$ is convergent to $(0,1)$ in $\mathbb{R}^2$ because the first component $\{\frac{1}{n}\}$ goes to 0 and the second component $\frac{n+1}{n}$ goes to 1.

A normed vector space, which is a complete metric space w.r.t. the distance defined by its norm, is called a Banach space. Such spaces are very useful in many engineering models.

We recall now, in our particular case of the metric space $(\mathbb{R}^m, d)$, where d is defined in (1.2), the following basic notion.

DEFINITION 16. *Let* $\mathbf{a} = (a_1, a_2, ..., a_m)$ *be a fixed point in* $\mathbb{R}^m$ *and let* $r > 0$ *be a positive real number. The set* $B(\mathbf{a},r) = \{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x} - \mathbf{a}\| = d(\mathbf{x}, \mathbf{a}) < r\}$ *is called the open ball with centre at* $\mathbf{a}$ *and of radius* r. *The set*

$$B[\mathbf{a},r] = \{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x} - \mathbf{a}\| = d(\mathbf{x}, \mathbf{a}) \leq r\}$$

is said to be the closed ball with centre at $\mathbf{a}$ *and of radius* r (≥ 0).

For instance, if $m = 1$, $\mathbf{a} = a \in \mathbb{R}$ then $B(\mathbf{a},r) = (a - r, a + r)$, the usual open interval with centre at a and of length $2r$ (prove this!). In the same case, $B[\mathbf{a},r] = [a - r, a + r]$. If $m = 2$, $B(\mathbf{a},r)$ is the usual open (without boundary!) disc, with centre at the point $\mathbf{a} = (a_1, a_2)$ and of radius r. If $m = 3$, $B(\mathbf{a},r)$ is the common 3-D open (without boundary) ball (a full sphere!) with centre at $\mathbf{a} = (a_1, a_2, a_3)$ and of radius r. The closed ball $B[\mathbf{a},r]$ is exactly the full sphere of radius r and with centre at $\mathbf{a}$, which contains its boundary

$$S = \{(x, y, z) : (x - a_1)^2 + (y - a_2)^2 + (z - a_3)^2 = r^2\}.$$

This last surface S is usually called the sphere of centre $\mathbf{a}$ and of radius r.

Let D be an arbitrary subset of $\mathbb{R}^m$. A point $\mathbf{d}$ of D is said to be *interior* in D, if there is a small ball $B(\mathbf{d}, r)$, $r > 0$ centered at $\mathbf{d}$ such that $B(\mathbf{d}, r) \subset D$. All the interior points of D is a subset of D denoted by $IntD$, the interior of D. It can be empty. For instance, any finite set of points has an empty interior.

DEFINITION 17. *A subset* D *of* $\mathbb{R}^m$ *is said to be an open subset if for any* $\mathbf{a}$ *in* D *there is a small* $r > 0$ *such that the open ball* $B(\mathbf{a},r)$

with centre at $\mathbf{a}$ *and of radius* r *is completely contained in* D, *i.e.* $B(\mathbf{a},r) \subset D$. *A subset* E *of* $\mathbb{R}^m$ *is said to be closed if its complementary*

$$E^c \stackrel{def}{=} \mathbb{R}^m \setminus E \stackrel{def}{=} \{\mathbf{x} \in \mathbb{R}^m : \mathbf{x} \notin E\}$$

in $\mathbb{R}^m$ *is an open subset of* $\mathbb{R}^m$.

For instance, any point or any finite set of points are closed subsets of $\mathbb{R}^m$. If $m = 1$, the closed intervals are closed subsets of $\mathbb{R}$. Moreover, an open ball is an open set and a closed ball is a closed set (prove it for $m = 1, 2, 3!$). It is not difficult to prove that a subset D of $\mathbb{R}^m$ is open if and only if it is equal to its interior. The *boundary* $\mathcal{B}(D)$ of a subset D of $\mathbb{R}^m$ is by definition the collection of all the points $\mathbf{b}$ of $\mathbb{R}^m$ such that any ball $B(\mathbf{b}, r)$, centered at $\mathbf{b}$ and of radius $r > 0$ has common points with D and with the complementary $\mathbb{R}^m \setminus D$ of D. For instance, the boundary of the disc $\{(x, y) : x^2 + y^2 \leq 1\}$ is the circle $\{(x, y) : x^2 + y^2 = 1\}$ (prove it!). It is easy to see that D is closed if and only if it contains its boundary. The set $D \cup \mathcal{B}(D)$ is called the *closure* of D. It is exactly the union of all the limits of all convergent sequences which have their terms in D.

REMARK 17. *The set* $\mathcal{O}$ *of all the open subsets of* $\mathbb{R}^m$ *has the following basic properties:*

1) $\varnothing$, *the empty set, and the whole set* $\mathbb{R}^m$ *are considered to be in* $\mathcal{O}$.

2) If D_1, D_2, ..., D_k *are in* $\mathcal{O}$, *then their intersection* $\bigcap_{i=1}^{k} D_i$ *is also in* $\mathcal{O}$.

3) If $\{D_\alpha\}$ *is any family of open subsets in* $\mathcal{O}$, *then their union* $\bigcup_\alpha D_\alpha$ *is also in* $\mathcal{O}$, *i.e. it is also open. We propose to the reader to prove all of these properties and to state and prove the analogous properties for the set* $\mathcal{C}$ *of all the closed subsets of* $\mathbb{R}^m$. *Mathematicians say that a collection* $\mathcal{O}$ *of subsets of an arbitrary set* M, *which fulfil the properties 1), 2) and 3) from above, gives rise to a topology on* M. *For instance, in a metric space* (X, d), *the collection* $\mathcal{O}$ *of all the open subsets (the definition is the same like that for* $\mathbb{R}^m$*!) gives rise to the natural topology of a metric space of* X. *A set* M *with a topology* $\mathcal{O}$ *on it (a collection of subsets with the properties 1), 2) and 3)) is called a topological space and we write it as* $(M, \mathcal{O})$. *This notion is the most general notion which can describe a "distance" between two objects in* M. *For instance, if* $(M, \mathcal{O})$ *is a topological space and if* a *is a "point" (an element) of* M, *then an element* b *is said to be "closer" to* a *then the element* c, *if there are two "open" subsets* D *and* F *of* M *such that*

$a, b \in D$, $a, c \in F$ and $D \subset F$. Meditate on this fact in a metric space X, for instance in the usual case $X = \mathbb{R}$.

Now, if (X, d) is a metric space, the definition of an open ball $B(a, r)$ with centre at an element a of X and of radius $r > 0$ is similar to the definition of the same notion in $\mathbb{R}^m$. Namely,

$$B(a, r) = \{x \in X : d(x, a) < r\}.$$

In the same way, a subset D of X is said to be *open* in X if for any $a \in D$ there is an open ball $B(a, r) = \{x \in X : d(x, a) < r\}$, with centre at a and of radius $r > 0$, such that $B(a, r) \subset D$. A subset E of X is called a *closed set* if its complementary $D = X \setminus E$ in X is an open set of X.

THEOREM 51. *(a closeness criterion) A subset E of a metric space (X, d) (in particular of $X = \mathbb{R}^m$) is closed if and only if any sequence $\{x_n\}$ of elements in E, which is convergent to an element x of X, has its limit x also in E.*

PROOF. Let us assume that E is closed and let $\{x_n\}$ be a sequence of elements in E which is convergent to an element x of X. If x were not in E then, since $D = X \setminus E$ is open, we could find a ball $B(x, r)$ with $r > 0$, such that $B(x, r) \subset D$, i.e. $B(x, r) \cap E = \varnothing$, the empty set. But, since $x_n \to x$, i.e. $d(x_n, x) \to 0$, for n large enough, $d(x_n, x) < r$, or $x_n \in B(x, r)$. Since all the terms x_n are in E, we succeeded to find at least one element $x_n \in B(x, r) \cap E = \varnothing$, which is a contradiction. So, x itself must be in E.

Conversely, we suppose now that any sequence of elements of E which is convergent to an element x of X has its limit x in E. If E were not closed, $D = X \setminus E$ were not open. This means that there is at least one element y of D such that any small ball $B(y, \frac{1}{n})$ cannot be contained in D. Hence, for any natural number $n > 0$, one can find at least one element $y_n \in B(y, \frac{1}{n}) \cap E$ (why?). This means that $d(y_n, y) < \frac{1}{n}$ and that $y_n \in E$ for any $n = 1, 2, \ldots$. Since $y_n \to y$ (why?) and since E has the above property, we see that y must be also in E. But,... y was chosen to be in $D = X \setminus E$, so it cannot be in E! We have a new contradiction! So, we cannot suppose that D is not open, i.e. we are forced to say that E is closed and the theorem is completely proved. □

DEFINITION 18. *Let A be a nonempty subset of $\mathbb{R}^m$ (or of an arbitrary metric space (X, d)). By the closure $\overline{A}$ of A in $\mathbb{R}^m$ (or in X) we mean the set of the limits of all the convergent sequences with terms in A.*

In particular, any element a of A is in $\overline{A}$ (take the constant sequence $a, a, a, ...$,etc.). We can easily see that $\overline{A}$ is the least closed subset of X (in particular of $\mathbb{R}^m$) which contains A (use Theorem 51).

REMARK 18. *A is closed if and only if $A = \overline{A}$. The closure of the open ball $B(a, r)$ in a metric space (X, d) is exactly the closed ball $\overline{B}[a, r]$. The operation $A \rightsquigarrow \overline{A}$ has the following main properties: 1) $\overline{A \cap B} \subset \overline{A} \cap \overline{B}$, 2) $\overline{A \cup B} = \overline{A} \cup \overline{B}$, 3) $A \cup \mathcal{B}(A) = \overline{A}$, where $\mathcal{B}(A) = \{x \in X : B(x, r) \cap A \neq \varnothing$ and $B(x, r) \cap (X \setminus A) \neq \varnothing$ for any $r > 0\}$ is the boundary of A in X (prove all these statements!).*

We naturally extend the definition of a limit point for a subset A of $\mathbb{R}$ (see Definition 4) to a subset of an arbitrary metric space (X, d).

Let A be a nonempty subset of a metric space (X, d) (in particular of $\mathbb{R}^m$). An element x of X is said to be a limit point for A if there is a nonconstant sequence $\{x_n\}$ with terms in A which is convergent to x.

For instance, $(0, 0)$ is a limit point for the half-plane $\{(x, y) : y > 0\}$. But $(0, -0.0001)$ is not a limit point for the same subset in $X = \mathbb{R}^2$. The subset $\{(n, m) : n, m \in \mathbb{N}\}$ of $\mathbb{R}^2$ has no limit points. The set of all the limit points of a subset A of a metric space (X, d) together the subset A itself is exactly the closure $\overline{A}$ of A (why?). The set of all the limit points of the closed cube $C = [0, 1] \times [0, 1] \times [0, 1]$ is the cube C itself. But,...the set of all the limit points of an arbitrary closed subset is not always the set itself. For instance, the set of all limit points of a point a of X is the empty set (which is distinct of $\{a\}$). A sequence $\{x_n\}$ has exactly only one limit point x, if and only if the sequence has an infinite distinct values and it is convergent to x.

DEFINITION 19. *A nonempty subset A in a metric space (X, d) is said to be bounded if there is a "reference" element $c \in X$ and a positive real number M such that $d(c, x) < M$ for any element x of A.*

REMARK 19. *It appears that the definition depends on the choice of the "reference" element c, i.e. that the boundedness of A is a c-boundedness. In fact, the definition does not depend on the element c. Namely, if a subset A is bounded relative to an element c of X, it is bounded relative to any other element b of X. Indeed, $d(b, x) \leq d(b, c) + d(c, x) < d(b, c) + M$, which is a fixed positive number w.r.t. the variable element x of A. Hence, A is also b-bounded. In a normed space (see Definition 13) we take as a "reference" element c the element $c = 0$. Thus, A is bounded in a normed space $(X, \|.\|)$ if and only if there is a positive real number M such that $\|x\| < M$ for any x of A.*

Cesaro-Bolzano-Weierstrass Theorem (see Theorem 12) has an extension to $\mathbb{R}^m$ for any $m = 2, 3, ...$.

THEOREM 52. *(Bolzano-Weierstrass Theorem). Let A be a bounded and infinite subset of $\mathbb{R}^m$. Then A has at least one limit point in $\mathbb{R}^m$. In particular, any bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.*

PROOF. To understand easier the idea behind the formal proof of this theorem, we shall take the particular case $m = 2$ (the case $m = 1$ was considered in Theorem 12). So, A is an infinite (contains an infinite number of distinct elements) and bounded subset of $\mathbb{R}^2$. Any element of A is a couple (x, y), where $x, y \in \mathbb{R}$. Since A is bounded by a positive real number M, we can write $\|(x, y)\| \leq M$, for any pair (x, y) of A, or $\sqrt{x^2 + y^2} \leq M$. Thus, the projections of A on the coordinates axes, $A_1 = \{a_1 \in \mathbb{R} : \text{there is an } a_2 \in \mathbb{R} \text{ with } (a_1, a_2) \in A\}$ and $A_2 = \{b_2 \in \mathbb{R} : \text{there is a } b_1 \in \mathbb{R} \text{ with } (b_1, b_2) \in A\}$ are bounded in $\mathbb{R}$ (prove it and make a drawing!). Since A is infinite, at least one of A_1 or A_2 is infinite (why?). We suppose that A_1 is infinite. Let us apply now Cesaro-Bolzano-Weierstrass Theorem (Theorem 12) for the subset A_1 of $\mathbb{R}$. Hence, there is a limit point x_1 for A_1, i.e. there is a sequence $\{x_1^{(n)}\}$ of elements in A_1, which is convergent to x_1. Let us look now at the definition of A_1! For any $x_1^{(n)}$, $n = 1, 2, ...$, we can find an element $x_2^{(n)}$ in $\mathbb{R}$ such that the couple $(x_1^{(n)}, x_2^{(n)})$ is in A. In fact, the sequence $\{x_2^{(n)}\}$ is bounded and its terms belong to A_2 (why?). If A_2 is also infinite, applying again Cesaro-Bolzano-Weierstrass theorem to the subset $\{x_2^{(n)}\}$, we get a limit point x_2 of this last sequence. This means that we can find a subsequence $\{x_2^{(k_n)}\}$ of $\{x_2^{(n)}\}$ ($k_1 < k_2 < ...$) which is convergent to x_2. For any k_n, $n = 1, 2, ...$, we consider the term $x_1^{(k_n)}$ of the sequence $\{x_1^{(n)}\}$ just found above. We obtain a new sequence $\{(x_1^{(k_n)}, x_2^{(k_n)})\}$ of elements from A, which is convergent to the pair (x_1, x_2) (why?...because it is componentwise convergent!). Thus (x_1, x_2) is a limit point of A. What happens if A_2 is finite? Then, at least one term $x_2^{(l)}$ repeats itself of an infinite number of times. We suppose that for $h_1 < h_2 < ...$ one has that $x_2^{(h_n)} = x_2^{(l)}$, for any $n = 1, 2, ...$. So, the sequence $\{(x_1^{(h_n)}, x_2^{(h_n)})\}$, with terms in A, is convergent to $(x_1, x_2^{(l)})$, which becomes in this way a limit point for A. A question can arise here: why can we choose all the elements of the sequence $\{(x_1^{(h_n)}, x_2^{(h_n)})\}$ to be distinct one to each other? Because the sequence $\{x_1^{(n)}\}$ can be chosen from the beginning to contain only distinct elements (A_1 is infinite!). Hence, in both cases A has a limit point and the proof is completed. □

We shall see in future the fundamental importance of this theoretical result. A limit point is also called in the literature an *accumulation* point.

Since the bounded and closed subsets in a space of the form $\mathbb{R}^m$ are very useful in many applications, we shall call them *compact sets*. For instance, $[a,b]$, $\{(x,y) : x^2+y^2 \leq r^2\}$ and, generally, any closed balls, are all compact sets in their corresponding arithmetical spaces of the type $\mathbb{R}^m$. A finite union and any intersection of compact sets is again a compact set (prove it!). An infinite union of compact sets is not always a compact set (find a counterexample!). For instance $D = \{\frac{1}{n}\}$ is bounded but it is not closed because $\frac{1}{n} \to 0$ and 0 is not in D. So, D is not a compact set but,...its closure $\overline{D} = \{0\} \cup \{\frac{1}{n}\}$ is a compact subset in $\mathbb{R}$ (prove this!). Any finite set of points in $\mathbb{R}^m$ is a compact set (why?).

Now we give a useful characterization of compact sets in $\mathbb{R}^m$.

THEOREM 53. *A subset C of $\mathbb{R}^m$ is a compact set if and only if any sequence of C contains a convergent subsequence with its limit in C.*

PROOF. We suppose that C is a compact set in $\mathbb{R}^m$ and let $\{\mathbf{x}^{(n)}\}$ be a sequence with terms in C. If $\{\mathbf{x}^{(n)}\}$ has an infinite number of distinct elements, $A = \{\mathbf{x}^{(n)}\}$ being bounded ($A \subset C$ and C is bounded), we can apply Theorem 52 and find that there is a convergent subsequence $\{\mathbf{x}^{(k_n)}\}$ of $\{\mathbf{x}^{(n)}\}$. Since C is closed, the limit of $\{\mathbf{x}^{(k_n)}\}$ belongs to C (see Theorem 51). If $\{\mathbf{x}^{(n)}\}$ has only a finite number of distinct terms, one of them appears in an infinite number of places. So, we take the constant subsequence generated by it.

Conversely, we assume that C has the property indicated in the statement of the theorem. Let us prove firstly that C is bounded. If it were not bounded, for any $n = 1, 2, ...$ one can find a vector $\mathbf{a}_n$ in C such that $\|\mathbf{a}_n\| > n$. The hypothesis says that the sequence $\{\mathbf{a}_n\}$ has a convergent subsequence $\{\mathbf{a}_{k_n}\}$. Let $\mathbf{a} = \lim\limits_{n\to\infty} \mathbf{a}_{k_n}$ be the limit of the sequence $\{\mathbf{a}_{k_n}\}$. Then

$$k_n < \|\mathbf{a}_{k_n}\| \leq \|\mathbf{a}_{k_n} - \mathbf{a}\| + \|\mathbf{a}\| .$$

Taking limits in the extreme sides of these inequalities, we get: $\infty \leq \|\mathbf{a}\|$, a contradiction. Hence, C must be bounded. Let us prove now that C is closed by using again Theorem 51. For this, let $\{\mathbf{y}_n\} \to \mathbf{y}$ be a convergent to $\mathbf{y}$ sequence with elements in C and its limit $\mathbf{y}$ in $\mathbb{R}^m$. By the hypothesis on C, the sequence $\{\mathbf{y}_n\}$ has a subsequence $\{\mathbf{y}_{k_n}\}$ which is convergent to an element $\mathbf{z}$ of C. Since $\{\mathbf{y}_n\}$ is convergent to $\mathbf{y}$, any subsequence of $\{\mathbf{y}_n\}$ is also convergent to $\mathbf{y}$. Indeed, let us prove

for instance that $\mathbf{z} = \mathbf{y}$. For this, let us evaluate $d(\mathbf{z}, \mathbf{y})$, the distance between $\mathbf{z}$ and $\mathbf{y}$:

$$d(\mathbf{z}, \mathbf{y}) \leq d(\mathbf{z}, \mathbf{y}_{k_m}) + d(\mathbf{y}_{k_m}, \mathbf{y}_n) + d(\mathbf{y}_n, \mathbf{y}), \tag{1.6}$$

where m and n are arbitrary chosen. If we make $m, n \to \infty$ in this last inequality, we get that $d(\mathbf{z}, \mathbf{y}) = 0$, i.e. $\mathbf{z} = \mathbf{y}$ (why?). Here we just used the fact that a convergent sequence is also a Cauchy sequence, i.e. for m, n large enough, the distance $d(\mathbf{y}_m, \mathbf{y}_n)$ goes to zero. Now, since $\mathbf{z}$ is in C we get that $\mathbf{y}$ is also in C, i.e. C is closed and the theorem is proved. □

The above characterization of compact subsets of $\mathbb{R}^m$ leads us to the introduction of the notion of a compact subset in an arbitrary metric space (X, d). We say that a subset C of X is *compact* if any sequence of elements from C has a subsequence which is convergent to an element of C.

For instance, any convergent sequence $\{x_n\}$ in a metric space X, together with its limit x is a compact subset of X (prove it!). Thus, $C = \{x_n\} \cup \{x\}$ is a compact subset of X.

2. Continuous functions of several variables

Let A be a nonempty subset of $\mathbb{R}^n$, the "arithmetical" n-dimensional vector space and let $f : A \to \mathbb{R}$, be a function defined on A with values in $\mathbb{R}$. Since the variable $\mathbf{x} = (x_1, x_2, ..., x_n)$ is a vector determined by n free scalar quantities, $x_1, x_2, ..., x_n$, we say that our function is *a function of n variables*. If $n \geq 2$, we say that f is a function of *"several" variables*. Since the values of f are scalars (real numbers), we say that f is a *scalar function of n variables*. A map $\mathbf{f} : A \to \mathbb{R}^m$ is called a *vector function of n variables*. This time, the values of $\mathbf{f}$ are m-dimensional vectors. Hence $\mathbf{f}(\mathbf{x}) = (y_1, y_2, ..., y_m)$ and we see that the numbers $y_1, y_2, ..., y_m$ are themselves functions f_1, f_2,..., f_m of $\mathbf{x}$: $y_1 = f_1(\mathbf{x}), ..., y_m = f_m(\mathbf{x})$. These scalar functions f_1, f_2, ..., f_m, defined on A with values in $\mathbb{R}$ this time, are called the components of $\mathbf{f}$. We write this as: $\mathbf{f} = (f_1, f_2, ..., f_m)$ and interpret it as a "vector" of m-components (coordinates) $f_1, f_2, ..., f_m$. In applications $\mathbf{f}$ is also called a *vector field of n variables*. "Field" comes from "field of forces". For instance,

$$\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2, \mathbf{f}(x, y) = (xy, x - y)$$

is a vector field in plane ($\mathbb{R}^2$) of 2 variables. Its components are $f_1(x, y) = xy$ and $f_2(x, y) = x - y$. We can give its image in some points. For instance, we can translate the vector $\mathbf{f}(2, 3) = (2 \cdot 3, 2 - 3) = (6, -1)$ at the point $(2, 3)$ and so we get "the image" of $\mathbf{f}$ at $(2, 3)$. In this way

we can fill the whole plane $\mathbb{R}^2$ with vectors (forces), i.e. we get a "field" of forces on the whole plane. If $n = 1$, the image of a vector field $\mathbf{f} : A \to \mathbb{R}^m$ ($A \subset \mathbb{R}$) is a "curve" in $\mathbb{R}^m$. For instance, $\mathbf{f}(t) = (R\cos t, R\sin t)$, $t \in [0, 2\pi)$ has as image in the plane $\mathbb{R}^2$ the usual circle of radius R and with centre at the origin $(0, 0)$. We say that the two components of $\mathbf{f}$, $f_1(t) = R\cos t$ and $f_2(t) = R\sin t$ are the parametric equations of this circle. One also write this as: $x = R\cos t$, $y = R\sin t$, $t \in [0, 2\pi)$. We can also interpret the image of a vector field $\mathbf{f} : [0, T] \to \mathbb{R}^m$ ($m = 2$ or $m = 3$) as the *trajectory* of a moving point

$$M(f_1(t), f_2(t), ..., f_m(t))$$

where t measures the "time" between the starting moment (usually $t = 0$) and the ending moment $t = T$. For instance, $\mathbf{f}(t) = (t, t^2)$, $t \in A = [0, 10]$, is a parabolic trajectory, along the arc of the parabola $y = x^2$, $x \in [0, 10]$. The new vector field

$$\mathbf{f}'(t) = (f_1'(t), f_2'(t), ..., f_m'(t))$$

(the componentwise derivative), associated to the vector field

$$\mathbf{f}(t) = (f_1(t), f_2(t), ..., f_m(t)), t \in [0, T],$$

is called the *velocities field* of the field $\mathbf{f}$.

In order to describe the "breaking" phenomena at a given point $\mathbf{a} =(a_1, a_2, ..., a_n)$ of $\mathbb{R}^n$, we need to see what happens with the values of a vector function (which describes our phenomenon) $\mathbf{f} : A \to \mathbb{R}^m$, whenever we becomes closer and closer to $\mathbf{a}$. For this, $\mathbf{a}$ must be a limit point of the definition domain A. We have to study the convergence of the sequence of vectors $\{\mathbf{f}(\mathbf{x}^{(n)})\}$ in $\mathbb{R}^m$, whenever the sequence $\{\mathbf{x}^{(n)}\}$, with terms in A, converges to $\mathbf{a}$ in the metric space $\mathbb{R}^n$. The most convenient situation is that when all the values $\{\mathbf{f}(\mathbf{x}^{(n)})\}$, for all the sequences $\{\mathbf{x}^{(n)}\}$, which are convergent to $\mathbf{a}$, become closer and closer to one and the same vector $\mathbf{L}$ from $\mathbb{R}^m$. This is why we give now the following definition.

DEFINITION 20. *Let A be a subset of $\mathbb{R}^n$ and let $\mathbf{a} =(a_1, a_2, ..., a_n)$ be a limit point of A. We say that $\mathbf{L} \in \mathbb{R}^m$ is the limit of a vector function $\mathbf{f} : A \to \mathbb{R}^m$ at the point $\mathbf{a}$ (write $\mathbf{L} = \lim\limits_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x})$), if for every sequence $\{\mathbf{x}^{(n)}\}$, $\mathbf{x}^{(n)} \neq \mathbf{a}$, $\mathbf{x}^{(n)} \in A$, which is convergent to the vector $\mathbf{a}$, one has that the sequence of images $\{\mathbf{f}(\mathbf{x}^{(n)})\}$ of $\{\mathbf{x}^{(n)}\}$ through $\mathbf{f}$ is convergent to $\mathbf{L}$. If such an $\mathbf{L}$ exists, independently on the choice of the sequence $\{\mathbf{x}^{(n)}\}$, we say that $\mathbf{f}$ has limit $\mathbf{L}$ at $\mathbf{a}$. This limit $\mathbf{L}$ depends only on $\mathbf{f}$ and on $\mathbf{a}$.*

If there is such a common limit $\mathbf{L}$, this is unique, because the limit of a sequence in a metric space is unique (if it exists!).

For instance, let us compute $\lim\limits_{(x,y)\to(-1,2)} f(x,y)$, where

$$f(x,y) = xy + x^2 + \ln(x^2+y^2).$$

Let us take a sequence $\{(x_n,y_n)\}$ which is convergent to $(-1,2)$. This means that $x_n \to -1$ and $y_n \to 2$ (see Theorem 50). But we know that the "taking limit" operation is compatible with the multiplication, addition and with the logarithm function (we say that ln is continuous!) (see also Theorem 14). Hence,

$$f(x_n,y_n) = x_n y_n + x_n^2 + \ln(x_n^2+y_n^2)$$

will be convergent to

$$(-1)\cdot 2 + (-1)^2 + \ln((-1)^2+2^2) = -1 + \ln 5.$$

We see that this limit is independent on the starting sequence (x_n,y_n) which tends to $(-1,2)$. Thus, for any sequence (x_n,y_n) which is convergent to $(-1,2)$,

$$\lim_{(x_n,y_n)\to(-1,2)} f(x_n,y_n) = -1 + \ln 5.$$

In fact, we see that for any sequence (x_n,y_n) which is convergent to $(-1,2)$,

$$\lim_{(x_n,y_n)\to(-1,2)} f(x_n,y_n) = f(-1,2).$$

This happens, because any elementary function of several variables is "continuous" (see the bellow definition) on its definition domain.

DEFINITION 21. *Let A be a subset of $\mathbb{R}^n$ and let $\mathbf{a}=(a_1,a_2,...,a_n)$ be a point of A. We say that the vector function $\mathbf{f}: A \to \mathbb{R}^m$ is continuous at the point $\mathbf{a}$, if for every sequence $\{\mathbf{x}^{(n)}\}$ of A, $\mathbf{x}^{(n)} \neq \mathbf{a}$ and which is convergent to the vector $\mathbf{a}$, one has that the sequence of the images $\{\mathbf{f}(\mathbf{x}^{(n)})\}$ of $\{\mathbf{x}^{(n)}\}$ through $\mathbf{f}$ is convergent to $\mathbf{f}(\mathbf{a})$, the value of $\mathbf{f}$ at $\mathbf{a}$. We say that $\mathbf{f}$ is continuous on the set A if $\mathbf{f}$ is continuous at any point of A.*

We see that $\mathbf{f}$ is continuous at a point $\mathbf{a}$ if and only if it has a limit $\mathbf{L}$ at $\mathbf{a}$ and this $\mathbf{L}$ is equal to $\mathbf{f}(\mathbf{a})$, the value of $\mathbf{f}$ at the point $\mathbf{a}$. The above definition is in accordance with the engineers perception of approximation processes. Let us suppose that $\mathbf{f}$ describes a physical phenomenon P and we are interested in the variation of this phenomenon around a fixed "point" (vector) $\mathbf{a}$. Let us take a neighboring point $\mathbf{z}$ of $\mathbf{a}$ and let us approximate $\mathbf{z}$ by $\mathbf{a}$. In this case, can we approximate $\mathbf{f}(\mathbf{z})$ by $\mathbf{f}(\mathbf{a})$? Or, can we consider that P is "almost the same" at $\mathbf{z}$ like

at **a**?. We can do this if **f** is continuous at **a**. Otherwise, we cannot do such approximations. We must be very careful for instance, in the case of earthquake models around the so called "singular" points (see the example bellow). Now we think that the reader is convinced that the continuity notion is important in modelling the physical phenomena. It is not difficult to prove that all the elementary functions and their compositions are continuous functions. In the following we supply with an example in which we shall see that the case of vector fields of several variables (for $n > 1$) is more complicated then the case of one variable. Let us see now if the following nonelementary (why?) function

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2}, \text{ if } x \neq 0, \text{ or } y \neq 0, \\ 0, \text{ if } x = 0 \text{ and } y = 0, \end{cases}$$

$f : \mathbb{R}^2 \to \mathbb{R}$, is continuous or not on the whole $\mathbb{R}^2$. If $(a,b) \neq (0,0)$, then $f(x,y) = \frac{xy}{x^2+y^2}$ on a small disc (not containing $(0,0)$) with centre at (a,b) (and a small radius). Since the restriction of f to this last disc is an elementary function, f is continuous at (a,b). What happens at $(0,0)$? If the function f were continuous at $(0,0)$ then, for any sequence (x_n, y_n) which tends to $(0,0)$ (i.e. $x_n \to 0$ and $y_n \to 0$), we should have that $f(x_n, y_n) \to f(0,0) = 0$. Let us take a nonzero real number r and let $\{x_n\}$ be an arbitrary sequence of nonzero real numbers which is convergent to 0. Take now $y_n = rx_n$ for any $n = 1, 2, ...$ This means that all the pairs (x_n, y_n) are on the line $y = rx$ (its slope is r) and that the sequence $\{(x_n, y_n)\}$ is convergent to $(0,0)$. But

$$f(x_n, y_n) = \frac{rx_n^2}{x_n^2 + r^2 x_n^2} = \frac{r}{1+r^2} \neq 0.$$

So the function f is not continuous at $(0,0)$. Moreover, since the limit

$$\lim_{(x_n,y_n)\to(0,0)} f(x_n, y_n) = \frac{r}{1+r^2}$$

is dependent on the slope r of the line $y = rx$, on which we have chosen our sequence (x_n, y_n), we see that the function f has no limit at $(0,0)$. Hence, we cannot extend f "by continuity" at $(0,0)$ with no real value. Such a point $(0,0)$ is called an *essential singular point for* f. This means that if we become closer and closer to $(0,0)$ on different sequences $\{(x_n, y_n)\}$, we obtain an infinite number of distinct values for the limit $\lim_{(x_n,y_n)\to(0,0)} f(x_n, y_n)$ (as we just saw above!).

The following criterion reduces the study of the limit or of the continuity of a vector function $\mathbf{f} : A \to \mathbb{R}^m$ at a point $\mathbf{a} \in A$, where A is an open subset of $\mathbb{R}^m$ and $\mathbf{f} = (f_1, f_2, ..., f_m)$, to the study of the same properties for the scalar functions $f_1, f_2, ..., f_m$.

THEOREM 54. *With these last notation, 1)* $\mathbf{f} = (f_1, f_2, ..., f_m)$ *has the limit* $\mathbf{L} = (L_1, L_2, ..., L_m)$ *at the point* $\mathbf{a}$ *if and only if every component function* f_j *has the limit* L_j *at the same point* $\mathbf{a}$, *for* $j = 1, 2, ...$ *and 2)* $\mathbf{f}$ *is continuous at the point* $\mathbf{a}$ *if and only if every component function* f_j *is continuous at* $\mathbf{a}$.

PROOF. Everything comes from the fact that the convergence in the normed spaces $\mathbb{R}^m$ is a componentwise convergence (see Theorem 50). Indeed, let us assume that $\mathbf{f} = (f_1, f_2, ..., f_m)$ has the limit $\mathbf{L} = (L_1, L_2, ..., L_m)$ at $\mathbf{a}$. Hence, for any sequence $\{(\mathbf{x}^{(n)})\}$ which is convergent to $\mathbf{a}$, one gets that $\lim \mathbf{f}(\mathbf{x}^{(n)}) = \mathbf{L}$, i.e. $\lim f_j(\mathbf{x}^{(n)}) = L_j$ for $j = 1, 2, ...$ (we just applied the "componentwise" principle). The existence is included here! (why?). Conversely, if for any $j = 1, 2, ...$, the limit $\lim f_j(\mathbf{x}^{(n)}) = L_j$ exists, then the limit $\lim \mathbf{f}(\mathbf{x}^{(n)}) = \mathbf{L}$ exists and $\mathbf{L} = (L_1, L_2, ..., L_m)$. We add the fact that $\mathbf{f} = (f_1, f_2, ..., f_m)$ is continuous at $\mathbf{a}$ if and only if

$$\mathbf{L} = (L_1, L_2, ..., L_m) = \mathbf{f}(\mathbf{a}) = (f_1(\mathbf{a}), f_2(\mathbf{a}), ..., f_m(\mathbf{a})),$$

or if and only if $f_j(\mathbf{a}) = L_j$ for any $j = 1, 2, ...$. But this means exactly the continuity of every f_j at $\mathbf{a}$ for $j = 1, 2, ...$. □

Using this last continuity test, we can easily decide if a vector function is continuous or not. For instance,

$$\mathbf{f}(x, y, z) = (x, 2x + y, 2x + 3y - 2z)$$

is continuous on $\mathbb{R}^3$ because all the scalar component functions

$$f_1(x, y, z) = x, f_2(x, y, z) = 2x + y$$

and $f_3(x, y, z) = 2x + 3y - 2z$ are polynomial functions so, they are all continuous on $\mathbb{R}^3$.

REMARK 20. *The existence of a limit at a point and the continuity at a point are "local" properties. They are defined "around" a given point* $\mathbf{a}$. *If we fix a n-D continuous curve* $\gamma : [a, b] \to A \subset \mathbb{R}^n$ *and if* $\mathbf{a} = \gamma(t_0)$ *is a point "on* γ*" (it is in the image of* γ*), we say that a vector function* $\mathbf{f} = (f_1, f_2, ..., f_m)$, *defined on A with values in* $\mathbb{R}^m$ *is continuous at* $\mathbf{a}$ *along the curve* γ *if the composed function* $\mathbf{f} \circ \gamma : [a, b] \to \mathbb{R}^m$ *(a new curve in* $\mathbb{R}^m$*) is continuous at* t_0. *This means that if we take any sequence of points* $\{\mathbf{x}^{(n)}\}$ *in A (is considered to be opened!) on* γ *(*$\mathbf{x}^{(n)} = \gamma(t_n)$*), which becomes closer and closer to* $\mathbf{a}$, *then* $\lim \mathbf{f}(\mathbf{x}^{(n)}) = \mathbf{f}(\mathbf{a})$. *For instance,*

$$f(x, y) = \begin{cases} \frac{xy}{x^2+y^2}, \text{ if } x \neq 0, \text{ or } y \neq 0, \\ 0, \text{ if } x = 0 \text{ and } y = 0, \end{cases}$$

$f : \mathbb{R}^2 \to \mathbb{R}$, is not continuous at $\mathbf{a} = (0,0)$, but it is continuous at $(0,0)$ along the both axes of coordinates. It has limits along any other fixed line $y = rx$ which is passing through $(0,0)$, but the limits are not the same! (see the above commentaries on this example). It is possible to construct a function of two variables which is continuous on $\mathbb{R}^2$ except the origin, where it has limit 0 along any line which passes through $(0,0)$, but it has no limit at $(0,0)$ (find such a function!).

THEOREM 55. *The composition between two continuous functions is also a continuous function.*

PROOF. Let A be an open subset of $\mathbb{R}^p$, let B be another open subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to B$, $\mathbf{g} : B \to \mathbb{R}^m$ be two continuous functions on their definition domains. The theorem says that the composed function $\mathbf{h} : A \to \mathbb{R}^m$, $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$, i.e. $\mathbf{h}(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x}))$ for any $\mathbf{x} \in A$, is also a continuous function on A. For proving this, let us take a point $\mathbf{a} \in A$ and an arbitrary sequence $\{\mathbf{x}^{(n)}\}$ in A which is convergent to $\mathbf{a}$ w.r.t. the distance of $\mathbb{R}^p$. Since $\mathbf{f}$ is continuous on A, in particular, it is also continuous at $\mathbf{a}$. So, the sequence $\{\mathbf{f}(\mathbf{x}^{(n)})\}$ is convergent to $\mathbf{f}(\mathbf{a})$. Now, since $\mathbf{g}$ is continuous on B, in particular, it is continuous at the point $\mathbf{f}(\mathbf{a})$ of B. Hence, the sequence $\{\mathbf{g}(\mathbf{f}(\mathbf{x}^{(n)}))\}$ tends to $\mathbf{g}(\mathbf{f}(\mathbf{a})) = \mathbf{h}(\mathbf{a})$ and so, $\mathbf{h}(\mathbf{x}^{(n)}) = \mathbf{g}(\mathbf{f}(\mathbf{x}^{n}))$ is convergent to $\mathbf{h}(\mathbf{a})$. This means that the composed function $\mathbf{h}$ is continuous at $\mathbf{a}$. Since $\mathbf{a}$ was arbitrary chosen in A, we have that $\mathbf{h}$ is continuous on the whole A. □

This theorem is very useful, because almost all the functions commonly used in applications are compositions of elementary functions and these last ones are continuous on their definitions domains. For instance,

$$f(x,y) = \cos\left[\frac{x + \sin xy}{1 + \ln(x^2 + y^2)}\right]$$

is defined on $\mathbb{R}^2 \backslash \gamma$, where γ is the circle: $x^2 + y^2 = \frac{1}{e}$, where $e = 2.71...$. Here f is the composition between the following continuous functions:

$$x \rightsquigarrow \cos x, (x,y) \rightsquigarrow \frac{x}{y}, y \neq 0, (x,y) \rightsquigarrow x + y, (x,y) \rightsquigarrow xy,$$

$$x \rightsquigarrow \sin x \text{ and } x \rightsquigarrow \ln x,\ x > 0$$

(prove everything slowly!). The same theorem is used to prove that the set of all continuous functions defined on the same set A (open, closed, etc.) is a real infinite dimensional (contains polynomials!) vector space (prove it!).

3. Continuous functions on compact sets

Let A be an arbitrary nonempty subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to \mathbb{R}^m$ be a continuous function (on the whole A). Let D be an open subset of $\mathbb{R}^n$ which is contained in A. Here is a question: "Is always the image $\mathbf{f}(D)$ of D through $\mathbf{f}$ open in $\mathbb{R}^m$? We shall see by simple examples that the answer is no! Let us take, for instance, $D = (0,1)$ and $f(x) = 3$ for any x in $(0,1)$. Since the set $\{3\}$ is closed in $\mathbb{R}$ (why?), $f(D)$ is not open. Let now E be an open subset of $\mathbb{R}^m$ and $\mathbf{f}^{-1}(E) = \{\mathbf{x} \in A : \mathbf{f}(\mathbf{x}) \in A\}$, the preimage of E in A. We say that a subset B of A *is open in* A if it is the intersection between A and an open subset D of $\mathbb{R}^n$, i.e. $B = A \cap D$. For instance, $B = (0,1]$ is not open in $\mathbb{R}$ (why?), but it is open in $A = [-1,1]$ because, $D = (0,3)$, which is open in $\mathbb{R}$, intersected with A is exactly B.

THEOREM 56. *With the definitions and notation given above,* $\mathbf{f} : A \to \mathbb{R}^m$ *is continuous if and only if* $\mathbf{f}^{-1}(E)$ *is open in* A *for any open subset of* $\mathbb{R}^m$, *i.e. if* $\mathbf{f}$ *carries back the open subsets of* $\mathbb{R}^m$ *into open subsets of* A.

PROOF. a) We assume that $\mathbf{f} : A \to \mathbb{R}^m$ is continuous and that E is an open subset of $\mathbb{R}^m$. To prove that $\mathbf{f}^{-1}(E)$ is open in A it is equivalent to prove that $C = A \backslash \mathbf{f}^{-1}(E)$ is closed in A, i.e. for any convergent sequence $\{\mathbf{x}^{(n)}\}$ of elements in C, convergent to an element $\mathbf{x}$ of A (pay attention!), one has that $\mathbf{x}$ is also in C. If it were not in C, $\mathbf{f}(\mathbf{x}) \in E$. Since E is open in $\mathbb{R}^m$, there is a small ball $B(\mathbf{f}(\mathbf{x}),r)$, with center at $\mathbf{f}(\mathbf{x})$ and of radius $r > 0$, which is contained in E. Since $\mathbf{x}^{(n)} \to \mathbf{x}$, and since $\mathbf{f}$ is continuous, one has that $\mathbf{f}(\mathbf{x}^{(n)})$ is convergent to $\mathbf{f}(\mathbf{x})$. So, there is at least one $\mathbf{x}^{(n_0)}$ with $\mathbf{f}(\mathbf{x}^{(n_0)})$ in $B(\mathbf{f}(\mathbf{x}),r)$, i.e. in E. So, $\mathbf{x}^{(n_0)}$ is in $\mathbf{f}^{-1}(E)$, a contradiction, because we have chosen the sequence $\{\mathbf{x}^{(n)}\}$ to have all its terms in C, i.e. not in $\mathbf{f}^{-1}(E)$.

b) We suppose now that $\mathbf{f}$ carries back the open subsets of $\mathbb{R}^m$ into open subsets of A. Let us prove that $\mathbf{f}$ is continuous at an arbitrary fixed point $\mathbf{z}$. For this, let $\{\mathbf{z}^{(n)}\}$ be a sequence in A which is convergent to $\mathbf{z} \in A$. We assume that $\{\mathbf{f}(\mathbf{z}^{(n)})\}$ is not convergent to $\mathbf{f}(\mathbf{z})$. Then, there is a small ball $B(\mathbf{f}(\mathbf{z}),r)$ in $\mathbb{R}^m$ such that an infinite number $\{\mathbf{f}(\mathbf{z}^{(k_n)})\}$, $n = 1, 2, ...$, of the terms of the sequence $\{\mathbf{f}(\mathbf{z}^{(n)})\}$ are outside of $B(\mathbf{f}(\mathbf{z}),r)$. Since $B(\mathbf{f}(\mathbf{z}),r)$ is an open subset in $\mathbb{R}^m$, following the last hypothesis, we get that the set $D = \mathbf{f}^{-1}(B(\mathbf{f}(\mathbf{z}),r))$ is an open subset of A which contains $\mathbf{z}$ (why?). Let $B(\mathbf{z},r')$, $r' > 0$ be a small ball with centre in $\mathbf{z}$ such that $G = B(\mathbf{z},r') \cap A \subset D$ (since D is open in A). All the terms of the subsequence $\{\mathbf{z}^{(k_n)}\}$ are not in G, in particular they are not in $B(\mathbf{z},r')$. But this last conclusion contradicts the fact that

$\mathbf{z}^{(n)} \to \mathbf{z}$. Thus, our assumption that $\{\mathbf{f}(\mathbf{z}^{(n)})\}$ is not convergent to $\mathbf{f}(\mathbf{z})$ is false and so, $\mathbf{f}$ is continuous at $\mathbf{z}$. Since this $\mathbf{z}$ was arbitrary chosen, we get that $\mathbf{f}$ is continuous at all the points of A. □

The following result is very useful in many situations of this course. It appears as a direct consequence of the above theorem.

THEOREM 57. *Let A be an open subset of $\mathbb{R}^n$, let $\mathbf{a}$ be a fixed point of A and let $f : A \to \mathbb{R}$ be a continuous function on A such that $f(\mathbf{a}) > 0$. Then there is an open ball $B(\mathbf{a}, r) \subset A$, $r > 0$, with the property that $f(\mathbf{x}) > 0$ for every $\mathbf{x}$ in $B(\mathbf{a}, r)$.*

PROOF. Take $\varepsilon > 0$ such that $f(\mathbf{a}) - \varepsilon > 0$ and take the open subset $Y = (f(\mathbf{a}) - \varepsilon, f(\mathbf{a}) + \varepsilon)$ of $\mathbb{R}$. Since f is continuous, $X = f^{-1}(Y)$ is an open subset of A which contains $\mathbf{a}$. So, there is a small ball $B(\mathbf{a}, r)$ such that $B(\mathbf{a}, r) \subset X$, i.e. $f(\mathbf{x}) \in Y$ for any $\mathbf{x}$ in $B(\mathbf{a}, r)$. But, for such $\mathbf{x}$ we have that $f(\mathbf{x}) > f(\mathbf{a}) - \varepsilon > 0$ and the proof is done. □

REMARK 21. *In the same way one can prove that $\mathbf{f} : A \to \mathbb{R}^m$ is continuous if and only if $\mathbf{f}$ carries back the closed subsets of $\mathbb{R}^m$ into closed subsets of A (define this notion by analogy!). To prove this, one can use the last theorem 56.*

Not always a continuous function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ carries a closed set of $\mathbb{R}^n$ in a closed set of $\mathbb{R}^m$. For instance, $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{1+x^2}$, carries the closed set $[0, \infty)$ into $(0, 1]$, which is not closed more. It is interesting to see that the closed set $[0, \infty)$ in unbounded. If one tries to substitute it with a closed and bounded interval, for the same function, we shall not succeed at all to find like an image a non closed set! Why? Because of the following basic result:

THEOREM 58. *Let C be a compact (closed and bounded) subset of $\mathbb{R}^n$ and let $\mathbf{f} : C \to \mathbb{R}^m$ be a continuous function. Then, the image $\mathbf{f}(C)$ of C, in $\mathbb{R}^m$, is also a compact subset there (in $\mathbb{R}^m$). Moreover, if $m = 1$, $\sup \mathbf{f}(C) = \mathbf{f}(\mathbf{z}_M)$ and $\inf \mathbf{f}(C) = \mathbf{f}(\mathbf{z}_m)$, where $\mathbf{z}_M$, $\mathbf{z}_m$ are in C.*

PROOF. We need to prove that: a) $\mathbf{f}(C)$ is bounded and, b) $\mathbf{f}(C)$ is closed. The ideas used for proving this theorem are exactly the same like those used in the particular case ($m = 1, n = 1$) of Theorem 32. We take them again here.

a) We assume that $\mathbf{f}(C)$ is not bounded. This means that for every $n = 1, 2, ...$, one can find a point $\mathbf{x}^{(n)}$ in C such that $\left\|\mathbf{f}(\mathbf{x}^{(n)})\right\| > n$ (why?). Since C is a compact subset in $\mathbb{R}^n$, we can find a convergent subsequence $\{\mathbf{x}^{(k_n)}\}$ to the point $\mathbf{x}$ of C (see Theorem 53). Since

$\mathbf{f} : C \to \mathbb{R}^m$ is continuous, the sequence $\{\mathbf{f}(\mathbf{x}^{(k_n)})\}$ is convergent to $\mathbf{f}(\mathbf{x})$. But $\left\|\mathbf{f}(\mathbf{x}^{(k_n)})\right\| > k_n$ and $k_n \to \infty$, so, the numerical sequence $\{\left\|\mathbf{f}(\mathbf{x}^{(k_n)})\right\|\}$ is unbounded (goes to ∞!). We shall see that this is a contradiction. Indeed,

$$\left\|\mathbf{f}(\mathbf{x}^{(k_n)})\right\| \leq \left\|\mathbf{f}(\mathbf{x}^{(k_n)}) - \mathbf{f}(\mathbf{x})\right\| + \|\mathbf{f}(\mathbf{x})\|.$$

If we take limits in this last inequality, we get: $\infty \leq 0 + \|\mathbf{f}(\mathbf{x})\|$, which is not possible! The contradiction appeared because we supposed that $\mathbf{f}(C)$ is unbounded. Hence, it is bounded, i.e. we just proved a).

b) We use now the closeness test (Theorem 51) for proving that $\mathbf{f}(C)$ is closed. Let us take for this a convergent sequence $\{\mathbf{f}(\mathbf{y}^{(n)})\}$, with terms in $\mathbf{f}(C)$ and with its limit $\mathbf{c}$ in $\mathbb{R}^m$. We have to prove that this $\mathbf{c}$ is also in $\mathbf{f}(C)$. Since C is a compact subset of $\mathbb{R}^n$, there is a subsequence $\{\mathbf{y}^{(h_n)}\}$ of the sequence $\{\mathbf{y}^{(n)}\}$ such that $\mathbf{y}^{(h_n)}$ is convergent to $\mathbf{y} \in C$. Since $\mathbf{f}$ is continuous, the sequence $\{\mathbf{f}(\mathbf{y}^{(h_n)})\}$ is convergent to $\mathbf{f}(\mathbf{y})$. But any subsequence of a convergent sequence is also convergent to the same limit of the whole sequence. Thus, $\mathbf{c} = \mathbf{f}(\mathbf{y})$ and so, $\mathbf{c} \in \mathbf{f}(C)$, what we wanted to prove. The other statements can be proved exactly in the same manner (see also Theorem 32). □

Let us give a nice application to this last result. We can assume that the surface of the Earth is closed and bounded in the 3-D space $\mathbb{R}^3$ (why?-you can take it for easy to be $S = \{(x, y, z) : x^2 + y^2 + z^2 = R^2\}$, ...a sphere of radius R, etc.; prove that S is closed and bounded!). At a fixed moment, to any point $M(x, y, z)$ from the Earth we associate its temperature $T(x, y, z)$ at that moment. Thus, we obtain a continuous function T defined on the compact surface of the Earth, with values in $\mathbb{R}$. Applying the above theorem, we always can find two points on the Earth in which the temperatures are extreme.

Let C be a compact (closed and bounded) subset of $\mathbb{R}^n$ and let $\mathbf{f} : C \to \mathbb{R}^m$ be a continuous function. Then, the norm $\|\mathbf{f}(C)\|$ of the image $\mathbf{f}(C)$ of C, in $\mathbb{R}$, is also a compact subset there (in $\mathbb{R}$). Moreover, $\sup \|\mathbf{f}(C)\| = \|\mathbf{f}(\mathbf{z})\|$ and $\inf \|\mathbf{f}(C)\| = \|\mathbf{f}(\mathbf{y})\|$, where $\mathbf{z}$ and $\mathbf{y}$ are in C. Firstly, the function

$$\mathbf{g} : \mathbb{R}^m \to \mathbb{R}, \mathbf{g}(\mathbf{x}) = \|\mathbf{x}\|,$$

is a continuous function. Indeed, let $\{\mathbf{x}^{(n)}\}$ be a sequence in $\mathbb{R}^m$, which is convergent to $\mathbf{x}$. Since $\left|\left\|\mathbf{x}^{(n)}\right\| - \|\mathbf{x}\|\right| \leq \left\|\mathbf{x}^{(n)} - \mathbf{x}\right\|$, we see that the sequence $\{\mathbf{g}(\mathbf{x}^{(n)}) = \{\left\|\mathbf{x}^{(n)}\right\|\}$ is convergent to $\|\mathbf{x}\|$, i.e. $\mathbf{g}$ is continuous. Secondly, let us consider the composition $\mathbf{g} \circ \mathbf{f} : C \to \mathbb{R}$ between the

continuous functions **f** and **g**. It is a continuous function (see Theorem 55) and we can apply the last theorem (do it slowly!).

REMARK 22. *The condition on the closeness of C in the above theorem (Theorem 58) is necessary as one can see in the example: $f:(0,1]\to\mathbb{R}$, $f(x)=\frac{1}{x}$; this function is continuous (prove it!), the interval $(0,1]$ is bounded, nonclosed and the image $f((0,1])=[1,\infty)$ is not bounded, so not a compact subset of $\mathbb{R}$. If C is closed but not bounded, its image through a continuous function f may be nonclosed and nonbounded at the same time. For instance, $C=[1,\infty)$, $f(x)=\frac{1}{x-1}$, so, $f(C)=(0,\infty)$, which is neither closed (it is open in $\mathbb{R}$), nor bounded. This theorem above is not true in general metric spaces. Because a compact subset C in a general metric space (X,d) is defined "by sequences". Namely, C is a compact subset of (X,d) if any sequence in C has a convergent subsequence with its limit also in C. This is not generally equivalent to "bounded and closed". The examples are two "exotic" and we do not give them here. In a metric space (X,d) we can introduce the "distance" between two compact subsets A and B of X. Namely,*

$$dist(A,B)=\inf\{d(a,b):a\in A,b\in B\}.$$

Since d is a continuous function this number $dist(A,B)$ is always finite and it is realized, i.e. there are a_0 in A and b_0 in B such that $dist(A,B)=d(a_0,b_0)$. For instance, the distance between the full square $A=[0,1]\times[1,2]$ and the disc $B=\{(x,y):(x-2)^2+y^2\le 1$ is $\sqrt{2}-1$ and it is realized at $a_0=(1,1)\in A$ and at $b_0=(2-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})$ (why?). It is easy to prove that the distance between two compact subsets A and B is realized on their boundaries (which are also compact subsets), i.e.

$$dist(A,B)=dist(\mathcal{B}(A),\mathcal{B}(B)).$$

Can you organize the set of all compact subsets of X as a metric space (with the distance function defined above)?

In practice, the above Theorem 58 can be applied to optimization problems. For instance, let us find the maximal and the minimal values of the function $f:[0,1]\times[0,2]\to\mathbb{R}$, $f(x,y)=x^4+y^4$. Since $C=[0,1]\times[0,2]$ is a compact subset in $\mathbb{R}^2$ (prove it!), Theorem 58 implies that its image is a compact subset of $\mathbb{R}$. So, $\sup f(C)=f(\mathbf{a})$ and $\inf f(C)=f(\mathbf{b})$. It is easy to see that $\mathbf{a}=(1,2)$ and $\mathbf{b}=(0,0)$ (the function is increasing relative to x and y, separately).

An useful notion in the integral computation (and not only!-see the bellow application) is the notion of "uniform continuity".

DEFINITION 22. *Let A be a nonempty subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to \mathbb{R}^m$ be a function defined on A with values in $\mathbb{R}^m$. We say that $\mathbf{f}$ is uniformly continuous on A if for any small quantity $\varepsilon > 0$, there is another small quantity $\delta_\varepsilon > 0$ (depending on ε) such that whenever we have two points $\mathbf{x}'$ and $\mathbf{x}''$ in A with the distance $\|\mathbf{x}' - \mathbf{x}''\|$ between them less then δ_ε, the distance $\left\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}'')\right\|$ between their images is less then ε.*

The word "uniform" reefers to the fact that here the continuity is not defined at a point, but on the whole A. Moreover, the variation $\left\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}'')\right\|$ of $\mathbf{f}(\mathbf{x})$ is uniform relative to the variation $\|\mathbf{x}' - \mathbf{x}''\|$ of $\mathbf{x}$. Thus, if we want that the variation of $\mathbf{f}(\mathbf{x})$ to be less than 0.001 ($\left\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}'')\right\| < 0.001$) in the case of an uniform continuous function $\mathbf{f}$, we can find a constant $\delta = \delta_{0.001} > 0$ such that anywhere $\mathbf{a}'$ and $\mathbf{a}''$ would be in A, with the distance between them less than this last constant δ, we are sure that the corresponding variation of $\mathbf{f}$, $\left\|\mathbf{f}(\mathbf{a}') - \mathbf{f}(\mathbf{a}'')\right\|$ is less then 0.001.

REMARK 23. *The notion of uniform continuity is stronger then the "simple" continuity. Indeed, let $\mathbf{f} : A \to \mathbb{R}^m$ be a uniformly continuous function on A and let $\mathbf{a}$ be a fixed point in A. We shall prove that $\mathbf{f}$ is continuous at $\mathbf{a}$. For this, let $\{\mathbf{a}^{(n)}\}$ be a convergent sequence to $\mathbf{a}$ in A. We want to prove that the sequence $\{\mathbf{f}(\mathbf{a}^{(n)})\}$ is convergent to $\mathbf{f}(\mathbf{a})$ by using only the definition of the convergence. In fact, we want to prove that the numerical sequence $\{d(\mathbf{f}(\mathbf{a}^{(n)}), \mathbf{f}(\mathbf{a}))\}$ tends to zero. Now we use the usually Definition 1. For this, let $\varepsilon > 0$ be a small positive real number. Since $\mathbf{f}$ is uniformly continuous, there is a $\delta_\varepsilon > 0$ such that whenever $\|\mathbf{x}' - \mathbf{x}''\| < \delta_\varepsilon$, one has that*

$$\left\|\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}'')\right\| < \varepsilon.$$

Let us take now $\mathbf{x}''$ to be $\mathbf{a}$ and $\mathbf{x}' = \mathbf{a}^{(n)}$, with $n \geq N$, this last N chosen such that $\left\|\mathbf{a}^{(n)} - \mathbf{a}\right\| < \delta_\varepsilon$. Thus,

$$\left\|\mathbf{f}(\mathbf{a}^{(n)}) - \mathbf{f}(\mathbf{a})\right\| < \varepsilon,$$

whenever $n \geq N$ and so, we have just proved that the sequence $\{\mathbf{f}(\mathbf{a}^{(n)})\}$ is convergent to $\mathbf{f}(\mathbf{a})$, i.e. $\mathbf{f}$ is continuous at an arbitrary chosen point $\mathbf{a}$.

But continuity does not always imply uniform continuity. For instance, $f(x) = \ln x$, $x \in (0, 1]$, is a continuous function and not a uniformly continuous one. Indeed, let the sequences $x'_n = \frac{1}{n}$ and $x''_n = \frac{1}{2n}$. It is clear that $|x'_n - x''_n| = \frac{1}{2n} \to 0$, but $|\ln x'_n - \ln x''_n| = \ln 2 \nrightarrow 0$.

Thus, if we take $\varepsilon < \ln 2$ in Definition 22, we can NEVER find a small $\delta_\varepsilon > 0$ such that for all pairs (x', x'') with $|x' - x''| < \delta_\varepsilon$ one has

$$|\ln x' - \ln x''| < \varepsilon < \ln 2.$$

To see this, let us take n_0 large enough such that

$$\left|x'_{n_0} - x''_{n_0}\right| = \frac{1}{2n_0} < \delta_\varepsilon.$$

For the pair (x'_{n_0}, x''_{n_0}),

$$\left|\ln x'_{n_0} - \ln x''_{n_0}\right| = \ln 2,$$

which is greater than ε, so the definition of the uniform continuity does not work for this function.

The next result says that for the functions defined on compact sets, continuity and uniform continuity coincide. Pay attention, in our case above $(0, 1]$ in not compact! This is way we could prove that $f(x) = \ln x$ is not uniformly continuous.

THEOREM 59. *Let C be a compact subset of $\mathbb{R}^n$ and let $\mathbf{f} : C \to \mathbb{R}^m$ be a continuous function defined on C. Then $\mathbf{f}$ is uniformly continuous on C.*

PROOF. We suppose on contrary, namely that $\mathbf{f}$ is not uniformly continuous on C. We must carefully negate the statement of Definition 22. Thus, there is an $\varepsilon_0 > 0$ such that for any small enough $\delta > 0$ there is at least one pair $(\mathbf{x}'_\delta, \mathbf{x}''_\delta)$ with elements in C such that $\|\mathbf{x}'_\delta - \mathbf{x}''_\delta\| < \delta$ and

$$\left\|\mathbf{f}(\mathbf{x}'_\delta) - \mathbf{f}(\mathbf{x}''_\delta)\right\| \geq \varepsilon_0.$$

In particular, let us take for these δ, $\delta_k = \frac{1}{k}$ for $k = 1, 2, \ldots$. Like above, for such δ_k, $k = 1, 2, ...$, one can find two sequences $\{\mathbf{x}'^{(k)}\}$ and $\{\mathbf{x}''^{(k)}\}$ with $\left\|\mathbf{x}'^{(k)} - \mathbf{x}''^{(k)}\right\| < \frac{1}{k}$ and

$$\left\|\mathbf{f}(\mathbf{x}'^{(k)}) - \mathbf{f}(\mathbf{x}''^{(k)})\right\| \geq \varepsilon_0 > 0.$$

Since C is a compact set, we can find two subsequences: $\{\mathbf{x}'^{(k_t)}\}$ of $\{\mathbf{x}'^{(k)}\}$ and $\{\mathbf{x}''^{(k_t)}\}$ of $\{\mathbf{x}''^{(k)}\}$ (why can we take the same k_t for both subsequences?) such that these both subsequences are convergent to the same limit $\mathbf{y} \in C$ because

$$\left\|\mathbf{x}'^{(k_t)} - \mathbf{x}''^{(k_t)}\right\| < \frac{1}{k_t} \to 0.$$

Since $\mathbf{f}$ is continuous, one has that the both sequences $\{\mathbf{f}(\mathbf{x}'^{(k_t)})\}$ and $\{\mathbf{f}(\mathbf{x}''^{(k_t)})\}$ are convergent to the same limit $\mathbf{f}(\mathbf{y})$. So the distance between the corresponding terms becomes smaller and smaller as $n \to \infty$,

i.e.

$$\left\|\mathbf{f}(\mathbf{x}'^{(k_t)}) - \mathbf{f}(\mathbf{x}''^{(k_t)})\right\| \to 0,$$

a contradiction, because $\left\|\mathbf{f}(\mathbf{x}'^{(k_t)}) - \mathbf{f}(\mathbf{x}''^{(k_t)})\right\|$ is always greater or equal to ε_0. Thus, our assumption on the nonuniform continuity of $\mathbf{f}$ is false. Hence, $\mathbf{f}$ is uniformly continuous. □

This result is very useful in practice. For instance, the function $f(x) = \ln x$ is uniform continuous on any closed interval $[a, b] \subset (0, \infty)$. Indeed, $[a, b]$ is a compact subset in the definition domain $(0, \infty)$ of f, f is continuous on $[a, b]$ and so we can apply the above Theorem 59.

EXAMPLE 13. *Let C be a 3D-object ($C \subset \mathbb{R}^3$), bounded and containing its boundary ∂C, like usually in practice. We know that C is closed if and only if it contains its boundary ∂C. Let us assume that at any point $M(x, y, z)$ of C we have a density $f(x, y, z)$. It is commonly to suppose that the density function $f : C \to \mathbb{R}$ is a continuous function. The above theorem and our hypotheses on C say that f is uniformly continuous. We cannot practically work with this function because nobody gives it us in advance. But we can perform some measurements. How do we perform such measurements $f(x_i, y_i, z_i)$, $i = 1, 2, ..., n$, such that if we chose a point $M(x, y, z)$ in C, we can find i_0 with*

$$|f(x, y, z) - f(x_{i_0}, y_{i_0}, z_{i_0})| < \varepsilon$$

(this is a small positive real number which controls the error, for instance $\varepsilon = 1/1000$). Since our function is uniformly continuous, there is a small $\delta > 0$ such that whenever the distance between two points $\mathbf{x}' = (x', y', z')$ and $\mathbf{x}'' = (x'', y'', z'')$ of C is less than this δ, we have that

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

It remains to us to divide the body C into subbodies C_i, $i = 1, 2, ..., n$, such that $C = \bigcup_{i=1}^{i=n} C_i$ and the diameters

$$\omega_i = \sup\{\|\mathbf{x}' - \mathbf{x}''\| : \mathbf{x}', \mathbf{x}'' \in C_i\}$$

of C_i are less then δ. Let us choose now a fixed point $M_i(x_i, y_i, z_i)$ in each C_i for $i = 1, 2, ..., n$. Then the approximation

$$f(x, y, z) \approx f(x_i, y_i, z_i)$$

is a good one if $M(x, y, z) \in C_i$. This means that

$$|f(x, y, z) - f(x_i, y_i, z_i)| < \varepsilon.$$

Thus, we can perform measurements of the density function values only at some arbitrarily chosen points M_i in each C_i.

We give here a very useful result, in a more general setting (define and prove things slowly!).

THEOREM 60. *Let X and Y be two compact metric spaces (recall that a metric space is compact if any sequence of it has at least one convergent subsequence) and let $f : X \to Y$ be a continuous bijection from X on Y. Let $g : Y \to X$ be its inverse. Then g is also continuous.*

PROOF. Let us prove that g carries back closed subsets of X into closed subsets of Y (see Remark 21). Let C be a closed subset of X and let $E = g^{-1}(C) = f(C)$. Since X is compact, C is also compact (prove it!). Since f is continuous, $E = f(C)$ is compact, so E itself is closed in Y (prove it!). Hence, g is continuous. □

COROLLARY 7. *Let f be a strictly monotone continuous function which carries the interval $[a, b]$ onto the interval $[c, d]$ (see also the next section, Darboux' theorem). Then f is inversable and its inverse g is also continuous.*

PROOF. Since f is strictly monotone it is one-to-one (injective). Since both intervals are compact metric spaces, we simply apply the previous result. Here, "onto" means surjectivity!. □

4. Continuous functions on connected sets

Let A be a subset of $\mathbb{R}^n$. A *continuous curve* in A is a vector continuous function $\boldsymbol{\gamma} : I \to A$, defined on an interval I, finite or not, opened or not, closed or not. In fact, we think of the image $\boldsymbol{\gamma}(I)$ of the interval I through $\boldsymbol{\gamma}$. Let $M(x_1, x_2, ..., x_n)$ be a point in A. We say that $\boldsymbol{\gamma}$ passes through M if there is t_0 in I such that $\boldsymbol{\gamma}(t_0) = M$.

DEFINITION 23. *We say that the subset A of $\mathbb{R}^n$ is connected if any two points M_1 and M_2 of A can be connected by a continuous curve, i.e. if there is a continuous function $\boldsymbol{\gamma} : I \to A$ and $t_1, t_2 \in I$ such that $\boldsymbol{\gamma}(t_1) = M_1$ and $\boldsymbol{\gamma}(t_2) = M_2$. This means that $\boldsymbol{\gamma}$ passes through M_1 and M_2.*

REMARK 24. *An interval I of $\mathbb{R}$ is a subset of $\mathbb{R}$ with the following property: if $a, b \in I$ and x is between a and b ($a \leq x \leq b$), then x is also in I. In $\mathbb{R}$, the connected subsets are exactly the intervals of $\mathbb{R}$. Indeed, let I be a connected subset of $\mathbb{R}$, let $a, b \in I$ and let x with $a \leq x \leq b$. Since I is connected, let $\boldsymbol{\gamma} : J \to I$ be a continuous curve which connect a and b. This means that there are t_1 and t_2 in J such that $\boldsymbol{\gamma}(t_1) = a$ and $\boldsymbol{\gamma}(t_2) = b$. We can restrict $\boldsymbol{\gamma}$ to the interval $[t_1, t_2] \subset J$ and apply Darboux property for the continuous function $\boldsymbol{\gamma}$ (see Theorem 33). Hence $x = \boldsymbol{\gamma}(t_3)$, where $t_3 \in [t_1, t_2]$. So $x \in I$;*

thus I is an interval. Conversely, let I be an interval in $\mathbb{R}$ and let x_1, $x_2 \in I$. Let $\boldsymbol{\gamma} : [x_1, x_2] \to I$ be the identity mapping. This is obviously a continuous curve which connect x_1 and x_2.

THEOREM 61. *Let A be a connected subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to \mathbb{R}^m$ be a continuous mapping defined on A with values in $\mathbb{R}^m$. Then the image $\mathbf{f}(A)$ of $\mathbf{f}$ in $\mathbb{R}^m$ is also a connected subset of $\mathbb{R}^m$.*

PROOF. Let $\mathbf{f}(\mathbf{x})$ and $\mathbf{f}(\mathbf{y})$ be two points in $\mathbf{f}(A)$, $\mathbf{x}, \mathbf{y} \in A$. Since A is connected, there is a continuous curve $\boldsymbol{\gamma} : I \to A$ and two points $a, b \in I$ (an interval in $\mathbb{R}$) such that $\boldsymbol{\gamma}(a) = \mathbf{x}$ and $\boldsymbol{\gamma}(b) = \mathbf{y}$. Now, the composition $\mathbf{f} \circ \boldsymbol{\gamma} : I \to \mathbb{R}^m$ is a continuous curve with $(\mathbf{f} \circ \boldsymbol{\gamma})(a) = \mathbf{f}(\mathbf{x})$ and $(\mathbf{f} \circ \boldsymbol{\gamma})(b) = \mathbf{f}(\mathbf{y})$. Thus $\mathbf{f}(A)$ is a connected subset of $\mathbb{R}^m$. □

This is a fundamental result in different practical exercises. For instance, let

$$S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq R^2\}$$

be the $3D$-ball of radius R with centre at origin. Let $f : S \to \mathbb{R}$ be the functions which associates to any point $M(x, y, z)$ the sum of these coordinates, namely

$$f(x, y, z) = x + y + z.$$

Let us find the image of S through f. Since S is connected (in fact S is a convex subset of $\mathbb{R}^3$, i.e. for any pair of points L, P of S, the segment $[L, P]$ is contained in S) and since f is continuous, its image in $\mathbb{R}$ is a connected subset (see Theorem 61), i.e. it is an interval (see Remark 24). In fact, this image is a closed and bounded interval because S is a compact set (way?) and f is continuous. So it is of the form $[m, M]$ where $m = \inf f(S)$ and $M = \sup f(S)$. To find m and M is not an easy task. We only remark that the points where it is realized the greatest and the smallest values must be on the boundary ∂S of S, namely where $x^2 + y^2 + z^2 = R^2$ (otherwise, if a point $H(a, b, c)$ of extremum, say a maximum, was inside the ball, not on the boundary ∂S, then we can gently increase (or decrease) one of the values a, b, or c, such that the new point L obtained in this way belongs to the ball and, in it the function f has a greater value then the value of f in H). In a later section (Conditional extremum points) we shall see how to compute m and M.

The above theorem is helpful in proving the following useful result (this result provides the basis of for different algorithms for solving algebraic equations).

THEOREM 62. *Let $f : [a, b] \to \mathbb{R}$ be a continuous function such that $f(a) \cdot f(b) < 0$. Then, there is a point c in (a, b) such that $f(c) = 0$.*

This means that the equation $f(x) = 0$ has at least one solution in the interval $[a, b]$.

PROOF. The set $f([a, b])$ is an interval (see Theorem 61 and Remark 24) which contains $f(a)$ and $f(b)$. Since $f(a) \cdot f(b) < 0$, the numbers $f(a)$ and $f(b)$ have distinct signs. Since $f([a, b])$ is an interval and since 0 is between $f(a)$ and $f(b)$, 0 must be also in $f([a, b])$. This means that there is a c in $[a, b]$ such that $f(c) = 0$. Since $f(a) \cdot f(b) < 0$, this c cannot be neither a nor b, so $c \in (a, b)$. □

REMARK 25. *In fact, the statement of this last theorem is equivalent with the statement of Darboux Theorem 33. Let us prove for instance that the above last theorem implies Darboux Theorem 33. Let* $m = \inf_{x\in[a,b]} f(x) = f(x_1)$ *(see Weierstrass Theorem 32) and* $M = \sup_{x\in[a,b]} f(x) = f(x_2)$. *Let choose a number* $\lambda \in (m, M)$ *and let consider the auxiliary continuous function* $g(x) = f(x) - \lambda$. *Let us take now the interval* $[x_1, x_2]^{\pm}$ *(here* $\pm$ *means that* $[x_1, x_2]^{\pm} = [x_1, x_2]$ *if* $x_1 < x_2$ *and* $[x_1, x_2]^{\pm} = [x_2, x_1]$ *if* $x_2 < x_1$; *if* $x_1 = x_2$ *our function is constant and one has nothing to prove). Since* $g(x_1) \cdot g(x_2) < 0$ *(if one of the factors is equal to* 0 *we also have nothing to prove more!), Theorem 62 says that there exists a number* $c \in (a, b)$ *such that* $g(c) = 0$, *i.e.* $f(c) = \lambda$ *and Darboux Theorem is proved. Conversely is very easy (prove it!).*

We can use Theorem 62 in order to find approximative solutions for an equation $f(x) = 0$ in an interval $[a, b]$, on which the function f is continuous (find a counterexample to this theorem in the case when f is not continuous). We also assume that $f(a) \cdot f(b) < 0$. Let us divide the segment $[a, b]$ into two equal parts and chose that one $[a_1, b_1]$ for which $f(a_1) \cdot f(b_1) < 0$ (if $f(a_1) = 0$ or $f(b_1) = 0$, $c = a_1$ or $c = b_1$ and we stop the process). Let us repeat the same with the subinterval $[a_1, b_1]$ instead of $[a, b]$, and so on. If we cannot find a_n or b_n, $n = 1, 2, ...$, such that $f(a_n) = 0$ or $f(b_n) = 0$, the solution c is (the unique point) in the intersection $\bigcap_{n=1}^{\infty} [a_n, b_n]$ (why?). So, for a small error indicator $\varepsilon > 0$, if we take n_0 such that $\frac{b-a}{2^{n_0}} < \varepsilon$, then the approximation $c \approx a_{n_0}$ (or $c \approx b_{n_0}$) lead us to an error less then ε (why?). This is in fact the description of a very known algorithm in Computer Science for constructing approximative solutions for a large class of equations.

5. The Riemann's sphere

In Fig.6.3 we have a sphere S of radius $R > 0$ and with center at the origin $O(0, 0, 0)$. Its equation is

$$x^2 + y^2 + z^2 = R^2 \tag{5.1}$$

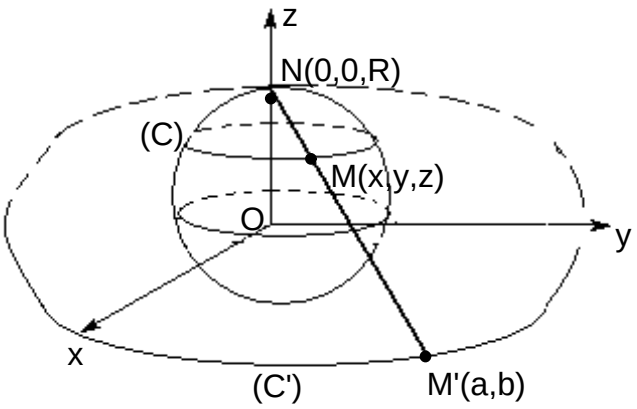


Fig. 6.3

We know that the subset

$$S = \{(x, y, z) : x^2 + y^2 + z^2 = R^2\}$$

is a compact subset of $\mathbb{R}^3$ (it is closed and bounded, why?). Since B. Riemann used this model for explaining the "compactification" of the usual complex plane $\mathbb{C}$ (identified here with the coordinate plane xOy), we call S the *Riemann sphere*.We call the point $N(0, 0, R)$, the *north pole* of S (see Fig.6.3). Let us associate to any point $M(x, y, z)$ of the sphere S, the point $M'(a, b, 0)$ in the plane xOy $(= \mathbb{C})$, obtained by intersecting the line NM with the plane xOy (see Fig.6.3). Since for N we cannot associate in this way a point in xOy, we say that there is a one to one correspondence between $S \setminus \{N\}$ and $\mathbb{C}$. Let us denote by $f : S \setminus \{N\} \to \mathbb{C}$, the mapping $M \rightsquigarrow M'$, or $f(M) = M'$. It is not so easy to express a and b as functions of x, y, z. If we think of a sequence $\{M_n\}$ of points on S, which is convergent in $\mathbb{R}^3$ to M, it is easy to see that the sequence $\{M'_n\}$ is convergent to M' in $\mathbb{C}$. So f is a continuous function on $S \setminus \{N\}$. As in the case of the "compactification" of $\mathbb{R}$ by adding of the symbols $\{\pm\infty\}$ (since in $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$ any sequence has at least one convergent subsequence-why?-it is a compact metric space!)) we take a symbol "$\underline{\infty}$" outside $\mathbb{C}$ and consider $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\underline{\infty}\}$ with some obvious algebraic operations: $x + \underline{\infty} = \underline{\infty} + x = \underline{\infty}$, $x \in \mathbb{C}$,

$|\underline{\infty}| = \infty$ (this is the symbol $+\infty$ from $\overline{\mathbb{R}}$), etc. If we extend now the function f to the whole sphere S by putting $f(N) = \underline{\infty}$, we obtain a bijection between the Riemann sphere and $\widehat{\mathbb{C}}$. We say that a sequence $\{z_n\}$ of $\widehat{\mathbb{C}}$ is convergent to $\underline{\infty}$ if $|z_n| \to \infty \in \overline{\mathbb{R}}$. So this f is invertible and f^{-1} is also continuous. In particular $\widehat{\mathbb{C}}$ is a compact metric space, the least compact metric space which contains $\mathbb{C}$ (why?). This is why one can also call $\widehat{\mathbb{C}}$ the Riemann sphere. For instance, a "ball" with centre at $\underline{\infty}$ is the exterior of an usual closed ball with centre at O and of radius $r > 0$: $\{(x, y, z) : x^2 + y^2 + z^2 > r^2\}$. The notion of Riemann sphere is very important when we work with functions of complex variable. Intuitively, $\underline{\infty}$ can be realized as the circumference of a "circle" with center at $O \in \mathbb{C}$ and of an infinite radius. So, the fundamental "ε-neighborhoods" of $\underline{\infty}$ are of the form $\{z \in \mathbb{C} : |z| > R\}$, where R is any positive (usually large) real number. We finally remark that the metric structure on S is that one induced from $\mathbb{R}^3$.

6. Problems

1. Say if the following sets are open, closed, bounded, compact or connected. In each case, compute their closure and their boundaries. Draw them carefully!

a)

$$\{(x, y) : x^2 + y^2 < 9\};$$

b)

$$\{(x, y) : x^2 + y^2 > 9\};$$

c)

$$\{(x, y) : x^2 + y^2 = 5\};$$

d)

$$\{(x, y) : x \in [0, 1); y \in (1, 2]\};$$

e)

$$\{(x, y) : x + y = 3\};$$

f)$\{(q, 0) : q \in \mathbb{Q}\}$; g)$\{(0, \frac{1}{n}) : n = 1, 2, ...\}$; h)$\{(x, y) : y^2 = 2x, x \in [0, 1)\}$; i)

$$\{(\frac{1}{n}, \frac{1}{n}) : n = 1, 2, ...\};$$

j)

$$\{(x, y, z) : x + y + z \leq 3; x, y, z \in [0, \infty)\}$$

k)

$$\{(x, y, z) : x \in [-1, 1], y \in (0, 4], z \in (-3, 5]\}$$

l) $\{z \in \mathbb{C} : |z - 2i| < 3\}$; m)$\{z \in \mathbb{C} : |2z + 3| \leq 6\}$; n)

$$\{z \in \mathbb{C} : |z + 3 - 2i| > 4\};$$

o)

$$\{z \in \mathbb{C} : z = x + iy, x = 2, y \leq 3\};$$

p)

$$\{z \in \mathbb{C} : 2 < |z - 2| \leq 4\};$$

q)

$$\{z \in \mathbb{C} : |z - 3 + 2i| > 2\};$$

r)

$$\{f \in C[0, 2] : \|f\| < 2\};$$

s)

$$\{f \in C[0, 2\pi] : \|f\| \geq 3\};$$

u)

$$\{f \in C[0, 2\pi] : \|f - \sin x\| < 0.3\}$$

v)

$$\{f \in C[-3.3] : g - \frac{1}{10} \leq f < g + \frac{1}{10},$$

where $g(x) = x$, $g(x) = -x$, or $g(x) = x^2\}$; w)

$$\{f \in C[0, 1] : 2 < \|f - g\| < 4\},$$

where $g(x) = x$; y)$D = \{(x, y) : \ln(x^2+y^2-4)/(x+2y)$ is well defined$\}$.

2. Compute the limits of the following sequences:

a)

$$\mathbf{x}^{(n)} = \left(\frac{1}{2n+1}, \frac{2n-1}{3n+4}, (1 + \frac{4}{n})^{2n}\right);$$

b)

$$\mathbf{x}^{(n)} = \left(\frac{\sqrt{n} - 1}{\sqrt[3]{n} - \sqrt[3]{n-1}}, \frac{n \sin \frac{1}{n}}{1+n}\right);$$

c)

$$z_n = \frac{3 + 2in}{n + 2i}, i = \sqrt{-1};$$

d) $z_n = \left(1 + \frac{i+1}{n}\right)^n$; e) $z_n = \exp\left(in + \frac{i}{n}\right)$;

3. Starting with the definition of continuity and of uniform continuity, determine what of the following functions are continuous and what are uniformly continuous.

a) $f(x) = \sin x$, $x \in [0, \pi]$;

b)

$$f(x, y) = (x + y, \frac{1}{xy}), x \in [1, 2], y \in [3, 4];$$

c) $f(x,y,z) = x - y$, where $x^2 + y^2 + z^2 = 4$; d) $f(x) = \frac{1}{x}$, $x \in (0,2]$.

4. Some of the following limits exist, some do not exist. Say (and prove!) which of them exist and compute them in the affirmative situation.

a) $\lim_{(x,y)\to(0,0)} \frac{x^3+y^3+1}{2x^3+3y^3+2}$; b) $\lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{xy+1}-1}$;

c) $\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^2}$ (Hint: $\frac{xy}{x^2+y^2} \leq \frac{1}{2}$, etc.);

d)

$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{|x|+|y|}$$

(Hint: $\frac{x}{|x|+|y|}, \frac{y}{|x|+|y|} \leq 1$, etc.); e) $\lim_{(x,y)\to(0,0)} \frac{x^3+y^3}{x^2+y^2}$; f) $\lim_{x\to 0} \frac{|x|}{x}$; g) $\lim_{x\to 0} \frac{\exp(-|x|)-1}{x}$; h) $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$;

i)

$$\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^4}$$

(Hint: use $(\frac{1}{n}, 0)$ and $(\frac{1}{n^2}, \frac{1}{n})$);

5. Compute, if you can, the following directional limits:

a) $\lim_{x\to 0, y=mx} \frac{xy}{x^2+y^2}$; b) $\lim_{x\to 0, y=mx} \frac{2x^3y}{x^6+y^2}$;

c)

$$\lim_{x\to\infty, y=mx} \frac{y}{x} \exp(-(x+y));$$

d)

$$\lim_{(x,y)\to(1,0), x^2+y^2=1} xy \exp(x^2+y^2).$$

6. Compute:

$$\lim_{(x,y,z)\to\mathbf{0}} \left(\frac{1}{x^2+y^2+1}, 1+xyz, \cos(x+y+z) \right)$$

and explain everything you did, step by step (small steps!).

7. Study the continuity of the following functions:

a)

$$f : \mathbb{R} \to \mathbb{R}, f(x) = 1,$$

if $x \in \mathbb{Q}$ and $f(x) = 0$, if $x \notin \mathbb{Q}$ (Dirichlet's function);

b)

$$f : \mathbb{R} \to \mathbb{R}, f(x) = x,$$

if $x \in \mathbb{Q}$, and $f(x) = -x$, if $x \notin \mathbb{Q}$;

c)

$$f : \mathbb{R} \to \mathbb{R}, f(x) = \exp(-x),$$

if $x \leq 0$ and $f(x) = \sin x$, if $x > 0$;

d)
$$f : \mathbb{R}^2 \to \mathbb{R}^2, f(x, y) = (x, 0);$$
e)
$$f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = d((x, y), (0, 0)) = \sqrt{x^2 + y^2};$$
f)
$$f : \mathbb{R}^2 \to \mathbb{R}^2, f(x, y) = \left(\frac{xy}{x^2 + y^2}, xy \right),$$
if $(x, y) \neq (0, 0)$ and $f(0, 0) = (0, 0)$;

g)
$$f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = xy \frac{x^2 - y^2}{x^2 + y^2},$$
if $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$;

h)
$$f : \mathbb{R}^2 \to \mathbb{R}, f(x, y) = \frac{\sin(x^3 + y^3)}{x^2 + y^2},$$
if $(x, y) \neq (0, 0)$ and $f(0, 0) = 0$.

8. Prove that $f(x) = x^2$ is uniformly continuous on $[0, 1]$, but it is not on the whole $\mathbb{R}$ (Hint: use $x_n = \sqrt{n}$, $x_{n+1} - x_n \to 0$, but $f(x_{n+1}) - f(x_n) = 1 \nrightarrow 0$).

9. Prove that $f(x) = \frac{1}{x^2}$ is uniformly continuous on $[1, 2]$, but not on $\mathbb{R}$.

10. Let (X, d) be a metric space. Prove that, for any fixed a in X, the mapping $f_a(x) = d(x, a)$ is a uniformly continuous function defined on X with values in $\mathbb{R}$.

11. Let $f : A \to \mathbb{R}$, $f(x, y, z) = x + y + z$, where
$$A = \{(x, y, z) \in \mathbb{R}^3 : 1 \leq x^2 + y^2 + z^2 \leq 4\}.$$
Prove that $f(A)$ is a closed interval in $\mathbb{R}$. Find it.

12. Do the same for
$$f(x, y) = x + y, x \in [1, 2], y \in [1, 2].$$

CHAPTER 7

Partial derivatives. Differentiability.

1. Partial derivatives. Differentiability.

Let A be an open subset in $\mathbb{R}$, a a fixed point in A and let $f : A \to \mathbb{R}$ be a function defined on A with values in $\mathbb{R}$. Let $B(a,r) = (a-r, a+r)$, $r > 0$, be a small ball (an open interval in our particular case) of radius r and with centre a, which is contained in A. Let h be a small quantity such that $a+h \in B(a,r)$. We call this h an "*increment*" of a in $B(a,r)$ (or in A if one takes h with $a+h \in A$). The difference $f(a+h)-f(a)$ is called the increment of f at a, corresponding to the increment h of a. So, here appears a new function $\varphi_{a,f}(h) = f(a+h)-f(a)$. This new function depends on a and on f. It is defined in a small ball, $(-\varepsilon, \varepsilon)$, which contains 0 as its centre and of radius ε, (at most r (why?)). The description of this last function is important in the case we want to evaluate the variation of a phenomenon around a given point a. For instance, if a worker has his salary a and if his salary increases with h, what is the increment $f(a+h)-f(a)$ of his family educational level? We say that the increment $f(a+h)-f(a)$ is *approximately linear around* a, if

$$f(a+h) - f(a) = \lambda(a,f)\cdot h + h\cdot \omega_{a,f}(h), \tag{1.1}$$

where $\omega_{a,f}$ is a function of h defined on $(-\varepsilon,\varepsilon)$, $\omega_{a,f}(0) = 0$ and $\omega_{a,f}(h) \to 0$, when $h \to 0$ (i.e. $\omega_{a,f}$ is continuous at 0). Here $\lambda(a,f)$ is a real number which depend on f and on a.

The birth of differential calculus began with the following result.

THEOREM 63. *With the above notation and hypotheses, the increment of f is approximately linear around a if and only if f is differentiable at a and, in this case $f'(a) = \lambda(a,f)$. Thus,*

$$f(a+h) - f(a) = f'(a)\cdot h + h\cdot \omega_{a,f}(h). \tag{1.2}$$

Hence,

$$f(a+h) - f(a) \approx f'(a)\cdot h$$

and the error $h\cdot\omega_{a,f}(h)$ is a zero $0(h)$ of h, i.e.

$$\lim_{h\to 0} \frac{h\cdot\omega_{a,f}(h)}{h} = 0.$$

PROOF. Let us divide by h the equality (1.1) and make $h \to 0$. We obtain that the limit

$$\lim_{h\to 0}\frac{f(a+h)-f(a)}{h} = \lambda(a,f).$$

So, if the increment $f(a+h)-f(a)$ is approximately linear around a, f is differentiable at a and $f'(a) = \lambda(a,f)$. Conversely, let us assume that f is differentiable at a. Then, if one construct

$$\omega_{a,f}(h) = \frac{f(a+h)-f(a)}{h} - f'(a), \tag{1.3}$$

it is easy to verify that this function $\omega_{a,f}$ is continuous at 0 and it is zero at $h=0$ (do it!). If we take now for $\lambda(a,f)$ the number $f'(a)$, and for $\omega_{a,f}$ the function constructed in (1.3), we obtain the formula (1.1), i.e. the increment of f is approximately linear around a. □

Let us evaluate the increment of $f(x) = -x^2+3x-7$ at $a=10$ if the increment h of a is 0.5. We simply apply formula (1.2) and find

$$f(10+0.5)-f(10) = f'(10)\cdot 0.5 + 0.5\cdot\omega_{f,10}(0.5) \approx -8.5.$$

DEFINITION 24. *With the above notation, the linear mapping* $df(a) : \mathbb{R}\to\mathbb{R}$, *defined by*

$$df(a)(h) = f'(a)\cdot h,$$

is called the first differential of f *at* a. *This one exists if and only if the first derivative* $f'(a)$ *of* f *at* a *exists (why?).*

Thus,

$$df(a)(h) \approx f(a+h)-f(a),$$

i.e. the value $df(a)(h)$ of the first differential of f at a, computed in the increment h of a, is approximative equal to the corresponding increment

$$f(a+h)-f(a)$$

of f at a.

Before extending the notion of a differential to a vector function we need some other simpler notion.

Let A be an open subset of $\mathbb{R}^n$, $\mathbf{f} : A \to \mathbb{R}^m$, a vector function of n variables, defined on A with values in the normed (or metric) space $\mathbb{R}^m$ and $\mathbf{a} = (a_1, a_2, ..., a_n)$ a point in A. We write $\mathbf{f} = (f_1, f_2, ..., f_m)$, where $f_1, f_2, ..., f_m$ are the m scalar component functions of $\mathbf{f}$. For the moment we take $m=1$ and write $\mathbf{f} = f$, like a scalar function (with values in $\mathbb{R}$). Let us fix a variable x_j $(j = 1, 2, ..., n)$ of the variable vector

$$\mathbf{x} = (x_1, x_2, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n).$$

For this fixed j, let us define a "partial function" φ_j of f at $\mathbf{a}$. For this we fix all the other variables $x_1, x_2, ..., x_{j-1}, x_{j+1}, ..., x_n$ (except x_j) by putting

$$x_1 = a_1, x_2 = a_2, ..., x_{j-1} = a_{j-1}, x_{j+1} = a_{j+1}, ..., x_n = a_n$$

and let us leave free the variable x_j in

$$f(\mathbf{x}) = f(x_1, x_2, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n),$$

i.e. we define

$$\varphi_j(t) = f(a_1, a_2, ..., a_{j-1}, t, a_{j+1}, ..., a_n), \tag{1.4}$$

where t runs over the projection $pr_j(A)$ of A along the Oj-axis, where

$$pr_j(x_1, x_2, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n) = x_j$$

DEFINITION 25. *With the above notation, if the function φ_j is differentiable at $t = a_j$, one says that f has a partial derivative $\varphi_j'(a_j)$ with respect to the variable x_j at $\mathbf{a}$ and we denote this last one by $\frac{\partial f}{\partial x_j}(\mathbf{a})$. The mapping $\mathbf{x} \rightsquigarrow \frac{\partial f}{\partial x_j}(\mathbf{x})$, $\mathbf{x} \in A$, is called the partial derivative of f with respect to x_j.*

Practically, if we want to compute the partial derivative of a scalar function f of n variables

$$x_1, x_2, ..., x_{j-1}, x_j, x_{j+1}, ..., x_n,$$

with respect to x_j, we think of the other variables

$$x_1, x_2, ..., x_{j-1}, x_{j+1}, ..., x_n$$

like being constants (parameters, or "inactivated" variables) and we perform the usual differential laws on the "active" variable x_j. If $n = 1$, we usually denote x_1 by x. If $n = 2$, we usually denote x_1 by x and x_2 by y. If $n = 3$, we usually denote x_1 by x, x_2 by y and x_3 by z. For instance, let

$$f(x, y) = \sin^2(x^3 + y^3)$$

be defined on $\mathbb{R}^2$ and let $\mathbf{a} = (0, \sqrt[3]{\frac{\pi}{2}})$ be the fixed point at which we want to compute the partial derivatives of f (with respect to x and to y respectively). Let us use the definition to compute $\frac{\partial f}{\partial x}(\mathbf{a})$. In our case,

$$\varphi_1(t) = \sin^2(t^3 + \frac{\pi}{2})$$

and

$$\varphi_1'(t) = 2\sin(t^3 + \frac{\pi}{2}) \cdot \cos(t^3 + \frac{\pi}{2}) \cdot 3t^2$$

(we just used the chain rule for computing the derivative of a composed function of one variable). Now,

$$\frac{\partial f}{\partial x}((0,\sqrt[3]{\frac{\pi}{2}})) = \varphi_1'(0) = 0.$$

Let us compute now

$$\frac{\partial f}{\partial y}((x,y)) = 2\sin(x^3+y^3)\cdot\cos(x^3+y^3)\cdot 3y^2 \tag{1.5}$$

Here, we simply considered that the initial function depended only on y and we looked at x like to a constant. If we want to compute $\frac{\partial f}{\partial y}((0,\sqrt[3]{\frac{\pi}{2}}))$, we simply make $x=0$ and $y=\sqrt[3]{\frac{\pi}{2}}$ in the general expression (1.5) of $\frac{\partial f}{\partial y}((x,y))$. Thus, $\frac{\partial f}{\partial y}((0,\sqrt[3]{\frac{\pi}{2}}))$ is also 0. Since both partial derivatives of f at $(0,\sqrt[3]{\frac{\pi}{2}})$ are zero, we say that this last point is a *stationary (or critical) point.*

If f is a function defined on an open subset A of $\mathbb{R}^n$ which has partial derivatives with respect to all its variables at a point $\mathbf{a}$, we define the gradient vector of f at $\mathbf{a}$ by the formula:

$$\operatorname{grad} f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), ..., \frac{\partial f}{\partial x_n}(\mathbf{a})\right).$$

We say that $\mathbf{a}$ is a critical (stationary) point for f if $\operatorname{grad} f(\mathbf{a}) = \mathbf{0}$. The gradient is the direct generalization of the notion of "velocity".

We know from any course of "Linear Algebra" that a mapping $\mathbf{T}: \mathbb{R}^n \to \mathbb{R}^m$ is said to be a linear mapping if $\mathbf{T}(\mathbf{x}+\mathbf{y}) = \mathbf{T}(\mathbf{x}) + \mathbf{T}(\mathbf{y})$ and $\mathbf{T}(\alpha\mathbf{x}) = \alpha\mathbf{T}(\mathbf{x})$ for any $\mathbf{x}, \mathbf{y}$ in $\mathbb{R}^n$ and α in $\mathbb{R}$. For instance, if $T: \mathbb{R} \to \mathbb{R}$ is linear, then $T(x) = xT(1)$ for any $x \in \mathbb{R}$. Hence, $T(x) = \lambda x$ ($\lambda = T(1)$!) for any x in $\mathbb{R}$. If $T: \mathbb{R}^n \to \mathbb{R}$ is linear then, by taking

$$\mathbf{x} = (x_1, x_2, ..., x_n) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n,$$

where $\mathbf{e}_1 = (1,0,0,...,0)$, $\mathbf{e}_2 = (0,1,0,...,0)$, $..., \mathbf{e}_n = (0,0,0,...,0,1)$, we get that

$$T(\mathbf{x}) = x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + ... + x_nT(\mathbf{e}_n) = \lambda_1x_1 + \lambda_2x_2 + ... + \lambda_nx_n,$$

where $\lambda_i = T(\mathbf{e}_i)$ for any $i = 1, 2, ..., n$. It is easy to see that if $T_1, T_2, ..., T_m$ are the component functions of $\mathbf{T}$, then $\mathbf{T}$ is a linear mapping if and only if all the component functions $T_1, T_2, ..., T_m$ of $\mathbf{T}$ are linear (prove it!).

THEOREM 64. *Any linear mapping $\mathbf{T}: \mathbb{R}^n \to \mathbb{R}^m$ is a continuous vector function of n variables.*

PROOF. It is sufficient to prove that any component function T_i, $i = 1, 2, ..., n$ of $\mathbf{T}$ is continuous (see Theorem 54). This means that we can reduce ourselves to the case of $m = 1$, i.e. to the case of a scalar function $T : \mathbb{R}^n \to \mathbb{R}$. Let

$$\{e_1 = (1, 0, 0, ..., 0), e_2 = (0, 1, 0, ..., 0), ..., e_n = (0, 0, 0, ..., 0, 1)\}$$

be the canonical basis of $\mathbb{R}^n$. This means that any vector $\mathbf{x} = (x_1, x_2, ..., x_n)$ can be uniquely represented as:

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n.$$

Let us denote

$$\alpha_1 = T(\mathbf{e}_1), \alpha_2 = T(\mathbf{e}_2), ..., \alpha_n = T(\mathbf{e}_n).$$

These are fixed real numbers. Hence,

$$T(\mathbf{x}) = T((x_1, x_2, ..., x_n)) = x_1\alpha_1 + ... + x_n\alpha_n.$$

If

$$\mathbf{x}^{(m)} = (x_1^{(m)}, x_2^{(m)}, ..., x_n^{(m)}) \to \mathbf{x} = (x_1, x_2, ..., x_n),$$

when $m \to \infty$, then,

$$x_1^{(m)} \to x_1, x_2^{(m)} \to x_2, ..., x_n^{(m)} \to x_n,$$

when $m \to \infty$ (componentwise convergence). Thus,

$$T(\mathbf{x}^{(m)}) = x_1^{(m)}\alpha_1 + x_2^{(m)}\alpha_2 + ... + x_n^{(m)}\alpha_n \to x_1\alpha_1 + ... + x_n\alpha_n$$

which is just $T(\mathbf{x})$. Hence, T is a continuous mapping. □

REMARK 26. *Let us define the associated matrix of*

$$\mathbf{T} = (T_1, T_2, ..., T_m)$$

by $a_{ij} = T_i(\mathbf{e}_j)$ *for* $i = 1, 2, ..., m$ *and* $j = 1, 2, ..., n$. *So the matrix* $A = (a_{ij})$ *is a* $m \times n$ *matrix with entries in* $\mathbb{R}$. *If we compute now*

$$\|\mathbf{T}(\mathbf{x})\|^2 = T_1(\mathbf{x})^2 + T_2(\mathbf{x})^2 + ... + T_m(\mathbf{x})^2 =$$

$$\left(\sum_{i=1}^n x_i a_{1i}\right)^2 + \left(\sum_{i=1}^n x_i a_{2i}\right)^2 + ... + \left(\sum_{i=1}^n x_i a_{mi}\right)^2 \leq$$

$$\leq \sum_{i=1}^n x_i^2 \sum_{i=1}^n a_{1i}^2 + \sum_{i=1}^n x_i^2 \sum_{i=1}^n a_{2i}^2 + ... + \sum_{i=1}^n x_i^2 \sum_{i=1}^n a_{mi}^2 = \|\mathbf{x}\|^2 \|A\|^2,$$

where we recall that

$$\|A\| = \sqrt{\sum_{j=1}^m \sum_{i=1}^n a_{ji}^2}.$$

Thus,

$$\|\mathbf{T}(\mathbf{x})\| \leq \|A\| \, \|\mathbf{x}\| \, . \tag{1.6}$$

From here we can easily directly prove the continuity of $\mathbf{T}$ *(do it!).*

Now, we come back to the definition of the linear approximation of the increment $f(x+h)-f(x)$ of a function f around a point a, in a general situation.

DEFINITION 26. *(Frechet) Let* D *be an open subset of* $\mathbb{R}^n$ *and let* $\mathbf{a}$ *be a fixed point in* D*. Let* $f : D \to \mathbb{R}$ *be a function defined on* D *with values in* $\mathbb{R}$*. We say that* f *is differentiable at* $\mathbf{a}$ *if there is a linear mapping* $T_{\mathbf{a}} = T : \mathbb{R}^n \to \mathbb{R}$ *and a continuous scalar function* $\varphi(\mathbf{h})$ *which is continuous at* $\mathbf{0} = (\underbrace{0, 0, ..., 0}_{n-times})$*, defined on a small ball* $B(\mathbf{0},r) \subset \mathbb{R}^n$*,* $r > 0, \varphi(\mathbf{0}) = 0$ *with* $\lim\limits_{\mathbf{h}\to\mathbf{0}} \frac{\varphi(\mathbf{h})}{\|\mathbf{h}\|} = 0$*, such that*

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = T(\mathbf{h}) + \varphi(\mathbf{h}). \tag{1.7}$$

This means that the increment $f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})$ *can be linearly approximated by the linear mapping* T *(which depend on* $\mathbf{a}$ *and on* f*) around the point* $\mathbf{a}$ *up to a function* $\varphi(\mathbf{h})$ *which is a zero of* $\mathbf{h}$ *(*$0(\mathbf{h})$*) of order* 1 *(*$\lim\limits_{\mathbf{h}\to\mathbf{0}} \frac{\varphi(\mathbf{h})}{\|\mathbf{h}\|} = 0$*). The linear mapping* T *is called the (first) differential of* f *at* $\mathbf{a}$*. We write it as* $df(\mathbf{a})$*. Hence, formula (1.7) becomes*

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = df(\mathbf{a})(\mathbf{h}) + \varphi(\mathbf{h}). \tag{1.8}$$

REMARK 27. *It is clear that* f *is differentiable at* $\mathbf{a}$ *if and only if there is a linear function* $T : \mathbb{R}^n \to \mathbb{R}$ *such that the following limit exists and it is zero:*

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - T(\mathbf{h})}{\|\mathbf{h}\|} = 0. \tag{1.9}$$

Indeed, if (1.9) is true, then $\varphi(\mathbf{h}) = f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})-T(\mathbf{h})$ *is continuous at* $\mathbf{0}$ *and its value at* $\mathbf{0}$ *is* 0*. If it were not continuous at* $\mathbf{0}$*, there would be an* $\varepsilon > 0$ *such that*

$$|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - T(\mathbf{h})| > \varepsilon$$

for any small values of $\mathbf{h} \to \mathbf{0}$*. So,*

$$\frac{|f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - T(\mathbf{h})|}{\|\mathbf{h}\|} > \frac{\varepsilon}{\|\mathbf{h}\|} \to \infty,$$

when $\mathbf{h} \to \mathbf{0}$*. Hence (1.9) could not be true, a contradiction!*

Shortly saying, f is differentiable at $\mathbf{a}$ if it can be "well" approximated on a small neighborhood of $\mathbf{a}$ by a formula of the following type:

$$f(\mathbf{a}+\mathbf{h}) \approx f(\mathbf{a})+T(\mathbf{h}), \tag{1.10}$$

where T is a linear mapping and $\mathbf{h}$ is a small increment of $\mathbf{a}$. This last interpretation is very useful in Physics and in Engineering when a phenomenon is "linearized".

The next big problem is how to compute this T in language of f and $\mathbf{a}$. But, first of all, let us use only the definition and the remark above to "guess" the differentials for some simple functions. For instance, if f has only one variable, we find again Definition 24. If f is a constant function, then $df(\mathbf{a})$ is the zero linear mapping (prove this!). The first differential of a linear mapping $T : \mathbb{R}^n \to \mathbb{R}$ is T itself (why?). In particular, the i-th projection $pr_i : \mathbb{R}^n \to \mathbb{R}$,

$$pr_i(h_1, h_2, ..., h_i, ..., h_n) = h_i,$$

is differentiable and its differential pr_i is denoted by dx_i, or dx, dy, dz in the $3D$-case. So

$$dy(1,2,-3)(3,1,-7) = 1, dz(a_1,a_2,a_3)(-2,3,5) = 5$$

for any $\mathbf{a} = (a_1, a_2, a_3)$.

THEOREM 65. *If f is differentiable at $\mathbf{a} \in D$, where D is an open subset of $\mathbb{R}^n$, then f is continuous at $\mathbf{a}$. This means that the property of differentiability is stronger then the property of continuity.*

PROOF. Let $\{\mathbf{a}^{(n)}\}$ be a sequence of vectors in $\mathbb{R}^n$ which is convergent to $\mathbf{a}$ and let $\mathbf{h}^{(n)} = \mathbf{a}^{(n)} - \mathbf{a}$ $(\to \mathbf{0})$. Then

$$f(\mathbf{a} + \mathbf{h}^{(n)}) = f(\mathbf{a}) + df(\mathbf{a})(\mathbf{h}^{(n)}) + \varphi(\mathbf{h}^{(n)})$$

(see (1.8)). Since $df(\mathbf{a})$ is a linear mapping, it is continuous (see Theorem 64), so

$$\lim_{n\to\infty} df(\mathbf{a})(\mathbf{h}^{(n)}) = 0.$$

Since $\lim_{\mathbf{h}\to\mathbf{0}} \frac{\varphi(\mathbf{h})}{\|\mathbf{h}\|} = 0$, one has that $\lim_{n\to\infty} \varphi(\mathbf{h}^{(n)}) = 0$ (why?). Hence,

$$f(\mathbf{a} + \mathbf{h}^{(n)}) \to f(\mathbf{a}),$$

when $n \to \infty$. □

THEOREM 66. *The linear mapping $T = df(\mathbf{a})$ is uniquely determined by f and $\mathbf{a}$.*

PROOF. The proof of this result is implicitely included in the statement of the next theorem (see Theorem (67). However, we give here another proof.

If there was another one U such that

$$f(\mathbf{a}+\mathbf{h})-f(\mathbf{a}) = U(\mathbf{h})+\varphi_1(\mathbf{h}), \tag{1.11}$$

where $\varphi_1(\mathbf{0})=0$, φ_1 is continuous at $\mathbf{0}$ and $\lim\limits_{\mathbf{h}\to\mathbf{0}}\frac{\varphi_1(\mathbf{h})}{\|\mathbf{h}\|}=0$, we can write that

$$T(\mathbf{h})+\varphi(\mathbf{h}) = U(\mathbf{h})+\varphi_1(\mathbf{h})$$

for all $\mathbf{h}$ in a small ball centered at origin. Moreover,

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{(T-U)(\mathbf{h})}{\|\mathbf{h}\|}=\lim_{\mathbf{h}\to\mathbf{0}}\frac{\varphi_1(\mathbf{h})-\varphi(\mathbf{h})}{\|\mathbf{h}\|}=0. \tag{1.12}$$

We want to prove that for any $\mathbf{x}$ in $\mathbb{R}^n$ one has $T(\mathbf{x})=U(\mathbf{x})$. We assume contrary, namely that there is a $\mathbf{x}_0$ such that $(T-U)(\mathbf{x}_0)\neq 0$. If $t>0$ is small, then $t\mathbf{x}_0$ is small, i.e. it is close to $\mathbf{0}$, because $\|t\mathbf{x}_0\|=t\,\|\mathbf{x}_0\|\to 0$, when $t\to 0$, $t>0$. Let us come back to (1.12) and write

$$\lim_{t\to 0}\frac{(T-U)(t\mathbf{x}_0)}{\|t\mathbf{x}_0\|}=\lim_{t\to 0}\frac{t\cdot(T-U)(\mathbf{x}_0)}{t\cdot\|\mathbf{x}_0\|}=0.$$

So, $(T-U)(\mathbf{x}_0)=0$ and we just obtained a contradiction. Hence, there is no $\mathbf{x}_0$ with $(T-U)(\mathbf{x}_0)\neq 0$ and so $T\equiv U$. □

Thus, if we find a method to compute $T=df(\mathbf{a})$, this T is unique. It depends only on f and on $\mathbf{a}$.

THEOREM 67. *If f is differentiable at $\mathbf{a}$, then all the partial derivatives $\frac{\partial f}{\partial x_1}$, $\frac{\partial f}{\partial x_2}$, ..., $\frac{\partial f}{\partial x_n}$ exists at $\mathbf{a}$ and*

$$df(\mathbf{a})(h_1,h_2,...,h_n)=\frac{\partial f}{\partial x_1}(\mathbf{a})h_1+\frac{\partial f}{\partial x_2}(\mathbf{a})h_2+...+\frac{\partial f}{\partial x_n}(\mathbf{a})h_n, \tag{1.13}$$

or, using the projection $pr_j=dx_j$ notation (see Remark 27), we get

$$df(\mathbf{a})=\frac{\partial f}{\partial x_1}(\mathbf{a})dx_1+\frac{\partial f}{\partial x_2}(\mathbf{a})dx_2+...+\frac{\partial f}{\partial x_n}(\mathbf{a})dx_n. \tag{1.14}$$

Moreover, if f is of class C^1 on a ball $B(\mathbf{a},r)$, for a small $r>0$, i.e. if $f\in C^1(B(\mathbf{a},r))$ (this means that f has partial derivatives with respect to all variables $x_1,x_2,...,x_n$ and all of these are continuous on $B(\mathbf{a},r)$), then f is differentiable at $\mathbf{a}$ and formula (1.14) works.

PROOF. We suppose that f is differentiable at $\mathbf{a}$ and let $T=df(\mathbf{a})$ be its differential at $\mathbf{a}$. We know from Linear Algebra or from the proof of Theorem 64 that

$$T(h_1,h_2,...,h_n)=\lambda_1h_1+\lambda_2h_2+...+\lambda_nh_n,$$

where $\lambda_1,\ \lambda_2,\ ...,\lambda_n$ are fixed real numbers (recall that $\lambda_i = T(\mathbf{e}_i)$, where $\mathbf{e}_i$ is the i-th vector of the canonical basis of $\mathbb{R}^n$, etc.). Let us chose now a j in $\{1,2,...,n\}$, let us take $\gamma > 0$, close to 0 and let us also take

$$\mathbf{h} = (0,0,...,0,\underbrace{\gamma}_{j},0,...,0)$$

in formula (1.9). We get

$$\lim_{\gamma\to 0}\frac{f(a_1,a_2,...,a_{j-1},a_j+\gamma,a_{j+1},...,a_n)-f(\mathbf{a})-\gamma\lambda_j}{\gamma} = 0.$$

Since this limit exists, the partial derivative with respect to j exists and, from this last formula we get that $\frac{\partial f}{\partial x_j}(\mathbf{a}) = \lambda_j$, for any $j \in \{1,2,...,n\}$. Hence,

$$T(h_1,h_2,...,h_n) = \frac{\partial f}{\partial x_1}(\mathbf{a})h_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})h_2 + ... + \frac{\partial f}{\partial x_n}(\mathbf{a})h_n$$

and the first part of the statement is completely proved.

Let us now assume that f is of class C^1 on a ball $B(\mathbf{a},r), r > 0$. Let us take the following linear mapping $T : \mathbb{R}^n \to \mathbb{R}$:

$$T(h_1,h_2,...,h_n) = \frac{\partial f}{\partial x_1}(\mathbf{a})h_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})h_2 + ... + \frac{\partial f}{\partial x_n}(\mathbf{a})h_n.$$

Let us prove that this T is indeed the differential of f at $\mathbf{a}$. To be easier, let us also assume that $n = 2$. Then, we want to prove that

$$\lim_{h_1,h_2\to 0}\frac{f(a_1+h_1,a_2+h_2)-f(a_1,a_2)-T(h_1,h_2)}{\|\mathbf{h}\|} = 0. \tag{1.15}$$

Let us write:

$$f(a_1+h_1,a_2+h_2)-f(a_1,a_2) = f(a_1+h_1,a_2+h_2)-f(a_1,a_2+h_2)$$

$$+f(a_1,a_2+h_2)-f(a_1,a_2). \tag{1.16}$$

Now, let us consider the function

$$\varphi_1(t) = f(t,a_2+h_2), t \in [a_1,a_1+h_1]^{\pm}$$

and let us apply to it Lagrange's formula:

$$f(a_1+h_1,a_2+h_2)-f(a_1,a_2+h_2) = \frac{\partial f}{\partial x_1}(c_1,a_2+h_2)\cdot h_1, \tag{1.17}$$

where $c_1 \in [a_1,a_1+h_1]^{\pm}$. Let us do the same for $f(a_1,a_2+h_2)-f(a_1,a_2)$ by considering the function

$$\varphi_2(t) = f(a_1,t), t \in [a_2,a_2+h_2]^{\pm}.$$

We get

$$f(a_1, a_2 + h_2) - f(a_1, a_2) = \frac{\partial f}{\partial x_2}(a_1, c_2) \cdot h_2, \tag{1.18}$$

where $c_2 \in [a_2, a_2+h_2]^{\pm}$. Let us come back in (1.16) with the expressions of (1.17) and (1.18). So,

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) - T(h_1, h_2)$$

$$= \left[\frac{\partial f}{\partial x_1}(c_1, a_2 + h_2) - \frac{\partial f}{\partial x_1}(a_1, a_2)\right] h_1 + \left[\frac{\partial f}{\partial x_2}(a_1, c_2) - \frac{\partial f}{\partial x_2}(a_1, a_2)\right] h_2. \tag{1.19}$$

Since the function f is of class C^1 in a small neighborhood of $\mathbf{a} = (a_1, a_2)$, one has that:

$$\left|\frac{\partial f}{\partial x_1}(c_1, a_2 + h_2) - \frac{\partial f}{\partial x_1}(a_1, a_2)\right| \to 0,$$

when $\mathbf{h} \to \mathbf{0}$ i.e. $h_1 \to 0$ and $h_2 \to 0$ and

$$\left|\frac{\partial f}{\partial x_2}(a_1, c_2) - \frac{\partial f}{\partial x_2}(a_1, a_2)\right| \to 0,$$

when $\mathbf{h} \to \mathbf{0}$. Since

$$\frac{|h_1|}{\|\mathbf{h}\|}, \frac{|h_2|}{\|\mathbf{h}\|} \leq 1,$$

one has that the limit in (1.15) is zero (do this slowly, step by step!). Hence, f is differentiable at $\mathbf{a}$ and its differential has the usual form:

$$df(\mathbf{a}) = \frac{\partial f}{\partial x_1}(\mathbf{a})dx_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})dx_2.$$

For an arbitrary n the proof is similar, but the writing is more complicated. □

This last theorem is very useful in computations. For instance, let $f : \mathbb{R}^3 \to \mathbb{R}$ be defined by

$$f(x, y, z) = \ln(1 + x^2 + y^4 + z^6).$$

All the partial derivatives

$$\frac{\partial f}{\partial x} = \frac{2x}{1 + x^2 + y^4 + z^6}, \frac{\partial f}{\partial y} = \frac{4y^3}{1 + x^2 + y^4 + z^6}$$

and

$$\frac{\partial f}{\partial z} = \frac{6z^5}{1 + x^2 + y^4 + z^6}$$

exist and are continuous on the whole $\mathbb{R}^3$, in particular around the point $(1,-1,2)$. Applying the last theorem (see Theorem 67) we see that f is differentiable at $(1,-1,2)$ and

$$df(1,-1,2)=\frac{\partial f}{\partial x}(1,-1,2)dx+\frac{\partial f}{\partial y}(1,-1,2)dy+\frac{\partial f}{\partial z}(1,-1,2)dz=$$

$$=\frac{2}{67}dx-\frac{4}{67}dy+\frac{192}{67}dz.$$

Recall a basic fact: $df(1,-1,2)$ is NOT a number, but a linear mapping from $\mathbb{R}^3$ to $\mathbb{R}$. For instance,

$$df(1,-1,2)(3,-4,0)=$$

$$=\frac{2}{67}dx(3,-4,0)-\frac{4}{67}dy(3,-4,0)+\frac{192}{67}dz(3,-4,0)=$$

$$=\frac{2}{67}\cdot 3-\frac{4}{67}\cdot(-4)+\frac{192}{67}\cdot 0=\frac{22}{67}.$$

This last one is a real number because $df(1,-1,2):\mathbb{R}^3\to\mathbb{R}$ is a linear mapping.

We want now to extend the notion of differentiability from scalar functions of n variables to vector functions.

DEFINITION 27. *Let $\mathbf{f}:D\to\mathbb{R}^m$ be a vector function with its components $(f_1,f_2,...,f_m)$, defined on an open subset D of $\mathbb{R}^n$ with values in $\mathbb{R}^m$. We say that f is differentiable at $\mathbf{a}\in D$ if all its components $f_1,f_2,...,f_m$ are differentiable at $\mathbf{a}$ like scalar functions. Moreover, if $\mathbf{h}=(h_1,h_2,...,h_n)$ is a vector in $\mathbb{R}^n$ and if*

$$df_i(\mathbf{a})(\mathbf{h})=a_{i1}h_1+a_{i2}h_2+...+a_{in}h_n,$$

where

$$a_{i1}=\frac{\partial f_i}{\partial x_1}(\mathbf{a}),a_{i2}=\frac{\partial f_i}{\partial x_2}(\mathbf{a}),...,a_{in}=\frac{\partial f_i}{\partial x_n}(\mathbf{a}),$$

then the matrix

$$J_{\mathbf{a},\mathbf{f}}=(a_{ij}=\frac{\partial f_i}{\partial x_j}(\mathbf{a})),$$

with m rows and n columns is called the Jacobi (or jacobian) matrix of f at $\mathbf{a}$. The linear mapping $\mathbf{T}:\mathbb{R}^n\to\mathbb{R}^m$ defined by the jacobian matrix $J_{\mathbf{a},\mathbf{f}}$ (with respect to the canonical bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively) is called the differential of $\mathbf{f}$ at $\mathbf{a}$. We write $\mathbf{T}=d\mathbf{f}(\mathbf{a})$. The determinant $|J_{\mathbf{a},\mathbf{f}}|$ of $J_{\mathbf{a},\mathbf{f}}$, in the particular case $n=m$, is said to be the jacobian of $\mathbf{f}$ at $\mathbf{a}$.

For instance,

$$\mathbf{f} : D \to \mathbb{R}^2, D = \{(x, y, z) \in \mathbb{R}^3 : x > 0, y > 0, z > 0\},$$

defined by

$$\mathbf{f}(x, y, z) = \left(\frac{1}{xyz}, xyz\right)$$

is differentiable at any point $\mathbf{a} =(a, b, c)$ of D because its components

$$f_1(x, y, z) = \frac{1}{xyz}$$

and

$$f_2(x, y, z) = xyz$$

have this last property (why?). Since

$$df_1(\mathbf{a}) = -\frac{1}{a^2bc}dx - \frac{1}{ab^2c}dy - \frac{1}{abc^2}dz$$

and

$$df_2(\mathbf{a}) = bc \cdot dx + ac \cdot dy + ab \cdot dz,$$

the jacobian matrix of $\mathbf{f}$ at $\mathbf{a}$ is the 2×3 matrix

$$\begin{pmatrix} -\frac{1}{a^2bc} & -\frac{1}{ab^2c} & -\frac{1}{abc^2} \\ bc & ac & ab \end{pmatrix}.$$

For instance, if $a = 1, b = 1$ and $c = -2$, we get the numerical matrix

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{4} \\ -2 & -2 & 1 \end{pmatrix}.$$

Now, if we want to compute the value of $df(1, 1, -2) : \mathbb{R}^3 \to \mathbb{R}^2$ at the point $(3, 4, -5)$, from Linear Algebra or from the remark 26, we get

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{4} \\ -2 & -2 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \\ -5 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} + \frac{4}{2} + \frac{5}{4} \\ -6 - 8 - 5 \end{pmatrix} = \begin{pmatrix} \frac{19}{4} \\ -19 \end{pmatrix},$$

so $df(1, 1, 2)(3, 4, -5) = (\frac{19}{4}, -19)$.

REMARK 28. *One can prove that* $\mathbf{f} : D \to \mathbb{R}^m$ *is differentiable at a point* $\mathbf{a} \in D \subset \mathbb{R}^n$ *if and only if there is a linear mapping* $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^m$ *which depends on* $\mathbf{a}$ *such that the following limit exists and is equal to zero:*

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|\mathbf{f}(\mathbf{a} + \mathbf{h}) - \mathbf{f}(\mathbf{a}) - \mathbf{T}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0. \tag{1.20}$$

We recall that

$$\|\mathbf{f}(\mathbf{a}+\mathbf{h})-\mathbf{f}(\mathbf{a})-\mathbf{T}(\mathbf{h})\| = \sqrt{\sum_{i=1}^{m}\left[f_i(\mathbf{a}+\mathbf{h})-f_i(\mathbf{a})-T_i(\mathbf{a})\right]^2}$$

and everything reduces to the scalar component functions, for which we know this result.

This above statement is equivalent to say that the increment

$$\mathbf{f}(\mathbf{a}+\mathbf{h})-\mathbf{f}(\mathbf{a})$$

of our vector function $\mathbf{f}$ *at* $\mathbf{a}$, *corresponding to the increment* $\mathbf{h}$ *of* $\mathbf{a}$, *can be "well" approximated by the value of the liner function* $\mathbf{T}$ *at* $\mathbf{h}$ *(do this slowly, step by step!). The uniqueness of the above* $\mathbf{T}$ *is obvious because its components are uniquely defined, being the differentials of some scalar functions, the components of* $\mathbf{f}$.

EXERCISE 1. *Let* $\mathbf{f},\mathbf{g}: D\to\mathbb{R}^m$, *be two differentiable functions on* D *(at any point of* D*), where* D *is an open subset in* $\mathbb{R}^n$ *and let* λ *be a real number. Then:* $\mathbf{f}+\mathbf{g}$, $\mathbf{f}-\mathbf{g}$, $\mathbf{fg}$ *(only for* $m=1$*)* $\frac{\mathbf{f}}{\mathbf{g}}$ *(only for* $m=1$ *and* $\mathbf{g}(\mathbf{a})\neq\mathbf{0}$*),* $\lambda\mathbf{f}$, *are also differentiable on* D *and*

a)

$$d(\mathbf{f}+\mathbf{g})(\mathbf{a}) = d\mathbf{f}(\mathbf{a})+d\mathbf{g}(\mathbf{a});$$

b)

$$d(\mathbf{f}-\mathbf{g})(\mathbf{a}) = d\mathbf{f}(\mathbf{a})-d\mathbf{g}(\mathbf{a});$$

c)

$$d(fg)(\mathbf{a}) = g(\mathbf{a})\cdot df(\mathbf{a})+f(\mathbf{a})\cdot dg(\mathbf{a});$$

d)

$$d(\frac{f}{g}) = \frac{g(\mathbf{a})\cdot df(\mathbf{a})-f(\mathbf{a})\cdot dg(\mathbf{a})}{g(\mathbf{a})^2};$$

e) $d(\lambda\mathbf{f})=\lambda\cdot d\mathbf{f}$ *for* $\lambda\in\mathbb{R}$.

In c) and d) f, g are only scalar functions!

2. Chain rules

Let A, B be two open subsets of $\mathbb{R}$ and let a be a point in A. Let $f: A\to B$ be a function defined on A with values in B such that f is differentiable at a. Let $g: B\to\mathbb{R}$ be a differentiable function at $f(a)$. Then the composed function $g\circ f: A\to\mathbb{R}$ is differentiable at a and

$$(g\circ f)'(a) = g'(f(a))\cdot f'(a)$$

(the simplest chain rule!). Indeed,

$$\lim_{x\to a}\frac{g(f(x))-g(f(a))}{x-a}=$$

$$=\lim_{f(x)\to f(a)}\frac{g(f(x))-g(f(a))}{f(x)-f(a)}\cdot\lim_{x\to a}\frac{f(x)-f(a)}{x-a}=g'(f(a))\cdot f'(a).$$

So $(g\circ f)'(a)$ exists and is exactly $g'(f(a))\cdot f'(a)$. In particular, if f is invertible and f^{-1} is differentiable at $b=f(a)$ then, from $f^{-1}(f(x))=x$, we get $f^{-1\prime}(b)\cdot f'(a)=1$, i.e. $f^{-1\prime}(b)=\frac{1}{f'(a)}$, or $(f^{-1})'(f(a))=\frac{1}{f'(a)}$.

We want now to generalize this simple chain rule to vector functions. Let us start with a simpler case, namely, let us take a "curve" $\mathbf{f}:A\to B$, $\mathbf{f}=(f_1,f_2,...,f_n)$, where A is an open subset in $\mathbb{R}$ and B is an open subset in $\mathbb{R}^n$. Let $g:B\to\mathbb{R}$ be a differential function at $\mathbf{b}=\mathbf{f}(a)$ and let us assume that $\mathbf{f}$ is differentiable at a. Let $h=g\circ\mathbf{f}:A\to\mathbb{R}$ be the composition between g and $\mathbf{f}$, i.e. the restriction of g to the n-D "curve" $\mathbf{f}$ (to the image of $\mathbf{f}$ in the common language!). Then, the following result is fundamental in applications.

THEOREM 68. *(differentiation along a curve) With the above notation and hypotheses,*

$$(g\circ\mathbf{f})'(a)=\frac{\partial g}{\partial x_1}(\mathbf{f}(a))\cdot f_1'(a)+\frac{\partial g}{\partial x_2}(\mathbf{f}(a))\cdot f_2'(a)+... \tag{2.1}$$

$$...+\frac{\partial g}{\partial x_n}(\mathbf{f}(a))\cdot f_n'(a).$$

For $n=1$ we find again the above formula $(g\circ f)'(a)=g'(f(a))\cdot f'(a)$.

PROOF. To be easier we take the particular case $n=2$ and we assume that $\mathbf{f}$ and g are functions of class C^1 on A and B respectively. Whenever we write limit of something or the derivative of a function, be sure that we implicitly prove that this limit or this derivative exists (prove this slowly in what follows!).

In this case, $h(x)=g(f_1(x),f_2(x))$ for any $x\in A$. So,

$$h'(a)=\lim_{x\to a}\frac{h(x)-h(a)}{x-a}=\lim_{x\to a}\frac{g(f_1(x),f_2(x))-g(f_1(a),f_2(a))}{x-a}=$$

$$=\lim_{x\to a}\frac{g(f_1(x),f_2(x))-g(f_1(a),f_2(x))}{x-a}+ \tag{2.2}$$

$$\lim_{x\to a}\frac{g(f_1(a),f_2(x))-g(f_1(a),f_2(a))}{x-a}.$$

Let us consider the first limit in (2.2) and let us apply Lagrange's formula (see Corollary 5) for the mapping $t \to g(f_1(t), f_2(x))$ on the interval $[a, x]$ (or $[x, a]$ if $x < a$). We get

$$g(f_1(x), f_2(x)) - g(f_1(a), f_2(x)) = \frac{\partial g}{\partial x_1}(f_1(c), f_2(x)) \cdot f_1'(c) \cdot (x - a),$$

where c is between a and x. Here we used our chain formula for $n = 1$ (where?-explain!). Coming back to the first limit in (2.2) and using the fact that $\frac{\partial g}{\partial x_1}$, f_1' and f_2 are continuous, we get:

$$\lim_{x \to a} \frac{g(f_1(x), f_2(x)) - g(f_1(a), f_2(x))}{x - a} = \lim_{x \to a} \frac{\partial g}{\partial x_1}(f_1(c), f_2(x)) \cdot f_1'(c) =$$

$$= \frac{\partial g}{\partial x_1}(f_1(a), f_2(a)) \cdot f_1'(a).$$

We take now the second limit in (2.2) and apply Lagrange's formula for the mapping $t \to g(f_1(a), f_2(t))$ on the same interval $[a, x]$. We get

$$g(f_1(a), f_2(x)) - g(f_1(a), f_2(a)) = \frac{\partial g}{\partial x_2}(f_1(a), f_2(s)) \cdot f_2'(s)) \cdot (x - a),$$

where s is a number between a and x. Since $\frac{\partial g}{\partial x_2}$, f_2 and f_2' are continuous (by our restrictive hypothesis in the present proof!), we obtain that

$$\lim_{x \to a} \frac{g(f_1(a), f_2(x)) - g(f_1(a), f_2(a))}{x - a} = \lim_{x \to a} \frac{\partial g}{\partial x_2}(f_1(a), f_2(s)) \cdot f_2'(s))$$

$$= \frac{\partial g}{\partial x_2}(f_1(a), f_2(a)) \cdot f_2'(a)),$$

thus our formula (2.1) is completely proved for $n = 2$. □

The statement of the theorem is true without these restrictions made here, but the proof is more sophisticated.

If the curve $\mathbf{f} : \mathbb{R} \to \mathbb{R}^3$ is a line which passes through the point $M_0(x_0, y_0, z_0)$ and having the direction of the versor

$$\mathbf{u} = (\cos\alpha, \cos\beta, \cos\gamma)$$

(these cosines are usually called the directional cosines of the line), i.e. $\mathbf{f}(t) = (x_0 + t\cos\alpha, y_0 + t\cos\beta, z_0 + t\cos\gamma)$, then, the above derivative

$$(g \circ \mathbf{f})'(0) = \frac{\partial g}{\partial x_1}(x_0, y_0, z_0)\cos\alpha + \frac{\partial g}{\partial x_2}(x_0, y_0, z_0))\cos\beta +$$

$$+ \frac{\partial g}{\partial x_3}(x_0, y_0, z_0)\cos\gamma = \langle \operatorname{grad} g(M_0), \mathbf{u} \rangle ,$$

(a scalar product!) is called the *directional derivative of g at the point M_0 along the versor* $\mathbf{u}$.

For instance, if $\mathbf{u} = (1, 0, 0)$, we get the partial derivative of g at M_0 with respect to x_1, etc.

We can now immediately extend the formula (2.1) for the case of a vector function $\mathbf{g} : B \to \mathbb{R}^m$, $\mathbf{g} = (g_1, g_2, ..., g_m)$. Thus, for any fixed $j \in \{1, 2, ..., m\}$, one has

(2.3)
$$(g_j \circ \mathbf{f})'(a) = \frac{\partial g_j}{\partial x_1}(\mathbf{f}(a)) \cdot f_1'(a) + \frac{\partial g_j}{\partial x_2}(\mathbf{f}(a)) \cdot f_2'(a) + ... + \frac{\partial g_j}{\partial x_n}(\mathbf{f}(a)) \cdot f_n'(a).$$

If we use now the matrix language, formula (2.3) becomes

(2.4)
$$\begin{pmatrix} (g_1 \circ \mathbf{f})'(a) \\ (g_2 \circ \mathbf{f})'(a) \\ . \\ . \\ . \\ (g_m \circ \mathbf{f})'(a) \end{pmatrix} =$$

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{f}(a)) & \frac{\partial g_1}{\partial x_2}(\mathbf{f}(a)) & . & . & . & \frac{\partial g_1}{\partial x_n}(\mathbf{f}(a)) \\ \frac{\partial g_2}{\partial x_1}(\mathbf{f}(a)) & \frac{\partial g_2}{\partial x_2}(\mathbf{f}(a)) & . & . & . & \frac{\partial g_2}{\partial x_n}(\mathbf{f}(a)) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \frac{\partial g_m}{\partial x_1}(\mathbf{f}(a)) & \frac{\partial g_m}{\partial x_2}(\mathbf{f}(a)) & . & . & . & \frac{\partial g_m}{\partial x_n}(\mathbf{f}(a)) \end{pmatrix} \cdot \begin{pmatrix} f_1'(a) \\ f_2'(a) \\ . \\ . \\ . \\ f_n'(a) \end{pmatrix}.$$

Up to now our function $\mathbf{f}$ was a function of one variable t. Let us make the last generalization and consider a vectorial function $\mathbf{f}$ of p variables $t_1, t_2, ..., t_p$ defined on an open subset A of $\mathbb{R}^p$. So we have the following composition: $A \xrightarrow{\mathbf{f}} B \xrightarrow{\mathbf{g}} \mathbb{R}^m$. We denote by $\mathbf{h} = \mathbf{g} \circ \mathbf{f} : A \to \mathbb{R}^m$ and preserve the notation $\mathbf{x} = (x_1, x_2, ..., x_n)$ for a point (vector!) in $\mathbb{R}^n$. Thus,

$$\mathbf{f}(t_1, t_2, ..., t_p) = (f_1(t_1, t_2, ..., t_p), f_2(t_1, t_2, ..., t_p), ..., f_n(t_1, t_2, ..., t_p))$$

and

$$\mathbf{g}(x_1, x_2, ..., x_n) = (g_1(x_1, x_2, ..., x_n), ..., g_m(x_1, x_2, ..., x_n)).$$

Let now $\mathbf{a}$ be a fixed point of A, $\mathbf{a} = (a_1, a_2, ..., a_p)$ and $\mathbf{b} = \mathbf{f}(\mathbf{a})$. We assume that $\mathbf{f}$ and $\mathbf{g}$ are differentiable at $\mathbf{a}$ and at $\mathbf{b}$ respectively.

THEOREM 69. *(chain rule theorem) With these notation and hypotheses, the composed function* $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ *is differentiable at* $\mathbf{a}$ *and one has the following relation between the corresponding jacobian matrices :*

(2.5)
$$J_{\mathbf{a},\mathbf{g}\circ\mathbf{f}} = J_{\mathbf{b},\mathbf{g}} \cdot J_{\mathbf{a},\mathbf{f}}.$$

This is the most sophisticated chain rule. Moreover, in this case, Linear Algebra says that

$$d(\mathbf{g}\circ\mathbf{f})(\mathbf{a}) = d\mathbf{g}(\mathbf{b})\circ d\mathbf{f}(\mathbf{a}), \tag{2.6}$$

this last composition being the composition between the corresponding linear mappings.

PROOF. Formula (2.6) is a direct consequence of formula (2.5) and the basic result of Linear Algebra which says that there is an isomorphic bijection between the $m\times n$ matrices and the linear mapping $T:\mathbb{R}^n\to\mathbb{R}^m$. This bijection carries the product between two matrices into the composition of the corresponding linear mappings. Hence, it remains us to prove formula (2.5). We shall see that this formula is a pure generalization of formula (2.4). Indeed, let us fix $i\in\{1,2,...,p\}$ and let us consider the mapping

$$\boldsymbol{\varphi}^{(i)}: A_i\to B, \boldsymbol{\varphi}^{(i)}=(\varphi_1^{(i)},\varphi_2^{(i)},...,\varphi_n^{(i)})$$

defined by

$$t\rightsquigarrow \mathbf{f}(a_1,a_2,...,a_{i-1},t,a_{i+1},...,a_p).$$

It is defined on the i-th projection $A_i=pr_i(A)$ of A (which is again open-why?). Let us denote $\mathbf{h}^{(i)}=\mathbf{g}\circ\boldsymbol{\varphi}^{(i)}$ and let us write formula (2.4) for it:

$$\begin{pmatrix}(g_1\circ\boldsymbol{\varphi}^{(i)})'(a_i)\\(g_2\circ\boldsymbol{\varphi}^{(i)})'(a_i)\\.\\.\\.\\(g_m\circ\boldsymbol{\varphi}^{(i)})'(a_i)\end{pmatrix}=$$

$$\begin{pmatrix}\frac{\partial g_1}{\partial x_1}(\boldsymbol{\varphi}^{(i)}(a_i)) & \frac{\partial g_1}{\partial x_2}(\boldsymbol{\varphi}^{(i)}(a_i)) & . & . & . & \frac{\partial g_1}{\partial x_n}(\boldsymbol{\varphi}^{(i)}(a_i))\\ \frac{\partial g_2}{\partial x_1}(\boldsymbol{\varphi}^{(i)}(a_i)) & \frac{\partial g_2}{\partial x_2}(\boldsymbol{\varphi}^{(i)}(a_i)) & . & . & . & \frac{\partial g_2}{\partial x_n}(\boldsymbol{\varphi}^{(i)}(a_i))\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ . & . & . & . & . & .\\ \frac{\partial g_m}{\partial x_1}(\boldsymbol{\varphi}^{(i)}(a_i)) & \frac{\partial g_m}{\partial x_2}(\boldsymbol{\varphi}^{(i)}(a_i)) & . & . & . & \frac{\partial g_m}{\partial x_n}(\boldsymbol{\varphi}^{(i)}(a_i))\end{pmatrix}.$$

$$\cdot \begin{pmatrix} \left[\varphi_1^{(i)}\right]'(a_i) \\ \left[\varphi_2^{(i)}\right]'(a_i) \\ \cdot \\ \cdot \\ \cdot \\ \left[\varphi_n^{(i)}\right]'(a_i) \end{pmatrix}. \tag{2.7}$$

We now see that

$$(g_j \circ \boldsymbol{\varphi}^{(i)})'(a_i) = \frac{\partial h_j}{\partial t_i}(\mathbf{a})$$

for any $j = 1, 2, ..., m$ and $i = 1, 2, ..., p$. Here $\mathbf{h} = (h_1, h_2, ..., h_m)$ are the components of the composed function $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$.

Another remark is that

$$\frac{\partial g_j}{\partial x_k}(\boldsymbol{\varphi}^{(i)}(a_i)) = \frac{\partial g_j}{\partial x_k}(\mathbf{f}(\mathbf{a}))$$

and $\left[\varphi_j^{(i)}\right]'(a_i) = \frac{\partial f_j}{\partial t_i}(\mathbf{a})$. But, if we substitute all of these in formula (2.7), we get exactly formula (2.5) from the statement of the theorem. □

REMARK 29. *It is possible to prove the chain rule theorem, namely the formula (2.6), in a not so long "upgrading" way. But that proof (see* [**Nik**]*, or* [**Pal**]*) is more abstract, more elaborated and not so natural. Our proof here is not so general, but it follows the natural historical way, from a "simpler" to a "more complicated" case.*

Let us take an usual situation and let us apply formula (2.5) to it. Let A and B be two open subsets of $\mathbb{R}^2$ and let $(x, y) \rightsquigarrow (u(x, y), v(x, y))$ be a differentiable (at any point of A) vector function defined on A with values in B. Let $f(u, v)$ be a differentiable function defined on B with values in $\mathbb{R}$. Here we also use u and v for the coordinates of a free vector in $B \subset \mathbb{R}^2$. The only connection between u, v and the functions of two variables $u(x, y)$ and $v(x, y)$ respectively, is that the variable u and v are substituted with two functions $u(x, y)$ and $v(x, y)$ respectively, in variables x and y. For instance, $u = x + y$, $v = xy$ and $f(x + y, xy)$. This is a new function in x and y. Here, $u(x, y) = x + y$ and $v(x, y) = xy$. This abuse of notation is still working for more then 200 years and it did not caused any damage in science. Let $h(x, y) = f(u(x, y), v(x, y))$ be the composition between f and the first function $(x, y) \to (u(x, y), v(x, y))$. This new function is also denoted by f, i.e. the notation $f(x, y) = f(u(x, y), v(x, y))$ produce no confusion for an

working mathematician (another abuse, which is not indicated to be used by a beginner!). The function h is also differentiable on A and

$$\left(\frac{\partial h}{\partial x}(a,b) \quad \frac{\partial h}{\partial y}(a,b)\right) =$$

$$\left(\frac{\partial f}{\partial u}(u(a,b),v(a,b)) \quad \frac{\partial f}{\partial v}(u(a,b),v(a,b))\right) \cdot \begin{pmatrix} \frac{\partial u}{\partial x}(a,b) & \frac{\partial u}{\partial y}(a,b) \\ \frac{\partial v}{\partial x}(a,b) & \frac{\partial v}{\partial y}(a,b) \end{pmatrix}.$$

Let us normally write this formula:

(2.8)

$$\frac{\partial h}{\partial x}(a,b) = \frac{\partial f}{\partial u}(u(a,b),v(a,b))\frac{\partial u}{\partial x}(a,b) + \frac{\partial f}{\partial v}(u(a,b),v(a,b))\frac{\partial v}{\partial x}(a,b),$$

$$\frac{\partial h}{\partial y}(a,b) = \frac{\partial f}{\partial u}(u(a,b),v(a,b))\frac{\partial u}{\partial y}(a,b) + \frac{\partial f}{\partial v}(u(a,b),v(a,b))\frac{\partial v}{\partial y}(a,b),$$

How do we recall these useful formulas? For this, write again $h(x,y) = f(u(x,y),v(x,y))$. To find $\frac{\partial h}{\partial x}$, we look at the variables u and v of f and observe where x is. If x appears in $u = u(x,y)$, we take the partial derivative of f w.r.t. u and multiply it by the partial derivative of u w.r.t. x. Here is a "chain": $f \to u \to x$. So we get $\frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial x}$. If x also appears in $v = v(x,y)$, we consider the chain $f \to v \to x$ and obtain $\frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial x}$. Since x appears both (if it is the case!) in u and in v, we must superpose both "effects" (add them!) and finally obtain:

$$\frac{\partial h}{\partial x} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial x} + \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial x}. \tag{2.9}$$

The corresponding points at which we compute these partial derivatives are easy to be find. If we change x with y in (2.9) we get the second essential formula of (2.8):

$$\frac{\partial h}{\partial y} = \frac{\partial f}{\partial u} \cdot \frac{\partial u}{\partial y} + \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial y}. \tag{2.10}$$

EXAMPLE 14. *In the Cartesian plane $\{O;\mathbf{i},\mathbf{j}\}$, we consider a heating source in the origin $O(0,0)$. The temperature $f(x,y)$ at the point $M(x,y)$ verifies the following equation (a partial differential equation of order 1− a PDE-1):*

$$y\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial y} = 0.$$

It says that at any point $M(x,y)$ the "gradient" vector

$$gradf = \left(\frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)\right)$$

of the temperature is perpendicular to the normal vector of the position vector $\overrightarrow{OM} = x\mathbf{i} + y\mathbf{j}$, *at the point* $M(x, y)$. *Hence, gradf is colinear to* $\overrightarrow{OM}$. *Let us change the variables* x *and* y *with* $u = x$ *and* $v = x^2 + y^2$. *The new function* $h(u, v)$ *is connected to* f *by the rule:*

$$f(x, y) = h(x, x^2 + y^2).$$

So,

$$\frac{\partial f}{\partial x} = \frac{\partial h}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial h}{\partial v}\frac{\partial v}{\partial x} = \frac{\partial h}{\partial u} + 2x\frac{\partial h}{\partial v}$$

and

$$\frac{\partial f}{\partial y} = \frac{\partial h}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial h}{\partial v}\frac{\partial v}{\partial y} = 2y\frac{\partial h}{\partial v}.$$

Hence,

$$0 = y\frac{\partial f}{\partial x} - x\frac{\partial f}{\partial y} = y\frac{\partial h}{\partial u} + 2xy\frac{\partial h}{\partial v} - 2xy\frac{\partial h}{\partial v} = y\frac{\partial h}{\partial u}.$$

Hence, whenever $y \neq 0$, $\frac{\partial h}{\partial u} = 0$ *is the equation in the new function* h. *So* h *is a function of* $v = x^2 + y^2$, *the square of the distance up to origin. Thus, the temperature is constant at all the points which are of the same circle of radius* $r > 0$. *We say that the level curves* ($f(x, y) =$ *constant) of the temperature are all the concentric circles with center at* O.

We must apply the "spirit" of the formulas (2.5) or (2.10), not the formulas themselves. For instance, let

$$\mathbf{f}(x, y, z) = (\sin(x^2 + y^2), \cos(2z^2), x^2 + y^2 + z^2).$$

Then,

$$\frac{\partial \mathbf{f}}{\partial x} = (2x\cos(x^2 + y^2), 0, 2x), \frac{\partial \mathbf{f}}{\partial y} = (2y\cos(x^2 + y^2), 0, 2y)$$

and

$$\frac{\partial \mathbf{f}}{\partial z} = (0, -4z\sin(2z^2), 2z).$$

If we want to compute $\frac{\partial \mathbf{f}}{\partial x}(1, -1, 7)$ we simply put $x = 1, y = -1$ and $z = 7$ in the expression of $\frac{\partial \mathbf{f}}{\partial x}$. So,

$$\frac{\partial \mathbf{f}}{\partial x}(1, -1, 7) = (2\cos 2, 0, 2).$$

Here $\cos 2$ means the cosinus of two radians.

EXAMPLE 15. *Let $M(x(t), y(t), z(t))$, t is time, $t \in (a, b)$, $a \geq 0$, be a moving point of mass $m = 5Kg$ on the curve*

$$\Gamma : x = x(t), y = y(t), z = z(t).$$

Let

$$\mathbf{v}(t) = (x'(t), y'(t), z'(t))$$

and

$$\mathbf{w}(t) = (x''(t), y''(t), z''(t))$$

be the velocity and the acceleration respectively. We assume that the kinetic energy

$$T = \frac{5}{2}\left\{[x'(t)]^2 + [y'(t)]^2 + [z'(t)]^2\right\}$$

does not depend on time, i.e. $T'(t) \equiv 0$. Let us use the chain rule to make the computation in this last equality:

$$T'(t) = 5\left\{[x'(t)]\,[x''(t)] + [y'(t)]\,[y''(t)] + [z'(t)]\,[z''(t)]\right\} = 0,$$

i.e. the scalar (inner) product between $\mathbf{v}$ and $\mathbf{w}$ is equal to zero. In this case, the acceleration is perpendicular on the velocity. This restriction is very useful in physical considerations.

DEFINITION 28. *A subset K of $\mathbb{R}^n$ is said to be a conic subset if for any $\mathbf{x}$ in K and any $t \in \mathbb{R}$, one has that $t\mathbf{x} \in K$ (see Fig.7.1).*

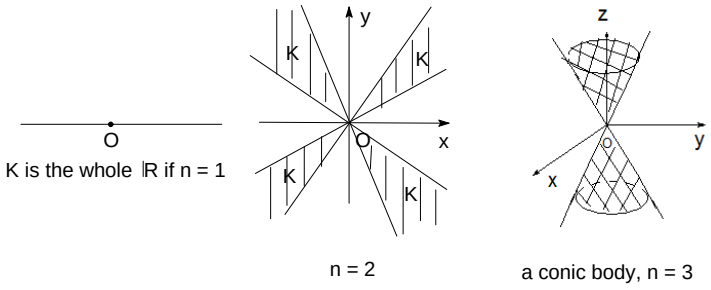


Fig. 7.1

For instance,

$$K = \mathbb{R}^n, K = \{(x, y) \in \mathbb{R}^2 : y = mx\},$$

where m is a fixed parameter (real number)},

$$K = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = z^2\}$$

are conic subsets (prove it!).

DEFINITION 29. *Let $f : K \to \mathbb{R}$, be a function defined on a conic subset $K \subset \mathbb{R}^n$ with values in $\mathbb{R}$ and let α be a fixed real number. We say that f is homogeneous of degree α if*

$$f(tx_1, tx_2, ..., tx_n) = t^\alpha f(x_1, x_2, ..., x_n), \tag{2.11}$$

for any $\mathbf{x} = (x_1, x_2, ..., x_n)$ in K and for any t in $\mathbb{R}_+$.

For instance, the distance to origin function

$$d(x, y, z) = \sqrt{x^2 + y^2 + z^2}$$

is a homogeneous function of degree 1. Indeed,

$$d(tx, ty, tz) = \sqrt{(tx)^2 + (ty)^2 + (tz)^2} = t\sqrt{x^2 + y^2 + z^2} = td(x, y, z).$$

L. Euler introduced these functions when he studied the mechanics of a moving point in plane. For $\alpha = 0$, we simply call these functions *homogeneous*. Euler discovered a very useful property for homogeneous functions. In the following we consider a generalization of the Euler's result.

THEOREM 70. *(Euler formula for homogeneous functions) Let K be a conic open subset in $\mathbb{R}^n$ and let f be a function of class C^1 on K, which is homogeneous of degree α. Then,*

$$x_1 \frac{\partial f}{\partial x_1}(\mathbf{x}) + x_2 \frac{\partial f}{\partial x_2}(\mathbf{x}) + ... + x_n \frac{\partial f}{\partial x_n}(\mathbf{x}) = \alpha \cdot f(\mathbf{x}). \tag{2.12}$$

PROOF. By the definition of a homogeneous function (Definition 29), we may look at the formula (2.11) and differentiate everything w.r.t. t (here we use the chain rule...explain slowly this...)

$$x_1 \frac{\partial f}{\partial x_1}(t\mathbf{x}) + x_2 \frac{\partial f}{\partial x_2}(t\mathbf{x}) + ... + x_n \frac{\partial f}{\partial x_n}(t\mathbf{x}) = \alpha t^{\alpha-1} \cdot f(\mathbf{x}).$$

We now make $t = 1$ in this last formula and obtain Euler formula (2.12). □

If $\alpha = 0$, i.e. if our function is homogeneous, Euler formula can be written as

$$\langle \mathbf{x}, \operatorname{grad} f(\mathbf{x}) \rangle = 0. \tag{2.13}$$

Here $\langle , \rangle$ is the (inner) scalar product in $\mathbb{R}^n$. This last formula (2.13) says that at any point $\mathbf{x}$ of the trajectory of a moving point in $\mathbb{R}^n$, the gradient (a generalization of the velocity for n variables!) of f is perpendicular on the position vector $\mathbf{x}$. For instance, we know that the temperature $T(x, y)$ in any point (x, y) of the plane $\mathbb{R}^2$ is the same for all the points of an arbitrary line $y = mx$, where m runs freely on $\mathbb{R}$. This means (in mathematical language) that $T(tx, ty) = T(x, y)$ for

any $(x, y) \in \mathbb{R}^2$ and any t in $\mathbb{R}_+$ (why?). So, the temperature is a homogeneous function and we can write the Euler's formula for $\alpha = 0$, i.e. $\langle \mathbf{x}, \operatorname{grad} T(\mathbf{x}) \rangle = 0$, where $\mathbf{x} = (x, y)$ and

$$\operatorname{grad} T(x, y) = \left(\frac{\partial T}{\partial x}(x, y), \frac{\partial T}{\partial y}(x, y) \right).$$

Finally we get the following PDE of order 1 :

$$x \frac{\partial T}{\partial x}(x, y) + y \frac{\partial T}{\partial y}(x, y) = 0,$$

i.e. in any point the gradient of the temperature is perpendicular on the position vector (x, y).

In exercises, one usually asks to verify Euler's formula for a given homogeneous function f. For instance, let us verify Euler's formula for $f(x, y, z) = xyz + 3x^3 + y^3$. We do not know yet if the function f is homogeneous and, if it is so, we also do not know the homogeneity degree of it. Let us put instead of x, y and z, tx, ty, and tz respectively:

$$f(tx, ty, tz) = t^3(xyz + 3x^3 + y^3) = t^3 f(x, y, z).$$

Thus, our function is homogeneous of degree 3. So we have to verify the following formula:

$$x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} + z \frac{\partial f}{\partial z} = 3f. \tag{2.14}$$

Indeed, $\frac{\partial f}{\partial x} = yz + 9x^2$; $\frac{\partial f}{\partial y} = xz + 3y^2$ and $\frac{\partial f}{\partial z} = xy$. Substituting in (2.14), we get:

$$x(yz + 9x^2) + y(xz + 3y^2) + zxy = 3(xyz + 3x^3 + y^3) = 3f.$$

Hence, we just verified Euler's formula for our particular function.

3. Problems

1. Compute the following partial derivatives:

a)

$$f(x, y) = \sqrt{x^2 + y^2}; \frac{\partial f}{\partial x}(1, 1), \frac{\partial^2 f}{\partial x \partial y}(1, 1).$$

b)

$$f(x, y) = \sqrt{\sin^2 x + \sin^2 y}; \frac{\partial f}{\partial x}(\frac{\pi}{4}, 0), \frac{\partial f}{\partial y}(\frac{\pi}{4}, \frac{\pi}{4}).$$

c)

$$f(x, y) = \ln(x + y^2 - 1); \frac{\partial f}{\partial x}(1, 1), \frac{\partial^2 f}{\partial y^2}(1, 1).$$

d)

$$f(x,y) = x\exp(xy); \frac{\partial^2 f}{\partial x \partial y}(1,0), \frac{\partial^2 f}{\partial x^2}(1,0), \frac{\partial^2 f}{\partial y^2}(1,0).$$

e)

$$f(x,y) = x^{\ln y}(x>0, y>0), \frac{\partial f}{\partial x}(e,e), \frac{\partial f}{\partial y}(e,e), \frac{\partial^2 f}{\partial x \partial y}(e,e).$$

f)

$$f(x,y,z) = x^{y^z}(x>0, y>0), \operatorname{grad} f(1,1,1).$$

g)

$$f(x,y) = \arctan xy, \frac{\partial^3 f}{\partial y \partial x^2}(1,1), \frac{\partial^3 f}{\partial x \partial y^2}(1,1), \frac{\partial^3 f}{\partial x^3}(1,1).$$

h)

$$f(x,y) = \arcsin(\frac{x}{y}), \frac{\partial^2 f}{\partial y \partial x}(1,2).$$

2. Prove that the following functions verify the indicated equations:

a)

$$z(x,y) = xy\Phi(x^2-y^2); xy^2\frac{\partial z}{\partial x} + x^2 y\frac{\partial z}{\partial y} = (x^2+y^2)z.$$

b)

$$z(x,y) = x\Phi(x^2-y^2); \frac{1}{x}\frac{\partial z}{\partial x} + \frac{1}{y}\frac{\partial z}{\partial y} = \frac{z}{y^2}.$$

c)

$$u(x,y) = \arctan\frac{y}{x}; \Delta u \overset{def}{=} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

d)

$$u(x,t) = \Phi(x-at) + \Psi(x+at); \frac{\partial^2 u}{\partial t^2} - a^2\frac{\partial^2 u}{\partial x^2} = 0$$

(the wave equation).

e)

$$z(x,y) = x\Phi(\frac{y}{x}) + \Psi(\frac{y}{x}); x^2\frac{\partial^2 z}{\partial x^2} + 2xy\frac{\partial^2 z}{\partial x \partial y} + y^2\frac{\partial^2 z}{\partial y^2} = 0.$$

f)

$$u(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}; \Delta u \overset{def}{=} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0.$$

Hint: Let us denote $r = \sqrt{x^2+y^2+z^2}$. Then, $\frac{\partial u}{\partial x} = -\frac{1}{r^2}\cdot\frac{\partial r}{\partial x}$, etc.

3. Show that the Euler's formula is true for the following homogeneous functions:

a) $f(x,y)=\frac{x+y}{x-y}$;

b)

$$f(x,y,z)=\sqrt{x}+\sqrt{y}+\sqrt{z};$$

c)

$$f(x,y,z)=\sqrt{x^2+y^2+z^2};$$

d) $f(x,y,z)=\frac{x}{y}\exp(\frac{x}{z})$.

4. Prove that the following function

$$f(x,y)=\begin{cases}\frac{xy}{\sqrt{x^2+y^2}},\ \text{for } (x,y)\neq(0,0)\\ 0,\ \text{if } x=0 \text{ and } y=0\end{cases}$$

is continuous, has partial derivatives, but it is not differentiable at $(0,0)$ (Hint: $\frac{|xy|}{\sqrt{x^2+y^2}}\leq|y|$, so

$$\lim_{x\to0,y\to0}\frac{xy}{\sqrt{x^2+y^2}}=0,\frac{\partial f}{\partial x}(0,0)=\frac{\partial f}{\partial y}(0,0)=0.$$

If it was differentiable at $(0,0)$ one has that

$$f(h_1,h_2)-f(0,0)=\frac{\partial f}{\partial x}(0,0)h_1+\frac{\partial f}{\partial y}(0,0)h_2+\omega(h_1,h_2), \tag{3.1}$$

where $\omega(0,0)=0$, ω is continuous at $(0,0)$ and

$$\lim_{x\to0,y\to0}\frac{\omega(x,y)}{\sqrt{x^2+y^2}}=0.$$

But, from (3.1), one has that $\omega(x,y)=\frac{xy}{\sqrt{x^2+y^2}}$ and so one would have that

$$\lim_{x\to0,y\to0}\frac{xy}{x^2+y^2}=0.$$

However, this last limit does not exist at all!!).

CHAPTER 8

Taylor's formula for several variables.

1. Higher partial derivatives. Differentials of order k.

Let $\frac{\partial f}{\partial x}$ be the partial derivative with respect to x of a function $f : A \to \mathbb{R}$, where A is an open subset in $\mathbb{R}^2$. $(x, y) \rightsquigarrow \frac{\partial f}{\partial x}(x, y)$ is a new function of two variables x and y. If this new function has a partial derivative $\frac{\partial}{\partial x}(\frac{\partial f}{\partial x})(a, b)$ w.r.t. x, at a point (a, b), we denote it by $\frac{\partial^2 f}{\partial x^2}(a, b)$ and say " d two f over d x two at $(a,\ b)$". If the same function $(x, y) \rightsquigarrow \frac{\partial f}{\partial x}(x, y)$ has a partial derivative $\frac{\partial}{\partial y}(\frac{\partial f}{\partial x})(a, b)$ w.r.t. y, at a point (a, b), we write it as $\frac{\partial^2 f}{\partial y \partial x}(a, b)$ and call it the mixed derivative of f at (a, b). What do we mean by $\frac{\partial^3 f}{\partial x \partial y^2}$ (say "d three f over d x d y two"; pay attention to the fact that 3 from ∂^3 is equal to the sum between 1 and 2, from ∂x and ∂y^2 respectively). In general, let $f : A \to \mathbb{R}$, $f(x_1, x_2, ..., x_n)$ be a function of n variables, defined on an open subset A of $\mathbb{R}^n$, such that it is k_n-times differentiable with respect to x_n, i.e. $\frac{\partial^{k_n} f}{\partial x_n^{k_n}}$ exists on A. If this new function

$$\mathbf{x} = (x_1, x_2, ..., x_n) \rightsquigarrow \frac{\partial^{k_n} f}{\partial x_n^{k_n}}(\mathbf{x})$$

is k_{n-1}-times differentiable with respect to x_{n-1}, the new obtained function

$$\mathbf{x} \rightsquigarrow \frac{\partial^{k_{n-1}}}{\partial x_{n-1}^{k_{n-1}}} \left(\frac{\partial^{k_n} f}{\partial x_n^{k_n}} \right) (\mathbf{x})$$

is denoted by $\frac{\partial^{k_n + k_{n-1}} f}{\partial x_{n-1}^{k_{n-1}} \partial x_n^{k_n}}$. And so on. We finally obtain the function $\frac{\partial^{k_n + k_{n-1} + ... + k_1} f}{\partial x_1^{k_1} ... \partial x_{n-1}^{k_{n-1}} \partial x_n^{k_n}}$. The order of variables $x_1,\ x_2, ..., x_n$ in the denominator can be changed, but then we may obtain another new function. For instance, if $f(x, y, z) = x^4 y^3 z^5$, then $\frac{\partial^5 f}{\partial y^2 \partial x^2 \partial z}$ can be successively computed. First of all we compute

$$g_1 = \frac{\partial f}{\partial z} = 5x^4 y^3 z^4.$$

Then we compute

$$g_2 = \frac{\partial g_1}{\partial x} = \frac{\partial^2 f}{\partial x \partial z} = 20x^3y^3z^4.$$

Now we compute

$$g_3 = \frac{\partial g_2}{\partial x} = \frac{\partial^3 f}{\partial x^2 \partial z} = 60x^2y^3z^4.$$

Then we consider

$$g_4 = \frac{\partial g_3}{\partial y} = \frac{\partial^4 f}{\partial y \partial x^2 \partial z} = 180x^2y^2z^4.$$

Finally,

$$g_5 = \frac{\partial g_4}{\partial y} = \frac{\partial^5 f}{\partial y^2 \partial x^2 \partial z} = 360x^2yz^4.$$

And this last one is our final result.

$\frac{\partial^{k_n+k_{n-1}+\ldots+k_1} f}{\partial x_1^{k_1} \ldots \partial x_{n-1}^{k_{n-1}} \partial x_n^{k_n}}$ is said to be the partial $k = k_n + k_{n-1} + \ldots + k_1$ derivative of f, k_n-times w.r.t. x_n, k_{n-1}-times w.r.t. $x_{n-1}, \ldots$, and k_1-times w.r.t. x_1. The mapping $f \rightsquigarrow \frac{\partial f}{\partial x_j}$ is also denoted by $D_{x_j} f$. This D_{x_j} is called the partial differential operator w.r.t. the variable x_j. So, $f \rightsquigarrow \frac{\partial^2 f}{\partial x_i \partial x_j}$ is the composition $D_{x_i} \circ D_{x_j}$ applied to f. In general, a mapping defined on a set of functions is called not a function more, but an *operator*. We also put $D_{x_i x_j}$ instead of $D_{x_i} \circ D_{x_j}$. Such an operator is called a *differential operator*. In general, the operators D_{x_i} and D_{x_j} do not commute if $i \neq j$. This means that there are examples of functions f and points $\mathbf{a}$ for which $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) \neq \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})$. Following [**Pal**], p. 145, we consider

$$f(x,y) = \begin{cases} xy\frac{x^2-y^2}{x^2+y^2}, \text{ if } (x,y) \neq (0,0) \\ \\ 0, \text{ if } x = 0, y = 0. \end{cases} \tag{1.1}$$

It is not difficult to prove that $\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1$, but $\frac{\partial^2 f}{\partial x \partial y}(0,0) = 1$ (do it step by step and explain everything!). Hence, in this case we cannot commute the order of derivation!

Let A be an open subset of $\mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a function of n variable defined on A. We say that f is of class C^2 on A if all the partial derivatives of order two, $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})$, exist and are continuous, at any point $\mathbf{a}$ of A. The following theorem gives us a sufficient condition under which the change of order of derivation has no influence on the final result.

THEOREM 71. *(Schwarz' Theorem) Let $f : A \to \mathbb{R}$ be a function of class C^2 on A. Then*

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

for any point $\mathbf{a}$ of A and for any pair (i, j). This means that for such a function (of class C^2 on A) we can commute the order of derivation.

PROOF. One can reduce everything to the two variables case (why?). Moreover, we can take an open ball (disc) $B(\mathbf{a}, r)$, $r > 0$, $\mathbf{a} = (a_1, a_2)$, included in A and consider f defined on this ball $B(\mathbf{a}, r)$. Let $\{(x_n, y_n)\}$ be a sequence of points in $B(\mathbf{a}, r)$ which converges to $\mathbf{a}$. For a fixed natural number n let us consider the segments $[a_1, x_n]$ and $[a_2, y_n]$ in $B(\mathbf{a}, r)$. Let

$$R(x_n, y_n) = f(x_n, y_n) - f(x_n, a_2) - f(a_1, y_n) + f(a_1, a_2) \tag{1.2}$$

and let $g(t) = f(t, y_n) - f(t, a_2)$, $t \in [a_1, x_n]$. Let us apply Lagrange's theorem (see Corollary 5) to function g on $[a_1, x_n]$:

$$g(x_n) - g(a_1) = g'(c_n) \cdot (x_n - a_1),$$

where $c_n \in [a_1, x_n]$. But

$$g(x_n) - g(a_1) = R(x_n, y_n)$$

and

$$g'(c_n) = \frac{\partial f}{\partial x}(c_n, y_n) - \frac{\partial f}{\partial x}(c_n, a_2).$$

So,

$$R(x_n, y_n) = \left[\frac{\partial f}{\partial x}(c_n, y_n) - \frac{\partial f}{\partial x}(c_n, a_2)\right](x_n - a_1).$$

Now we apply again Lagrange's theorem to the function

$$u \to \frac{\partial f}{\partial x}(c_n, u),$$

where $u \in [a_2, y_n]$. Hence,

$$R(x_n, y_n) = \frac{\partial^2 f}{\partial y \partial x}(c_n, d_n) \cdot (x_n - a_1)(y_n - a_2), \tag{1.3}$$

where $d_n \in [a_2, y_n]$. Now we take a new function

$$h(t) = f(x_n, t) - f(a_1, t),$$

$t \in [a_2, y_n]$ and observe that

$$R(x_n, y_n) = h(y_n) - h(a_2).$$

Let us apply Lagrange's theorem to h on $[a_2, y_n]$:

$$R(x_n, y_n) = h'(e_n) \cdot (y_n - a_2), \tag{1.4}$$

where $e_n \in [a_2, y_n]$. But $h'(e_n) = \frac{\partial f}{\partial y}(x_n, e_n) - \frac{\partial f}{\partial y}(a_1, e_n)$ so, applying again Lagrange's theorem to the function:

$$v \to \frac{\partial f}{\partial y}(v, e_n),$$

where $v \in [a_1, x_n]$, we get:

$$h'(e_n) = \frac{\partial^2 f}{\partial x \partial y}(s_n, e_n) \cdot (x_n - a_1),$$

where $s_n \in [a_1, x_n]$. Hence,

$$R(x_n, y_n) = \frac{\partial^2 f}{\partial x \partial y}(s_n, e_n) \cdot (x_n - a_1)(y_n - a_2). \tag{1.5}$$

Comparing the formulas (1.3) and (1.5), we get:

$$\frac{\partial^2 f}{\partial y \partial x}(c_n, d_n) = \frac{\partial^2 f}{\partial x \partial y}(s_n, e_n). \tag{1.6}$$

Since the functions $\frac{\partial^2 f}{\partial y \partial x}$ and $\frac{\partial^2 f}{\partial x \partial y}$ are continuous on A, since $\{c_n\}, \{s_n\} \to a_1$ and since $\{d_n\}, \{e_n\} \to a_2$ (why?), from formula (1.6), we get:

$$\frac{\partial^2 f}{\partial y \partial x}(a_1, a_2) = \frac{\partial^2 f}{\partial x \partial y}(a_1, a_2).$$

Hence, the proof of the theorem is complete. □

In (1.1)

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1 \neq \frac{\partial^2 f}{\partial x \partial y}(0,0) = 1,$$

because $\frac{\partial^2 f}{\partial y \partial x}$ is not continuous at $(0,0)$. Indeed,

$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = \begin{cases} \frac{x^6 - y^6 - 9x^2y^4 - 15x^4y^2}{(x^2+y^2)^3}, \text{ if } (x,y) \neq (0,0) \\ -1, \text{ if } x = 0, y = 0. \end{cases},$$

and this last function has no limit at $(0,0)$. This is because, if we take an arbitrary m and consider (x,y) with $y = mx$, we get that

$$\lim_{x \to 0, y = mx} \frac{x^6 - y^6 - 9x^2y^4 - 15x^4y^2}{(x^2+y^2)^3} = \frac{1 - 25m^6}{(1+m^2)^3},$$

which is dependent on m. So, the limit at $(0,0)$ is not a unique number. It depends on the direction on which we come to $(0,0)$. All of these happen because the function

$$\frac{x^6 - y^6 - 9x^2y^4 - 15x^4y^2}{(x^2+y^2)^3}$$

is homogeneous of degree 0 (make clear this for yourself!)

In engineering, the case of functions of class C^2 is mostly frequent, thus we assume in the following that the order of derivation does not matter. For instance, $f(x,y) = 4x^3y^2 + 2x^2y$ is of class C^∞ on $\mathbb{R}^2$ (why?). In particular, it is of class C^2 because C^∞ means that f has partial derivatives of any order (so these derivatives are continuous-why?). Schwarz' theorem says that

$$\frac{\partial^2 f}{\partial x \partial y}(a,b) = \frac{\partial^2 f}{\partial y \partial x}(a,b)$$

for any point (a,b) in $\mathbb{R}^2$. Indeed,

$$\frac{\partial^2 f}{\partial x \partial y}(a,b) = \frac{\partial}{\partial x}(\frac{\partial f}{\partial y})(a,b) = \frac{\partial}{\partial x}(8x^3y + 2x^2) \mid_{(a,b)} =$$

$$= 24x^2y + 4x \mid_{(a,b)} = 24a^2b + 4a$$

and

$$\frac{\partial^2 f}{\partial y \partial x}(a,b) = \frac{\partial}{\partial y}(\frac{\partial f}{\partial x})(a,b) = \frac{\partial}{\partial y}(12x^2y^2 + 4xy) \mid_{(a,b)} =$$

$$= 24x^2y + 4x \mid_{(a,b)} = 24a^2b + 4a.$$

Sometimes is more convenient to change the order of derivation. For instance, $f(x,y) = y\ln(x^2+y^2+1)$ is of class C^∞ on $\mathbb{R}^2$ (why?). In order to compute $\frac{\partial^2 f}{\partial x \partial y}$ it is easier to compute $\frac{\partial^2 f}{\partial y \partial x}$ i.e. to compute firstly $\frac{\partial f}{\partial x} = \frac{2xy}{x^2+y^2+1}$, and secondly

$$\frac{\partial}{\partial y}\left(\frac{2xy}{x^2+y^2+1}\right) = \frac{2x(x^2+y^2+1) - 2y \cdot 2xy}{(x^2+y^2+1)^2} = \frac{2x^3 - 2xy^2 + 2x}{(x^2+y^2+1)^2},$$

then to compute firstly

$$\frac{\partial f}{\partial y} = \ln(x^2+y^2+1) + \frac{2y^2}{x^2+y^2+1}$$

and secondly

$$\frac{\partial}{\partial x}\left[\ln(x^2+y^2+1) + \frac{2y^2}{x^2+y^2+1}\right]$$

(why?-count the number of operations and their difficulties in each case!).

The following notion will be very helpful in the applications of the differential calculus.

DEFINITION 30. *Let A be an open subset in $\mathbb{R}^n$ and let $\mathbf{a} = (a_1, a_2, ..., a_n)$ be a fixed point (vector) in A. Let f be a function of class C^2 on A, $f : A \to \mathbb{R}$. The symmetric matrix*

$$H_{f,\mathbf{a}} = (s_{ij}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) \right), i = 1, 2, ..., n; j = 1, 2, ..., n$$

is called the Hessian matrix of f at $\mathbf{a}$. The quadratic form $d^2 f(\mathbf{a})$ defined on $\mathbb{R}^n$, relative to its canonical basis

$$\{\mathbf{e}_1 = (1, 0, 0, ..., 0), \mathbf{e}_2 = (0, 1, 0, ..., 0), ..., \mathbf{e}_n = (0, 0, 0, ...0, 1)\}$$

(see a Linear Algebra course!) with values in $\mathbb{R}$,

$$d^2 f(\mathbf{a})(h_1, h_2, ..., h_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) h_i h_j. \tag{1.7}$$

is called the second differential of f at $\mathbf{a}$. Its matrix is exactly the Hessian matrix of f at $\mathbf{a}$. For instance, if f is a function of 2 variables, $x_1 = x$, $x_2 = y$ and $\mathbf{a} = (a, b)$, then formula (1.7) becomes

$$d^2 f(a, b)(h_1, h_2) = \frac{\partial^2 f}{\partial x^2}(a, b) h_1^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(a, b) h_1 h_2 + \frac{\partial^2 f}{\partial y^2}(a, b) h_2^2. \tag{1.8}$$

If we introduce the projection functions $dx_i(h_1, h_2, ..., h_n) = h_i$ for $i = 1, 2, ..., n$, we get a more compact formula for (1.7)

$$d^2 f(\mathbf{a}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) dx_i dx_j. \tag{1.9}$$

Here, $dx_i dx_j$ is the product between the two linear mappings $dx_i, dx_j : \mathbb{R}^n \to \mathbb{R}$, i.e.

$$dx_i dx_j(\mathbf{h}) = dx_i(\mathbf{h}) \cdot dx_j(\mathbf{h}) = h_i h_j,$$

where $\mathbf{h} = (h_1, h_2, ..., h_n)$. For two variables we get

$$d^2 f(a, b) = \frac{\partial^2 f}{\partial x^2}(a, b) dx^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(a, b) dx dy + \frac{\partial^2 f}{\partial y^2}(a, b) dy^2, \tag{1.10}$$

where dx^2 is $dx \cdot dx$ and not $d(x^2)$ which is equal to $2xdx$ (why?). The same for dy^2... . The analogous formula for a function of 3 variables $f(x, y, z)$ is

$$d^2 f(a, b, c) = \frac{\partial^2 f}{\partial x^2}(a, b, c) dx^2 + \frac{\partial^2 f}{\partial y^2}(a, b, c) dy^2 + \frac{\partial^2 f}{\partial z^2}(a, b, c) dz^2 +$$

$$+ 2 \frac{\partial^2 f}{\partial x \partial y}(a, b, c) dx dy + 2 \frac{\partial^2 f}{\partial x \partial z}(a, b, c) dx dz + 2 \frac{\partial^2 f}{\partial y \partial z}(a, b, c) dy dz. \tag{1.11}$$

For instance, let us compute the second differential for

$$f(x,y,z)=2x^3+3xy^2z+z^3$$

at the point $(-1,2,3)$. First of all we compute

$$\frac{\partial^2 f}{\partial x^2}(x,y,z)=\frac{\partial}{\partial x}(\frac{\partial f}{\partial x})(x,y,z)=\frac{\partial}{\partial x}(6x^2+3y^2z)=12x.$$

So, $\frac{\partial^2 f}{\partial x^2}(-1,2,3)=-12$. It is easy to find

$$\frac{\partial^2 f}{\partial y^2}(-1,2,3)=-18,\frac{\partial^2 f}{\partial z^2}(-1,2,3)=18,$$

$$\frac{\partial^2 f}{\partial x\partial y}(-1,2,3)=36,\frac{\partial^2 f}{\partial x\partial z}(-1,2,3)=12,\frac{\partial^2 f}{\partial y\partial z}(-1,2,3)=-12.$$

Now we use (1.11) and find

(1.12)

$$d^2f(-1,2,3)=-12dx^2-18dy^2+18dz^2+72dxdy+24dxdz-24dydz,$$

i.e. we have a quadratic form in 3 variables dx,dy,dz. Clearer, this last quadratic form is

$$g(X,Y,Z)=-12X^2-18Y^2+18Z^2+72XY+24XZ-24YZ.$$

Now, if we substitute X with dx, Y with dy and Z with dz, we get (1.12).

Let us compute the value of this last function

$$d^2f(-1,2,3):\mathbb{R}^3\to\mathbb{R}$$

at the point $(2,-3,-4)$. Since

$$dx^2(2,-3,-4)=2^2=4,dy^2(2,-3,-4)=(-3)^2=9,$$

$$dz^2(2,-3,-4)=(-4)^2=16,dxdy(2,-3,-4)=2\cdot(-3)=-6,$$

$$dxdz(2,-3,-4)=2\cdot(-4)=-8,dydz(2,-3,-4)=(-3)(-4)=12,$$

we finally obtain

$$d^2f(-1,2,3)(2,-3,-4)=-12\cdot4-18\cdot9+18\cdot16+72\cdot(-6)+$$

$$+24\cdot(-8)-24\cdot12=-12\cdot4+7\cdot18+24(-18-8-12)$$

$$=-12\cdot4+7\cdot18+24\cdot(-38)=-12(4+76)+7\cdot18=6(-139)=-834.$$

Now, let us look carefully at the formulas (1.13), (1.7) and (1.9). We introduce some symbolic operations in order to find a unitary and

general formula. We called $\frac{\partial}{\partial x_j}$ a differential operator. By definition, we multiply two such operators $\frac{\partial}{\partial x_j}$ and $\frac{\partial}{\partial x_i}$ by a simple composition:

$$\frac{\partial}{\partial x_j} \cdot \frac{\partial}{\partial x_i} \stackrel{def}{=} \frac{\partial^2}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j} \circ \frac{\partial}{\partial x_i}.$$

For instance,

$$\left(\frac{\partial}{\partial x} \cdot \frac{\partial}{\partial y}\right)(3x^2 + 5xy^3) = \frac{\partial}{\partial x}(\frac{\partial}{\partial y}(3x^2 + 5xy^3)) = \frac{\partial}{\partial x}(15xy^2) = 15y^2.$$

Moreover,

$$df(a,b) = \frac{\partial f}{\partial x}(a,b)dx + \frac{\partial f}{\partial y}(a,b)dy$$

can be written as an operator "on f" at an arbitrary point (which will not appear)

$$d = \frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy,$$

This is also called a differential operator. How do we multiply two such operators?

$$\left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)\left(\frac{\partial}{\partial z}dz + \frac{\partial}{\partial w}dw\right) =$$

$$\stackrel{def}{=} \frac{\partial^2}{\partial x \partial z}dxdz + \frac{\partial^2}{\partial y \partial z}dydz + \frac{\partial^2}{\partial x \partial w}dxdw + \frac{\partial^2}{\partial y \partial w}dydw.$$

This means that whenever we multiply operators we just compose them and whenever we multiply linear mappings we just multiply them as functions. These last are always coefficients of differential operators. For instance

$$\left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 = \frac{\partial^2}{\partial x^2}dx^2 + 2\frac{\partial^2}{\partial x \partial y}dxdy + \frac{\partial^2}{\partial y^2}dy^2. \tag{1.13}$$

Hence,

$$d^2 f(a,b) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 (f)(a,b),$$

with this last notation. We observe that in (1.13) one has a binomial formula of the type $(a+b)^2 = a^2 + 2ab + b^2$ (with the above indicated multiplication between differential operators). If we multiply again by $\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy$ the both sides in (1.13) we easily get

$$\left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^3 = \frac{\partial^3}{\partial x^3}dx^3 + 3\frac{\partial^3}{\partial x^2 \partial y}dx^2 dy + 3\frac{\partial^3}{\partial x \partial y^2}dxdy^2 + \frac{\partial^3}{\partial y^3}dy^3,$$

i.e. the analogous formula of $(a+b)^3 = a^3 + 3a^2 b + 3ab^3 + b^3$.

DEFINITION 31. *(the differential of order k) In general, if a function f of n variables, $f : A \to \mathbb{R}$, is of class C^k on A, i.e. it has all partial differentials of the type*

$$\frac{\partial^k f}{\partial x_1^{k_1}\partial x_2^{k_2}...\partial x_n^{k_n}}(\mathbf{a})$$

(where k is a fixed natural number, $k > 0$ and $k_1, k_2, ..., k_n$ are natural numbers such that $k = k_1 + k_2 + ... + k_n$ and $0 \leq k_1, k_2, ..., k_n \leq n$), at any point $\mathbf{a}$ of A, the k-th differential of f at $\mathbf{a}$ is by definition

$$d^k f(\mathbf{a}) = \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + ... + \frac{\partial}{\partial x_n}dx_n\right)^k (f)(\mathbf{a}). \tag{1.14}$$

For instance, if $n = 2$, $x_1 = x$, $x_2 = y$ and $\mathbf{a} =(a, b)$, then this last formula becomes

(1.15)

$$d^k f(a,b) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^k (f)(a,b) = \sum_{i=0}^{k}\binom{k}{i}\frac{\partial^k f}{\partial x^{k-i}\partial y^i}(a,b)dx^{k-i}dy^i,$$

where $\binom{k}{i} = \frac{k!}{i!(k-i)!}$ is the combination of k objects taken i. The analogy with the binomial formula

$$(a+b)^k = \sum_{i=0}^{k}\binom{k}{i}a^{k-i}b^i$$

is now clear.

Let us compute

$$d^4 f(1,-1) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^4 (f)(1,-1)$$

for $f(x, y) = x^5 + xy^4$. For $k = 4$ formula (1.15) becomes

$$\left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^4 (f)(1,-1) = \binom{4}{0}\frac{\partial^4 f}{\partial x^4}(1,-1)dx^4 +$$

$$\binom{4}{1}\frac{\partial^4 f}{\partial x^3\partial y}(1,-1)dx^3dy + \binom{4}{2}\frac{\partial^4 f}{\partial x^2\partial y^2}(1,-1)dx^2dy^2 +$$

$$\binom{4}{3}\frac{\partial^4 f}{\partial x\partial y^3}(1,-1)dxdy^3 + \binom{4}{4}\frac{\partial^4 f}{\partial y^4}(1,-1)dy^4.$$

Now, everything reduces to the computation of the mixed partial derivatives.

$$\frac{\partial^4 f}{\partial x^4}(1,-1) = 120, \frac{\partial^4 f}{\partial x^3\partial y}(1,-1) = 0, \frac{\partial^4 f}{\partial x^2\partial y^2}(1,-1) = 0,$$

$$\frac{\partial^4 f}{\partial x \partial y^3}(1,-1) = -24, \frac{\partial^4 f}{\partial x \partial y^3}(1,-1) = -24, \frac{\partial^4 f}{\partial y^4}(1,-1) = 24.$$

Hence,

$$\left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^4 (f)(1,-1) = 120dx^4 - 96dxdy^3 + 24dy^4.$$

If we want to compute the value of this last differential at $(2,3)$ for instance, we obtain

$$120 \cdot 2^4 - 96 \cdot 2 \cdot 3^3 + 24 \cdot 3^4 = -1320.$$

Let us now compute

$$d^2 f(1,1,0) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy + \frac{\partial}{\partial z}dz\right)^2 (f)(1,1,0)$$

for $f(x,y,z) = x^2+y^2+xz+yz$. To be easier, let us recall the elementary algebraic formula:

$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc.$$

Using the above multiplicity between operators, etc., we get

$$d^2 f(1,1,0) = \frac{\partial^2 f}{\partial x^2}(1,1,0)dx^2 + \frac{\partial^2 f}{\partial y^2}(1,1,0)dy^2 +$$

$$\frac{\partial^2 f}{\partial z^2}(1,1,0)dz^2 + 2\frac{\partial^2 f}{\partial x \partial y}(1,1,0)dxdy + 2\frac{\partial^2 f}{\partial x \partial z}(1,1,0)dxdz +$$

$$2\frac{\partial^2 f}{\partial y \partial z}(1,1,0)dydz = 2dx^2 + 2dy^2 + 2dxdz + 2dydz.$$

If one wants to compute $d^2 f(1,1,0)(3,4,5)$ we get

$$d^2 f(1,1,0)(3,4,5) = 2 \cdot 3^2 + 2 \cdot 4^2 + 2 \cdot 3 \cdot 5 + 2 \cdot 4 \cdot 5 = 120.$$

Since

$$(a_1 + a_2 + ... + a_n)^m = \sum_{k_1+k_2+...+k_n=m, k_i \in \mathbb{N}} \frac{m!}{k_1!k_2!...k_n!} a_1^{k_1} a_2^{k_2} ... a_n^{k_n},$$

one has the following definition of the m-th differential of f at a point $\mathbf{a} \in A$:

$$d^m f(\mathbf{a}) = \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + ... + \frac{\partial}{\partial x_n}dx_n\right)^m$$

$$= \sum_{k_1+k_2+...+k_n=m, k_i \in \mathbb{N}} \frac{m!}{k_1!k_2!...k_n!} \frac{\partial^m f}{\partial x_1^{k_1} \partial x_2^{k_2} ... \partial x_n^{k_n}} dx_1^{k_1} x_2^{k_2} ... dx_n^{k_n},$$

where in these last two sums $k_1, k_2, ..., k_n$ take all the natural values under the restriction $k_1 + k_2 + ... + k_n = m$.

2. Chain rules in two variables

During the mathematical modeling process of the physical phenomena, usually one must find functions $z = z(x, y)$ which verify an equality of the following form (a partial differential equation of order 2, i.e. a PDE):

$$A(x,y)\frac{\partial^2 z}{\partial x^2}(x,y) + 2B(x,y)\frac{\partial^2 z}{\partial x \partial y}(x,y) + C(x,y)\frac{\partial^2 z}{\partial y^2}(x,y)$$

$$+E\left(x, y, z(x,y), \frac{\partial z}{\partial x}(x,y), \frac{\partial z}{\partial y}(x,y)\right) = 0, \tag{2.1}$$

where A, B, C, E are continuous functions of the indicated free variables. Relative to E we must add that it is a continuous function $E(X, Y, Z, U, V)$ of 5 free variables, where instead of X, Y, Z, U, V, we put $x, y, z(x, y)$, $\frac{\partial z}{\partial x}(x, y)$ and $\frac{\partial z}{\partial y}(x, y)$ respectively. In order to find all the functions $z(x, y)$ of class C^2 on a fixed plane domain D, which verifies (2.1) we change the "old" variables x, y with new ones $u = u(x, y)$ and $v = v(x, y)$ respectively (functions of the firsts) such that some of the new "coefficients" A, B, or C to become zero. How do we find these new functions $u = u(x, y)$ and $v = v(x, y)$ is a problem which will be considered in another course. Our problem here is how to write the partial derivatives,

$$\frac{\partial^2 z}{\partial x^2}(x,y), \frac{\partial^2 z}{\partial x \partial y}(x,y), \frac{\partial^2 z}{\partial y^2}(x,y), \frac{\partial z}{\partial x}(x,y), \frac{\partial z}{\partial y}(x,y)$$

as functions of u and v. The transition from the "old" variables to the "new" ones u and v are realised by a "change of variables" function $\mathbf{F}(x, y) = (u(x, y), v(x, y))$ such that $\mathbf{F}$ is invertible and of class C^1 on its definition domain. Moreover, its inverse $\mathbf{G} = \mathbf{F}^{-1}$ is also a function (in variables u and v) of class C^1 (see also the section "Change of variables"). Let $\overline{z}$ be the composed function $z \circ \mathbf{G}$. Hence, $z = \overline{z} \circ \mathbf{F}$, or

$$\overline{z}(u(x,y), v(x,y)) = z(x,y).$$

The chain rules formulas (2.9) and (2.10) supply us with formulas for $\frac{\partial z}{\partial x}(x, y)$ and $\frac{\partial z}{\partial y}(x, y)$:

(2.2)

$$\frac{\partial z}{\partial x}(x,y) = \frac{\partial \overline{z}}{\partial u}(u(x,y), v(x,y))\frac{\partial u}{\partial x}(x,y) + \frac{\partial \overline{z}}{\partial v}(u(x,y), v(x,y))\frac{\partial v}{\partial x}(x,y),$$

and

(2.3)
$$\frac{\partial z}{\partial y}(x,y) = \frac{\partial \overline{z}}{\partial u}(u(x,y),v(x,y))\frac{\partial u}{\partial y}(x,y) + \frac{\partial \overline{z}}{\partial v}(u(x,y),v(x,y))\frac{\partial v}{\partial y}(x,y).$$

Let us use these formulas to find a similar formula for $\frac{\partial^2 z}{\partial x \partial y}(x,y)$. For this, let us denote by $g(x,y)$ and by $h(x,y)$ the new functions of x and y obtained in (2.3)

$$g(x,y) \stackrel{def}{=} \frac{\partial \overline{z}}{\partial u}(u(x,y),v(x,y))$$

and

$$\frac{\partial \overline{z}}{\partial v}(u(x,y),v(x,y)) \stackrel{def}{=} h(x,y).$$

Let us compute $\frac{\partial g}{\partial x}(x,y)$ and $\frac{\partial h}{\partial x}(x,y)$ by using the formula (2.2) with g instead of z and h instead of z respectively:

$$\frac{\partial g}{\partial x}(x,y) = \frac{\partial}{\partial u}\left(\frac{\partial \overline{z}}{\partial u}(u(x,y),v(x,y))\right)\frac{\partial u}{\partial x}(x,y)+$$

(2.4)
$$\frac{\partial}{\partial v}\left(\frac{\partial \overline{z}}{\partial u}(u(x,y),v(x,y))\right)\frac{\partial v}{\partial x}(x,y) = \frac{\partial^2 \overline{z}}{\partial u^2}(u(x,y),v(x,y))\frac{\partial u}{\partial x}(x,y)+$$
$$\frac{\partial^2 \overline{z}}{\partial v \partial u}(u(x,y),v(x,y))\frac{\partial v}{\partial x}(x,y).$$

and

$$\frac{\partial h}{\partial x}(x,y) = \frac{\partial}{\partial u}\left(\frac{\partial \overline{z}}{\partial v}(u(x,y),v(x,y))\right)\frac{\partial u}{\partial x}(x,y)+$$

(2.5)
$$\frac{\partial}{\partial v}\left(\frac{\partial \overline{z}}{\partial v}(u(x,y),v(x,y))\right)\frac{\partial v}{\partial x}(x,y) = \frac{\partial^2 \overline{z}}{\partial u \partial v}(u(x,y),v(x,y))\frac{\partial u}{\partial x}(x,y)+$$
$$\frac{\partial^2 \overline{z}}{\partial v^2}(u(x,y),v(x,y))\frac{\partial v}{\partial x}(x,y).$$

Let us come back to formula (2.3) and let us differentiate it (both sides) with respect to x. We get:

$$\frac{\partial^2 z}{\partial x \partial y}(x,y) = \frac{\partial g}{\partial x}(x,y)\frac{\partial u}{\partial y}(x,y) + g\frac{\partial^2 u}{\partial x \partial y}(x,y)+$$
$$\frac{\partial h}{\partial x}(x,y)\frac{\partial v}{\partial y}(x,y) + h\frac{\partial^2 v}{\partial x \partial y}(x,y).$$

If we take count of the formulas (2.4) and (2.5) we finally obtain:

$$\frac{\partial^2 z}{\partial x \partial y}(x,y) = \frac{\partial^2 \overline{z}}{\partial u^2}(u(x,y),v(x,y))\frac{\partial u}{\partial x}(x,y)\frac{\partial u}{\partial y}(x,y)+ \tag{2.6}$$

$$+\frac{\partial^2\overline{z}}{\partial u\partial v}(u(x,y),v(x,y))\left[\frac{\partial u}{\partial x}(x,y)\frac{\partial v}{\partial y}(x,y)+\frac{\partial u}{\partial y}(x,y)\frac{\partial v}{\partial x}(x,y)\right]+$$
$$+\frac{\partial^2\overline{z}}{\partial v^2}(u(x,y),v(x,y))\frac{\partial v}{\partial x}(x,y)\frac{\partial v}{\partial y}(x,y)+$$
$$+\frac{\partial\overline{z}}{\partial u}(u(x,y),v(x,y))\frac{\partial^2 u}{\partial x\partial y}(x,y)+\frac{\partial\overline{z}}{\partial v}(u(x,y),v(x,y))\frac{\partial^2 v}{\partial x\partial y}(x,y).$$

We can simply rewrite this formula as:

$$\frac{\partial^2 z}{\partial x\partial y}=\frac{\partial^2\overline{z}}{\partial u^2}\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}+\frac{\partial^2\overline{z}}{\partial u\partial v}\left[\frac{\partial u}{\partial x}\frac{\partial v}{\partial y}+\frac{\partial u}{\partial y}\frac{\partial v}{\partial x}\right]+$$
$$+\frac{\partial^2\overline{z}}{\partial v^2}\frac{\partial v}{\partial x}\frac{\partial v}{\partial y}+\frac{\partial\overline{z}}{\partial u}\frac{\partial^2 u}{\partial x\partial y}+\frac{\partial\overline{z}}{\partial v}\frac{\partial^2 v}{\partial x\partial y}.$$

If in this formula, we formally put x instead of y we get another useful formula:

$$\frac{\partial^2 z}{\partial x^2}=\frac{\partial^2\overline{z}}{\partial u^2}\left(\frac{\partial u}{\partial x}\right)^2+2\frac{\partial^2\overline{z}}{\partial u\partial v}\frac{\partial u}{\partial x}\frac{\partial v}{\partial x}+\frac{\partial^2\overline{z}}{\partial v^2}\left(\frac{\partial v}{\partial x}\right)^2+$$
$$\frac{\partial\overline{z}}{\partial u}\frac{\partial^2 u}{\partial x^2}+\frac{\partial\overline{z}}{\partial v}\frac{\partial^2 v}{\partial x^2}. \tag{2.7}$$

If here, in this last formula, we put y instead of x, we get the last useful chain rule formula:

$$\frac{\partial^2 z}{\partial y^2}=\frac{\partial^2\overline{z}}{\partial u^2}\left(\frac{\partial u}{\partial y}\right)^2+2\frac{\partial^2\overline{z}}{\partial u\partial v}\frac{\partial u}{\partial y}\frac{\partial v}{\partial y}+\frac{\partial^2\overline{z}}{\partial v^2}\left(\frac{\partial v}{\partial y}\right)^2+$$
$$\frac{\partial\overline{z}}{\partial u}\frac{\partial^2 u}{\partial y^2}+\frac{\partial\overline{z}}{\partial v}\frac{\partial^2 v}{\partial y^2}. \tag{2.8}$$

EXAMPLE 16. *(vibrating string equation) Let S be a one-dimensional elastic wire (infinite, homogeneous and perfect elastic) which vibrates freely, without an exterior perturbing force. It is considered to lay on the real line Ox. Let $y\geq 0$ be time and let $z(x,y)$ be the deflection of the string at the point M of coordinate x and at the moment y. If one write the D'Alembert equality, which makes equal the dynamic Newtonian force and the Hook elasticity force, we get a PDE of order 2 (the vibrating string equation):*

$$\frac{\partial^2 z}{\partial y^2}=a^2\frac{\partial^2 z}{\partial x^2}, \tag{2.9}$$

where $a>0$ is a constant depending on the density and on the elasticity modulus. In order to find all the functions $z=z(x,y)$ which verify the equality (2.9), i.e. to solve that equation, we must change the variables x and y with new ones $u=x-ay$ and $v=x+ay$ (see the Differential

Equations course). Let us use chain formulas (2.7) and (2.8) in order to change the variables in the equation (2.9):

$$\frac{\partial^2 z}{\partial x^2} = \frac{\partial^2 \overline{z}}{\partial u^2} + 2\frac{\partial^2 \overline{z}}{\partial u \partial v} + \frac{\partial^2 \overline{z}}{\partial v^2},$$

and

$$\frac{\partial^2 z}{\partial y^2} = \frac{\partial^2 \overline{z}}{\partial u^2}a^2 - 2\frac{\partial^2 \overline{z}}{\partial u \partial v}a^2 + \frac{\partial^2 \overline{z}}{\partial v^2}a^2.$$

If we substitute these expressions in (2.9) we finally get

$$\frac{\partial^2 \overline{z}}{\partial u \partial v} = 0. \tag{2.10}$$

But this last PDE of order 2 can easily be solved. From 2.10 we obtain: $\frac{\partial}{\partial u}\left(\frac{\partial \overline{z}}{\partial v}\right) = 0$, *i.e.* $\frac{\partial \overline{z}}{\partial v}$ *is only a function* $h(v)$. *Hence,*

$$\overline{z}(u,v) = \int h(v)dv = f(v) + g(u)$$

(why?), where f *and* g *are two arbitrary functions of class* C^2 *on some open real subsets. Coming back to* x *and* y *we finally get the "general solution" of the vibrating string equation:*

$$z(x,y) = f(x+ay) + g(x-ay).$$

Other examples in which we use higher chain rules (here "higher" means $2 > 1$*!) will appear in the section "Change of variables".*

3. Taylor's formula for several variables

In Theorem 44 we obtained an approximation of a function of one variable, of class C^{m+1} on an ε-neighborhood $(a-\varepsilon, a+\varepsilon)$ of a fixed point a, with a polynomial (the Taylor's polynomial) of degree m (m is a fixed natural number). We also estimated the error in this approximative process. We write again this classical and fundamental formula and try to generalize it to the case of a function of n variables.

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(n)}(a)}{n!}(x-a)^n + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{3.1}$$

where c is a number between x and a. Let us write again formula (3.1) by putting $h = x - a$, or $x = a + h$ and $c = a + t_* h$, where $t_* \in (0,1)$ ($t_* = \frac{c-a}{x-a}$, why?):

(3.2)

$$f(a+h) = f(a) + \frac{f'(a)}{1!}h + \frac{f''(a)}{2!}h^2 + ... + \frac{f^{(n)}(a)}{n!}h^n + \frac{f^{(n+1)}(a+t_* h)}{(n+1)!}h^{n+1}.$$

It is enough to generalize this formula for a scalar function of n variables because, if $\mathbf{f} = (f_1, f_2, ..., f_k)$ is a vector function with k components, we simply write the Taylor formula for any component, separately, i.e. we approximate componentwisely.

Let A be an open subset of $\mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a function of class C^{m+1} on A. Let $\mathbf{a} = (a_1, a_2, ..., a_n)$ be a fixed point of A and let $V = B(\mathbf{a}, r)$ be an n-dimensional open ball (see its definition in Chapter 6, Section 1) with centre at $\mathbf{a}$ and of radius $r > 0$ which is contained in A (why such thing is possible?). If a point $\mathbf{x} = (x_1, x_2, ..., x_n)$ is in the ball V, the whole segment

$$[\mathbf{a}, \mathbf{x}] = \{\mathbf{z} = \mathbf{a}+t(\mathbf{x} - \mathbf{a}) : t \in [0, 1]\}$$

is contained in V (why?-in general, a ball is a convex subset...prove it!). A subset C of $\mathbb{R}^n$ is said to be *convex* if whenever $\mathbf{a}$ and $\mathbf{b}$ are in C, the whole segment $[\mathbf{a}, \mathbf{b}]$ is contained in C.

THEOREM 72. *(Taylor's formula for n variables) With the above notation and hypotheses, for any $\mathbf{h} = (h_1, h_2, ..., h_n)$ small enough, such that $\mathbf{x} = \mathbf{a} + \mathbf{h} \in V$ ($\|\mathbf{h}\| < r$), one has the following Taylor's formula:*

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \frac{1}{1!}df(\mathbf{a})(\mathbf{h}) + \frac{1}{2!}d^2f(\mathbf{a})(\mathbf{h}) + ... + \frac{1}{m!}d^mf(\mathbf{a})(\mathbf{h}) + \frac{1}{(m+1)!}d^{m+1}f(\mathbf{c})(\mathbf{h}), \tag{3.3}$$

where $\mathbf{c} \in (\mathbf{a}, \mathbf{a} + \mathbf{h})$, i.e. $\mathbf{c} = \mathbf{a}+t_*\mathbf{h}$ for a $t_* \in (0, 1)$.

PROOF. ($n = 2$) Let

$$\mathbf{a} = (a_1, a_2), \mathbf{x} = (x_1, x_2), \mathbf{h} = (h_1, h_2), h_1 = x_1 - a_1, h_2 = x_2 - a_2.$$

The segment $[\mathbf{a}, \mathbf{x}]$ is the usual segment with ends $\mathbf{a}$ and $\mathbf{x}$ in the plane xOy (see Fig. 8.1). Let us restrict f to the segment $[\mathbf{a}, \mathbf{x}]$. This means that to any point $\mathbf{a}+t\mathbf{h}$, $t \in [0, 1]$ we assign the number $f(\mathbf{a}+t\mathbf{h})$. One obtains a mapping $t \rightsquigarrow f(\mathbf{a}+t\mathbf{h})$, denoted here by $g : [0, 1] \to \mathbb{R}$,

$$g(t) = f(\mathbf{a}+t\mathbf{h}) = f(a_1 + th_1, a_2 + th_2).$$

Let us denote by u_1 and u_2 the functions $u_1(t) = a_1 + th_1$ and respectively $u_2(t) = a_2 + th_2$. So, if

$$\mathbf{u}(t) = (a_1 + th_1, a_2 + th_2),$$

i.e. if $\mathbf{u} = (u_1, u_2)$, one has that $\mathbf{g} = \mathbf{f} \circ \mathbf{u}$. Here $\mathbf{u}$ is a continuous one-to-one mapping from $[0, 1]$ onto $[\mathbf{a}, \mathbf{x}]$. Since $\mathbf{u}$ is of class C^∞ on $[0, 1]$ (why?), we see that g is of class C^{m+1} on $[0, 1]$. Let us apply Mac

Laurin's formula (1.16) (or the general Taylor formula (3.1) with $a = 0$ and $x = 1$) for the function g :

(3.4)

$$g(1) = g(0) + \frac{1}{1!}g'(0) + \frac{1}{2!}g''(0) + ... + \frac{1}{m!}g^{(m)}(0) + \frac{1}{(m+1)!}g^{(m+1)}(t_*),$$

where $t_* \in (0, 1)$. Since $g(1) = f(\mathbf{a} + \mathbf{h})$ and $g(0) = f(\mathbf{a})$, one has only to prove that $g^{(k)}(0) = d^k f(\mathbf{a})(\mathbf{h})$ for any $k = 1, 2, ..., m + 1$. We can use mathematical induction to prove this. Here, we prove only that $g'(0) = df(\mathbf{a})(\mathbf{h})$ and that $g''(0) = d^2 f(\mathbf{a})(\mathbf{h})$. For this purpose we use the chain rules formulas and the definition of the differential of order k. Indeed,

$$g'(t) = \frac{\partial f}{\partial x_1}[u_1(t), u_2(t)] \cdot u_1'(t) + \frac{\partial f}{\partial x_2}[u_1(t), u_2(t)] \cdot u_2'(t). \tag{3.5}$$

Hence,

$$g'(0) = \frac{\partial f}{\partial x_1}(a_1, a_2) \cdot h_1 + \frac{\partial f}{\partial x_2}(a_1, a_2) \cdot h_2 = df(\mathbf{a})(\mathbf{h}).$$

Let us use the formula (3.5) to compute $g''(t)$:

$$g''(t) = \frac{\partial^2 f}{\partial x_1^2}[u_1(t), u_2(t)] \cdot [u_1'(t)]^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}[u_1(t), u_2(t)] \cdot u_1'(t) \cdot u_2'(t) +$$

$$\frac{\partial f}{\partial x_1}[u_1(t), u_2(t)] \cdot u_1''(t) + \frac{\partial^2 f}{\partial x_1 \partial x_2}[u_1(t), u_2(t)] \cdot u_1'(t) \cdot u_2'(t) +$$

$$\frac{\partial^2 f}{\partial x_2^2}[u_1(t), u_2(t)] \cdot [u_2'(t)]^2 + \frac{\partial f}{\partial x_2}[u_1(t), u_2(t)] \cdot u_2''(t).$$

Since $u_1''(t) = 0$ and $u_2''(t) = 0$, one has:

$$g''(0) = \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) \cdot h_1^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) \cdot h_1 \cdot h_2 + \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) \cdot h_2^2 = d^2 f(\mathbf{a})(\mathbf{h}).$$

If we take $\mathbf{c} = \mathbf{a} + t_*\mathbf{h}$, one gets the formula (3.3) for $n = 2$. □

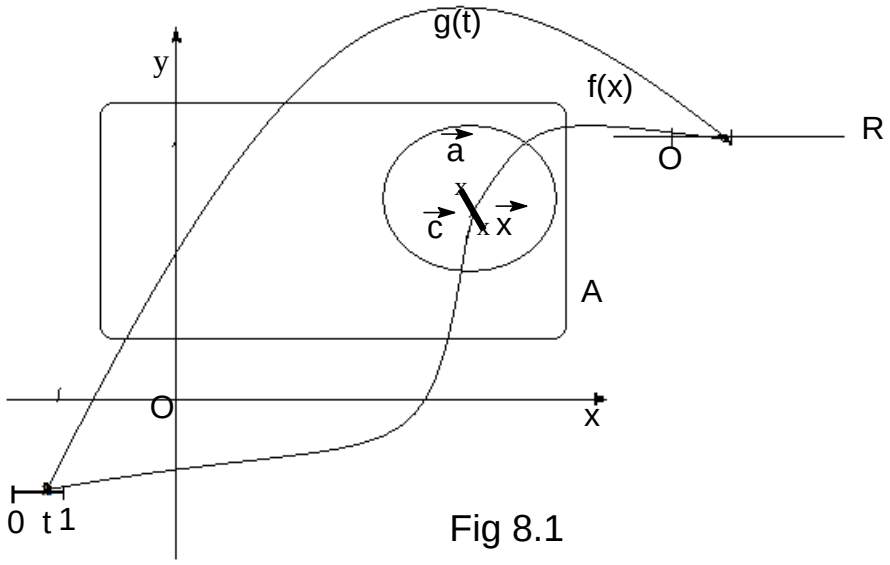


Fig 8.1

Let

$$P(x,y) = 2x^2y + 3xy^2 + x + y$$

be a polynomial of two variables x and y. Let us write $P(x,y)$ as a polynomial $Q(x-1, y+2)$, i.e.

$$P(x,y) = a_{00} + a_{10}(x-1) + a_{01}(y+2) + a_{20}(x-1)^2 + a_{11}(x-1)(y+2) +$$

$$a_{02}(y+2)^2 + a_{30}(x-1)^3 + a_{21}(x-1)^2(y+2) +$$
$$a_{12}(x-1)(y+2)^2 + a_{03}(y+2)^3.$$

We stop here because the "total" degree of $P(x,y)$ is $3 = 2+1$. We could find the coefficients a_{ij} by elementary tricks (do it!). However, let us use Taylor formula (3.3) with

$$\mathbf{a} = (1,-2), \mathbf{x} = (x,y), h_1 = x-1, h_2 = y+2,$$

etc. We have only to compute $dP(\mathbf{a})$, $d^2P(\mathbf{a})$ and $d^3P(\mathbf{a})$ (why not $d^4P(\mathbf{a})$?). So,

$$dP(\mathbf{a}) = \frac{\partial P}{\partial x}(\mathbf{a})dx + \frac{\partial P}{\partial y}(\mathbf{a})dy = (4xy + 3y^2 + 1) \mid_{(1,-2)} dx$$
$$+(2x^2 + 6xy + 1) \mid_{(1,-2)} dy = 5dx - 9dy$$

Thus,

$$dP(\mathbf{a})(\mathbf{h}) = 5(x-1) - 9(y+2).$$

Hence,

$$a_{00} = P(1,-2) = 7; a_{10} = 5; a_{01} = -9.$$

The coefficients a_{20}, a_{11} and a_{02} can be computed from the expression of $\frac{1}{2!}d^2P(\mathbf{a})(\mathbf{h})$. Namely,

$$\frac{\partial^2 P}{\partial x^2}(\mathbf{a}) = (4y) \mid_{(1,-2)} = -8, \frac{\partial^2 P}{\partial x \partial y}(\mathbf{a}) = (4x+6y) \mid_{(1,-2)} = -8$$

and $\frac{\partial^2 P}{\partial y^2}(\mathbf{a}) = 6x \mid_{(1,-2)} = 6$, i.e.

$$\frac{1}{2!}d^2P(\mathbf{a})(\mathbf{h}) = -4(x-1)^2 - 8(x-1)(y+2) + 3(y+2)^2$$

and so, $a_{20} = -4$, $a_{11} = -8$ and $a_{02} = 3$. In order to find a_{30}, a_{21}, a_{12} and a_{03} one must compute

$$\begin{aligned}\frac{1}{3!}d^3f(\mathbf{a})(\mathbf{h}) = \frac{1}{6}\Big[&\frac{\partial^3 P}{\partial x^3}(\mathbf{a})(x-1)^3 + 3\frac{\partial^3 P}{\partial x^2 \partial y}(\mathbf{a})(x-1)^2(y+2) \\ &+3\frac{\partial^3 P}{\partial x \partial y^2}(\mathbf{a})(x-1)(y+2)^2 + \frac{\partial^3 P}{\partial y^3}(\mathbf{a})(y+2)^3\Big] \\ &= 2(x-1)^2(y+2) + 3(x-1)(y+2)^2.\end{aligned}$$

Thus, $a_{30} = 0$; $a_{21} = 2$; $a_{12} = 3$ and $a_{03} = 0$. Finally one has:

$$\begin{aligned}P(x,y) = 7 + 5(x-1) - 9(y+2) - 4(x-1)^2 - 8(x-1)(y+2) + \\ +3(y+2)^2 + 2(x-1)^2(y+2) + 3(x-1)(y+2)^2.\end{aligned}$$

THEOREM 73. *(Lagrange's Theorem for many variables, or the Mean Value Theorem) Let $A \subset \mathbb{R}^n$ be an open subset of $\mathbb{R}^n$, let $\mathbf{a}$ be a point in A and let $V = B(\mathbf{a}, r) \subset A$, $r > 0$ be a ball with centre at $\mathbf{a}$ and of radius r. Let $f : A \to \mathbb{R}$, be a function of class C^1 defined on A. Then, for any $\mathbf{x}$ in X, there is a point $\mathbf{c}$ in $[\mathbf{a}, \mathbf{x}]$ such that:*

(3.6)

$$f(\mathbf{x}) - f(\mathbf{a}) = \frac{\partial f}{\partial x_1}(\mathbf{c})(x_1 - a_1) + ... + \frac{\partial f}{\partial x_n}(\mathbf{c})(x_n - a_n) = \langle \operatorname{grad} f(\mathbf{c}), \mathbf{h} \rangle,$$

i.e. the "increasing" $f(\mathbf{x}) - f(\mathbf{a})$ of f on the interval $[\mathbf{a}, \mathbf{x}]$ is equal to the scalar product between the gradient vector $\operatorname{grad} f(\mathbf{c})$ of f at a point $\mathbf{c}$ of the segment $[\mathbf{a}, \mathbf{x}]$, and the the vector $\mathbf{x} - \mathbf{a}$. If $\mathbf{x}$ is very close to $\mathbf{a}$, then we have an "affine" approximation of $f(\mathbf{x})$:

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \frac{\partial f}{\partial x_1}(\mathbf{a})(x_1 - a_1) + ... + \frac{\partial f}{\partial x_n}(\mathbf{a})(x_n - a_n), \tag{3.7}$$

or a linear approximation of $f(\mathbf{x}) - f(\mathbf{a})$:

(3.8)

$$f(\mathbf{x}) - f(\mathbf{a}) \approx \frac{\partial f}{\partial x_1}(\mathbf{a})(x_1 - a_1) + ... + \frac{\partial f}{\partial x_n}(\mathbf{a})(x_n - a_n) = \langle \operatorname{grad} f(\mathbf{a}), \mathbf{h} \rangle.$$

PROOF. It is sufficient to take $m = 0$ in the formula (3.3). □

From formula (3.7) we see that it is sufficient to know the gradient vector grad $f(\mathbf{a})$ of a function f at a point $\mathbf{a}$ and the value $f(\mathbf{a})$ of the same function at $\mathbf{a}$, in order to approximate the values of this functions in a neighborhood of $\mathbf{a}$. For instance, let us compute approximately $\sin 46^\circ \cos 1^\circ$. For this, let us consider the function of two variables $f(x, y) = \sin x \cos y$, the point $\mathbf{a} = (\frac{\pi}{4}, 0)$ and the point $\mathbf{x} = (\frac{\pi}{4} + \frac{\pi}{180}, \frac{\pi}{180})$. Then, formula (3.7) says that: $\sin 46^\circ \cos 1^\circ \approx \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} \cdot \frac{\pi}{180}$.

4. Problems

1. Compute df and $d^2 f$ for:

a)

$$f(x, y) = \sin(x^2 + y^2);$$

b)

$$f(x, y, z) = \sqrt{x^2 + y^2 + z^2};$$

c)

$$f(x, y) = \exp(xy)$$

at $(1, 1)$; find also $df(1, 1)(0, 1)$ and $d^2 f(1, 1)(0, 1)$.

2. Approximate $\Delta f = f(x, y) - f(x_0, y_0)$ by $df(x_0, y_0)(\Delta x, \Delta y)$, where $\Delta x = x - x_0$, $\Delta u = y - y_0$ and then compute:

a)

$$f(x, y) = x^{\ln y}$$

at the point $A(e + 0.1, 1 + 0.2)$;

b)

$$f(x, y) = \sqrt{x^2 + y^2}$$

at $A(4.001, 3.002)$;

c)

$$f(x, y) = x^y$$

at $A(1.02; 3.01)$.

3. Use Taylor's formula to approximate f by the Taylor polynomial T_n with Lagrange's remainder:

a)

$$f(x, y) = \ln(1 + x) + \ln(1 + y)$$

at $(0, 0)$, with T_4;

b)

$$f(x, y) = x^y$$

at $(1, 1)$, with T_3 and compute approximately $(1.1)^{1.2}$;

c)

$$f(x, y) = (\exp x) \sin y$$

at $(0,0)$ with T_2;

d)

$$f(x,y,z) = x^3 + y^3 + z^3 - 3xyz$$

at $(1,1,1)$, with T_2.

4. Write

$$P(x,y) = 2x^3 - 3x^2y + 2y^3 + 9x^2 - 3y + 6x + 3$$

as $Q(x+1, y-1)$.

5. Compute approximately $(0.95)^{2.01}$; Hint: take

$$g(x,y) = y^x$$

around $A(2,1)$ and use T_2.

6. Compute $d^2f(0,0,0)$ for

$$f(x,y,z) = x^2 + y^3 + z^4 - 2xy^2 + 3yz - 5x^2z^2.$$

7. Compute $d^3f(0,0)(0,0)$ for

$$f(x,y) = \cos(3x + 2y).$$

8. Prove that

$$u(x,t) = \frac{1}{2a\sqrt{\pi t}} \exp\left(-\frac{(x-b)^2}{4a^2t}\right)$$

verify the "heat equation": $\frac{\partial u}{\partial t}(x,t) = a^2 \frac{\partial^2 u}{\partial x^2}(x,t)$.

9. Use Taylor's formula to justify the following approximations:

a)

$$\frac{\cos x}{\cos y} \approx 1 - \frac{x^2 - y^2}{2}$$

around $(0,0)$;

b)

$$\arctan \frac{x+y}{1+xy} \approx x + y,$$

around $(0,0)$;

c)

$$\ln(1+x) \cdot \ln(1+y) \approx xy,$$

around $(0,0)$.

10. Find $df(1,-2)(2,3)$; $d^2f(1,-2)(2,3)$ and $d^3f(1,-2)(2,3)$ for

$$f(x,y) = x^3 + 2x^2y.$$

CHAPTER 9

Contractions and fixed points

1. Banach's fixed point theorem

Let (X, d) be a metric space, i.e. a set X with a distance function d on it. This function d associates to any pair (x, y) of elements of X a nonnegative real number $d(x, y)$ with the following properties:

i) $d(x, y) = 0$ if and only if $x = y$.

ii) $d(x, y) = d(y, x)$ for any x, y in X and

iii) $d(x, z) \leq d(x, y) + d(y, z)$ for any x, y, z in X (the *triangle inequality*).

This triangle inequality can be generalized and one obtains the *polygon inequality*:

$$(1.1)\quad d(x_0, x_n) \leq d(x_0, x_1) + d(x_1, x_2) + d(x_2, x_3) + ... + d(x_{n-1}, x_n).$$

for any finite sequence $\{x_0, x_1, x_2, ..., x_n\}$ of X. It can be easily proved if we use mathematical induction on n. For $n = 1$, or 2, it is clear. Suppose $n > 2$ and assume that the polygon inequality is true for any sequence of $k \leq n$ elements of X. Let us prove it for a sequence of $n+1$ elements $\{x_0, x_1, x_2, ..., x_n\}$. Thus,

$$(1.2)\quad d(x_0, x_{n-1}) \leq d(x_0, x_1)+d(x_1, x_2)+d(x_2, x_3)+...+d(x_{n-2}, x_{n-1}).$$

Now,

$$d(x_0, x_n) \leq d(x_0, x_{n-1}) + d(x_{n-1}, x_n) \leq$$

$$[d(x_0, x_1) + d(x_1, x_2) + d(x_2, x_3) + ... + d(x_{n-2}, x_{n-1})] + d(x_{n-1}, x_n).$$

and the proof of (1.1) is done.

We just met many examples of metric spaces: $(\mathbb{R}, d(x, y) = |x - y|)$, $(\mathbb{C}, d(z, w) = |z - w|)$, $(\mathbb{R}^n, d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|)$, $C[a, b] = \{f : [a, b] \to \mathbb{R},\ f \text{ continuous}\}$ with

$$d(f, g) = \|f - g\| = \sup\{|f(x) - g(x)| : x \in [a, b]\},$$

etc. All of these metric spaces are *complete metric spaces*, i.e. metric spaces (X, d) with the property that any Cauchy sequence has a limit in X. Not all metric spaces are complete. For instance, $X = (0, 1]$ with the same distance like that of $\mathbb{R}$ is not complete, because the sequence

$\{\frac{1}{n}\}$ is a Cauchy sequence in X but it has no limit in X (why?). It is easy to see that a subset Y of a metric space (X,d) is complete relative to the same distance like that of X if and only if it is closed in X (prove it!).

DEFINITION 32. *(contraction) Let (X,d) be a metric space. A function $f: X \to X$ is said to be a contraction on X if there is a number $\lambda \in (0,1)$ such that*

$$d(f(x), f(y)) \leq \lambda d(x,y) \tag{1.3}$$

for any x, y in X. This number λ is called the (contraction) coefficient of f.

For instance, $f: [0,1] \to [0,1]$, $f(x) = 0.5x$ is a contraction of coefficient 0.5 (prove it!). But $g: \mathbb{R} \to \mathbb{R}$, $g(x) = 2x$, is not a contraction on $\mathbb{R}$ but,...it is a contraction on $[0, 0.44]$ (prove it!).

Any contraction on X is a uniformly continuous function on X (why?). The same result is true even λ is an arbitrary positive real number. In this more general case we say that f is a *Lipschitzian function on X*.

THEOREM 74. *Let A be a convex subset of $\mathbb{R}^n$ (if $\mathbf{a}$ and $\mathbf{b}$ are in A, then the whole segment $[\mathbf{a}, \mathbf{b}]$ is in A). Let $f: A \to A$ be a function of class C^1 on A such that all the partial derivatives of f are bounded by a number of the form λ/n. where $\lambda \in (0,1)$. Then f is a contraction of coefficient λ on A.*

PROOF. Let us take $\mathbf{a}, \mathbf{b}$ in A and let us write Taylor's formula for $m = 0$ ($\mathbf{b} = \mathbf{a} + \mathbf{h}$):

(1.4)

$$f(\mathbf{b}) - \mathbf{f}(\mathbf{a}) = \frac{\partial f}{\partial x_1}(\mathbf{c}) \cdot (b_1 - a_1) + \frac{\partial f}{\partial x_2}(\mathbf{c}) \cdot (b_2 - a_2) + ... + \frac{\partial f}{\partial x_n}(\mathbf{c}) \cdot (b_n - a_n),$$

where $\mathbf{c}$ is a point on the segment $[\mathbf{a}, \mathbf{b}]$ and $\mathbf{a} = (a_1, a_2, ..., a_n)$, $\mathbf{b} = (b_1, b_2, ..., b_n)$.

So,

$$d(f(\mathbf{a}), f(\mathbf{b})) = \|f(\mathbf{b}) - f(\mathbf{a})\| \leq \left| \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{c}) \right| \|\mathbf{a} - \mathbf{b}\|$$

$$\leq \left[\sum_{i=1}^{n} \left| \frac{\partial f}{\partial x_i}(\mathbf{c}) \right| \right] \|\mathbf{a} - \mathbf{b}\| \leq \lambda d(\mathbf{a}, \mathbf{b}).$$

Thus, our function is a contraction. □

For instance, $f(x) = \frac{1}{5}x^3$ is a contraction on $[0,1]$, because $|f'(x)| = \frac{3}{5}|x^2| \leq \frac{3}{5}$ on $[0,1]$.

THEOREM 75. *(Banach's fixed point theorem) Let (X, d) be a complete metric space and let $f : X \to X$ be a contraction of coefficient $\lambda \in (0, 1)$. Then there is a unique element x in X such that $f(x) = x$ (a fixed point for f). This unique fixed point x of f on X can be obtained by the following method (the successive approximates method). Start with an arbitrary element x_0 of X and recurrently construct: $x_1 = f(x_0)$, $x_2 = f(x_1), ..., x_n = f(x_{n-1}), ...$ Then, the sequence $\{x_n\}$ is convergent to this fixed point x. Moreover, if we approximate x by x_n, the error $d(x, x_n)$ can be evaluated by the following formula*

$$d(x, x_n) \leq d(x_1, x_0) \cdot \frac{\lambda^n}{1 - \lambda}. \tag{1.5}$$

PROOF. It is sufficient to prove that $\{x_n\}$ is a Cauchy sequence (why?-remember that X is complete so, $x_n \to x$, then use the continuity of f in the recurrence relation-take limits and find $x = f(x)$). Let us evaluate the distance between the terms of the sequence $\{x_n\}$ by using the contraction formula (1.3).

$$d(x_2, x_1) = d(f(x_1), f(x_0)) \leq \lambda d(x_1, x_0),$$

$$d(x_3, x_2) = d(f(x_2), f(x_1)) \leq \lambda d(x_2, x_1) \leq \lambda^2 d(x_1, x_0),$$

and so on, up to a general relation (use mathematical induction if you want!):

$$d(x_{n+1}, x_n) \leq \lambda^n d(x_1, x_0). \tag{1.6}$$

Now,

(1.7)

$$d(x_{n+p}, x_n) \leq d(x_{n+p}, x_{n+p-1}) + d(x_{n+p-1}, x_{n+p-2}) + ... + d(x_{n+1}, x_n)$$

comes from applying of the polygon inequality (1.1). If in (1.7) we introduce the formula from (1.6), we get:

$$d(x_{n+p}, x_n) \leq (\lambda^{n+p-1} + \lambda^{n+p-2} + ... + \lambda^n) d(x_1, x_0)$$

$$\leq \lambda^n (1 + \lambda + \lambda^2 + ...) d(x_1, x_0) = \frac{\lambda^n}{1 - \lambda} d(x_1, x_0). \tag{1.8}$$

Since $\frac{\lambda^n}{1-\lambda} \to 0$, independently on p, the sequence $\{x_n\}$ is a Cauchy sequence. Since (X, d) is complete, this sequence has a limit $x = \lim x_n$. Making $p \to \infty$ in (1.8) we get the desired estimation of the error:

$$d(x, x_n) \leq \frac{\lambda^n}{1 - \lambda} d(x_1, x_0).$$

(why $d(x_{n+p}, x_n) \to d(x, x_n)$ if $p \to \infty$? Prove it!). Since $x_n = f(x_{n-1})$ and since f is continuous, one has that $x = f(x)$. This fixed point x is unique. Indeed, if $x = f(x)$ and $y = f(y)$, then

$$d(x, y) = d(f(x), f(y)) \leq \lambda d(x, y),$$

or

$$d(x, y) \cdot [\lambda - 1] \geq 0.$$

Since $\lambda \in (0, 1)$ and since $d(x, y) \geq 0$, the unique possibility is that $d(x, y) = 0$, i.e. $x = y$. □

The Banach's fixed point theorem has many applications. For instance, it can be used to find approximate solutions for equations and system of equations (linear or not!).

Take for example the polynomial

$$P(x) = x^3 - x^2 + 2x - 1$$

and let us search for a solution of the equation $P(x) = 0$ in the interval $X = [0, 1]$. The equation $x^3 - x^2 + 2x - 1 = 0$ can also be written as:

$$\frac{x^2 + 1}{x^2 + 2} = x. \tag{1.9}$$

Let us prove that $f(x) = \frac{x^2+1}{x^2+2}$ is a contraction on $[0, 1]$. Indeed, $f'(x) = \frac{2x}{(x^2+2)^2}$ and

$$\left| \frac{2x}{(x^2 + 2)^2} \right| \leq \frac{1}{2}$$

(why?) on $[0, 1]$. Applying Theorem 74

we get that f is a contraction of coefficient $\lambda = \frac{1}{2}$. So, the equation (1.9) has a unique solution a in $[0, 1]$. Let us find it approximately with "two exact decimals". Formula (1.5) says that:

$$|a - x_n| \leq \left(\frac{1}{2}\right)^n \cdot \frac{2}{1} |x_1 - x_0| = \left(\frac{1}{2}\right)^{n-1} |x_1 - x_0| .$$

Let us take $x_0 = 0$. Then $x_1 = f(x_0) = \frac{1}{2}$. Thus,

$$|a - x_n| \leq \frac{1}{2^n}.$$

If we force with $\frac{1}{2^n} \leq \frac{1}{10^2}$, we get $n = 7$. Hence, the true solution a is approximately equal to

$$x_7 = (f \circ f \circ f \circ f \circ f \circ f \circ f)(0) = f(f(f(f(f(f(f(0))))))).$$

This last number can be easily find by using a cyclic instruction in a computer language, like Pascal or C++. The committed error is less then 0.01.

2. Problems

1. Using the Banach's Fixed Point Theorem, find approximate solutions with the error $\varepsilon = 10^{-2}$ for the following equations:

a) $x^3 + x - 5 = 0$; b) $x^3 - \sin x = 3$; c) $x = \frac{\pi}{3\sqrt{3}} \cos x$.

2. Which of the following mappings are contractions? Study the fixed points of them.

a) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x$; b) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^7$; c) $f : \mathbb{C} \to \mathbb{C}$, $f(z) = z^4$;

d) $f : \mathbb{C} \to \mathbb{C}$, $f(z) = z^2 + z + 1$; e) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{5}x + 3$;

f) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \frac{1}{5} \arctan x$; g) $f : \mathbb{R} \to \mathbb{R}$, $f(x, y) = (\frac{1}{7}x, \frac{1}{8}y)$.

3. Try to find approximate solutions with 2 exact decimals for the following linear system of algebraic equations:

$$\begin{cases} 100x + 2y = 1 \\ 4x + 200y = 5 \end{cases}.$$

Hint: Write this system as:

$$\begin{cases} 0.01 - 0.02y = x \\ 0.025 - 0.02x = y \end{cases}.$$

Prove that the vector function $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$, defined by the formula, $\mathbf{f}(x, y) = (0.01 - 0.02y, 0.025 - 0.02x)$ is a contraction of coefficient $0.02 \times \sqrt{2} < 1$. Then apply the Banach's Fixed Point Theorem. At the end, compare the approximate result with the exact one!

4. What is the particularity of the system from Problem 3? Can we apply the Banach's Fixed Point Theorem to all the linear systems?

CHAPTER 10

Local extremum points

1. Local extremum points for many variables

Let A be an open subset of $\mathbb{R}^n$ and let $f : A \to \mathbb{R}$ be a scalar function defined on A. We say that $\mathbf{a} = (a_1, a_2, ..., a_n)$ is a *local maximum (minimum) point* of f if there is a small open ball $B(\mathbf{a}, r) \subset A$, $r > 0$, such that $f(\mathbf{x}) \leq f(\mathbf{a})$ ($f(\mathbf{x}) \geq f(\mathbf{a})$) for any $\mathbf{x}$ in $B(\mathbf{a}, r)$. Local maxima and local minima are referred to as *local extrema*. A local maximum point or a local minimum point is called an extremum point.

REMARK 30. *Let A be an open subset of $\mathbb{R}^n$ and let i be a fixed natural number in the set $\{1, 2, ..., n\}$. Then the i-th projection $pr_i(A)$ of A is the set of all $t \in \mathbb{R}$ such that there is an*

$$\mathbf{x} = (x_1, x_2, ..., x_{i-1}, t, x_{i+1}, ..., x_n)$$

in A with t at the i-th position. It is also an open subset of $\mathbb{R}$. Indeed, take $t_0 \in pr_i(A)$ and take $\mathbf{a}$ in A such that $\mathbf{a} = (a_1, ..., a_{i-1}, t_0, a_{i+1}, ..., a_n)$. Since A is open, there is a ball $B(\mathbf{a}, r) \subset A$ with $r > 0$. We prove that the 1-D ball $(t_0 - r, t_0 + r)$ is contained in $pr_i(A)$. It is in fact the i-th projection of $B(\mathbf{a}, r)$. For this, let $u \in (t_0 - r, t_0 + r)$, i.e. $|u - t_0| < r$. It is easy to see that

$$\mathbf{v} = (a_1, a_2, ..., a_{i-1}, u, a_{i+1}, ..., a_n) \in B(\mathbf{a}, r) \subset A.$$

Thus

$$u = pr_i(\mathbf{v}) \in pr_i(A).$$

So $pr_i(A)$ is also open in $\mathbb{R}$.

THEOREM 76. *(Fermat's theorem for many variables) Let A be an open subset of $\mathbb{R}^n$ and let $\mathbf{a} \in A$ be an extremum point of a function $f : A \to \mathbb{R}$, defined on A with values in $\mathbb{R}$. If f has partial derivatives $\frac{\partial f}{\partial x_j}(\mathbf{a})$, $j = 1, 2, ..., n$ at $\mathbf{a}$, then all of these are zero, i.e. any extremum point $\mathbf{a}$ of f is a stationary (critical) point for f. This means that $\mathbf{a}$ is a root of the vector equation: $\operatorname{grad} f(\mathbf{x}) = \mathbf{0}$, i.e. $\operatorname{grad} f(\mathbf{a}) = \mathbf{0}$, or $df(\mathbf{a}) = 0$, if this last one exists.*

PROOF. Let us fix an i in $\{1, 2, ..., n\}$ and let us define a function of one variable $g_i : (a_i - r, a_i + r) \to \mathbb{R}$ by the formula:

$$g_i(t) = f(a_1, ..., a_{i-1}, t, a_{i+1}, ..., a_n).$$

Here $r > 0$ is the radius of a small ball $B(\mathbf{a}, r)$ which is contained in A (see the above discussion). Assume that $\mathbf{a}$ is a local maximum point for f. We can take r to be small enough such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for any $\mathbf{x}$ in the ball $B(\mathbf{a},r)$ (why?). If $u \in (a_i - r, a_i + r)$, then

$$\mathbf{v} = (a_1, a_2, ..., a_{i-1}, u, a_{i+1}, ..., a_n) \in B(\mathbf{a}, r)$$

so,

$$g_i(u) = f(a_1, ..., a_{i-1}, u, a_{i+1}, ..., a_n) \leq$$
$$\leq f(a_1, ..., a_{i-1}, a_i, a_{i+1}, ..., a_n) = g_i(a_i).$$

This means that a_i is a local maximum for the function g_i. We use now Fermat's theorem 35 for the one variable function g_i at the point a_i. Thus, $g_i'(a_i) = 0$. But

$$g_i'(t) = \frac{\partial f}{\partial x_i}(a_1, ..., a_{i-1}, t, a_{i+1}, ..., a_n).$$

Hence, $g_i'(a_i) = \frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$, for any $i = 1, 2, ..., n$ and the proof of the theorem is complete. □

The Fermat's theorem says that for the class of differential functions f defined on an open subset A of $\mathbb{R}^n$, the local extremum points must be searched between the critical points, i.e. between the points $\mathbf{a}$ which are zeros for the gradient of f. For instance, for $f(x, y) = x^4 + y^4$, the gradient of f is $\operatorname{grad} f = (4x^3, 4y^3)$. So, one has only one point $(0, 0)$ which makes zero this gradient. Since $0 = f(0, 0) \leq x^4 + y^4$, for any $x, y \in \mathbb{R}$, the point $(0, 0)$ is a "global" minimum point for f. It is easy to see that for the function $h(x, y) = x^2 - y^2$, the point $(0, 0)$ is a critical point, but it is neither a local minimum, nor a local maximum point for f, because, in any neighborhood of $(0, 0)$ the function $h(x, y)$ has positive and negative values (why?). So we need a criterion to distinguish the local extremum points between the critical points. We recall that a quadratic form in n variables $X_1, X_2, ..., X_n$ is a homogeneous polynomial function $g(X_1, X_2, ..., X_n)$ of degree two of these n independent variables,

$$g(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j,$$

where $a_{ij} = a_{ji}$ for all $i, j \in \{1, 2, ..., n\}$, i.e. if its associated $n \times n$ matrix (a_{ij}) is symmetric. Here this last matrix is considered with entries in $\mathbb{R}$. We say that the quadratic form g is *positive definite* if

$g(x_1, x_2, ..., x_n) \geq 0$ for any real numbers $x_1, x_2, ..., x_n$ and, it is zero if and only if all of these numbers are zero. For instance,

$$g(X, Y) = X^2 + XY + Y^2$$

is positive definite. Assume contrary, namely we could find $(x, y) \neq (0, 0)$, say $y \neq 0$, such that

$$g(x, y) = x^2 + xy + y^2 < 0.$$

Let us divide by y^2 and put $t = x/y$. We get $t^2 + t + 1 < 0$, which is false because

$$t^2 + t + 1 = (t + 1/2)^2 + 3/4$$

cannot be negative for ever (why?). Moreover, if $x^2 + xy + y^2 = 0$ and if $(x, y) \neq (0, 0)$, then we obtain $t^2 + t + 1 = 0$ for $t = x/y$ or $t = y/x$. But the equation $Z^2 + Z + 1 = 0$ has no real root!

We say that the quadratic form g is *negative definite* if

$$g(x_1, x_2, ..., x_n) \leq 0$$

for any real numbers $x_1, x_2, ..., x_n$ and, it is zero if and only if all of these numbers are zero. For instance,

$$g(X, Y) = -X^2 - XY - Y^2$$

is negative definite (prove it!). If a quadratic form is negative definite or positive definite, we say that it is *definite*. If it is neither positive definite, nor negative definite, we say that it is *nondefinite*. For instance, $g(X, Y) = X^2$ is a quadratic form which is nondefinite because, for $x = 0$ and any $y \neq 0$, it is zero! A basic result in the theory of quadratic forms (see any serious course in Linear Algebra!) gives us a criterion which says when a quadratic form is positive definite, negative definite, or nondefinite. The point is to consider the principal minors

$$\Delta_1 = a_{11}, \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, ..., \Delta_n = \begin{vmatrix} a_{11} & a_{12} & . & . & a_{1n} \\ a_{21} & a_{22} & . & . & a_{2n} \\ . & & & & . \\ . & & & & . \\ a_{n1} & a_{n2} & . & . & a_{nn} \end{vmatrix},$$

of the matrix (a_{ij}).

THEOREM 77. *(Sylvester's criterion) A quadratic form*

$$g(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} X_i X_j$$

is positive definite if and only if

$$\Delta_1 > 0, \Delta_2 > 0, \Delta_3 > 0, ..., \Delta_n > 0.$$

It is negative definite if and only if

$$\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 > 0, ..., (-1)^n \Delta_n > 0.$$

If none of these both conditions are fulfilled, the quadratic form g is nondefinite.

For instance,

$$g(x, y, z) = x^2 + y^2 - z^2$$

is nondefinite because $\Delta_1 = 1 > 0$, $\Delta_2 = 1 > 0$ and $\Delta_3 = -1 < 0$.

Now, we are ready to prove our above announced criterion for distinguishing the local extremum points between all the critical points.

THEOREM 78. *(The Decision Theorem) Let $f : A \to \mathbb{R}$ be a function of class C^2 (it has continuous partial derivatives of second order on A) defined on an open subset A of $\mathbb{R}^n$. Let $\mathbf{a} \in A$ be a critical point of f and let*

$$g(h_1, h_2, ..., h_n) = d^2 f(\mathbf{a})(h_1, h_2, ..., h_n)$$

be the second differential of f at the point $\mathbf{a}$. It is in fact the quadratic form

$$g(h_1, h_2, ..., h_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) h_i h_j.$$

i) Assume that $d^2 f(\mathbf{a})$ is not identical to zero and that $d^2 f(\mathbf{a})$ is a negative definite quadratic form. Then $\mathbf{a}$ is a local maximum point for f.

ii) Assume that $d^2 f(\mathbf{a})$ is not identical to zero and that $d^2 f(\mathbf{a})$ is a positive definite quadratic form. Then $\mathbf{a}$ is a local minimum point for f.

Let k be the first natural number such that f is of class C^k on A and $d^k f(\mathbf{a})$ is not identical to zero.

iii) If k is even and if

$$d^k f(\mathbf{a})(h_1, h_2, ..., h_n) < 0$$

for any $h_1, h_2, ..., h_n$ not all zero, then $\mathbf{a}$ is local maximum point for f.

iv) If k is even and if

$$d^k f(\mathbf{a})(h_1, h_2, ..., h_n) > 0$$

for any $h_1, h_2, ..., h_n$ not all zero, then $\mathbf{a}$ is local minimum point for f. If k is odd and $d^k f(\mathbf{a}) \neq 0$, then $\mathbf{a}$ is not a local extremum point.

PROOF. Let us denote by $\mathbf{h}$ the variable vector $(h_1, h_2, ..., h_n)$ and let us write Taylor's formula (3.3) for $m = 1$. We get:

$$f(\mathbf{a}+\mathbf{h})-f(\mathbf{a}) = \frac{1}{2}d^2f(\mathbf{c_h})(\mathbf{h}), \tag{1.1}$$

where $\mathbf{c_h}$ is a point on the segment $[\mathbf{a}, \mathbf{a}+\mathbf{h}]$ and $\|\mathbf{h}\| < r$, with $r > 0$, a sufficiently small real number such that $B(\mathbf{a}, r) \subset A$ and. Here $df(\mathbf{a}) = 0$ because $\mathbf{a}$ was considered to be a critical point. Since $d^2f(\mathbf{x})$ is continuous as a function of $\mathbf{x}$ ($d^2f(\mathbf{x})(\mathbf{h}) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})h_ih_j$) and the second order derivatives are continuous by our hypothesis!), eventually in a smaller ball $B(\mathbf{a}, r')$ with centre at $\mathbf{a}$ and of radius $r' \leq r$, one has that the sign of $d^2f(\mathbf{x})(\mathbf{h})$, $\mathbf{x} \in B(\mathbf{a}, r')$, is the same like the sign of $d^2f(\mathbf{a})(\mathbf{h})$ (why?). Hence, the sign of the difference $f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})$ is the same with the sign of $d^2f(\mathbf{a})(\mathbf{h})$ for $\|\mathbf{h}\| < r'$. Now, the statements of the theorem becomes very clear. Indeed, let us consider for instance that the quadratic form $d^2f(\mathbf{a})$ is negative definite, i.e. $d^2f(\mathbf{a})(\mathbf{h}) < 0$ for any $\mathbf{h} \neq \mathbf{0}$. Then $d^2f(\mathbf{x})(\mathbf{h}) < 0$ for any $\mathbf{x}$ in a small ball $B(\mathbf{a}, r')$ like above and for any $\mathbf{h} \neq \mathbf{0}$. So, in (1.1), if we take $\mathbf{h}$ such that $\|\mathbf{h}\| < r'$, i.e. $\mathbf{x} = \mathbf{a}+\mathbf{h} \in B(\mathbf{a}, r')$, we get that $f(\mathbf{x}) \leq f(\mathbf{a})$ for any $\mathbf{x}$ in $B(\mathbf{a}, r')$, i.e. $\mathbf{a}$ is a local maximum point for f. To prove ii) we proceed in the same way (do it!).

To prove iii) and iv) we use the Taylor formula:

$$f(\mathbf{a}+\mathbf{h})-f(\mathbf{a}) = \frac{1}{k!}d^kf(\mathbf{c_h})(\mathbf{h})$$

and the fact that a homogenous polynomial $P(X_1, X_2, ..., X_n)$ of odd degree k can NEVER have a constant sign in a neighborhood of $\mathbf{0}$. If k is even and if $d^kf(\mathbf{a})(\mathbf{h}) < 0$ for any nonzero $\mathbf{h}$, there is a whole small ball $B(\mathbf{a}, \varepsilon)$ on which $d^kf(\mathbf{x})(\mathbf{h}) < 0$ for any nonzero $\mathbf{h}$. So, on such a ball, $f(\mathbf{a}+\mathbf{h})-f(\mathbf{a}) < 0$, i.e. $\mathbf{a}$ is a local maximum point for f, etc. □

Let us apply this theorem to the following problem. Let

$$f(x, y) = x^4 + y^4 - 4xy, f : \mathbb{R}^2 \to \mathbb{R}.$$

Let us find all the local extrema for f. First of all we find the critical points: $\frac{\partial f}{\partial x} = 4x^3-4y = 0$ and $\frac{\partial f}{\partial y} = 4y^3-4x = 0$ imply $x^9-x = 0$. So we find the following critical points: $M_1(0, 0)$, $M_2(1, 1)$ and $M_3(-1, -1)$. In order to apply Theorem 78 we need to compute the Hessian matrix of f, i.e. the matrix of the quadratic form d^2f, at every of the three critical points.

$$A = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}.$$

At M_1 the matrix is

$$\begin{pmatrix} 0 & -4 \\ -4 & 0 \end{pmatrix}.$$

Since $\Delta_1 = 0$, from Theorem 78 we obtain that M_1 is not a local extremum for f. At M_2 and M_3 the Hessian matrix is

$$\begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}.$$

So, $\Delta_1 = 12 > 0$ and $\Delta_2 = 144 - 16 = 128 > 0$. Thus, both M_2 and M_3 are local minimum points.

EXAMPLE 17. *(regression line) In the Cartesian xOy plane we consider n distinct points $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, ..., $M_n(x_n, y_n)$. We search for the "closest" line $y = ax + b$ (the regression line) with respect to this set of points. Here, the "distance" from the set $\{M_i\}$ up to the line $y = ax + b$ is the "square" distance distance:*

$$SD(a, b) = \sqrt{\sum_{i=1}^{n} [y_i - (ax_i + b)]^2}. \tag{1.2}$$

The "closest" line $y = ax + b$ is that one for which the nonnegative function $SD(a, b)$ is minimum. Thus, we must find the local minimum points for the two variable function $SD(a, b)$. Let us find the critical points by solving the 2×2 system:

$$\begin{cases} \frac{\partial SD}{\partial a} = 2\sum_{i=1}^{n} -x_i(y_i - ax_i - b) = 0 \\ \frac{\partial SD}{\partial b} = 2\sum_{i=1}^{n} -(y_i - ax_i - b) = 0 \end{cases}. \tag{1.3}$$

Let us write this system in the canonical way

$$\begin{cases} \left(\sum x_i^2\right) a + \left(\sum x_i\right) b = \sum x_i y_i \\ \left(\sum x_i\right) a + nb = \sum y_i \end{cases}. \tag{1.4}$$

If not all the points $\{M_i\}$ are on the same line (in this last case the regression line is obvious the line on which these points are!), the determinant of this system cannot be zero (use the Cauchy-Schwarz inequality from Linear Algebra, the equality special case!). So we have a unique solution (a_0, b_0) of this system. Let us prove that this point realize a minimum for the square distance function $SD(a, b)$. Indeed, the Hessian matrix of f is

$$\begin{pmatrix} 2\sum x_i^2 & 2\sum x_i \\ 2\sum x_i & 2n \end{pmatrix}.$$

In this case, $\Delta_1 = 2\sum x_i^2 > 0$ (otherwise all the points M_i would be on the Oy-axis) and $\Delta_2 = 4\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]$. In order to prove that

Δ_2 is greater than zero we consider in $\mathbb{R}^n$ the vectors $\mathbf{1} = (1, 1, ..., 1)$, $\mathbf{x} = (x_1, x_2, ..., x_n)$ and write the inequality Cauchy-Schwarz for them: $|\langle \mathbf{1}, \mathbf{x} \rangle| \leq \|\mathbf{1}\| \cdot \|\mathbf{x}\|$ or (by squaring) $\left(\sum x_i\right)^2 \leq n \sum x_i^2$. We know that equality appears if and only if the two vectors are collinear, i.e. if and only if $x_1 = x_2 = ... = x_n$. But this last case appears only if the points $\{M_i\}$ are on a vertical line and we just assumed that $\{M_i\}$ are not collinear. Hence, $\Delta_2 > 0$ and the point (a_0, b_0) is a local (in fact a global-why?) minimum for the square distance function SD.

The method described above is said to be the *least squares method* (LSM). It can be generalized to other classes of curves or surfaces.

Let us apply the LSM for the set of points $M_1(-1, 1)$, $M_2(0, 0)$, $M_3(1, 2)$ and $M_4(2, 3)$. To solve the system (1.4) we must compute $\sum x_i^2 = 6$, $\sum x_i = 2$, $\sum x_i y_i = 7$ and $\sum y_i = 6$. Then the system becomes:

$$\begin{cases} 6a + 2b = 7 \\ 2a + 4b = 6 \end{cases}.$$

We get $a = 4/5$ and $b = 11/10$. Hence, the regression line is $y = \frac{4}{5}x + \frac{11}{10}$.

2. Problems

1. Find the local extrema for:

a)

$$f(x, y, z) = x^2 + y^2 + z^2 - xy + x - 2z;$$

b)

$$f(x, y) = x^3 y^2 (6 - x - y), x > 0, y > 0;$$

c)

$$f(x, y) = (x - 2)^2 + (y + 7)^2$$

(try directly, without the above algorithm!);

d)

$$f(x, y) = xy(2 - x - y);$$

e)

$$f(x, y) = \ln(1 - x^2 - y^2);$$

f)

$$f(x, y) = x^3 + y^3 - 3xy;$$

g)

$$f(x, y) = x^4 + y^4 - 2x^2 + 4xy - 2y^2;$$

h)

$$f(x, y, z) = xyz(4a - x - y - z),$$

a, x, y and z are not zero.

2. Find α, β, γ such that

$$f(x, y) = 2x^2 + 2y^2 - 3xy + \alpha x + \beta y + \gamma$$

has a minimum equal to zero in $A(2, -1)$.

3. A price function is of the form

$$f(x, y) = x^2 + xy + y^2 - 3ax - 3by,$$

where a, b are constant numbers. Find a and b such that the minimum of f be the biggest possible.

4. Study the local extrema for $f(x, y) = x^4 + y^4 - x^2$.

CHAPTER 11

Implicitly defined functions

1. Local Inversion Theorem

Let $\mathbf{a}$ be a point in $\mathbb{R}^n$. By a (open) *neighborhood* A of $\mathbf{a}$ we mean any open subset A of $\mathbb{R}^n$ which contains the point $\mathbf{a}$. So, if A is a neighborhood of $\mathbf{a}$, then there is an open ball $B(\mathbf{a},r)$, centered at $\mathbf{a}$ and of radius $r > 0$ which is contained in A.

DEFINITION 33. *Let A and B be two open subsets of $\mathbb{R}^n$. A vector function $\mathbf{f} : A \to B$ is said to be a diffeomorphism between A and B if: i) $\mathbf{f}$ is a bijection; ii) $\mathbf{f}$ is of class C^1 on A and iii) $\mathbf{f}^{-1} : B \to A$ is of class C^1 on B.*

For instance, $f_a : \mathbb{R} \to \mathbb{R}$, $f_a(x) = x+a$ is a diffeomorphism because its inverse $g(x) = x-a$ is of class C^1 on $\mathbb{R}$. But the mapping $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^5$ is not a diffeomorphism because its inverse $g(x) = \sqrt[5]{x}$ is not differentiable at $x = 0$ (why?).

REMARK 31. *It is easy to see that the composition between two diffeomorphisms is also a diffeomorphism (prove it!).*

THEOREM 79. *Let $\mathbf{f} : A \to B$ be a diffeomorphism and let $\mathbf{a}$ be a point in A. Then the linear mapping $d\mathbf{f}(\mathbf{a}) : \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism of real vector spaces. In particular, the Jacobi matrix $J_{\mathbf{a},\mathbf{f}}$ of $\mathbf{f}$ at $\mathbf{a}$ is invertible and its determinant has a constant sign in a neighborhood of $\mathbf{a}$. This means that there is an open ball $B(\mathbf{a},r)$, $r > 0$, contained in A, such that $\det J_{\mathbf{x},\mathbf{f}} > 0$ (or $\det J_{\mathbf{x},\mathbf{f}} < 0$) for any $\mathbf{x} \in B(\mathbf{a},r)$. In fact, the sign of $\det J_{\mathbf{x},\mathbf{f}}$ is the same with the sign of $\det J_{\mathbf{a},\mathbf{f}}$ for any $\mathbf{x}$ in $B(\mathbf{a},r)$.*

PROOF. Let $\mathbf{g} : B \to A$ be the inverse of $\mathbf{f}$ and let $\mathbf{b} = \mathbf{f}(\mathbf{a})$. Then $\mathbf{g} \circ \mathbf{f} = \mathbf{1}_A$, the identity mapping defined on A. Now, Theorem 69 says that $J_{\mathbf{b},\mathbf{g}} \cdot J_{\mathbf{a},\mathbf{f}} = \mathbf{1}_{n\times n}$, the $n \times n$ identity matrix. Hence, the Jacobi matrix $J_{\mathbf{a},\mathbf{f}}$ is invertible, i.e. $df(\mathbf{a})$ is an isomorphism of real vector spaces (see the connections between the linear mappings and their corresponding matrices, w.r.t. a fixed basis in $\mathbb{R}^n$). Moreover, $\det J_{\mathbf{a},\mathbf{f}}$ cannot be zero (why?), say positive, for instance. Since $\mathbf{f}$ is a function of class C^1 on A, all the partial derivatives which appear as entries in the matrix

of $J_{\mathbf{x},\mathbf{f}}$ are continuous. Thus, the mapping $\mathbf{x} \rightsquigarrow \det J_{\mathbf{x},\mathbf{f}}$ (denoted here by T) is a continuous mapping on A, particularly at $\mathbf{a}$. Since $T(\mathbf{a}) > 0$, we state that there is at least one small positive real number $r > 0$ such that for any $\mathbf{x}$ in $B(\mathbf{a}, r)$ we have $T(\mathbf{x}) > 0$. Indeed, otherwise, we could construct a sequence $\{\mathbf{x}^m\}$ of elements in A which is convergent to $\mathbf{a}$ and for which $T(\mathbf{x}^m) \leq 0$, $m = 1, 2, ...$ The continuity of T would imply that $T(\mathbf{a}) \leq 0$, a contradiction! Hence, there is such a small ball $B(\mathbf{a}, r)$, $r > 0$ on which $T(\mathbf{x})$ is positive and the proof is complete. □

Thus, locally, around a fixed point $\mathbf{a}$, the differential $df(\mathbf{x})$ is invertible. We know that the increment $\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})$ of the function $\mathbf{f}$ at $\mathbf{a}$ can be well approximated by $df(\mathbf{a})(\mathbf{x} - \mathbf{a})$ (see Taylor's formula for many variables). A natural question arises: " Is $\mathbf{f}$ itself invertible in a neighborhood of $\mathbf{a}$?" If the function $\mathbf{f}$ describes a physical phenomenon, this means that this phenomenon can be reversible whenever we become closer and closer to the point $\mathbf{a}$ and, this is very important to be known in the engineering practice. The following result is fundamental in all pure and applied mathematics. It is a reverse result relative to the above theorem

THEOREM 80. *(Local Inversion Theorem) Let A be an open subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to \mathbb{R}^n$ be a function of class C^1 on A. Let $\mathbf{a}$ be a point in A such that* $\det J_{\mathbf{a},\mathbf{f}} \neq 0$. *Then there is a neighborhood U of $\mathbf{a}$, $U \subset A$, such that the restriction of $\mathbf{f}$ to U, $\mathbf{f}\mid_U : U \to V = \mathbf{f}(U)$, is a diffeomorphism. In particular,* $\det J_{\mathbf{x},\mathbf{f}} \neq 0$ *on U and if $\mathbf{g} : V \to U$ is the local inverse of $\mathbf{f}$ ($\mathbf{g} = (\mathbf{f}\mid_U)^{-1}$), then* $\det J_{\mathbf{f}(\mathbf{x}),\mathbf{g}} = \frac{1}{\det J_{\mathbf{x},\mathbf{f}}}$ *and* $J_{\mathbf{f}(\mathbf{x}),\mathbf{g}} = (J_{\mathbf{x},\mathbf{f}})^{-1}$.

PROOF. (only for $n = 1$. See a complete proof in Section 7 of this chapter) Let $\mathbf{f} = f$ and $\mathbf{a} = a \in A \subset \mathbb{R}$ be the usual notation in this restricted case. Now $\det J_{\mathbf{a},\mathbf{f}} = f'(a)$ (why?) and the hypotheses says that $f'(a)$ is not zero, say that $f'(a) > 0$. Since f' is continuous (f is of class C^1 on A), like in the proof of the above theorem, we can conclude that there is an open ball $U = B(a, r) = (a - r, a + r)$, $r > 0$, on which f' is positive, i.e. $f'(x) > 0$ for any x in U. This means that on this U our function f is strictly increasing. So, the restriction of f to U has an inverse $g : V = f(U) \to U$. Since f is continuous and strictly increasing, one can easily prove that $f^{-1} = g$ is continuous on V (prove it! or find by yourself a previous result from which this statement immediately comes!). We now prove that this function $g(y) = x$, where $y = f(x)$, is differentiable on V. Indeed, let $b = f(a)$ be a point in V and let $\{y_n = f(x_n)\}$ be a convergent sequence to b. Then $\{x_n = g(y_n)\}$ tends

to a (because of the continuity of g) and

$$\lim_{y_n \to b} \frac{g(y_n) - g(b)}{y_n - b} = \lim_{x_n \to a} \frac{x_n - a}{f(x_n) - f(a)} = \frac{1}{f'(a)}.$$

Thus, g is differentiable at b and $g'(b) = \frac{1}{f'(a)}$. □

EXAMPLE 18. *(Polar coordinates) Let $M(x, y)$ be a point in the Cartesian plane $\{O, \mathbf{i}, \mathbf{j}\}$ and let $\rho = \sqrt{x^2 + y^2}$be the distance from M up to the origin O. Let θ be the unique angle in $[0, 2\pi]$ such that $x = \rho \cos\theta$ and $y = \rho \sin\theta$ (prove that such an angle exists and that it is unique!-see Fig.10.1). Let us consider $A = (0, \infty) \times (0, 2\pi) \subset \mathbb{R}^2$ and $B = \mathbb{R}^2 \setminus \{[0, \infty) \times \{0\}\}$ in the same $\mathbb{R}^2$. Let $\mathbf{f} : A \to B$, $\mathbf{f}(\rho, \theta) = (\rho \cos\theta, \rho \sin\theta)$. It is easy to see that $\det J_{(\rho,\theta),\mathbf{f}} = \rho \neq 0$. It it easy to prove that this $\mathbf{f}$ is a diffeomorphism. The analytical expression of its inverse $\mathbf{f}^{-1}$ is not so simple (why?-find it!). The new "coordinates" (ρ, θ) are called the polar coordinates of M. For instance, the Cartesian equation of the circle $x^2 + y^2 = R^2$ may be simply written in polar coordinates like $\rho = R$!*

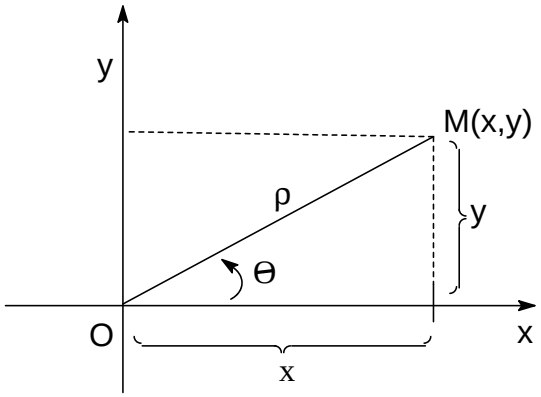


Fig. 10.1

DEFINITION 34. *(regular transformations) Let A be an open subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to \mathbb{R}^n$ be a mapping defined on A with values in $\mathbb{R}^n$. We say that $\mathbf{f}$ is a regular transformation at the point $\mathbf{a}$ of A if there is a neighborhood U of $\mathbf{a}$, $U \subset A$, such that the restriction of $\mathbf{f}$ to U give rise to a diffeomorphism $\mathbf{f}\,|_U \colon U \to V = \mathbf{f}(U)$. If $\mathbf{f}$ is regular at any point of A, we say that $\mathbf{f}$ is a regular transformation on A or that $\mathbf{f}$ is a local diffeomorphism on A.*

In particular, for a local diffeomorphism $\mathbf{f}$, one has that $\det J_{\mathbf{a},\mathbf{f}} \neq 0$ on A and, if in addition A is connected, then $\det J_{\mathbf{a},\mathbf{f}}$ has a constant sign

on A (why?). For instance, the polar coordinates transformation (see Example 18) is a regular transformation (prove it!). The composition between two regular transformations is again a regular transformation. Such transformations are "good" for engineers. They are locally sufficiently "smooth". This means that they do not produce "breaking" or "noncontinuous (broken) velocities", or "corners".

REMARK 32. *The local inversion theorem applied to the regular transformations gives rise to some basic properties of these last ones. For instance, a regular transformation* $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ *carries an open subset* A *of* $\mathbb{R}^n$ *into the open subset* $\mathbf{f}(A)$ *(why?). If* A *is a domain, i.e. if* A *is an open and a connected subset of* $\mathbb{R}^n$*, then* $\mathbf{f}(A)$ *is also a domain of* $\mathbb{R}^n$ *(why?). Moreover, the Jacobian* $\det J_{\mathbf{x},\mathbf{f}}$ *has the same sign on* A*, if* A *is a domain (try to prove it!).*

2. Implicit functions

What is the difference between the curves: 1) $C_1 = \{(x, y) \in \mathbb{R}^2 : y = \sqrt{1 - x^2}\}$ and 2) $C_2 = \{(x, y) : x^2 + y^2 = 1, y \geq 0\}$? They represent the same object, the half of the circle of radius 1, with centre at O, which is above the Ox-axis, but... the representations are distinct. In the first case we have an "explicit" representation, i.e. we can write $y = f(x)$, this means that we can write one variable as a known function of the other one. In the second case we have to compute y as a function of x from the "implicit" relation $x^2 + y^2 = 1$. In our case this can be done, but in other cases such an explicit computation cannot be done. For instance, it is very difficult to express y as a function of x if

$$(*)\ x^3 + 2y^3 - 3xy = 0.$$

But, if we knew that such an expression $y = f(x)$ exists (theoretically) in a neighborhood of a point on the curve, say $(1, 1)$, we can compute the "velocity" $f'(1)$, the "acceleration" $f''(1)$, $f'''(1)$, etc. Practically, we proceed as follows. Let us write again the implicit relation $(*)$ with $f(x)$ instead of y :

$$x^3 + 2f(x)^3 - 3xf(x) = 0$$

and let us differentiate it with respect to x :

$$(**)\ 3x^2 + 6f(x)^2 f'(x) - 3f(x) - 3xf'(x) = 0.$$

We see that always (does not matter the implicit relation is!) the first derivative $f'(x)$ appears to power 1, i.e. it can be "linearly" computed

from $(**)$:

$$f'(x)=\frac{f(x)-x^2}{2f(x)^2-x}. \tag{2.1}$$

If one put $x=1$ in (2.1) one obtains $f'(1)=0$. If we differentiate again formula (2.1) with respect to x, we get

$$f''(x)=\frac{-2f(x)^2f'(x)-4xf(x)^2-xf'(x)+4x^2f(x)f'(x)+f(x)+x^2}{\left[2f(x)^2-x\right]^2}.$$

If here we substitute $f'(x)$ with its expression from (2.1), we get the expression of $f''(x)$ only as an explicit function of x and of $f(x)$. Let us put now $x=1$ and we obtain $f''(1)$, etc.

In our above discussion we supposed that our equation can be uniquely solved with respect to y. But this is not always true. For instance, if $x^2+y^2=1$, then $y(x)=\pm\sqrt{1-x^2}$, so that in any neighborhood of $(1,0)$ we cannot find a UNIQUE function $y=y(x)$ such that $x^2+y(x)^2=1$. Hence, we cannot compute $y'(1)$, $y''(1)$, etc. This is why we need a mathematical result to precisely say when we have or not such a unique "implicit" function.

THEOREM 81. *($(1\leftrightarrow 1)$ Implicit Function Theorem) Let A be an open subset of $\mathbb{R}^2$ and let $F:A\to\mathbb{R}$ be a function of two variables which verifies the following properties at a fixed point (a,b) of A :*

i) F is a function of class C^1 on A.

ii) $F(a,b)=0$, i.e. (a,b) is a solution of the equation $F(x,y)=0$.

iii) $\frac{\partial F}{\partial y}(a,b)\neq 0$.

Then there is a neighborhood U of a, a neighborhood V of b with $U\times V\subset A$ and a unique function $f:U\to V$ such that:

1) $F(x,f(x))=0$ for all x in U.

2) $f(a)=b$.

3) f is of class C^1 on U and

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

for all x in U.

PROOF. We construct an auxiliary function

$$\mathbf{\Phi}=(\varphi_1,\varphi_2):A\to\mathbb{R}^2,\mathbf{\Phi}(x,y)=(x,F(x,y))$$

for all (x,y) in A. Thus, $\varphi_1(x,y)=x$ and $\varphi_2(x,y)=F(x,y)$. We are to apply the Local Inversion Theorem to this function $\mathbf{\Phi}$. Let us compute

the Jacobi matrix of $\mathbf{\Phi}$ at (a,b) :

$$J_{(a,b),\mathbf{\Phi}} = \begin{pmatrix} 1 & 0 \\ \frac{\partial F}{\partial x}(a,b) & \frac{\partial F}{\partial y}(a,b) \end{pmatrix}.$$

Since $\mathbf{\Phi}(a,b) = (a,0)$ and since $\det J_{(a,b),\mathbf{\Phi}} = \frac{\partial F}{\partial y}(a,b) \neq 0$, Local Inversion Theorem 80 says that there is an open neighborhood $U \times V$ of (a,b) and an open neighborhood $U \times W$ of $(a,0)$ (why can we take the same U?) such that the restriction $\mathbf{\Phi} \mid_{U\times V} : U \times V \to U \times W$ of Φ to $U \times V$ is a diffeomorphism. Let $\mathbf{\Psi} = (\psi_1, \psi_2) : U \times W \to U \times V$ the inverse of this diffeomorphism. Let us define $f(x) = \psi_2(x,0)$ for any x in U. It is clear that $f : U \to V$ is of class C^1 on U, $f(a) = b$ and for any x of U we have

$$\begin{aligned}(x,0) = \mathbf{\Phi}[\mathbf{\Psi}(x,0)] = \ & \mathbf{\Phi}[\psi_1(x,0), \psi_2(x,0)] \\ = \mathbf{\Phi}[x, f(x)] &= (x, F(x, f(x))),\end{aligned}$$

i.e. $F(x, f(x)) = 0$, for any x in U. The function $f : U \to V$ is of class C^1 on U because $\psi_2(X,Y)$ has continuous partial derivative with respect to X at any point of the form $(x,0)$ for any x in U. Let us differentiate totally with respect to x (this means that x is considered not only like "the first" partial free variable of $F(x,y)$, but even as an implicit hidden variable in $y = f(x)$) the relation $F(x, f(x)) = 0$:

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

thus

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

for any x in U. Since $\det J_{(x,y),\mathbf{\Phi}} \neq 0$ on $U \times V$ (why?) we get from

$$J_{(x,y),\mathbf{\Phi}} = \begin{pmatrix} 1 & 0 \\ \frac{\partial F}{\partial x}(x,y) & \frac{\partial F}{\partial y}(x,y) \end{pmatrix}$$

that $\frac{\partial F}{\partial y}(x, f(x)) \neq 0$ for any x in U.

If g was another function defined on an open neighborhood U_1 of a, which verifies the conditions 1), 2) and 3) then, on the neighborhood $U_2 = U \cap U_1$ we would have

$$\psi_2(x, F(x, g(x)) = g(x)$$

for any x in U_2, or $\psi_2(x,0) = g(x) = f(x)$ for any x in U_2. Hence, the uniqueness reefers to another smaller neighborhood of U on which f and g are equal. In some conditions, this uniqueness can be extended to the whole initial U or even to the whole $pr_x(A)$, the projection of A on the Ox-axis. □

Let us consider again the implicit equation

$$x^3 + 2y^3 - 3xy = 0$$

and let us study it around the solution $(1,1)$. Since $\frac{\partial F}{\partial y}(1,1) = 3 \neq 0$, the (1-1) Implicit Function Theorem says that there is a neighborhood U of $x = 1$, a neighborhood V of $y = 1$ and a function $f : U \to V$, of class C^1 on U, such that the points $\{(x, f(x)) : x \in U\}$ are on the plane curve $x^3 + 2y^3 - 3xy = 0$, i.e. $x^3 + 2f(x)^3 - 3xf(x) = 0$ for any x in U. Now, if we are sure on the existence of such a f, we can use different approximation methods to compute it (approximately!). The worst situation is when the conditions of the Implicit Function Theorem fail and we try to compute $y = f(x)$ approximately! Usually, in this last case one has more then one function $y = f(x)$ which verify our equation and during our approximate process we "jump" from one "branch" to another one, the obtained values for "$f(x)$" having a chaotic behavior. For instance, around the point $(1,0)$, the implicit solution of the equation $x^2+y^2 = 1$ with respect to y has two branches: $y = \sqrt{1-x^2}$ and $y = -\sqrt{1-x^2}$. This is because $\frac{\partial F}{\partial y}(1,0) = 0$ and the Implicit Function Theorem fails around the point $(1,0)$.

There are two directions for generalizations of this basic theorem. One reefers to increase the number of variables and the other to consider vector fields relations, i.e. a system of implicit equations. We do not prove these generalizations because these proofs do not contain new ideas and the "many" variables notation are too sophisticated.

THEOREM 82. *($(n \leftrightarrow 1)$ Implicit Function Theorem) Let A be an open subset of $\mathbb{R}^{n+1}$, let $(\mathbf{a}, b) = (a_1, a_2, ..., a_n, b)$ be a point of A and let $F : A \to \mathbb{R}$, $F(x_1, x_2, ..., x_n, y\ )$ be a function of $n+1$ variables which verifies the following conditions:*

i) F is of class C^1 on A, i.e. it has continuous partial derivatives with respect to each of its $n+1$ variable.

ii) $F(\mathbf{a}, b) = 0$.

iii) $\frac{\partial F}{\partial y}(\mathbf{a}, b) \neq 0$.

Then there is a neighborhood U of $\mathbf{a}$, a neighborhood V of b such that $U \times V \subset A$ and a unique function $f : U \to V$ such that:

1) $F[\mathbf{x}, f(\mathbf{x})] = 0$ for all $\mathbf{x}$ in U.

2) $f(\mathbf{a}) = b$.

3) f is of class C^1 on U and

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

for any $\mathbf{x}$ in U.

For a proof see [**FS**]. Let us take the following equation:

$$2x^3 + y^3 + 2z^3 - 5xyz = 0$$

and its solution $M(1,1,1)$ (prove this!). Since $\frac{\partial F}{\partial z}(1,1,1) = 1 \neq 0$, one can apply the last theorem and can write $z = z(x,y)$ around the point $(1,1)$. Let us compute $\frac{\partial^2 z}{\partial x \partial y}(1,1)$. The most practical way is to put $z = z(x,y)$ into our equation:

$$2x^3 + y^3 + 2z(x,y)^3 - 5xyz(x,y) = 0$$

and let us differentiate this with respect to x and to y :

$$6x^2 + 6z(x,y)^2 \frac{\partial z}{\partial x}(x,y) - 5yz(x,y) - 5xy\frac{\partial z}{\partial x}(x,y) = 0,$$

$$3y^2 + 6z(x,y)^2 \frac{\partial z}{\partial y}(x,y) - 5xz(x,y) - 5xy\frac{\partial z}{\partial y}(x,y) = 0.$$

From these equations we compute

$$\frac{\partial z}{\partial x}(x,y) = \frac{6x^2 - 5yz}{5xy - 6z^2},\ \frac{\partial z}{\partial y}(x,y) = \frac{3y^2 - 5xz}{5xy - 6z^2}. \tag{2.2}$$

Now,

$$\frac{\partial^2 z}{\partial x \partial y} = \frac{\partial}{\partial x}\left(\frac{3y^2 - 5xz(x,y)}{5xy - 6z(x,y)^2}\right) = \tag{2.3}$$

$$\frac{(-5z - 5x\frac{\partial z}{\partial x})(5xy - 6z^2) - (3y^2 - 5xz)(5y - 12z\frac{\partial z}{\partial x})}{(5xy - 6z^2)^2}.$$

We need to compute $\frac{\partial z}{\partial x}(1,1)$, so we must use formula (2.2) and find $\frac{\partial z}{\partial x}(1,1) = -1$ (because $z(1,1) = 1$). Come back to formula (2.3) and find $\frac{\partial^2 z}{\partial x \partial y}(1,1) = 34$.

We consider now many relations, i.e. instead of the scalar function F we take a vector function $\mathbf{F} = (F_1, F_2, ..., F_m) : A \to \mathbb{R}^m$, where A is an open subset in $\mathbb{R}^{n+m}$.

THEOREM 83. *Let A be an open subset of $\mathbb{R}^{n+m}$ and let*

$$(\mathbf{a}, \mathbf{b}) = (a_1, a_2, ..., a_n; b_1, b_2, ..., b_m)$$

be a point in A. Let $\mathbf{F} = (F_1, F_2, ..., F_m) : A \to \mathbb{R}^m$ be a function which verifies the following conditions:

i) $\mathbf{F}$ is a function of class C^1 on A.

ii) $\mathbf{F}(\mathbf{a},\mathbf{b})=\mathbf{0}$, *i.e.*

$$\begin{cases} F_1(a_1,a_2,...a_n;b_1,b_2,...,b_m)=0 \\ \qquad\qquad\qquad\qquad . \\ \qquad\qquad\qquad\qquad . \\ \qquad\qquad\qquad\qquad . \\ F_m(a_1,a_2,...a_n;b_1,b_2,...,b_m)=0 \end{cases}.$$

iii) For $\mathbf{F}(\mathbf{x},\mathbf{y})=\mathbf{F}(x_1,x_2,...,x_n;y_1,y_2,...,y_m)$, *we define the Jacobian matrix relative to* $\mathbf{y}=(y_1,y_2,...,y_m)$ *only, as follows:*

$$J_{\mathbf{y},\mathbf{F}}(\mathbf{x},\mathbf{y})=\begin{pmatrix} \frac{\partial F_1}{\partial y_1}(\mathbf{x},\mathbf{y}) & . & . & . & \frac{\partial F_1}{\partial y_m}(\mathbf{x},\mathbf{y}) \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \frac{\partial F_m}{\partial y_1}(\mathbf{x},\mathbf{y}) & . & . & . & \frac{\partial F_m}{\partial y_m}(\mathbf{x},\mathbf{y}) \end{pmatrix}$$

The condition is that $\det J_{\mathbf{y},\mathbf{F}}(\mathbf{a},\mathbf{b})\neq 0$. *This last determinant can be suggestively denoted by*

$$\det J_{\mathbf{y},\mathbf{F}}(\mathbf{a},\mathbf{b})=\frac{D(F_1,F_2,...,F_m)}{D(y_1,y_2,...,y_m)}(\mathbf{a},\mathbf{b}).$$

Then there is a neighborhood $U=U_1\times U_2\times ...\times U_n$ *of* $\mathbf{a}=(a_1,a_2,...,a_n)$, *a neighborhood* $V=V_1\times V_2\times ...\times V_m$ *of* $\mathbf{b}=(b_1,b_2,...,b_m)$, *such that* $U\times V\subset A$ *and a unique function* $\mathbf{f}=(f_1,f_2,...,f_m)$, $f_i:U\to V_i$, $i=1,2,...,m$, *with the following properties:*

1) $\mathbf{F}(\mathbf{x},\mathbf{f}(\mathbf{x}))=\mathbf{0}$ *for any* $\mathbf{x}$ *in* U.

2) $\mathbf{f}(\mathbf{a})=\mathbf{b}$.

3) $\mathbf{f}$ *is of class* C^1 *on* U *and*

$$\frac{\partial f_i}{\partial x_j}(\mathbf{x})=-\frac{\frac{D(F_1,F_2,...,F_m)}{D(y_1,y_2,...,y_{j-1},x_j,y_{j+1},...,y_m)}(\mathbf{x},\mathbf{f}(\mathbf{x}))}{\frac{D(F_1,F_2,...,F_m)}{D(y_1,y_2,...,y_m)}(\mathbf{x},\mathbf{f}(\mathbf{x}))}. \tag{2.4}$$

It is not necessarily to memorize this last cumbersome formula as we can see in the following example.

Let $(C):x^2+y^2-z^2=0$ be a conic surface and let $(E):x^2+2y^2+3z^2-4=0$ be an ellipsoid. Let $\gamma=(C)\cap(E)$ be the intersection curve of them. We see that the point $M(1,0,1)$ is on this curve. The question is if we can find a parametrization of the form

$$\gamma:\begin{cases} x=x(y) \\ \quad y \\ z=z(y) \end{cases},$$

i.e. if we can use y as a parameter for this curve in a neighborhood of M. This is equivalent to see if the following system of the implicit

functions $x = x(y)$ and $z = z(y)$ can be solved around M :

$$\begin{cases} F_1(y; x, z) = x^2 + y^2 - z^2 = 0, \\ F_2(y; x, z) = x^2 + 2y^2 + 3z^2 - 4 = 0. \end{cases} \tag{2.5}$$

Since all our functions are elementary ones, we need only to check the condition *iii)* of the theorem:

$$\frac{D(F_1, F_2)}{D(x, z)}(1, 0, 1) = \begin{vmatrix} \frac{\partial F_1}{\partial x}(1, 0, 1) & \frac{\partial F_1}{\partial z}(1, 0, 1) \\ \frac{\partial F_2}{\partial x}(1, 0, 1) & \frac{\partial F_2}{\partial z}(1, 0, 1) \end{vmatrix} = 16 \neq 0.$$

So, x and z can be seen like functions of y in a neighborhood of M. Let us compute the "velocity" and the "acceleration" at M, along the curve γ. For this, it is not necessarily to use the formula (2.4). Namely, let us put in (2.5) instead of x, $x(y)$ and instead of z, $z(y)$:

$$\begin{cases} x(y)^2 + y^2 - z(y)^2 = 0, \\ x(y)^2 + 2y^2 + 3z(y)^2 - 4 = 0. \end{cases}$$

Let us differentiate both equations with respect to the ONLY free variable y :

$$\begin{cases} 2x(y)x'(y) + 2y - 2z(y)z'(y) = 0, \\ 2x(y)x'(y) + 4y + 6z(y)z'(y) = 0. \end{cases}$$

This is an algebraic linear system in the variables $x'(y)$ and $z'(y)$. Solving it, we get

$$x'(y) = -\frac{5y}{4x(y)}, z'(y) = -\frac{y}{4z(y)}. \tag{2.6}$$

To find $x''(y)$ and $z''(y)$ we differentiate again in the formulas (2.6) and get:

$$x''(y) = -\frac{5}{4}\frac{x(y) - yx'(y)}{x(y)^2}, z''(y) = -\frac{1}{4}\frac{z(y) - yz'(y)}{z(y)^2} \tag{2.7}$$

Now, it is easy to find $x'(0) = 0$, $z'(0) = 0$, $x''(0) = -\frac{5}{4}$ and $z''(0) = -\frac{1}{4}$. Here is an example when the velocity is zero at a point M but the acceleration is not zero at the same point. Thus, one has a nonzero force at a stationary point!

3. Functional dependence

Let A be an open subset of $\mathbb{R}^n$ and let $f_1, f_2, ..., f_m$ be m functions defined on A with real values. We assume that each f_i is of class C^1 on A.

DEFINITION 35. *We say that $\{f_1, f_2, ..., f_m\}$ are functional dependent on A if one of them, say f_m is "a function" of the others*

$$f_1, f_2, ..., f_{m-1},$$

i.e. there is a function $\phi(y_1, y_2, ..., y_{m-1})$ of $m-1$ variables, of class C^1 on $\mathbb{R}^{m-1}$, such that

$$f_m(\mathbf{x}) = \phi[f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_{m-1}(\mathbf{x})],$$

for any $\mathbf{x}$ in A.

For instance,

$$f_1(x_1, x_2, x_3) = x_1 + x_2 + x_3, f_2(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3, \tag{3.1}$$

$$f_3(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2$$

are functional dependent because $f_3 = f_1^2 - 2f_2$. Thus, $\phi(y_1, y_2) = y_1^2 - 2y_2$.

We know from Linear Algebra that $f_1, f_2, ..., f_m$ are linear dependent if there are $\lambda_1, \lambda_2, ..., \lambda_m$ scalars, not all zero, such that

$$\lambda_1 f_1 + \lambda_2 f_2 + ... + \lambda_m f_m = 0, \tag{3.2}$$

i.e. $\lambda_1 f_1(\mathbf{x}) + \lambda_2 f_2(\mathbf{x}) + ... + \lambda_m f_m(\mathbf{x}) = 0$ for any $\mathbf{x}$ in A. Assume that $\lambda_m \neq 0$, divide the equality (3.2) by λ_m and compute f_m:

$$f_m = -\frac{\lambda_1}{\lambda_m} f_1 - \frac{\lambda_2}{\lambda_m} f_2 - ... - \frac{\lambda_{m-1}}{\lambda_m} f_{m-1}.$$

Hence, $f_1, f_2, ..., f_m$ are also functional dependent. Conversely it is not true. For instance, the functions f_1, f_2, f_3 from (3.1) are functional dependent but they are not linear dependent (prove it!). This shows that the notion of functional dependence from Analysis is more general then the notion of linear dependence from Linear Algebra.

THEOREM 84. *Let A be an open subset of $\mathbb{R}^n$ and let $f_1, f_2, ..., f_m : A \to \mathbb{R}$ be m function of class C^1 on A. If $\{f_1, f_2, ..., f_m\}$ are functional dependent on A, then the rank of the Jacobian matrix of $\mathbf{f} = (f_1, f_2, ..., f_m) : A \to \mathbb{R}^m$ is less than m.*

PROOF. Suppose that $f_m(\mathbf{x}) = \phi[f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_{m-1}(\mathbf{x})]$ for all $\mathbf{x}$ in A. Then,

$$\frac{\partial f_m}{\partial x_j} = \frac{\partial \phi}{\partial y_1}\frac{\partial f_1}{\partial x_j} + \frac{\partial \phi}{\partial y_2}\frac{\partial f_2}{\partial x_j} + ... + \frac{\partial \phi}{\partial y_{m-1}}\frac{\partial f_{m-1}}{\partial x_j}$$

for all $j = 1, 2, ..., n$. This means that the m-th row of the matrix $J_{\mathbf{x},\mathbf{f}}$ is a linear combination of the first $m-1$ rows, so the rank of the Jacobian matrix $J_{\mathbf{x},\mathbf{f}}$ is less than m (why?-see any Linear Algebra course). □

We say that $f_1, f_2, ..., f_m$ are *dependent at* $\mathbf{a}$, a point in A, if there is a neighborhood U of $\mathbf{a}$, $U \subset A$, such that $f_1, f_2, ..., f_m$ are dependent on U. If $f_1, f_2, ..., f_m$ are not dependent at $\mathbf{a}$, we say that they are *independent* at $\mathbf{a}$. If $f_1, f_2, ..., f_m$ are independent at any point of A, we say that $f_1, f_2, ..., f_m$ are *independent on* A.

THEOREM 85. *If the rank of* $J_{\mathbf{x},\mathbf{f}}$ *is equal to* m *for any* $\mathbf{x}$ *in* A, *then* $f_1, f_2, ..., f_m$ *are independent on* A.

PROOF. Suppose contrary, namely that there is a point $\mathbf{a}$ in A and a small neighborhood U of $\mathbf{a}$, such that $f_1, f_2, ..., f_m$ are dependent on U. Applying Theorem 84 we get that the rank of $J_{\mathbf{a},\mathbf{f}}$ is less than m. A contradiction! Thus, $f_1, f_2, ..., f_m$ are independent on A. □

We also have a reverse of the last two theorems.

THEOREM 86. *With the above notation and hypotheses, if* $m \leq n$, *if* $\mathbf{f} = (f_1, f_2, ..., f_m)$ *is of class* C^1 *on* A *and if for a fixed point* $\mathbf{a}$ *of* A *one has that the rank of* $J_{\mathbf{a},\mathbf{f}}$ *is less than* m, *then there is a neighborhood* U *of* $\mathbf{a}$, $U \subset A$, *and* s *functions from* $\{f_1, f_2, ..., f_m\}$, *say* $f_1, f_2, ..., f_s$, *which are independent on* U, *such that the other functions* $\{f_{s+1}, f_{s+2}, ..., f_m\}$ *are functional dependent on* $f_1, f_2, ..., f_s$ *on* U. *This means that there are* $m - s$ *functions* $\phi_1, \phi_2, ..., \phi_{m-s}$ *of class* C^1 *on* $\mathbb{R}^s$ *such that*

$$f_{s+1}(\mathbf{x}) = \phi_1(f_1(\mathbf{x}), ..., f_s(\mathbf{x})), ..., f_m(\mathbf{x}) = \phi_{m-s}(f_1(\mathbf{x}), ..., f_s(\mathbf{x}))$$

for all $\mathbf{x}$ *in* U.

The proof involves some more sophisticated tools and we send the interested reader to [**Pal**] or [**FS**]. Let us apply this last theorem in a more complicated example. Let

$$\begin{cases} f_1 = x_1x_3 + x_2x_4 \\ f_2 = x_1x_4 - x_2x_3 \\ f_3 = x_1^2 + x_2^2 - x_3^2 - x_4^2 \\ f_4 = x_1^2 + x_2^2 + x_3^2 + x_4^2 \end{cases}$$

be four functions of variables x_1, x_2, x_3, x_4. The Jacobian matrix of $\mathbf{f} = (f_1, f_2, f_3, f_4)$ at $\mathbf{a} = (1, 1, 0, 0)$ is

$$J_{\mathbf{a},\mathbf{f}} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 1 \\ 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \end{pmatrix}.$$

Since the rank of this matrix is 3 and a nonzero 3×3 determinant involves the first 3 rows, one sees that f_1, f_2, f_3 are functional independent at $\mathbf{a}$ and f_4 is a function of the others in a neighborhood of $\mathbf{a}$.

If we look carefully, we see that $f_4^2 = 4(f_1^2 + f_2^2) + f_3^2$, so f_1, f_2, f_3, f_4 are functional dependent on the whole $\mathbb{R}^4$.

4. Conditional extremum points

Sometimes we have to find the extremum points for a function f defined on a compact subset C of $\mathbb{R}^n$. For instance, let C be the closed ball

$$B[\mathbf{0}, 3] = \{(x, y, z) : x^2 + y^2 + z^2 \leq 9\},$$

centered at $\mathbf{0} = (0, 0, 0)$ and of radius 3. The problem of finding the extremum points of the function $f(x, y, z) = x + 2y + 3z$ defined on C can be divided into two parts. First of all we find the local extrema points of f defined only on the open set

$$B(\mathbf{0}, 3) = \{(x, y, z) : x^2 + y^2 + z^2 < 9\}$$

by using Fermat's theorem, then we consider only the points on the sphere $x^2 + y^2 + z^2 = 9$ and try to find the extremum points $M(x, y, z)$ of f, which verify this last supplementary condition (a constraint). This last problem is an example of a conditional extremum points problem.

The general method for solving such problems is the *"method of Lagrange's multipliers"*. In the following we shall describe this method.

Let A be an open subset of $\mathbb{R}^n$ and let f, g_1, g_2, ..., g_m $(m < n)$ be functions of class C^1 on A. We assume that g_1, g_2, ..., g_m are functional independent on A, particularly, if $\mathbf{g} = (g_1, g_2, ..., g_m)$, its Jacobian matrix $J_{\mathbf{x},\mathbf{g}}$ has the rank m at any point $\mathbf{x}$ of A. Let $S \subset A$ be the set of all solutions (in A) of the following system of equations:

$$\left\{ \begin{array}{l} g_1(x_1, x_2, ..., x_n) = 0 \\ \quad . \\ \quad . \\ \quad . \\ g_m(x_1, x_2, ..., x_n) = 0 \end{array} \right. , \tag{4.1}$$

These equations are called *constraints* or *supplementary conditions* for the variables $x_1, x_2, ..., x_n$.

DEFINITION 36. *We say that a point $\mathbf{a} = (a_1, a_2, ..., a_n)$ of S is a local conditional maximum point for f with the constraints (4.1) if there is a neighborhood U of $\mathbf{a}$, $U \subset A$, such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for any $\mathbf{x}$ in $U \cap S$. The notion of a local conditional minimum point with the same constraints, for the same function f, can be defined in the same manner.*

For instance, $(0, 0)$ is a local conditional minimum for $f(x, y) = x^2 + y$ defined on $\mathbb{R}$ with the constraint $y - x^2 = 0$. Indeed, $f(x, x^2) =$

$2x^2 \geq 0 = f(0,0)$ for any $x \in \mathbb{R}$. But $(0,0)$ is not a local extremum point for f.

Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_m)$ be a variable vector in $\mathbb{R}^m$. These new auxiliary variables $\lambda_1, \lambda_2, ..., \lambda_m$ are called *Lagrange's multipliers* and the new auxiliary function

$$\Phi(x_1, x_2, ..., x_n; \lambda_1, \lambda_2, ..., \lambda_m) = \Phi(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j g_j(\mathbf{x}) \tag{4.2}$$

is called *Lagrange's associated function.*

THEOREM 87. *(Lagrange's Theorem) Let us preserve all the above notation and hypotheses. Assume that* $\mathbf{a}$ *is a local conditional extremum point for* f*, with the constraints (4.1). Then there is a vector* $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, ..., \lambda_m^*)$ *in* $\mathbb{R}^m$ *such that the point*

$$(\mathbf{a}, \boldsymbol{\lambda}^*) = (a_1, a_2, ..., a_n; \lambda_1^*, \lambda_2^*, ..., \lambda_m^*)$$

is a critical (stationary) point for Lagrange's function Φ*, i.e.*

$grad\Phi(\mathbf{a}, \boldsymbol{\lambda}^*) = \mathbf{0}$.

PROOF. (for $n = 2$ and $m = 1$) Suppose that $\mathbf{a}$ is a local conditional maximum point for f. Since $g = g_1$ is functional independent, it cannot be a constant function, say $\frac{\partial g}{\partial x_2}(\mathbf{a}) \neq 0$. We can apply the Implicit Function Theorem and find a function $h : U_1 \to U_2$ of class C^1 on U_1, an appropriate neighborhood of a_1 (U_2 is a neighborhood of a_2), such that $h(a_1) = a_2$, $g(x_1, h(x_1)) = 0$ for all x_1 in U_1 and

$$h'(x_1) = -\frac{\frac{\partial g}{\partial x_1}(x_1, h(x_1))}{\frac{\partial g}{\partial x_2}(x_1, h(x_1))} \tag{4.3}$$

for all x_1 in U_1. We can assume that the neighborhood of $\mathbf{a}$, $U = U_1 \times U_2$ is sufficiently small such that $f(\mathbf{x}) \leq f(\mathbf{a})$ for any $\mathbf{x}$ in U. We define now a new function $D : U_1 \to \mathbb{R}$, $D(x_1) = f(x_1, h(x_1))$ for any x_1 in U_1. Since $D(x_1) \leq D(a_1)$, for all x_1 in U_1, we see that a_1 is a local maximum point for the function D. Use now Fermat's Theorem and find that $D'(a_1) = 0$, or that

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) + \frac{\partial f}{\partial x_2}(\mathbf{a}) \cdot h'(a_1) = 0.$$

Thus,

$$h'(a_1) = -\frac{\frac{\partial f}{\partial x_1}(\mathbf{a})}{\frac{\partial f}{\partial x_2}(\mathbf{a})}. \tag{4.4}$$

But the same $h'(a_1)$ can also be computed from the formula (4.3)

$$h'(a_1) = -\frac{\frac{\partial g}{\partial x_1}(a_1, a_2)}{\frac{\partial g}{\partial x_2}(a_1, a_2)}.$$

If we equals the both expression of $h'(a_1)$ we get

$$\frac{\partial f}{\partial x_1}(\mathbf{a})\frac{\partial g}{\partial x_2}(\mathbf{a}) - \frac{\partial f}{\partial x_2}(\mathbf{a})\frac{\partial g}{\partial x_1}(\mathbf{a}) = 0.$$

Let us put

$$\lambda^* \stackrel{def}{=} -\frac{\frac{\partial f}{\partial x_1}(\mathbf{a})}{\frac{\partial g}{\partial x_1}(\mathbf{a})} = -\frac{\frac{\partial f}{\partial x_2}(\mathbf{a})}{\frac{\partial g}{\partial x_2}(\mathbf{a})} \tag{4.5}$$

and let us write the Lagrange's auxiliary function for this "multiplier" λ^* :

$$\Phi(\mathbf{x},\lambda^*) = f(\mathbf{x}) + \lambda^* g(\mathbf{x}).‘$$

Let us compute the $\text{grad}\Phi(\mathbf{a},\lambda^*)$ by taking count of the value of λ^* from (4.5):

$$\begin{cases} \frac{\partial \Phi}{\partial x_1}(\mathbf{a},\lambda^*) = \frac{\partial f}{\partial x_1}(\mathbf{a}) + \lambda^* \frac{\partial g}{\partial x_1}(\mathbf{a}) = 0 \\ \frac{\partial \Phi}{\partial x_2}(\mathbf{a},\lambda^*) = \frac{\partial f}{\partial x_2}(\mathbf{a}) + \lambda^* \frac{\partial g}{\partial x_2}(\mathbf{a}) = 0 \\ \frac{\partial \Phi}{\partial \lambda_1}(\mathbf{a},\lambda^*) = g(\mathbf{a}) = 0, \text{ because } \mathbf{a} \in S. \end{cases}$$

Hence $\text{grad}\Phi(\mathbf{a},\lambda^*) = \mathbf{0}$ and the proof is complete. □

Look now at the function

$$\Phi(\mathbf{x}, \boldsymbol{\lambda}^*) = f(\mathbf{x}) + \sum_{j=1}^{m} \lambda_j^* g_j(\mathbf{x}),$$

where $\boldsymbol{\lambda}^* = (\lambda_1^*, \lambda_2^*, ..., \lambda_m^*)$ is the vector just constructed in Theorem 87. It is easy to see that $\mathbf{a}$ is a local conditional maximum (for instance!) for f if and only if $\mathbf{a}$ is an usual local maximum for the function $T(\mathbf{x}) = \Phi(\mathbf{x}, \boldsymbol{\lambda}^*)$. Thus, if we want do decide if a stationary point $(\mathbf{a}, \boldsymbol{\lambda}^*)$ of the Lagrange function is a conditional extremum point, we must consider the second differential of T at $\mathbf{a}$. But, in the expression of $d^2T(\mathbf{a})$ we must take count of the connections between dx_1, $dx_2, ..., dx_n$. These connections can be found by differentiating the equations 4.1:

$$\begin{cases} \frac{\partial g_1}{\partial x_1}(\mathbf{a})dx_1 + ... + \frac{\partial g_1}{\partial x_n}(\mathbf{a})dx_n = 0 \\ \cdot \\ \cdot \\ \cdot \\ \frac{\partial g_m}{\partial x_1}(\mathbf{a})dx_1 + ... + \frac{\partial g_m}{\partial x_n}(\mathbf{a})dx_n = 0 \end{cases}.$$

Since the rank of the Jacobi matrix $J_{\mathbf{a},\mathbf{g}}$ is $m < n$, this linear system in the unknown quantities dx_1, dx_2, ..., dx_n has an infinite number of solutions. Namely, say that the last $n - m$ unknowns $dx_{m+1}, ..., dx_n$ remain free and the others dx_1, dx_2, ..., dx_m can be linearly expressed as functions of the last $n-m$. Thus, the differential $d^2\Phi(\mathbf{a}, \boldsymbol{\lambda}^*)$ becomes a quadratic form in $n-m$ free variables. The sign of this last one must be considered in any discussion about the nature of the point $\mathbf{a}$.

Let us find the points of the compact $x^2 + y^2 \leq 1$ in which the function $f(x, y) = (x-1)^2+(y-2)^2$ has the maximum and the minimum values. Let us find firstly the local extrema inside the disc: $x^2+y^2 \leq 1$.

$$\frac{\partial f}{\partial x} = 2(x-1) = 0, \frac{\partial f}{\partial y} = 2(y-2) = 0.$$

So the critical point is $M(1, 2)$. But this point is outside the disk, thus $M(1, 2)$ is not a local extremum point of f.

Let us consider now the local conditional problem:

$$\max(\min) f$$

with the restriction

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

The auxiliary Lagrange's function is

$$\Phi(x, y, \lambda) = f(x, y) + \lambda(x^2 + y^2 - 1).$$

Let us find its critical points:

$$\begin{cases} \frac{\partial \Phi}{\partial x} = 2(x-1) + 2\lambda x = 0 \\ \frac{\partial \Phi}{\partial y} = 2(y-2) + 2\lambda y = 0 \\ \frac{\partial \Phi}{\partial \lambda} = x^2 + y^2 - 1 = 0 \end{cases}.$$

Solve this system and find $x = \frac{1}{\lambda+1}$ and $y = \frac{2}{\lambda+1}$ (why λ cannot be -1?), $\lambda_1 = \sqrt{5} - 1$, $x_1 = \frac{1}{\sqrt{5}}$, $y_1 = \frac{2}{\sqrt{5}}$ and $\lambda_2 = -\sqrt{5} - 1$, $x_2 = -\frac{1}{\sqrt{5}}$, $y_1 = -\frac{2}{\sqrt{5}}$. Let us denote $M_1(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})$ and $M_2(-\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}})$. In order to see the nature of these critical points, let us find the expression of the second differential of $\Phi(x, y, \lambda)$ for a constant parameter λ. We find

$$d^2\Phi(x, y, \lambda) = (2 + 2\lambda)dx^2 + (2 + 2\lambda)dy^2.$$

Since $xdx + ydy = 0$, then $dy = -\frac{x}{y}dx$, so,

$$d^2\Phi(x, y, \lambda) = (2 + 2\lambda)(1 + \frac{x^2}{y^2})dx^2.$$

For $\lambda_1 = \sqrt{5} - 1$, we get that M_1 is a local conditional minimum. For $\lambda_2 = -\sqrt{5} - 1$, we obtain that M_2 is a local conditional maximum.

Hence, the global maximum of f on the compact subset $\{(x,y) : x^2 + y^2 \leq 1\}$ is $f\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = 6 + 3\sqrt{2}$. Its global minimum is $6 - 3\sqrt{2}$.

Let us consider now a practical problem of conditional extremum. Let us find the distance between the line $x - y = 5$ and the parabola $y = x^2$. Let $L(x_1, y_1)$ be a running point on the line and let $P(x_2, y_2)$ be a running point on the parabola. The square $f(x_1, x_2, y_1, y_2) = (x_1 - x_2)^2 + (y_1 - y_2)^2$ of the distance between two such points must be minimum and the constraints are

$$g_1(x_1, x_2, y_1, y_2) = x_1 - y_1 - 5 = 0$$

and

$$g_2(x_1, x_2, y_1, y_2) = x_2^2 - y_2 = 0.$$

The Lagrange's function is

$$\begin{aligned} \Phi(x_1, x_2, y_1, y_2; \lambda_1, \lambda_2) &= (x_1 - x_2)^2 + (y_1 - y_2)^2 + \\ &\quad + \lambda_1(x_1 - y_1 - 5) + \lambda_2(x_2^2 - y_2). \end{aligned}$$

If we solve the 4×4 algebraic system $\mathrm{grad}\Phi = \mathbf{0}$, we get $x_1 = \frac{23}{8}$, $y_1 = -\frac{17}{8}$, $x_2 = \frac{1}{2}$, $y_2 = \frac{1}{4}$ and the corresponding distance is $\frac{19}{4\sqrt{2}}$.

5. Change of variables

What is the plane curve $xy = 2$? We know that an equation of the form $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ is a hyperbola. If we introduce two new variables X and Y such that $x = \frac{1}{\sqrt{2}}X - \frac{1}{\sqrt{2}}Y$ and $y = \frac{1}{\sqrt{2}}X + \frac{1}{\sqrt{2}}Y$, we introduce in fact a new cartesian coordinate system XOY which is obtained from xOy by a rotation of 45° in the direct sense (see Fig.10.2).

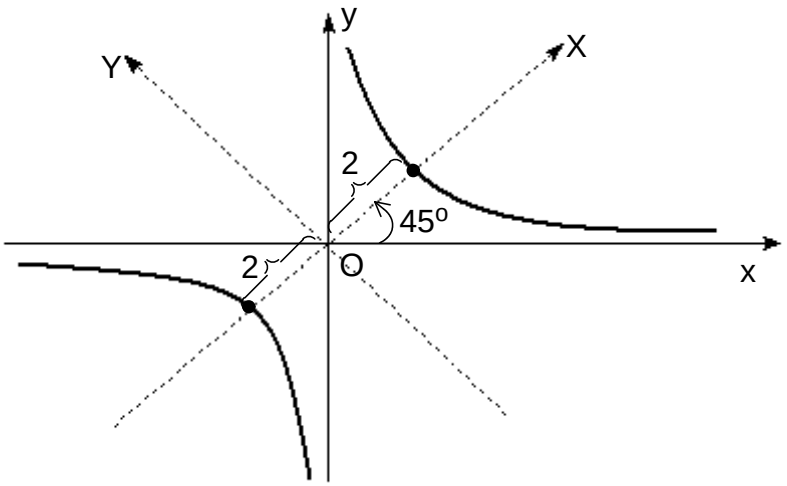


Fig. 10.2

Our initial curve $xy = 2$ becomes $X^2 - Y^2 = 4$, i.e. we have an usual hyperbola with $a = b = 2$ relative to the new cartesian coordinate system XOY.

The moral is that sometimes is better to change the old cartesian coordinate system i.e. to change the old variables $x_1, x_2, ..., x_n$ with another new ones $y_1, y_2, ..., y_n$ which are functions of the first ones:

$$\left\{\begin{array}{l} y_1 = y_1(x_1, x_2, ..., x_n) \\ \qquad\qquad . \\ \qquad\qquad . \\ \qquad\qquad . \\ y_n = y_n(x_1, x_2, ..., x_n) \end{array}\right. . \tag{5.1}$$

Here we forced the notation. The function of n variables which defines the new variable y_1 is also denoted by y_1, etc.

DEFINITION 37. *Let D, Ω be two open subsets of $\mathbb{R}^n$ and let $\mathbf{f} : D \to \Omega$ be a diffeomorphism of class C^k on D, i.e. $\mathbf{f}$ is a bijection, it is of class C^k on D and its inverse $\mathbf{f}^{-1}$ is also of class C^k on Ω. Usually, $k = 1$ or 2. We call such a $\mathbf{f}$ a change of variables of class C^k.*

If we write

$$\mathbf{f}(x_1, x_2, ..., x_n) = (y_1(x_1, x_2, ..., x_n), ..., y_n(x_1, x_2, ..., x_n)),$$

we have a representation like (5.1) for the vector function $\mathbf{f}$. We also call such a representation a change of variables. We represent the inverse

of $\mathbf{f}$ by:

$$\left\{\begin{array}{l} x_1 = x_1(y_1, y_2, ..., y_n) \\ \quad . \\ \quad . \\ \quad . \\ x_n = x_n(y_1, y_2, ..., y_n) \end{array}\right. . \tag{5.2}$$

In fact, we solved the system (5.1) and we computed $x_1, x_2, ..., x_n$ as functions of $y_1, y_2, ..., y_n$. For instance, if $y_1 = x_1 + x_2$ and $y_2 = 2x_1 - x_2$, then $x_1 = \frac{1}{3}(y_1 + y_2)$ and $x_2 = \frac{1}{3}(2y_1 - y_2)$.

If one considers an expression like

$$E(x_1, x_2, ..., x_n, g(x_1, x_2, ..., x_n), \frac{\partial g}{\partial x_j}, \frac{\partial^2 g}{\partial x_j \partial x_i}, ...),$$

the problem is to find an appropriate change of variables of the form (5.2) such that the new expression in the new variables $y_1, y_2, ..., y_n$ has a simpler form. Thus, the "old" function $g(x_1, x_2, ..., x_n)$ becomes a "new" function $\overline{g}(y_1, y_2, ..., y_n)$. The relations between these two functions are

$$\overline{g}(y_1, y_2, ..., y_n) = g(x_1(y_1, y_2, ..., y_n), ..., x_n(y_1, y_2, ..., y_n)) \tag{5.3}$$

and

$$g(x_1, x_2, ..., x_n) = \overline{g}(y_1(x_1, x_2, ..., x_n), ..., y_n(x_1, x_2, ..., x_n)). \tag{5.4}$$

Now, the problem is to express the partial derivatives

$$\frac{\partial g}{\partial x_j}(x_1, x_2, ..., x_n), \frac{\partial^2 g}{\partial x_j \partial x_i}(x_1, x_2, ..., x_n), ...$$

only in language of the partial derivatives of the new function

$\overline{g}(y_1, y_2, ..., y_n)$. This is an easy job if we know to manipulate the chain rules. For instance, if $\mathbf{x} = (x_1, x_2, ..., x_n)$ and $\mathbf{y} = (y_1, y_2, ..., y_n)$, from (5.4) one has:

$$\frac{\partial g}{\partial x_i}(\mathbf{x}) = \frac{\partial \overline{g}}{\partial y_1}(\mathbf{y}) \cdot \frac{\partial y_1}{\partial x_i}(\mathbf{x}) + ... + \frac{\partial \overline{g}}{\partial y_n}(\mathbf{y}) \cdot \frac{\partial y_n}{\partial x_i}(\mathbf{x}),$$

$i = 1, 2, ..., n$. To have "everything" in $y_1, y_2, ..., y_n$ we finally put instead of x_1, $x_1(y_1, y_2, ..., y_n), ...$, instead of x_n, $x_n(y_1, y_2, ..., y_n)$.

For instance, let us make the substitution (change of variables) $x = \exp(t)$ in the following Euler's equation:

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} = 0, x > 0.$$

First of all recall the differential notation: $y = y(x)$, $y'(x) = \frac{dy}{dx}$ (since $dy = y'(x)dx$) and $y''(x) = \frac{d^2y}{dx^2}$ (since $d^2y = y''(x)dx^2$-see the formula for the second differential!). Let us denote by $\overline{y}(t) = y(\exp(t))$. Since $y(x) = \overline{y}(\ln x)$, one has that

$$\frac{dy}{dx} = \frac{d\overline{y}}{dt} \cdot \frac{dt}{dx} = \frac{d\overline{y}}{dt} \cdot \frac{1}{x}, i.e. \frac{d}{dx} = \frac{d}{dt} \cdot \exp(-t).$$

Let us compute

$$\frac{d^2y}{dx^2} = \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d}{dx}\left(\frac{d\overline{y}}{dt} \cdot \exp(-t)\right) = \frac{d}{dt}\left(\frac{d\overline{y}}{dt} \cdot \exp(-t)\right) \cdot \exp(-t).$$

Applying the rule of the differential of a product, we get:

$$\frac{d^2}{dx^2} = \left(\frac{d^2}{dt^2} - \frac{d}{dt}\right) \cdot \exp(-2t).$$

Substituting in the initial equation, we get $\frac{d^2\overline{y}}{dt^2} = 0$, i.e. $\overline{y} = C_1 t + C_2$, where C_1, C_2 are arbitrary constants. Thus, $y(x) = C_1 \ln x + C_2$ and we just found the general solution of the initial differential equation.

6. The Laplacian in polar coordinates

The polar coordinates ρ, θ were introduced in Example 18. The "linear operator" Δ, the Laplacian, carries functions $u(x, y)$ of class C^2, defined on a fixed domain $D \subset \mathbb{R}^2$ into continuous functions:

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, i.e. \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

For instance, in order to solve the famous Laplace equation, $\Delta u = 0$, which appears in many applications, we sometimes need to write the operator Δ in polar coordinates ρ and θ. We know that

$$\begin{cases} x = \rho\cos\theta \\ y = \rho\sin\theta \end{cases},$$

where $\rho \in (0, \infty)$ and $\theta \in [0, 2\pi)$. The Jacobian of this transformation is $\det J_{(\rho,\theta),\mathbf{g}} = \rho \neq 0$, where $\mathbf{g}(\rho, \theta) = (\rho\cos\theta, \rho\sin\theta)$. Let us denote by $\overline{u}(\rho, \theta) = u(\rho\cos\theta, \rho\sin\theta)$, the new function in the new variables ρ and θ. Let us denote by $\rho = \rho(x, y)$ and by $\theta = \theta(x, y)$ the coordinates of the inverse function $\mathbf{g}^{-1}$. Thus,

$$u(x, y) = \overline{u}(\rho(x, y), \theta(x, y)).$$

Hence,

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial \overline{u}}{\partial \rho}\frac{\partial \rho}{\partial x} + \frac{\partial \overline{u}}{\partial \theta}\frac{\partial \theta}{\partial x} \\ \frac{\partial u}{\partial y} = \frac{\partial \overline{u}}{\partial \rho}\frac{\partial \rho}{\partial y} + \frac{\partial \overline{u}}{\partial \theta}\frac{\partial \theta}{\partial y} \end{cases} \tag{6.1}$$

These last relations can be represented in a matrix form

$$\begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{pmatrix} = \begin{pmatrix} \frac{\partial \rho}{\partial x} & \frac{\partial \theta}{\partial x} \\ \frac{\partial \rho}{\partial y} & \frac{\partial \theta}{\partial y} \end{pmatrix} \begin{pmatrix} \frac{\partial \overline{u}}{\partial \rho} \\ \frac{\partial \overline{u}}{\partial \theta} \end{pmatrix}. \tag{6.2}$$

Since $\mathbf{g} \circ \mathbf{g}^{-1} =$ the identity mapping, we have that

$$\begin{pmatrix} \frac{\partial \rho}{\partial x} & \frac{\partial \theta}{\partial x} \\ \frac{\partial \rho}{\partial y} & \frac{\partial \theta}{\partial y} \end{pmatrix}^{trans} = \left[J_{(\rho,\theta),\mathbf{g}}\right]^{-1} = \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix}^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\frac{\sin\theta}{\rho} & \frac{\cos\theta}{\rho} \end{pmatrix}.$$

Let us come back to formula 6.2 and find:

$$\begin{pmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\frac{\sin\theta}{\rho} \\ \sin\theta & \frac{\cos\theta}{\rho} \end{pmatrix} \begin{pmatrix} \frac{\partial \overline{u}}{\partial \rho} \\ \frac{\partial \overline{u}}{\partial \theta} \end{pmatrix}.$$

Let us write this formula in a nonmatriceal form:

$$\begin{cases} \frac{\partial u}{\partial x} = \frac{\partial \overline{u}}{\partial \rho}\cos\theta - \frac{\partial \overline{u}}{\partial \theta}\frac{\sin\theta}{\rho} \\ \frac{\partial u}{\partial y} = \frac{\partial \overline{u}}{\partial \rho}\sin\theta + \frac{\partial \overline{u}}{\partial \theta}\frac{\cos\theta}{\rho} \end{cases}. \tag{6.3}$$

Let us use now these formulas and the chain rules formulas 2.7, 2.8 to compute $\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$:

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 \overline{u}}{\partial \rho^2}\cos^2\theta - 2\frac{\partial^2 \overline{u}}{\partial \rho \partial \theta}\frac{\sin\theta\cos\theta}{\rho} + \frac{\partial^2 \overline{u}}{\partial \theta^2}\frac{\sin^2\theta}{\rho^2} + \frac{\partial \overline{u}}{\partial \rho}\frac{\sin^2\theta}{\rho} + 2\frac{\partial \overline{u}}{\partial \theta}\frac{\sin\theta\cos\theta}{\rho^2},$$

$$\frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 \overline{u}}{\partial \rho^2}\sin^2\theta + 2\frac{\partial^2 \overline{u}}{\partial \rho \partial \theta}\frac{\sin\theta\cos\theta}{\rho} + \frac{\partial^2 \overline{u}}{\partial \theta^2}\frac{\cos^2\theta}{\rho^2} + \frac{\partial \overline{u}}{\partial \rho}\frac{\cos^2\theta}{\rho} - 2\frac{\partial \overline{u}}{\partial \theta}\frac{\sin\theta\cos\theta}{\rho^2}.$$

Hence, the formula for the Laplacian in polar coordinates is:

$$\Delta u = \frac{\partial^2 \overline{u}}{\partial \rho^2} + \frac{1}{\rho^2}\frac{\partial^2 \overline{u}}{\partial \theta^2} + \frac{1}{\rho}\frac{\partial \overline{u}}{\partial \rho}.$$

This formula will be used later in the course of partial differential equations with direct applications in Engineering.

7. A proof for the Local Inversion Theorem

Here we present a complete proof for the Local Inversion Theorem (see Theorem 80). We prefer an elementary longer proof then a shorter sophisticated one. Let us state again this basic result.

THEOREM 88. *Let A be an open subset of $\mathbb{R}^n$ and let $\mathbf{f} : A \to \mathbb{R}^n$ be a function of class C^1 on A. Let $\mathbf{a}$ be a point in A such that the Jacobian determinant* $\det J_{\mathbf{a},\mathbf{f}} \neq 0$. *Then there are two open sets $X \subset A$ and $Y \subset \mathbf{f}(A)$ and a uniquely determined function $\mathbf{g}$ with the following properties:*

i) $\mathbf{a} \in A$ *and* $\mathbf{f}(\mathbf{a}) \in Y$,

ii) $Y = \mathbf{f}(X)$,

iii) $\mathbf{g} : Y \to X$, $\mathbf{g}(Y) = X$ *and* $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x}$ *for any* $\mathbf{x}$ *in* X,

iv) $\mathbf{g}$ *is of class* C^1 *on* Y *and the restriction of* f *to* X, $f\mid_X: X \to Y$ *is a diffeomorphism with* $\mathbf{g} = (\mathbf{f}\mid_X)^{-1}$. *Particularly,*

$$J_{\mathbf{f}(\mathbf{x}),\mathbf{g}} = (J_{\mathbf{x},\mathbf{f}})^{-1}$$

and

$$\det J_{\mathbf{f}(\mathbf{x}),\mathbf{g}} = \frac{1}{\det J_{\mathbf{x},\mathbf{f}}}.$$

PROOF. STEP 1. First of all let us remark that if $(h_{ij}(\mathbf{x})), i, j = 1, 2, ..., n$ are n^2 continuous functions defined on A, such that

$\det[h_{ij}(\mathbf{a})] \neq 0$, then there is a small closed ball $B[\mathbf{a}, r]$ with centre at $\mathbf{a}$ and of radius $r > 0$, $B[\mathbf{a}, r] \subset A$ with the property that whenever we take n^2 points $\{\mathbf{x}_{ij}\}$ in $B[\mathbf{a}, r]$, one has that $\det[h_{ij}(\mathbf{x}_{ij})] \neq 0$. Indeed, let us define a continuous function of n^2 variables on the product $\underbrace{A \times A \times ... \times A}_{n^2-\text{times}}$:

$$D(\mathbf{X}_{11}, \mathbf{X}_{12}, ..., \mathbf{X}_{1n}, ..., \mathbf{X}_{n1}, \mathbf{X}_{n2}, ..., \mathbf{X}_{nn}) = \det[h_{ij}(\mathbf{X}_{ij})].$$

Since $D(\mathbf{a}, \mathbf{a}, ..., \mathbf{a}) = \det(h_{ij}(\mathbf{a}))$ is not zero, say $D(\mathbf{a}, \mathbf{a}, ..., \mathbf{a}) > 0$, one can find a small ball $B(\mathbf{a}, r') \subset A$, $r' > 0$, on which

$$D(\mathbf{x}_{11}, \mathbf{x}_{12}, ..., \mathbf{x}_{nn}) = \det(h_{ij}(\mathbf{x}_{ij})) > \mathbf{0}$$

for every $\mathbf{x}_{ij}$ in $B(\mathbf{a}, r')$ (see Theorem 57). If one takes any r, $0 < r < r'$, then $\det(h_{ij}(\mathbf{x}_{ij})) > \mathbf{0}$ for any arbitrary n^2 elements $\{\mathbf{x}_{ij}\}$ in $B[\mathbf{a}, r]$. In our case, $\det J_{\mathbf{a},\mathbf{f}} = \left(\det \frac{\partial f_i}{\partial x_j}(\mathbf{a})\right) \neq 0$, where $\mathbf{f} = (f_1, f_2, ..., f_n)$. Hence, we can find a small closed ball $W = B[\mathbf{a}, r] \subset A$, $r > 0$, on which $\left(\det \frac{\partial f_i}{\partial x_j}(\mathbf{x}_{ij})\right) \neq 0$ for any n^2 elements $\mathbf{x}_{ij}$ in W.

STEP 2. Let us prove now that the restriction of $\mathbf{f}$ to W is one-to-one. Suppose that $\mathbf{x}$ and $\mathbf{z}$ are in W such that $\mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{z})$. This means that for every $i = 1, 2, ..., n$ one has that $f_i(\mathbf{x}) = f_i(\mathbf{z})$. Let us apply the Lagrange theorem (see Theorem 73) on the segment $[\mathbf{x}, \mathbf{z}]$:

$$0 = f_i(\mathbf{x}) - f_i(\mathbf{z}) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\mathbf{c}^{(i)}) \cdot (x_j - z_j), \tag{7.1}$$

where $\mathbf{c}^{(i)}$ is a point on the segment $[\mathbf{x}, \mathbf{z}]$ and $\mathbf{x} = (x_1, x_2, ..., x_n)$, $\mathbf{z} = (z_1, z_2, ..., z_n)$. Since the segment $[\mathbf{x}, \mathbf{z}]$ is contained in W (why?), all $\mathbf{c}^{(i)}$, $i = 1, 2, ..., n$, are contained in W and so, $\det\left(\frac{\partial f_i}{\partial x_j}(\mathbf{c}^{(i)})\right) \neq 0$. Hence, the homogeneous linear system

$$0 = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\mathbf{c}^{(i)}) \cdot (x_j - z_j),$$

$i = 1, 2, ..., n$, in the unknowns $x_1 - z_1, x_2 - z_2, ..., x_n - z_n$, has only the trivial solution, i.e. $x_1 = z_1, ..., x_n = z_n$ or $\mathbf{x} = \mathbf{z}$. Thus, $\mathbf{f}$ is one-to-one on $W = B[\mathbf{a}, r]$.

STEP 3. Let us prove now that the image $\mathbf{f}(Z)$ of $Z = B(\mathbf{a}, r)$, the interior of W, is an open subset of $\mathbb{R}^n$. Indeed, let us define the continuous function $g : \partial Z \to \mathbb{R}$ (here $\partial Z = W \setminus Z$ is the boundary of Z):

$$g(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\|,$$

for $\mathbf{x} \in \partial Z$. Since ∂Z is a compact subset of $\mathbb{R}^n$ (prove it!) and since $\mathbf{f}$ is one-to-one (see STEP 2), the minimum value m of g on ∂Z is > 0 (why?). Let us denote by $T = B(\mathbf{f}(\mathbf{a}), \frac{m}{2})$ and let us prove that this open ball T is contained in $\mathbf{f}(Z)$. For this, let $\mathbf{y}$ be a fixed element in T and let us define the following continuous function:

$$h(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|$$

for any $\mathbf{x}$ in W. Let us see that the absolute minimum of h cannot be attained on the boundary ∂Z. Indeed, since

$$h(\mathbf{a}) = \|\mathbf{f}(\mathbf{a}) - \mathbf{y}\| < \frac{m}{2},$$

one has that $\min h(\mathbf{x}) < \frac{m}{2}$. But, if $\mathbf{x} \in \partial \mathbf{Z}$, we have

$$\begin{aligned} h(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\| &\geq \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\| - \|\mathbf{f}(\mathbf{a}) - \mathbf{y}\| \\ &> g(\mathbf{x}) - \frac{m}{2} \geq \frac{m}{2}, \end{aligned}$$

i.e. $h(\mathbf{x}) > \frac{m}{2}$ for any $\mathbf{x}$ in ∂Z. Hence, let $\mathbf{c}$ be in Z such that

$$h(\mathbf{c}) = \min\{h(\mathbf{x}) : \mathbf{x} \in W\}.$$

This $\mathbf{c}$ also realizes the absolute minimum for

$$h^2(\mathbf{x}) = \|\mathbf{f}(\mathbf{x}) - \mathbf{y}\|^2 = \sum_{r=1}^{n} [f_r(\mathbf{x}) - y_r]^2.$$

Then Fermat's theorem says that:

$$\frac{\partial}{\partial x_k} \left\{ \sum_{r=1}^{n} [f_r(\mathbf{x}) - y_r]^2 \right\} = 2 \sum_{r=1}^{n} [f_r(\mathbf{x}) - y_r] \cdot \frac{\partial f_r}{\partial x_k}(\mathbf{x})$$

is zero at $\mathbf{c}$, i.e.

$$\sum_{r=1}^{n} \frac{\partial f_r}{\partial x_k}(\mathbf{c}) \cdot [f_r(\mathbf{c}) - y_r] = 0$$

for every $k = 1, 2, ..., n$. This is again a homogenous linear system in the unknowns $\{f_r(\mathbf{c}) - y_r\}_r$ with a nonzero determinant. Hence, we have only the trivial solution, i.e. $f_r(\mathbf{c}) = y_r$ for every $r = 1, 2, ..., n$. Thus, $\mathbf{f}(\mathbf{c}) = \mathbf{y}$ and so $\mathbf{y} \in \mathbf{f}(Z)$. But, the same type of reasoning can

be done for any other $\mathbf{b} = \mathbf{f}(\mathbf{e})$, where $\mathbf{e} \in Z$ and $\mathbf{b} \in \mathbf{f}(Z)$. Namely, we take a sufficiently small open ball $B(\mathbf{e}, r'') \subset B(\mathbf{a}, r)$ and we repeat the above reasoning for $B(\mathbf{e}, r'')$ instead of $B(\mathbf{a}, r)$. We find that

$$T' = B(\mathbf{b}, \frac{m'}{2}) \subset \mathbf{f}(B(\mathbf{e}, r'')) \subset \mathbf{f}(Z)$$

for the minimum m' of the function

$$\mathbf{x} \to \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{e})\|,$$

defined on $\partial B(\mathbf{e}, r'')$. Hence, $\mathbf{f}(Z)$ is open in $\mathbb{R}^n$. Moreover, $\mathbf{f}$ carries an open subset X of Z into an open subset $\mathbf{f}(X)$ of $\mathbb{R}^n$ (why?).

STEP 4. Let now $Y = B(\mathbf{f}(\mathbf{a}), r')$ be an open ball centered at $\mathbf{f}(\mathbf{a})$ such that its closure $B[\mathbf{f}(\mathbf{a}), r']$ is included in $\mathbf{f}(Z)$ and let $X = \mathbf{f}^{-1}(Y) \cap Z$. It is clear that the restriction $\mathbf{f}\mid_X : X \to Y$ is a continuous bijection between X and Y. Let $\mathbf{g} : Y \to X$, $\mathbf{g}(\mathbf{y}) = \mathbf{x}$ be its inverse. Let $\overline{X}$ and $\overline{Y}$ be the topological closure of X and Y respectively. They both are compact subsets of $\mathbb{R}^n$ and $\mathbf{f}\mid_{\overline{X}}: \overline{X} \to \overline{Y}$ is also a bijection, because $\overline{X} \subset W$ and $\mathbf{f}$ is one-to-one on W (see STEP 1). Its inverse $(\mathbf{f}\mid_{\overline{X}})^{-1} : \overline{Y} \to \overline{X}$ is continuous (because $\mathbf{f}$ is continuous and $\overline{X}$ and $\overline{Y}$ are compact sets...it reverses closed subsets into closed subsets!). Since the restriction of $(\mathbf{f}\mid_{\overline{X}})^{-1}$ to Y is exactly $\mathbf{g}$ (why?), $\mathbf{g}$ is also a continuous mapping and $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x}$ for any $\mathbf{x}$ in X.

STEP 5. It remains us to prove that $\mathbf{g} = (g_1, g_2, ..., g_n)$ is of class C^1 on Y. We fix an $r = 1, 2, ..., n$ and we shall prove that $\frac{\partial g_j}{\partial y_r}$ exists at any fixed point $\mathbf{y}$ in Y and that they are continuous. Let $\mathbf{e}_r = (0, 0, ..., 0, 1, 0, ..., 0)$ be the r-th unit vector in $\mathbb{R}^n$ (with 1 at the r-th position!) and let us consider the difference quotient:

$$\frac{g_j(\mathbf{y} + t\mathbf{e}_r) - g_j(\mathbf{y})}{t}, \tag{7.2}$$

where t is a small real number such that $\mathbf{y} + t\mathbf{e}_r \in Y$ (Y is open). Let $\mathbf{x} = \mathbf{g}(\mathbf{y})$ and $\mathbf{x}' = \mathbf{g}(\mathbf{y} + t\mathbf{e}_r)$. Thus,

$$\mathbf{f}(\mathbf{x}') - \mathbf{f}(\mathbf{x}) = t\mathbf{e}_r$$

implies that

$$f_i(\mathbf{x}') - f_i(\mathbf{x}) = \begin{cases} 0, & \text{if } i \neq r, \\ t, & \text{if } i = r. \end{cases} \tag{7.3}$$

Let us apply Lagrange's theorem (see Theorem 73) for f_i on the segment $[\mathbf{x}, \mathbf{x}'] \subset Z$. We get:

$$0 \text{ or } 1 = \frac{f_i(\mathbf{x}') - f_i(\mathbf{x})}{t} = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\mathbf{d}^{(i)}) \cdot \frac{x_j' - x_j}{t}, \tag{7.4}$$

$i = 1, 2, ..., n$, where $\mathbf{d}^{(i)}$ is a point on the segment $[\mathbf{x}, \mathbf{x}'] \subset Z$. Since $\det\left[\frac{\partial f_i}{\partial x_j}(\mathbf{d}^{(i)})\right] \neq 0$, the linear system (7.4), in variables $\left\{\frac{x'_j - x_j}{t}\right\}_j$ has a unique solution (Cramer's rule):

$$\frac{x'_j - x_j}{t} = \frac{\Delta_j}{\Delta},$$

$j = 1, 2, ..., n$, where Δ and Δ_j are determinants with entries of the form $\frac{\partial f_i}{\partial x_j}(\mathbf{d}^{(i)})$, 0, or 1. When $t \to 0$, the determinant $\Delta \to J_{\mathbf{x},\mathbf{f}} \neq 0$ (why?), so

$$\left(\frac{\Delta_1}{\Delta}, \frac{\Delta_2}{\Delta}, ..., \frac{\Delta_n}{\Delta}\right) \to \left(\frac{\partial g_1}{\partial y_r}(\mathbf{y}), \frac{\partial g_2}{\partial y_r}(\mathbf{y}), ..., \frac{\partial g_n}{\partial y_r}(\mathbf{y})\right),$$

i.e all the partial derivatives $\frac{\partial g_j}{\partial y_r}(\mathbf{y})$ exist. Since their expressions involve only partial derivatives of the type $\frac{\partial f_i}{\partial x_j}(\mathbf{x})$ which are continuous, the function $\mathbf{g}$ is of class C^1 on Y and the proof of the Local Inversion Theorem is now complete. □

The proof is long, but elementary and very natural. Trying to understand this proof one remembers many basic things from previous chapters. Moreover, the proof itself reflects some of the indescribable Beauty of Mathematical Analysis.

8. The derivative of a function of a complex variable

Let A be an open subset of the complex plane $\mathbb{C}$. If we associate to any complex number $z = x + iy$ of A, where x, y are real numbers and $i = \sqrt{-1}$ is a fixed root of the equation $x^2 + 1 = 0$, another complex number $w = f(z)$, we say that the mapping $z \to f(z)$ is a function of a complex variable defined on A. Like in the case of a function of a real variable, we say that f has the limit L at the point $z_0 = x_0 + iy_0$ of A if for any sequence $\{z_n\}$, $n = 1, 2, ...$, of complex numbers $z_n = x_n + iy_n$, $x_n, y_n \in \mathbb{R}$, which tends to a, one has that $f(z_n) \to L$. If $L = f(z_0)$ we say that f is continuous at z_0. Let us assume that $f(x + iy) = u(x, y) + iv(x, y)$, where u and v are two real functions of two variables. One calls $u = \operatorname{Re} f$, the real part of f and $v = \operatorname{Im} f$, the imaginary part of f. It is not difficult to see that f is continuous at $z_0 = x_0 + iy_0$ if and only if u and v are continuous at (x_0, y_0). Let us define the derivative of a function f of a complex variable z at a fixed point z_0. We say that f is differentiable at z_0 if

the following limit exists and is finite:

$$\lim_{z\to z_0}\frac{f(z)-f(z_0)}{z-z_0}=f'(z_0). \tag{8.1}$$

We denoted its value by $f'(z_0)$ and we call it the derivative of f at z_0. For instance, $(z^2)'=2z$, because

$$\lim_{z\to z_0}\frac{z^2-z_0^2}{z-z_0}=\lim_{z\to z_0}(z+z_0)=2z_0.$$

Generally speaking, the usual differential rules of the functions of a real variable also works for functions of a complex variable. For instance, $(f+g)'=f'+g'$, $(\alpha f)'=\alpha f'$, $(fg)'=f'g+fg'$, $\left(\frac{f}{g}\right)'=\frac{f'g-fg'}{g^2}$, $(f\circ g)'(z)=f'(g(z))\cdot g'(z)$, $(\sin z)'=\cos z$, $(\exp(z))'=\exp(z)$, etc. Many formulas in complex function theory (the theory of functions of a complex variable) can be easily proved by using the following fundamental result.

THEOREM 89. *(Identity Theorem) Let A be a subset of complex numbers with at least one limit point and let f and g be two differentiable complex functions defined on a complex domain B (it is open and connected) which contains A. Assume that f and g are equal at any point of A. Then f and g are identical, this means that $f(z)=g(z)$ for all z of B.*

For a proof of this basic result see any book of complex function theory (see for instance [**ST**]). Let us use this result to compute the derivative of $\exp(z)=\sum_{n=0}^{\infty}\frac{z^n}{n!}$, $z\in\mathbb{C}$. Let us denote by $g(z)$ the derivative of $\exp(z)$. Since for any real number x one has that $\exp(x)'=\exp(x)$, we have that $g(x)=\exp(x)$ for any x in $\mathbb{R}$. But all the point of $\mathbb{R}$ are limit points so, $g(z)=\exp(z)$. Here we tacitly used another basic result of complex function theory.

THEOREM 90. *If a complex function $f:A\to\mathbb{C}$, where A is a complex domain, is differentiable on A, then it has derivatives of any order on A, i.e. it is of class C^∞ on A.*

Following an analogous theory like the Weierstrass theory for the real series of functions, we can prove that $\exp(z)$ is a differential function. Hence, its derivative $g(z)$ is also differentiable on $\mathbb{C}$. This is why we could apply Theorem 89 for the complex function $\exp(z)$.

What can we say about the two variables real functions $u=\operatorname{Re}f$ and $v=\operatorname{Im}f$ if f is differentiable at a point z_0?

THEOREM 91. *(Cauchy-Riemann relations) If the function $f(x+iy)=u(x,y)+iv(x,y)$ is differentiable at a point $z_0=x_0+iy_0$, then the*

two variables real functions u and v have partial derivatives at (x_0, y_0) and between them we have the following relations (the Cauchy-Riemann relations):

$$\frac{\partial u}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0),\ \frac{\partial u}{\partial y}(x_0, y_0) = -\frac{\partial v}{\partial x}(x_0, y_0) \tag{8.2}$$

Moreover, $f'(z_0) = \frac{\partial u}{\partial x}(x_0, y_0) + i\frac{\partial v}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0) - i\frac{\partial u}{\partial y}(x_0, y_0)$.

PROOF. If f is differentiable at the point z_0 the following limit exists:

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} = f'(z_0).$$

This means that for any sequence (x_n, y_n) which converges to (x_0, y_0) (in $\mathbb{R}^2$) one has that

$$\lim_{x_n \to x_0, y_n \to y_0} \frac{u(x_n, y_n) - u(x_0, y_0) + i[v(x_n, y_n) - v(x_0, y_0)]}{x_n - x_0 + i(y_n - y_0)} = f'(z_0). \tag{8.3}$$

Firstly take here $y_n = y_0$ for any $n = 1, 2, ...$ We get

$$\frac{\partial u}{\partial x}(x_0, y_0) + i\frac{\partial v}{\partial x}(x_0, y_0) = f'(z_0). \tag{8.4}$$

Secondly, let us consider in (8.3) $x_n = x_0$ for any $n = 1, 2, ...$ We find

$$\frac{1}{i}\left[\frac{\partial u}{\partial y}(x_0, y_0) + i\frac{\partial v}{\partial y}(x_0, y_0)\right] = f'(z_0) \tag{8.5}$$

Comparing (8.3) and (8.5) we get the Cauchy-Riemann relations (8.2). □

The Cauchy-Riemann relations imply that the real and the imaginary part of a differentiable complex function are harmonic functions, i.e. they are solutions of the Laplace equation:

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{8.6}$$

and

$$\Delta v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0$$

(prove it!).

Let $f = u + iv$ be a complex function differentiable on a complex open subset A and let $\mathbf{F}(x, y) = (v(x, y), u(x, y))$ be its associated field of plane forces. By definition, the curl (the rotational) of $\mathbf{F}$ is the 3-D vector field $\operatorname{curl} \mathbf{F} = (0, 0, \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y})$. Since $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$ on A, one sees that $\operatorname{curl} \mathbf{F} = \mathbf{0}$ i.e. the vector field $\mathbf{F}$ is irrotational. By definition, the

divergence of $\mathbf{F}$ is $\operatorname{div}\mathbf{F} = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}$. But this last one is 0 because of the second Cauchy-Riemann relation.

Moreover, if one know one of the two functions u or v, one can determine the other up to a complex constant, such that the couple (u, v) be the real and the imaginary part respectively of a differentiable complex function f. Indeed, suppose we know u and we want to find v from the Cauchy-Riemann relations:

$$\frac{\partial v}{\partial x}(x, y) = -\frac{\partial u}{\partial y}(x, y) \tag{8.7}$$

and

$$\frac{\partial v}{\partial y}(x, y) = \frac{\partial u}{\partial x}(x, y) \tag{8.8}$$

From (8.7) we can write

$$v(x, y) = -\int \frac{\partial u}{\partial y}(x, y)dx + C(y).$$

We prove that we can determine the unknown function $C(y)$ up to a constant term. Let us come to the relation (8.8) with this last expression of v. Here we use the famous Leibniz formula on the differential of an integral with a parameter (see the Integral calculus in any course of Analysis):

$$\frac{\partial u}{\partial x}(x, y) = -\int \frac{\partial^2 u}{\partial y^2}(x, y)dx + C'(y).$$

From (8.6) we find

$$\frac{\partial u}{\partial x}(x, y) = \int \frac{\partial^2 u}{\partial x^2}(x, y)dx + C'(y) = \frac{\partial u}{\partial x}(x, y) + K(y) + C'(y), \tag{8.9}$$

where $C(y)$ and $K(y)$ are functions of y. From (8.9) we get

$$C'(y) = -K(y).$$

Therefore, always one can find the function $C(y)$, and so the function $v(x, y)$ up to a real constant c. Hence, we can determine the function $f = u + iv$ up to a purely imaginary constant ic.

For instance, let us consider $u(x, y) = x^2 - y^2$ and let us find f (if it is possible! It is, because u is a harmonic function!-this is the only thing we used above!). The Cauchy-Riemann relations become:

$$\frac{\partial v}{\partial x}(x, y) = 2y$$

and

$$\frac{\partial v}{\partial y}(x, y) = 2x$$

Let us integrate the first equality with respect to x

$$v(x,y) = 2xy + C(y),$$

where $C(y)$ is a constant function with respect to x but,...it can depend on y! Come now to the second relation and find

$$2x = 2x + C'(y),$$

so, $C'(y) = 0$, i.e. $C(y)$ does not depend on y. It is a pure constant c. Hence, $v(x,y) = 2xy + c$ and $f(z) = x^2 - y^2 + i(2xy + c) = (x+iy)^2 + ic$, where c is a real arbitrary constant.

Let us now come back to formula (8.1) and consider an arbitrary smooth curve γ which passes through z_0. Let us take z very close to z_0 but on the curve γ. So, we can approximate:

$$\frac{f(z) - f(z_0)}{z - z_0} \approx f'(z_0) \tag{8.10}$$

Hence,

$$|f(z) - f(z_0)| \approx |z - z_0|\,|f'(z_0)| =$$
$$|z - z_0| \sqrt{\left[\frac{\partial u}{\partial x}(x_0, y_0)\right]^2 + \left[\frac{\partial v}{\partial x}(x_0, y_0)\right]^2}.$$

So, the length of the segment $[f(z_0), f(z)]$ is proportional to the length of the segment $[z_0, z]$. The "dilation" coefficient

$$\lambda = \sqrt{\left[\frac{\partial u}{\partial x}(x_0, y_0)\right]^2 + \left[\frac{\partial v}{\partial x}(x_0, y_0)\right]^2}$$

does not depend on the curve on which z becomes closer and closer to z_0.

Let us recall that any complex number z can be uniquely written as: $z = r\exp(i\alpha)$, where $\alpha \in [0, 2\pi)$. This angle α is called the argument of z. From the formula (8.10) we get

$$\arg\left[f(z) - f(z_0)\right] \approx \arg(z - z_0) + \arg f'(z_0). \tag{8.11}$$

Here we assume that $f'(z_0) \neq 0$. Formula (8.11) says that in a small neighborhood of z_0 our differentiable function preserve the angle between two curves which pass through z_0 (why?). So, we can locally approximate the action of a differentiable function by a rotation of angle $\arg f'(z_0)$, followed by a "dilation"(or a "contraction") of coefficient $|f'(z_0)|$. We assume that $f'(z_0) \neq 0$. Otherwise, the transformation $z \to f(z)$ is almost constant around z_0. A transformation of the complex plane into itself with this last two properties is called a conformal transformation. These are very important in some engineering applications (hydraulics, fluid mechanics, electricity, etc.).

If we write the plane transformation $z \to f(z)$ as

$$(x, y) \to (u(x, y), v(x, y)),$$

where $f(z) = u + iv$, the Jacobian determinant of this at (x_0, y_0) is

$$\begin{vmatrix} \frac{\partial u}{\partial x}(x_0, y_0) & \frac{\partial u}{\partial y}(x_0, y_0) \\ \frac{\partial v}{\partial x}(x_0, y_0) & \frac{\partial v}{\partial y}(x_0, y_0) \end{vmatrix} =$$

$$\left[\frac{\partial u}{\partial x}(x_0, y_0)\right]^2 + \left[\frac{\partial v}{\partial x}(x_0, y_0)\right]^2 = |f'(z_0)|^2 .$$

Here we used again the Cauchy-Riemann relations. If we want that our transformation $z \to f(z)$ to be locally invertible around the point z_0, we must assume that $f'(z_0) \neq 0$ (see the Local Inversion Theorem). In this last case, this transformation is locally a conformal transformation, i.e. it preserves the angles (with their directions) and it changes the lengthens with the same "velocity" around the point z_0.

9. Problems

1. Find $y'(x)$ if $y = 1+y^x$. Why we cannot perform this computation for the points on the curve $xy^{x-1} = 1, y > 0$?

2. Compute $\frac{dy}{dx}$ and $\frac{d^2y}{dx^2}$, if $y = x + \ln y$, $y \neq 1$.

3. If $z = z(x, y)$ and

$$x^3 + 2y^3 + z^3 - 3xyz - 2y + 3 = 0,$$

find dz and d^2z.

4. Find $\inf f$ and $\sup f$ for:

a)

$$f(x, y) = x^3 + 3xy^2 - 15x - 12y;$$

b)

$$f(x, y) = xy$$

with $x + y - 1 = 0$;

c)

$$f(x, y, z) = x^2 + y^2 + z^2$$

with $ax + by + cz - 1 = 0$ (What this means?);

5. Find the distance from $M(0, 0, 1)$ to the curve $\{y = x^2\} \cap \{z = x^2\}$.

6. Find the distance between the line $3x + y - 9 = 0$ and the ellipse $\frac{x^2}{4} + \frac{y^2}{9} - 1 = 0$.

7. Compute the velocity and the acceleration on the circle

$$\{x^2 + y^2 + z^2 = a^2\} \cap \{x + y + z = a\}$$

by using a parametrization of the type: $x = x$, $y = y(x)$, $z = z(x)$.

8. Are the functions

$$u = (x+y+z)^2, v = 3x - y + 3z, w = x^2 + xy + yz + zx$$

independent at $(0,0,0)$?

9. Change the variables in the following expressions:

a)

$$(1-x^2)\frac{d^2y}{dx^2} - x\frac{dy}{dx} + \omega y = 0,$$

$x = \cos t$;

b)

$$x^2\frac{\partial^2 z}{\partial x^2} - y^2\frac{\partial^2 z}{\partial y^2} = 0, u = xy, v = \frac{x}{y};$$

c) $\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2$, $x = \rho\cos\theta$, $y = \rho\sin\theta$;

10. Find all Φ such that $u = \Phi(x+y)$ and $v = \Phi(x)\Phi(y)$ be dependent on $\mathbb{R}^2$.

11. Prove that the following complex functions are differentiable and find their derivatives. Take a point z_0 and study the geometrical behavior of the transformation $z \to f(z)$ around this point z_0.

a) $f(z) = 3z + 2$; b) $f(z) = 2iz + 3$; c) $f(z) = \frac{1}{z}$, $|z| > 1$;

d) $f(z) = \exp(iz)$; e) $f(z) = z^3 + 2$, $z \neq 0$; g) $f(z) = z\sin z$;

Bibliography

[A] T. M. Apostol, *Mathematical Analysis*, Narosa Publishing House, India, 2002.

[Dem] B. Demidovich, *Problems in Mathematical Analysis,* Mir Publishers, Moscow, 1989.

[DOG] C. Drăguşin, O. Olteanu, M. Gavrilă, *Mathematical Analysis. Theory and Applications* (Romanian), Vol. I and Vol. II, Matrix Rom, Bucharest, 2006, 2007.

[EP] E. Popescu, *Mathematical Analysis (Differential Calculus)* (Romanian), Matrix Rom, Bucharest, 2006.

[FS] P. Flondor, O. Stănăşilă, *Lectures in Mathematical Analysis* (Romanian), All Publishers, Bucharest, 1993.

[GG] G. Groza, *Numerical Analysis* (Romanian), Matrix Rom, 2005.

[JJ] J. Jost, *Postmodern Analysis, Springer, 2002.*

[La] S. Lang, *Calculus of several variables*, Springer Verlag, 1996.

[Nik] S. M. Nikolsky, *A course of Mathematical Analysis*, Vol. I, II, Mir Publishers, Moskow, 1981.

[Pal] G. Păltineanu, *Mathematical Analysis. Differential Calculus* (Romanian), AGIR Publishers, Bucharest, 2002.

[Pro] *** *Problems in Mathematical Analysis* (Romanian), Department of Math. and Computer Science, TUCIB, Matrix Rom, Bucharest, 2002.

[ST] A. Sveshnikov, A. Tikhonov, *The Theory of Functions of a Complex Variable,* Mir Publishers, 1978.

[R] W. Rudin, *Principles of Mathematical Analysis,* McGraw-Hill, N.Y., 1964.